Bruno de Finetti (Ed.)

Mathematical Optimisation in Economics

CIME Summer Schools

38

L´Aquila, Italy 1965

Springer

FONDAZIONE CIME ROBERTO CONTI

Bruno de Finetti (Ed.)

Mathematical Optimisation in Economics

Lectures given at a Summer School of the
Centro Internazionale Matematico Estivo (C.I.M.E.),
held in L´Aquila, Italy,
August 29-September 7, 1965

Springer

FONDAZIONE
CIME
ROBERTO CONTI

C.I.M.E. Foundation
c/o Dipartimento di Matematica "U. Dini"
Viale Morgagni n. 67/a
50134 Firenze
Italy
cime@math.unifi.it

ISBN 978-3-642-11038-2 e-ISBN: 978-3-642-11039-9
DOI:10.1007/978-3-642-11039-9
Springer Heidelberg Dordrecht London New York

Reprint of the 1st ed. C.I.M.E., Ed. Cremonese, Roma 1966
With kind permission of C.I.M.E.

Printed on acid-free paper

Springer.com

CENTRO INTERNAZIONALE MATEMATICO ESTIVO

(C.I.M.E.)

3 ciclo - L''Aquila 30 agosto - 7 settembre 1965

"METHODS OF MATHEMATICAL OPTIMIZATION IN ECONOMICS"

Coordinatore : Bruno de Finetti

SPUNTI PER UNA DISCUSSIONE

Je veux seulement prèsenter quelques reflexions :elles me sont suggérées par des questions reliant entre eux les trois cours que nous suivons, d'autant que les sujets et les tons en sont très différents. Mais je considère comme une circonstance particulièrement heureuse et stimulante que ces cours aient des inspirations si éloignées : fait qui découlait déjà de la differente personnalité des trois orateurs et qui s'est bien révélé dès la première journée par le choix du sujet et l'orientation de chacun d'entre eux. Et nous voyons se développer, jour après jour, trois différents aspects de l'étude mathématique de l'économie. De l'un des côtés extrèmes, M. Guilbaud nous montre, avec son esprit de finesse et sa profondeur, les tâches d'un criticisme mathématique qui se préoccupe de séparer soigneusement ce qui revient à l'économie ou à la mathématique, de séparer les axiomes, les notions et les théories des interprétations empiriques de la realite économique. De l'autre côté, M. Frisch, prend comme point de départ un criticisme sociologique, portant sur la distinction de buts et d'intérêts entre l'économie descriptive et l'économie normative, entre modèles de développement et modèles de décision, entre limitations thechnologiques non directement modifiables et limitations administratives qui en principe seraient modifiables, et entre les différentes façons de prendre des décisions, notamment entre les procédés démocratiques et les autres. Pour lui, la mathematique est l'outil qui permet d'indiquer à l'humanité la voie pour surmonter les difficultés économiques provenant de l'insuffisante connaissance des faits et des possibilités et de l'absece de rationalité et de coordination entre les décisions. Enfin, M. Morishima nous montre les types plus poussés de formalisation des études d'économie mathématique par des modèles fort complexes qui témoignent des efforts modernes (auxquels il a apporté des contributions personnelles universellement appréciées) pour maitriser un ensemble toujours plus riche d'aspects du problème de l'economie.

On voit bien par cette comparaison combien le champ d'application de la mathématique à l'économie est vaste : les études particulières apparaissent - pourrait-on dire- comme des îles separées par des lacunes qu'il faudrait bien combler, ou dont il faudrati au moins se rendre compte afin que l'attention portée aux questions particulières ne fassent pas perdre de vue l'ensemble.

Ces mots ne signifient nullement une critique à ce qui a été fait dans ce domaine ; ils veulent souligner simplement qu'il y a beaucoup à faire encore, qu'il faut continuer. Continuer, reflêchir, reviser sans cesse, c'est d'ailleurs l'impératif éternel de toute science ; pour ce qui est de l'économie, cette importance d'une réélaboration continuelle a été dûment soulignée comme un des mérites de Pareto dans un article de journal que par hasard j'ai lu tout juste hier : "egli non cessò mai di meditare sulle conclusioni già raggiunte, di metterle in dubbio e di cercar di superarle" (Leo Valiani, "L'Espresso",5/9/65).

C'est dans cet esprit que je voudrais développer quelques commentaires qui pourraient etre décrits, sans doute trop superficiellement, comme criticisme du criticisme mathématique de M. Guilbaud et criticisme du criticime sociologique de M. Frisch ; quelques considérations concernant le cours de M. Morishima pourront entrer aussi dans ce cadre.

Je suis tout à fait d'accord avec M. Frisch sur la façon, telle que je l'ai rappelée, de poser les questions ; si j'attache beaucoup d'intérêts et d'importance à l'étude de l'économie et à l'application de la mathématique à l'économie, ce n'est pas seulement pour accroître le patrimoine de travaux scientifiques, mais surtout parce qu'il me semble que la vie économique est fort mal organisée , et les possibilités créées par le progrès très imparfaitement exploitées pour le bienêtre des populations, dans les différents régimes et systèmes économiques existants. Il semble bien sûr qu'il suffirait de modifier ces limitations que M. Frisch dit "administratives" pour améliorer la situation dans le monde, et une analyse mathématique est une prémisse nécessaire à ces améliorations.

Mais l'aspect mathématique est quand même étroitement lié à l'aspect social et à l'aspect moral (s'il est permis de prononcer un mot qui semblait devoir être écarté) . A qui reviendra le droit et la tâche de prendre les decisions, de pondérer les preférences des différents individus, des générations actuelles et futures ? La réponse de M. Frisch, en faveur de la méthode démocratique, est sans doute la plus séduisante, mais ce mot "démocratique", ne semble pas épuiser la question.

La question qui demeure ouverte est celle de la répartition des champs et des pouvoirs de décision ; la répartition en champs en est l'aspect technologique (et il faut bien souhaiter que la voie tracée par M. Frisch, avec la notion de "canal", ouvre la possibilité d'individualiser un critère de division rationnalle qui évitera les défauts des divisions par industries, par régions, par entreprises, par individus) ; la répartition des pouvoirs en est l'aspect sociologique, et il faut bien s'assurer qu'elle conduit à un résultat "équitable" (c'est un mot mal définissable, mais on ne peut pas s'en passer et il n'y 'a rien de mieux définissable qui semble pouvoir le remplacer.)

La Théorie classique disait qu'il suffirait que chaque individu soit libre de choisir ses actions d'après son intérêt personnel, et l'on voit que cela est dangereux. Mais si les choix de la collectivité sont décidés par les individus qui en font partie, par exemple par une règle de majorité, cela pourrait être aussi dangereux, si le vote de chacun était déterminé par ce même intérêt : la décision collective pourrait être incohérente, ou myope, ou préjudiciable aux intérêts d'une minorité. Il faudrait bien - mais de quelle façon ? - garantir que la décision représente un compromis dans la signification la meilleure du mot, comme M. Frisch le préconise : la situation la meilleure (la seule réellement bonne) serait celle où chaque membre de la collectivité aurait la même connaissance des exigences de tous les autres et le même soin d'en tenir compte en vue d'une décision à prendre de la façon la meilleure,en moyenne,pour tous,comme dans une famille (peut-être dans une famille idéale qui n'existe pas) . Pourrait-on donner aux peuples une éducation civique de cette sorte? et sinon, devrait-on limiter la liberté de choix (par exemple, si la population sous-estimait les necessités du futur, ou des dangers survenant à long terme etc...) et qui devrait en décider ?

La connexion que je viens de souligner entre l'aspect mathématique de l'individualisation de partitions de canaux ou autres choses suffisamment peu interdépendantes au point de vue technologique afin d'être acceptables comme champs de décisions séparés, et l'aspect social et moral concernant les formes d'intervention des volontés individuelles, est bien un exemple de connexion qui existe toujours entre tous les aspects de tout problème et de toute theorie de l'economie.

Cela ne veut pas dire que les distinctions ne sont pas admissibles, ni qu'elles sont inutiles ; il faut seulement se garder d'aller trop loin. Les distinctions sont utiles comme artifice pour fixer l'attention sur un aspect indépendamment des autres momentanément négligés, mais non pour l'ériger en quelque chose d'autonome et d'autosuffisant. Cette exigence, unique avec ses deux faces opposees, de distinguer et de connecter les différents aspects, est présente à chaque pas.

De ce fait, je suis parfaitement d'accord aussi avec M. Guilbaud lorsqu'il souligne l'intérêt qu'il y a à se demander ce qui ressort aux mathématiques, quels sont les axiomes ou hypothèses dont dépendent logiquement des conclusions, etc.... Tout cela est fort utile, pas seulement pour le mathématicien qui a et doit avoir le souci de préciser les passages dont il est responsable en les distinguant des prémisses reçues de l'exterieur comme point de départ et des interprétations qui seront données à l'extérieur comme point d'arrivée, mais aussi pour l'économiste qui doit en juger et en faire application avec sagesse.

Et je suis d'autant plus d'accord avec lui qu'il souligne - comme il vient de le faire ce matin(0) - la liaison que ces aspects mathématiques ont avec les questions sociales et morales dont il nous a promis de parler, en faisant allusion à ces grands esprits : Condorcet , Rousseau, Cournot - avec lesquels on sent bien qu'il est familiarisé. Supprimer ces liaisons ce serait déssécher nos études.

(0) Ces allusions ne se trouvent pas dans le texte réproduit ici .

Mais, malheureusement, l'effet de l'emploi de la mathématique n'est pas toujours si bien interprété : la mathématique jouit en effet d'une réputation d'exactitude et d'infaillibilité si extraordinaire que tout ce qu'elle manipule semble souvent être considéré de ce fait transformé en or de vérité absolue, comme s'il avait été touché par le roi Midas. C'est de ce fait par exemple, que la notion de prix semble se présenter et s'imposer avec une interpretation intrinsèque et générale , donnant une signification à la "valeur" de quantités quelconques de marchandises, tandis que cela découle, au point de vue mathématique, de la simplification de considérer une fonction linéaire et de l'habitude à des faits " administratifs". En réalité la définition n'a de sens que dans le voisinage d'un point, et , en dehors de ce voisinage, il faudrait considérer seulement les variétés d'indifférence.

On peut illustrer cette remarque, qui s'applique à toute théorie plus perfectionnée, même en se bornant au schéma très primitif d'Isnard que M. Guilbaud nous a présenté. Le passage de la situation initiale à la finale pourrait bein se produire de maintes façons, soit par des échanges successifis à des prix différents, ou par des transferts gratuits. Chercher un prix unique sous-jacent pourrait ne pas dire grand-chose , et il pourrait même se faire que les prix d'équilibre dans la situation finale soient tels qu'ils empêchent la transaction, soit au point de vue d'un producteur, soit à celui des consommateurs. Si par exemple l'un des trois individus avait eu la chance de faire d'une récolte très bonne de son produit, il ne pourrait que le céder grauitement aux deux autres sans pouvoir se procurer les biens qu'ils produisent. Et, de l'autre côté, il n'y a pas moyen d'assurer les quantités indispensables des biens essentiels à la population sans recours, parfois, à un rationnement au lieu d'un fonctionnement du système de prix. Sur une échelle beaucoup plus large, il semble bien que c'est de faits de ce genre que dépendent les difficultés de certaines catégories (comme les agriculteurs) ou des pays sous-développés qui se trouvent en situation de faiblesse dans leurs rapports avec les catégories ou pays plus industrialisés (comme il est décrit par exemple par Hla Myint[1] et Galbraith[2], qui montrent la nécessité d'une "counter-balancing power"), mais dont je ne connais pas d'analyses mathématiques. Dans ce même ordre d'idées (ou de négligences) on néglige avec facilité la question des "économies extérieures" : des coûts que les conventions "administratives" ou traditionnelles ne font pas payer à celui qui en est la cause et qui en tire un profit.

C'est bien de situations de ce genre, qui donnent lieu à des problèmes, qu'il faudrait s'occuper surtout ; mais il semble cependant que c'est le contraire qui a plus d'attrait, notamment le souci de prouver rigoureusement que tout est "lapallissade" si l'on néglige ce qui ne l'est pas.

On considère, par exemple, des modèles, pour lesquels on démontre l'existence d'un équilibre, ou la validite de telle ou telle autre propriété. Naturellement, la définition même du modèle, ou des

---

[1] Hla Myint - An interpretation of economiç bacrkwardness, Oxford economic papers, Vol 6, n. 2, 1954, pp. 132-163; trad. ital. dans F. Caffé, Economisti moderni, Garzanti ed., 1962.

[2] John Galbraith - The affluent society, Hamish - Hamilton, London, 1958.

restrictions ou axiomes ultérieurement introduits, montrent bien que les conclusions sont valables sous des hypothèses particulières, en général très loin d'être réalistes, et souvent choisies expréssément pour arriver à cette conclusions. Par le dit effet de la réputation des mathématiques, il semble bien souvent, dans cette situations, que l'on pense que les conclusions sont mathématiquement vraies et donc telles même dans la vie pratique, sans se soucier des hypothèses de départ; tandis qu'il fallait examiner ce qui se passe lorsque ces restrictions ne sont pas valables.

On est ainsi conduit maintes fois à se souvenir du P. Saccheri, qui avait presque découvert les géométries non-euclidiennes, mais qui a fait tout effort possible pour s'empêcher d'y parvenir, en écartant comme déraisonnables les conséquences des hypothèses qui auraient caractérisé la géométrie de Bolyai et celle de Lobatchevski. Dans notre cas de l'économie, la chose est bien plus dangereuse, car elle revient à suggérer des visions trop limitées, simplifiées et optimistes sur l'ensemble des problèmes : ces insuffisances conduisent généralement à conclure avec M. Pangloss que tout va au mieux dans le meilleur des mondes possibles, ou, peut-être, que cela serait vrai si seulement on admettait le "laisser faire, laisser passer" comme dirait un M. Pangloss de droite, ou bien si seulement on acceptait de planifier comme dirait un M. Pangloss de gauche (qui pourrait penser - comme dans le cas mentionné par M. Frisch[1] - qu'il s'agit tout simplement de se donner le vecteur des consommations désirées

(1) Dans le memorandum sur la "Méthode des canaux" Oslo.

et de faire l'inversion d'une matrice). ( )

Même les exposés mathématiques semblent parfois vouloir suggérer l'idée que la mathématique a le pouvoir de rendre facile ce qui est complexe ; et cela est vrai dans les limites du raisonnable, mais si on les dépasse cela veut dire que l'on n'a pas affronté le problème: on l'a seulement eludé. Il ne faut pas se borner aux choses claires et commodes pour montrer qu'on peut les mathématiser de façon très claire et trèse commode, et prouver qu'il suffit de négliger ou d'écarter ce qui ne plait pas pour accepter ou justifier ce qui semble raisonnable; il faut pousser l'analyse jusqu'à considérer toute situation incommode, en voir les raisons, en chercher les principes pour corriger ce qu'il faudrait corriger.

Malheureusement, il semble bien que la tendance naturelle de toute chose c'est de se dérouler de la facon la plus mauvaise possible si on ne s'efforce pas de l'en empêcher : je crois que l'expérience vécue par nous tous à notre sièle confirme tout-à-fait celle de Candide au siècle de Voltaire. Et pour empêcher cette tendance du mieux que nous le pouvons, il ne suffit pas de choisir une formule magique ou autre, un shéma idéologique ou autre, mais d'extraire de tout point de vue, d'examiner et d'essayer ce qui peut être approprié dans les différents problèmes et situations. On ne peut pas penser, pour faire une image, qu'il suffit de confier tout à des dispositifs basés sur l'hypothèse que le frottement n'existe pas, ni à des

( ) Les considérations à ce sujet ont été développées davantage par l'auteur dans un article écrit à la suite du 1er Congrès de l'Econometric Society (Rome, sept. 1965) : B. de Finetti, " Econometristi allo spettroscopio", La rivista trimestrale, Roma, 1965.

méchanismes "parfaits" en espérant qu'aucune défaillance (qui pourrait causer des désastres) ne se produira jamais.

Lorsqu'on passe au cas des modèles plus complexes, il s'avère que, lorsqu'un modèle est à la mode, les hypothèses et les objectifs qu'il envisage apparaissent hors de discussion : par exemple, que le but de l'expansion la plus rapide possible de la production soit l'objectif naturel et que les hypothèses implicites sont des assumptions mathématiques dont il n'y a rien à dire. C'est avec plaisir que j'ai oui caractériser si bien, par M. Morishima, les hypothèses du modèle de von Neumann : un modèle où les travailleurs sont considérés comme des animaux élevés pour produire du travail et les capitalistes sont des automates au service de la production. Et c'est curieux : je viens d'apprendre ces jours-ci qu'une réalisation conforme à cette image a été exploitée, tout près d'ici, sur l'autre versant du Gran Sasso : il y a un élevage de singes dressés pour la récolte des olives, dans des endroits où la main d'oeuvre humaine couterait trop chère.

Mais quand même , le modèle de von Neumann est bien interessant, et c'est intéressant de remarquer, comme l'a fait M. Koopmans[1] , comment un modèle si peu réaliste dans son point de départ peut s'avérer puissant comme outil d'étude sous des hypothèses raisonnables ; j'attends avec intérêt de voir comment M. Morishima va rattacher ces questions au "Turnpike theorem" qu'il a annoncé nous exposer demain. Car c'est de cette façon qu'il est instructif de voir les choses, et non du point de vue formaliste qui néglige la considération

(1) Tjalling C. Koopmans, "Economic growth at a maximal rate", Qu. J. of Econ. LXXVIII (1964) 355-394.

des buts effectifs du bien-être de la collectivité.

Toutes les théoriès, les méthodes, les modèles peuvent être utiles, mais leur utilité n'est bien assurée et bien fondée sinon lorsqu'elles arrivent à faire partie d'une vision d'ensemble tendant à embrasser l'économie toute entière ; vision qui ne semble pouvoir se baser sur d'autre fondement que celle que l'on appelle "l'économie du bien-être" , ou , selon la dénomination la plus universellement connue, la "Welfare Economics" .

B. de Finetti

3 Sett. 1965

CENTRO INTERNAZIONALE MATEMATICO ESTIVO

(C.I.M.E.)

G. Th. GUILBAUD

QUELQUES REFLEXIONS MATHEMATIQUES
sur
LES EQUILIBRES ECONOMIQUES

Corso tenuto all'Aquila dal 30 agosto al 7 settembre 1965

QUELQUES REFLEXIONS MATHEMATIQUES
sur
LES EQUILIBRES ECONOMIQUES
par
G.. Th. Guilbaud

1 LE MODÈLE D'ISNARD

Selon Schumpeter (History of Economic Analysis) le premier auteur qui a tenté une définition mathématique de l'EQUILIBRE est ISNARD sans son Traité des Richesses (1781) "qui attend encore dans l'histoire de la science économique la situation qui lui est due" (celle de précurseur de Walras) .

Il s'agit de montrer comment les valeurs se constituent à partir des quantités disponibles et des besoins. Dans l'échange entre deux agents économiques, c'est clair : si l'on échange tant de litres de vin contre tant de mesures de blé , les termes mêmes de cet échange définissent le rapport en question. Ce qui importe c'est la généralisation, pour laquelle l'expression mathématique est nécessaire.

Commençons par un exemple numérique, à la manière ancienne. Trois personnes possèdent respectivement 20 setiers, 25 barils, 30 quintaux.

Tableau initial :

(Ti)

	I	II	III
S	20	--	--
B	--	25	--
Q	--	--	30

Imaginons qu'un échange préalable d'informations et un débat conduise les trois personnes à souhaiter (tels sont les "besoins") la réalisation du tableau final :

(Tf) :

	I	II	III
S	9	6	5
B	7	15	3
Q	2	4	24

Alors sont déterminés des prix, non pas comme des nombres absolus, mais des rapports au sens arithmétique du mot.

Soient en effet s, b, q, les valeurs unitaires pour les trois marchandises; il suffit d'écrire la balance des comptes :

$$20\ s = 9\ s + 7\ b + 2\ q$$
$$25\ b = 6\ s + 15\ b + 4\ q$$
$$30\ q = 5\ s + 3\ b + 24\ q$$

Système de trois équations qu'on "résoudra" même sans grandes connaissances d'algèbre et l'on trouve :

$$\frac{s}{12} = \frac{b}{14} = \frac{q}{17}$$

ce qui donne les proportions désirées entre les valeurs .

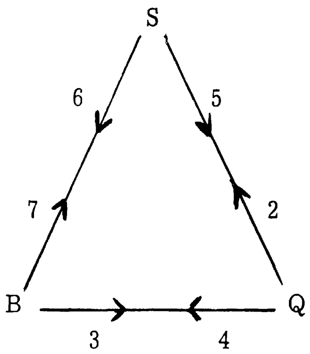
Deux données : (Ti) et (Tf) , c'est à dire la répartition des marchandises avant et après l'échange. C'est d'ailleurs seulement ce qu'on peut appeler leur différence (Tf - Ti) qui intervient :

-11	+ 7	+ 2
+ 6	- 10	+ 4
+ 5	+ 3	- 6

(dT = Tf - Ti)

et qui permet d'écrire l'équilibre des valeurs :

$$\begin{cases} 11\,s = 7\,b + 2\,q \\ 10\,b = 6\,s + 4\,q \\ 6\,q = 5\,s + 3\,b \end{cases}$$



Le réseau des échanges

$$\begin{cases} 6\,s + 5\,s = 7\,b + 2\,q \\ 7\,b + 3\,b = 6\,s + 4\,q \\ 4\,q + 2\,q = 5\,s + 3\,b \end{cases}$$

Comptabilité en valeur

Le fait mathématique à mettre en évidence : cette procédure réussit ! Réussit-elle toujours ?

D'abord le dire abstraitement : un tableau carré de nombres positifs (entiers dans notre exemple, en tous cas rationnels) , complété dans la diagonale par des nombres négatifs calculés par l'addition en colonne. Des équations s'en déduisent . Donnent-elles toujours une solution acceptable ?

Existence de multiplicaterus (positifs) qu'on puisse interpréter comme des prix .

Quelqu'un qui ne sait pas la démonstration comprend-il la thèse économique ? et qui sait des mathématiques comprend-il mieux les phénomènes économiques ?

<u>Description</u> mathématique du problème posé ;

On donne deux tableaux carrés, indiquant l'un la répartition initiale, l'autre la finale . Par exemple :

$$\begin{bmatrix} 11 & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 6 \end{bmatrix} \longrightarrow \begin{bmatrix} 0 & 7 & 2 \\ 6 & 0 & 4 \\ 5 & 3 & 0 \end{bmatrix}$$

et l'on demande de trouver des multiplicateurs (qui seront interprétés comme des <u>prix</u>) :

$$a, \; b, \; c, \; \ldots$$

tels que :

$$\begin{aligned} 11\,a &= 7\,b + 2\,c \\ 10\,b &= 6\,a + 4\,c \\ 6\,c &= 5\,a + 3\,b \end{aligned}$$

qu'on écrira plus commodément sous forme matricielle :

$$\begin{pmatrix} 11 & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 6 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 0 & 7 & 2 \\ 6 & 0 & 4 \\ 5 & 3 & 0 \end{pmatrix} \cdot \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

On demande, non pas seulement de résoudre, c'est à dire calculer effectivement, mais d'étudier en général la possibilité d'un tel calcul, en accordant une attention toute particulière au fait que, dans notre interprétation économique, les multiplicateurs seront des prix et doivent être positifs .

Le langage vectoriel est le plus commode ; il s'agit de trouver des multiplicateurs (scalaires) a, b, c, ... tels que :

$$a \begin{pmatrix} -11 \\ +6 \\ +5 \end{pmatrix} + b \begin{pmatrix} +7 \\ -10 \\ +3 \end{pmatrix} + c \begin{pmatrix} +2 \\ +4 \\ -6 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Le problème général est donc le suivant : étant donné un ensemble de vecteurs, existe-t-il une combinaison linéaire à coefficients positifs qui soit égale au vecteur zéro ?

2 ESPACES VECTORIELS ET CONES

Les colonnes de nombres (rationnels) que manipule notre comptabilité sont des vecteurs, c'est à dire des objets qui obéissent aux axiomes bien connus des espaces vectoriels (il y a un vecteur zéro, on sait additionner deux vecteurs, cette addition constitue l'ensemble en groupe abélien, il y a des homothéties, etc...) .

Etant donnée une collection quelconque (ici elle sera toujours finie) de vecteurs d'un tel espace, l'ensemble de toutes les combinaisons linéaires de ces vecteurs constitue un sous-espace vectoriel (ou varieté linéaire).Si parmi les combinaisons linéaires on sélectionne celles dont aucun coefficient n'est négatif on obtient une partie du susdit sous-espace, partie que l'on nomme cône.

Ainsi à partir d'une famille de vecteurs V_1, V_2, ..., V_n on constitue deux ensembles :

$L(V_1, V_2, \dots, V_n)$ ou $L(V_i)$ = espace engendré linéairement par les V_i

$K(V_1, V_2, \dots, V_n)$ ou $K(V_i)$ = cône engendré positivement par les V_i

On a évidemment : $K \subset L$, tout cône est une partie de l'espace corrispondant .

Mais il peut arriver qu'un cône soit non pas une vraie partie d'espace , mais un espace tout entier. Il est facile de construire des exemples. Ainsi les vecteurs :

$$\begin{pmatrix}1\\0\\0\end{pmatrix} \quad \begin{pmatrix}0\\1\\0\end{pmatrix} \quad \begin{pmatrix}0\\0\\1\end{pmatrix}$$

engendrent linéairement un espace . Adjoignons à ces trois vecteurs, un quatrième ;

$$\begin{pmatrix}-1\\-1\\-1\end{pmatrix}$$

Alors tout vecteur de l'espace peut être écrit comme combinaison linéaire positive de ces quatre vecteurs. Si les trois coordonnées sont positives, les trois premiers suffisent. Sinon, comme par exemple:

$$\begin{pmatrix}2\\-3\\4\end{pmatrix} = 2\begin{pmatrix}1\\0\\0\end{pmatrix} - 3\begin{pmatrix}0\\1\\0\end{pmatrix} + 4\begin{pmatrix}0\\0\\1\end{pmatrix}$$

on effectue le remplacement :

$$\begin{pmatrix}0\\1\\0\end{pmatrix} = -\begin{pmatrix}-1\\-1\\-1\end{pmatrix} - \begin{pmatrix}1\\0\\0\end{pmatrix} - \begin{pmatrix}0\\0\\1\end{pmatrix}$$

ce qui donne :

$$\begin{pmatrix} 2 \\ -3 \\ 4 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + 3 \begin{pmatrix} -1 \\ -1 \\ -1 \end{pmatrix} + 3 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + 3 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} + 4 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

$$= 5 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + 7 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} + 3 \begin{pmatrix} -1 \\ -1 \\ -1 \end{pmatrix}$$

dont tous les coeffiecients sont positifs .

Etablissons, en général , la proposition suivante :

Tout espace est un Cône

Soit L l'espace engendré linéairement par les vecteurs : $X_1, X_2, \ldots, X_n$. Adjoignons le vecteur X_o défini par :

$$X_o + X_1 + X_2 + \ldots + X_n = \text{Zéro}$$

Ce vecteur appartient évidemment à l'espace L .

Et considérons le cône K engendré par $X_o, X_1, \ldots, X_n$.

Il est clair que : $K \subset L$. Mais on a aussi : $L \subset K$;Soit en effet X un vecteur quelconque de L.

$X \in L$:signifie qu'il existe des coefficients P_i tels que :

$$X = \sum P_i \quad X_i$$

Si tous les P_i sont positifs, on a $X \in K$. Sinon considérons le plus petit coefficient : $- P_o$ (il est négatif) et écrivons

$$X = X + P_o (X_o + X_1 + X_2 + \ldots + X_n)$$

ou bien :

$$X = P_o X_o + (P_1 + P_o) X_1 + \ldots + (P_n + P_o) X_n$$

qui présente X comme combinaison linéaire des vecteurs : $X_o, X_1, \ldots, X_n$. Mais puisque :

$$- P_o \leq P_i \qquad \text{pour tous les } i \in \{1, \ldots, n\}$$

tous les coefficients $P_i + P_o$ sont positifs.
Donc $X \in K$.

On verra sans peine qu'on aurait pu remplacer la somme X_o par une combinaison linéaire quelconque à coefficients négatifs (et non forcément égaux à - 1) pourvu qu'aucun ne soit nul. Inversement, si un espace est positivement engendré par une collection de vecteurs, tout vecteur X de cet espace est égal à une combinaison positive (à coefficients non tous nuls) et il en est de même du vecteur - X, donc également de leur somme X - X = Zéro. Ainsi le vecteur Zéro peut être représenté comme combinaison linéaire positive (à coefficients non tous nuls). Ainsi:

La condition nécéssaire et suffisante pour que le cône K engendré positivement par une collection de vecteurs soit identique à l'espace engendré linéairement est qu'il existe une combinaison linéaire positive qui soit identiquement nulle (aucun coefficient ne ne doit être nul).

Nous retrouvons ici notre problème : on donne une collection de vecteurs, existe-t-il une combinaison linéaire positive qui soit identiquement nulle ? Et rappelons la signification "comptable" : étant donné un schéma d'échanges de biens, existe-t-il des prix assurant la balance des comptes (un équilibre en "valeur") ?

Il sera commode d'appeler "Cône aplati" un cône identique à l'espace engendré.

Notons au passage qu'il existe des cônes, qui sans être aplatis, contiennent des sous-espaces linéaires : on les nomme des "Coins" (anglais : wedges) . Un demi-plan, un angle plan, un angle dièdre en fournissent des illustrations élémentaires. Pour un coin il existe aussi une combinaison positive identiquement nulle, mais certains coefficients sont nuls. Enfin notons que parmi les coins figurent, comme cas extrêmes, les demi-espaces.

(Un cône qui n'est pas un coin peut être dit "saillant") .

Il peut être intéressant de présenter un algorithme permettant de décider si un système de vecteurs engendre positivement un vrai cône, ou bien s'il engendre tout l'espace .

Dans le premier cas, les théorèmes de séparation bien connus affirment l'existence d'un opérateur linéaire f tel que :

$$f(x_i) > 0$$

pour chacun des vecteurs x_i donnés. ...

On va donc, étant donné des vecteurs x_1, x_2, ... x_n, chercher par tâtonnements une forme f pour laquelle tous les résultats soient positifs.

Du même coup nous résoudrons le problème suivant : classification de vecteurs en deux classes .

Des vecteurs x, une forme f ; les fx sont des scalaires ; les vecteurs seront classés selon que fx est positif ou négatif .

Si l'on donne deux classes, les vecteurs x et les vecteurs y - il s'agit de trouver une forme f telle que les fx soient positifs et les fy négatifs .

Il revient au même (à cause de la linéarité) de chercher une f telle que fz soit toujours positif en rassemblant dans une même classe z : les (x) et les (-y) . .

Le problème principal est donc :

Etant donné un ensemble de vecteurs, trouver une forme qui, appliquée à ces vecteurs donne des scalaires tous positifs.

Schéma d'un algorithme : on essaie un f, s'il ne convient pas, c'est-à-dire s'il donne un résultat scalaire négatif pour quelqu'un des vecteurs soumis à l'épreuve, on doit modifier f . Cette modification sera, bien entendu, fonction du vecteur rebelle : il faut donc établir une correspondance entre vecteurs x et corrections Df , c'est-à-dire entre deux espaces vectoriels en dualité. Le plus simple sera de choisir une correspondance linéaire : mais faire correspondre linéairement à tout vecteur x une forme Df (c'est-à-dire une opération qui transforme tout vecteur v en un scalaire s), c'est finalement faire correspondre à une paire de vecteurs (x,v) un scalaire s. Et cette correspondance doit être bilinéaire, c'est-à-dire linéaire pour x comme pour v.

On va donc travailler dans un espace vectoriel muni d'une opération bilinéaire qui à toute paire de vecteurs fait correspondre un scalaire. Lorsque cette opération est symétrique et positive, on peut l'appeler "produit scalaire" - et elle donne à l'espace vectoriel une structure géométrique ("euclidienne" ou "préhilbertienne") ; on y trouve des longueurs (ou normes) par les carrés scalaires et des cosinus (ou coefficients de corrélation) grâce à la célèbre inégalité (de Cauchy, Schwarz et Buniakovski) :

$$(x.y)\,(x.y) \leq (x.x)\,(y.y)$$

Décrivons maintenant, sur un exemple très rudimentaire (choisi pour pouvoir effectuer tous les calculs à la main en peu de temps), un algorithme particulièrement simple .

3 ALGORITHME

Vecteurs donnés

$$\begin{aligned} x &= (0,\ 1,\ 0,\ 1) \\ y &= (1,\ \bar{1},\ \bar{1},\ 1) \\ z &= (\bar{1},\ 0,\ 0,\ \bar{1}) \\ w &= (0,\ \bar{1},\ \bar{1},\ 1) \end{aligned}$$

Critère : (si Tv est positif, conserver l'operateur T, sinon remplacer T par T + v)

	Essai	Réponse	Opérateur	Vecteurs acceptés
A = (0, 0, 0, 0)	Ax = 0	non	A + x = B	-
B = (0, 1, 0, 1)	By = 0	non	B + y = C	-
C = (1, 0, $\bar{1}$, 2)	Cz = $\bar{3}$	non	C + z = D	-
D = (0, 0, $\bar{1}$, 1)	Dw = 2	oui	D	w
	Dx = 1	oui	D	w, x
	Dy = 2	oui	D	w, x, y
	Dz = $\bar{1}$	non	D + z = E	-
E = ($\bar{1}$, 0, $\bar{1}$, 0)	Ew = 1	oui		w
	Ex = 0	non	E + x = F	-
F = ($\bar{1}$, 1, $\bar{1}$, 1)	Fy = 0	non	F + y = G	-
G = (0, 0, $\bar{2}$, 2)	Gz = $\bar{2}$	non	G + z = H	-
H = ($\bar{1}$, 0, $\bar{2}$, 1)	Hw = 3	oui		w
	Hx = 1	oui		w, x
	Hy = 2	oui		w, x, y
	Hz = 0	non	H + z = J	-
J = ($\bar{2}$, 0, $\bar{2}$, 0)	Jw = 2	oui		w
	Jx = 0	non	J + x = K	-
K = ($\bar{2}$, 1, $\bar{2}$, 1)	Ky = 0	non	K + y = L	-
L = ($\bar{1}$, 0, $\bar{3}$, 2)	Lz = $\bar{1}$	non	L + z = M	-
M = ($\bar{2}$, 0, $\bar{3}$, 1)	Mw =4	oui		w
	Mx = 1	oui		w, x
	My = 2	oui		w, x, y
	Mx = 1	oui		w, x, y, z
				(arrêt)

En recapitulant (et en ne conservant que les étapes de modification) :

A = 0		
B = A + x	(x) . y	= 0
C = B + y	(x + y) . z	= 3
D = C + z	(x + y + z) . z	= $\bar{1}$
E = D + z	(x + y + z + z) . x	= 0
F = E + x	(x + y + z + z + x) . y	= 0
G = F + y	(x + y + z + z + x + y) . z	= 2
H = G + z	(x + y + z + z + x + y + z) . z	= 0
J = H + z	(x + y + z + z + x + y + z + z) . x	= 0
K = J + x	(x + y + z + z + x + y + z + z + x) . y	= 0
L = K + y	(x + y + z + z + x + y + z + z + x + y) . z	= $\bar{1}$
M = L + z	(x + y + z + z + x + y + z + z + x + y + z):	

On a donc sélectionné une suite de vecteurs qu'on peut noter :

a

a + b

a + b + c

....

les vecteurs a, b, c, ... étant choisis (avec répétitions) parmi les vecteurs donnés x, y, z, w (en nombre fini) .

Mais l'on a :

$$2(a.b) = (a+b)^2 - a^2 - b^2$$

$$2(a+b.c) = (a+b+c)^2 - (a+b)^2 - c^2$$

$$2(a+b+c.d) = (a+b+c+d)^2 - (a+b+c)^2 - d^2$$

etc...

Si tous les premiers membres sont négatifs (comme c'est le cas) , il en résulte , par additon , que :

$$(a+b+c\ldots)^2 - (a^2+b^2+c^2 + \ldots)$$

est négatif.

Donc la suite sélectionnée est telle que :

carré de la somme $\leq$ somme des carrés.

On va voir que cela implique que la somme en question ne peut contenir une infinité de termes.

Soit S le vecteur somme en question et F un opérateur linéaire, (ou covecteur) . D'après l'inégalité de Cauchy et consorts :

$$(F . S)\ (F . S) \leq (F . F)\ (S . S)$$

Or :

$$F.S = F.a + F.b + F.c + \&c$$

et $$F.S \quad a.a + b.b + c.c + \&c$$

Donc : $$(Fa + Fb + \ldots)^2 \leq (F.F)\ (a.a + b.b + \ldots)$$

S'il existe un F tel que tous les produits Fa, Fb , ... soient positifs, il existe une borne :

$$Fa > \alpha > 0$$

$$Fb > \alpha > 0$$

etc...

donc $(Fa + F \ + \ldots)^2 > n^2 . \alpha^2$

si n est le nombre de terme de la somme.

D'autre part a.a, b.b, ... carrés scalaires sont positifs et puisque a, b, ... sont choisis dans un ensemble fini, ils sont bornés supérieurement

$$a.a < \beta \quad , \quad b.b < \beta \qquad \text{etc....}$$

Finalement : $$n^2 \alpha^2 < F^2 . n\beta$$

ce qui prouve que n est borné.

On est donc sûr que s'il existe un plan séparateur, (c'est à dire si nos vecteurs engendrent un cône qui n'est pas aplati) l'algorithme en trouvera un au bout d'un nombre d'opérations inférieur à

$$n_o = \text{entier de } \left(\frac{F^2\beta}{\alpha^2}\right)$$

Pour montrer ce qui se passe dans le cas contraire , on va pour terminer, traiter un exemple simple. Avec quatre vecteurs :

a = (-2, 0, 1)

b = (0, 2, -1)

c = (0, -2, -3)

d = (1, 0, +3)

O = (0, 0, 0)	a = (-2, 0, 1)	Oa = 0	O + a = P
P = (-2, 0, 1)	b = (0, 2, -1)	Pb = -1	P + b = Q
Q = (-2, 2, 0)	c = (0, -2, -3)	Qc = -4	Q + c = R
R = (-2, 0, -3)	d = (1, 0, 3)	Rd = -11	R + d = S
S = (-1, 0, 0)	a = (-2, 0, 1)	Sa = + 2	S = S
idem	b = (0, 2, -1)	Sb = 0	S + b = T
T = (-1, 2, -1)	c = (0, -2, -3)	Tc = -1	T + c = U
U = (-1, 0, -4)	d = (1, 0, 3)	Ud = -13	U + d = V
V = (0, 0, -1)	a = (-2, 0, 1)	Va = -1	V + a = W
W = (-2, 0, 0)	b = (0, 2, -1)	Wb = 0	W + b = X
X = (-2, 2, -1)	c = (0, -2, -3)	Xc = 0	X + c = Y
Y = (-2, 0, -4)	d = (1, 0, 3)	Yd = -14	Y + d = Z

Z = (-1, 0, -1)	a = (-2, 0, 1)	Za = 1	Z = Z
idem	b = (0, 2, -1)	Zb = 1	Z = Z
idem	c = (0, 2, -3)	Zc = 3	Z = Z
idem	d = (1, 0, 3)	Zd = -4	Z + d = A
A = (0, 0, 2)	a = (-2, 0, 1)	Aa = -2	A = A
idem	b = (0, 2, -1)	Ab = -2	A + b = B
B = (0, 2, 1)	c = (0, -2, -3)	Bc = -7	B + c = C
C = (0, 0, -2)	d = (1, 0, 3)	Cd = -6	C + d = D
D = (1, 0, 1)	a = (-2, 0, 1)	Da = -1	D + a = E
E = (-1, 0, 2)	b = (0, 2, -1)	Eb = -2	E + b = F
F = (-1, 2, 1)	c = (0, -2, -3)	Fc = -7	F + c = G
G = (-1, 0, -2)	d = (1, 0, 3)	Gd = -7	G + d = H
H = (0, 0, 1)	a = (-2, 0, 1)	Ha = 1	H = H
idem	b = (0, 2, -1)	Hb = -1	H + b = I
I = (0, 2, 0)	c = (0, -2, -3)	Ic = -4	I + c = J
J = (0, 0, -3)	d = (1, 0, 3)	Jd = -9	J + d = K
K = (1, 0, 0)	a = (-2, 0, 1)	Ka = -2	K + a = L
L = (-1, 0, 1)	b = (0, 2, -1)	Lb = -1	L + b = M
M = (-1, 2, 0)	c = (0, -2, -3)	Mc = -4	M + c = N
N = (-1, 0, -3)	d = (1, 0, 3)	Nd = -10	N + d = 0

Puisque chacun des opérateurs P, Q, R , . . . , N est une combination linéaire à coefficients entiers des vecteurs donnés , la dernière équation :

$$\vec{N} + d = \vec{0}$$

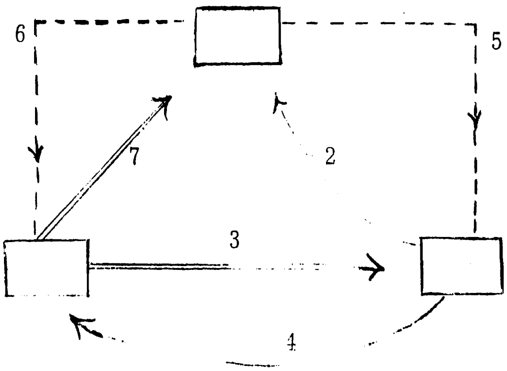
fournit une relation :

$$4a + 7b + 7c + 8d = 0$$

Le caractère cyclique est évidemment dû ici à l'emploi des entiers (et au fait que ces entiers soient bornés) .

4 TATONNEMENTS

Reprenons un schéma d'échanges :



Les quantités sont notées sur le diagramme; on n'oubliera pas qu'elles sont de nature variées (ce qui est symbolisé par les trois espèces de flèches) .

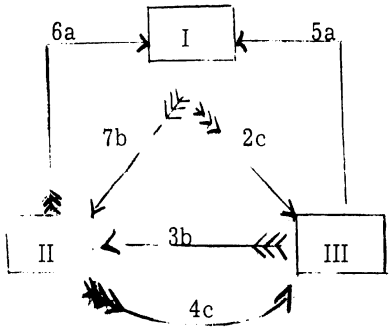
Il s'agit d'associer à un tel schèma de circulation des biens, un autre schéma de circulation "monétaire" (ou plutôt en "valeur") .
Ce nouveau schéma a même forme mais :

1°) Les flèches vont en sens contraires (le paiement est la contre-partie d'une livraison)

2°) elles sont toutes de même nature

3°) les valeurs numériques sont liées aux quantités par une loi de proportionnalité.

Voici donc, dans notre exemple, le schéma de la circulation des valeurs :



Une première étude a consisté à chercher quels sont les coefficients (a, b, c) - c'est à dire quel est le système des prix - qui assure la balance des comptes :

	Recettes		Dépenses
I)	6a + 5 a	=	7b + 2c
II)	7b + 3b	=	6a + 4c
III)	2c + 4c	=	5a + 3b

Mais on peut imaginer les "tatonnements" walrasiens : On commence par choisir des prix d'achat (a_o, b_o, c_o) qui déterminent les dépenses - et à chercher quels doivent être les prix de vente (a_1, b_1, c_1) qui assureraient les recettes adéquates :

$$\left|\begin{array}{l} 11a_1 = 7b_o + 2c_o \\ 10b_1 = 6a_o + 4c_o \\ 6c_1 = 5a_o + 3b_o \end{array}\right.$$

ce qui s'écrit aussi :

$$\begin{pmatrix} a_1 \\ b_1 \\ c_1 \end{pmatrix} = \begin{pmatrix} 0 & 7/11 & 2/11 \\ 6/10 & 0 & 4/10 \\ 5/6 & 3/6 & 0 \end{pmatrix} \begin{pmatrix} a_o \\ b_o \\ c_0 \end{pmatrix}$$

Le problème mathématique est donc le suivant : on effectue une transformation linéaire du vecteur des prix :

$$P_1 = T\,(P_o)$$

que se passe-t-il si l'on réitère la transformation, c'est à dire si l'on prend les prix de vente P_1 comme nouveaux prix d'achat ?

Par exemple $P_o = (1, 1, 1)$ donnera : $P_1 = (9/11, 10/10, 8/6)$ ou bien (puisqu'il suffit de connaitre les rapports de prix) $P_1 = (27, 33, 44)$ qui conduit à $P_2 = (145, 169, 195)$.

Si l'on part de $P_3 = (14, 17, 19)$, voisin de P_2, on trouve (approximativement) $P_4 = (14, 16, 20)$. Et l'on vérifie qu'on s'approche (mais asymptotiquement) de l'équilibre :

$$P_* = (12, 14, 17).$$

pour lequel on a

$$P_* = T(P_*)$$

(point fixe) .

5 MODÈLES MARKOVIENS

Un autre modèle de circulation monétaire conduit aux memes problèmes, sous une forme encore plus proche des premiers modèles de A.A. Markov.

Imaginons plusieurs agents économiques qui établissent leur plan de dépense selon le revenu de la période antérieure et des proportions (dites : propensions) fixes .

Soit pour fixer les idées, trois agents et les proportions suivantes:

	de A	de B	de C
vers A	0,40	0,10	0,05
vers B	0,15	0,30	0,70
vers C	0,45	0,60	0,25

Partons des revenus :

A	B	C
100	300	200

ce qui donne les dépenses :

40	30	10
15	90	140
45	180	50

Mais chaque dépense est une recette pour le partenaire . En additionnant les lignes on trouve pour revenus de la période suivante :

80 245 275

qui se transforme par un calcul analogue en :

70,25 278,00 251,75

et en recommençant :

68, 49	270, 16	261, 35

puis :

67, 48	274, 27	258, 25
67, 33	273, 18	259, 49
67, 22	273, 70	259, 08

etc...

En cherchant d'autre part une distribution qui serai invariante (ou "fixe") on trouve :

67, 20	273, 60	259, 20

Nous sommes encore en présence de même schéma : un vecteur V est transformé (linéairement) par l'opérateur T

$$V_1 = T\ V_0$$
$$V_2 = T\ V_1$$
$$V_3 = T\ V_2 \quad \text{etc...}$$

et l'on constate :

1°) il existe un vecteur fixe

$$V_* = T\ V_*$$

2°) V_n est d'autant plus voisin de V_* que n est grand

$$\lim\ (V_n) = V_*$$

Il s'agit alors de savoir quels sont les caractéres de l'opérateur T succeptibles d'assurer une telle convergence .

Markoff lui-même a commencé l'étude de type algébrique. On commencé par noter que l'on peut écrire :

$$V_n = T^n\ (V_o)$$

et l'on concentre l'attention sur la transformation composée T^n.
En écrivant $T^{n+1} = T^n T$, on vérifie que les lignes de la matrice (n+1) sont obtenues en calculant des moyennes sur les lignes de la matrice (n) .
On utilise alors le lemme des moyennes itérées pour montrer que chaque ligne de T^n tend vers l'uniformité.

Lemme: Soit un vecteur x de coordonnées x_i et une moyenne calculée avec les pondérations P_i . On écrira :

$$m(x) = \sum P_i \; x_i$$

Prenons une autre moyenne m' avec des coefficients P'_i .

$$m(x) - m'(x) = \sum (P_i - P'_i) \, x_i$$

mais $\sum P_i = \sum P'_i = 1$

donc les $(P_i - P'_i)$ ont une somme nulle ; séparons alors les termes positifs et les négatifs .

$$m(x) - m'(x) = \sum a^+_j \, x_j - \sum b^-_k \; x_k$$

avec : $\sum a^+_j = \sum b^-_k$

Désignons cette somme par e

$$e = \sum a^+ = \sum b^-$$

on a aussi bien : $e = \frac{1}{2} (\sum a^+ + \sum b^-)$

donc : $e = \frac{1}{2} \sum \left| P_i - P'_i \right|$

C'est un indicateur de l'écart entre les deux covecteurs P et P' - ou si l'on veut un indicateur de l'écart entre les deux opérations de moyennes m et m' .

On a donc :

$$m(x) - m'(x) = e \left[\frac{\sum a^+ x}{\sum a^+} - \frac{\sum b^- x}{\sum b^-} \right]$$

c'est à dire :

m - m' = e (une différence entre deux moyennes)

Mais si nous appelons d le diamètre de l'ensemble des x_i (c'est à dire la plus grande différence) , toute différence de moyennes est inférieure à d.

Finalement :

$$m - m' \leqslant e.d$$

Si maintenant, comme dans le cas markovien on prend n moyennes des coordonnées d'un même vecteur (ligne d'une matrice), et si l'on appelle d' le nouveau diamètre, on aura :

$$d' \leqslant d.E$$

E étant le plus grand écart.

Dans le cas des moyennes : $E \leqslant 1$

Si $E < 1$ on est sûr qu'en répétant l'opération d tendra vers zéro .C.q.f.d.

Dans l'exemple numérique cité plus haut, les pondérations sont :

P	P'	P"
0,40	0,10	0,05
0,15	0,30	0,70
0,45	0,60	0,25

l'écart entre P et P' se calcule en faisant d'abord la différence:

$$\begin{pmatrix} +30 \\ -15 \\ -15 \end{pmatrix} \qquad \text{d'où : } e_1 = 0,30$$

On calculera demême $e_2 = 0,40$, $e_3 = 0,55$

Le plus grand écart $E = 0,55$.

On peut donc affirmer que le diamètre décroît selon la loi :

$$d_{n+1} \leqslant (0,55)\ d_n$$

c'est à dire

$$d_n \leqslant (0,55)^n \ . \ d_o$$

ce qui est fort rapide .

6 ANALYSE SPECTRALE

La méthode précédente apparaît comme assez spéciale. Il convient d'élargir les points de vue et de rattacher l'étude de T^n à l'analyse spectrale.

L'ensemble des opérateurs à étudier :

$$(T^o = \text{identité},\ T^1 = T,\ T^2,\ T^3, \dots ,\ T^n\ ,\ \dots)$$

est un ensemble de vecteurs ; il est naturel d'envisager l'espace engendré, c'est à dire toutes les combinaisons :

$$a_o\ T^o + a_1\ T^1 + a_2\ T^2 + \dots + a_n\ T^n$$

c'est à dire encore les polynômes en T .

D'autre part, on élargira la recherche en étudiant non seulement les vecteurs invariants pour T, mais tous les espaces invariants. Soit un sous-espace à 1 dimension. Dire qu'il est invariant c'est dire qu'un vecteur x de cet espace est transformé en un vecteur de même espace c'est a dire homothétique :

$$Tx = \lambda x$$

oú encore $$(T - \lambda T^o)\, x = 0$$

Soit un sous -espace invariant de dimension deux. Un vecteur quelconque x de cet espace devient Tx = y, mais le transformé Ty étant dans le même espace on peut écrire :

$$Ty = \lambda y + \mu x$$

en éliminant y il vient :

$$T^2 x = \lambda Tx + \mu x$$

c'est à dire :

$$(T^2 - \lambda T - \mu T^o)\ x = 0$$

On généralise sans peine : à tout sous-espace de dimension d est associé un polynôme de degré d annulateur de tout vecteur de l'espace.

D'autre part la correspondance entre sous-espaces invariants et polynômes annulateurs se précise par : la relation d'inclusion entre sous-espaces se traduit par la relation de divisibilité entre polynômes.

Le polynôme correspondant à l'espace total est celui du théorème de Cayley-Hamilton, dit parfois caractéristique.

On est donc ramené à un problème d'algèbre concernant la décomposition d'un polynôme en produit . Dans le corps des complexes, la réduction est complète et chacun des facteurs

$$T - s\,T^{o}$$

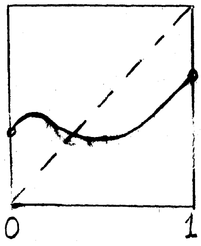
donne une valeur spectrale s.

Dans le cas markovien, une valeur spectrale est égale à l'unité (on a donc des vecteurs invariants) et les autres ont un module inférieur à l'unité. Bien entendu de telles analyses peuvent être étendues à des opérateurs de dimension infinie, moyennant les conditions de compacité traditionnelles.
(Voir : Bourbaki, Espaces vectoriels topologiques, appendice du chapitre 2 : Théorème de Markoff et Kakutani, fascicule XV, page 114) .

7 EQUILIBRES A LA COURNOT

En matière de théorie économique (en forme mathématique) le cadre linéaire n'est pas le mieux représenté au XIX siècle. Bien au contraire on a introduit très tôt (sans doute le premier fut Cournot) des fonctions "quelconques" (mais bien entendu continues dérivables, et représentables par une "courbe") . Déjà chez Cournot de nombreux équilibres sont déterminés "par l'intersection de deux courbes"; et plus tard on usera et abusera (après Marshall) de l'intersection de l'offre et de demande.

Pendant assez longtemps (jusqu'à l'époque de C. Jordan) on se contentait d'une "évidence" visuelle. Ainsi en joignant deux points pris chacun sur deux côtés opposés d'un carré, et ce par une courbe continue,il était "clair"que cette courbe devait rencontrer la diagonale. C'était dire que



l'équation $x=f(x)$ a au moins une solution quand $f(0)$ et $f(1)$ sont compris entre 0 et 1. De telles propositions sont "vraies", mais il faut les démontrer, pour comprendre pourquoi elles sont vraies.

Etant données une application continue d'un espace E en lui-même

$$E \xrightarrow{f} E$$

à quelles conditions peut-on assurer qu'il existe au moins un point invariant ou fixe :

$$x \in E, \quad x = f(x)$$

L'exemple précédemment donné fournit une image intuitive du cas où E est un segment réel : $(0 \leq x \leq 1)$.

C'est Brouwer qui a montré qu'on peut affirmer le même résultat pour un simplexe de dimension quelconque.

Les divers problèmes évoqués précédemment sont des cas particuliers : ainsi dans un processus markovien, on a une transformation T qui transforme une distribution P $(P_1 + P_2 + \ldots + P_n = 1,\ P_i \geqslant 0)$ en une distribution q $(\sum q_i = 1,\ q_i \geqslant 0)$. C'est bien une application d'un simplexe en lui-même. Bien sûr, dans les modèles markoviens, T est linéaire : mais l'assertion de Brouwer nous indique en quelque sorte que ce n'est pas la linéarité qui importe le plus.

On peut se demander à quoi bon chercher de si larges généralisations. L'important n'est-il pas de pouvoir calculer ?

C'est encore Cournot qu'il faut lire à ce sujet. Dans la préface de ses célèbres "Recherches sur les principes mathématiques de la Théorie des Richesses" (1838) il demande qu'on remarque "dans l'exposé des premières notions sur la concurrence et le concours des producteurs, certaines relations assez curieuses à les envisager sous le point de vue purement abstrait". Cournot avait certainement conscience d'avoir ouvert une voie nouvelle vers une théorie générals des actions humaines conjointes (concertées, concurrentes, etc...).

Il s'agit de deux, puis d'un nombre quelconque d'agents économiques dont les buts sont donnés (chacun, de son côté, cherche à rendre son revenu le plus grand possible). Par exemple le marchand de cuivre fixe le prix P_1 de son cuivre, le marchand de zinc fixe le prix P_2 du zinc, il en résulte une demande de cuivre et de zinc de la part des fabricants de laiton, qui, compte tenu des prix offerts par les deux fournisseurs peuvent calculer leur prix de revient, prévoir leurs ventes et passer les commandes en conséquence. Ainsi les ventes du cuivre dépendent des décisions des fabricats d'alliage, lesquelles à leur tour dépendent à la fois des décisions concernant le cuivre et le laiton. Le bénéfice B_i des marchands de cuivre sera fonction des deux prix :

$$B_1 = f \;(P_1, P_2)$$

et de même :

$$B_2 = g \;(P_1, P_2)$$

Chercher le maximum de B_1, en agissant seulement sur P_1, n'a de sens que si P_2 est connu. Mais inversement le calcul économique pour le zinc (maximum de B_2 en agissant sur P_2) suppose la connaissance de P_1. C'est le cercle logique typique dans les situations de ce genre : chacune des décisions dépende de l'autre. Mais c'est justement à ce point que l'analyse mathématique (celle qui n'a pas seulement pour objet le calcul numérique, précise Cournot), servira la recherche.

A chaque valeur éventuelle de P_2 correspondra, de par la finalité du premier agent, une valeur optimale de P_1 :

$$P_1 = \varphi(P_2)$$

et de même les intentions du second agent s'interprètent par une liaison :

$$P_2 = \psi(P_1)$$

D'où il suit, conclut (un peu brièvement) Cournot, que les valeurs <u>définitives</u> (c'est à dire d'équilibre) de P_1 et P_2 seront déterminées au moyen du système d'équations qui exprime les deux relations fonctionnelles précédentes.

En utilisant la figuration cartésienne dans laquelle toute situation possible (P_1, P_2) est figurée par un point, nous aurons deux courbes de réaction φ et ψ. Il s'agit de savoir si et comment ces deux courbes' "se coupent".

Bien entendu Cournot suppose φ et ψ continues et se contente d'une évidence pré-jordanienne. Après Cournot (citons par exemple: A. Marshall, Stackelberg) le nombre de modèles économiques dans lesquels un "équilibre" est défini par une intersection, est devenu considérable.

Mais il a fallu attendre le célèbre mémoire de Von Neumann (Über ein okonomisches gleichungsystem und eine Verallgemeinerung des Brouwer'schen Fixpunktsatzes, 1935 ; trad. Italienne : L'industria, 1952, n 1, PP. 1 - 28) pour songer à établir avec toute la rigueur topologique souhaitable l'existence de tels équilibres.

A quoi bon l'existence dira-t-on si l'on n'a point d'algorithm ? C'est que dans la pensée de Cournot, comme de ses continuateurs, l'analyse purement "qualitative" joue un rôle fondamental dans la théorie des phénomèmes sociaux .
Une classification des jeux par exemple selon l'existence ou non de l'équilibre est essentielle (voir Stratégies et décisions économiques, Paris, C.N.R.S. , 1954, pp. III - 14 et suivantes) .

8 <u>TOPOLOGIE DES POINTS FIXES</u>

1. Commençons par un cas fort simple : application d'un segment en lui-même .

Soit une fonction $y = f(x)$ telle que :

$$0 \leq x \leq 1 \quad \text{(segment fermé)} \implies y \in [0, 1]$$

L'intuition (graphique) conduit sans peine à distinguer trois classes de valeurs pour x.

1) tous les x tels que $0 < x < f(x)$
(ainsi que $x = 0$, si $f(0) > 0$)

2) $x < f(x) < 1$ (ainsi que $x = 1$, si $f(1) < 1$)

3) $x = f(x)$; c'est à dire les points fixes .

il s'agit de montrer que si la classe (3) est vide, alors f ne peut être continue .

Cette classification est en effet équivalent à une application de l'ensemble $0 \leq x \leq 1$ <u>sauf</u> les points fixes dans l'ensemble constitué par les deux points 0 et 1. Appelons cet ensemble le <u>bord</u> B du segment E. Nous avons donc une classification K :

$$E^* \xrightarrow{K} B$$

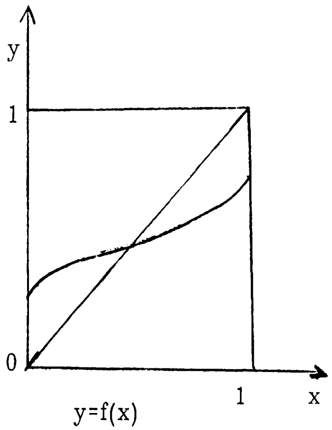
$$E^* = E - \{\text{fixes}\}$$

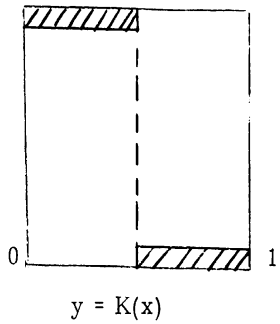
Or il est facile de voir que si f est continue, la fonction K l'est aussi.

Mais il est impossible de trouver une fonction continue définie sur E tout entier et qui ne prenne que deux valeurs. Donc E^* n'est pas E, c'est à dire que l'ensemble de points fixes n'est pas vide.

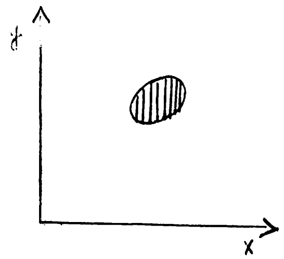
On pourrait donner une illustration géométrique de cette méthode de démonstration. Mais il faut bien comprendre qu'une telle figuration deviendra impossible pour les généralisations ultérieures. Voici cette illustration cependant :

On classe les points x selon que la courbe f(x) est au-dessus ou bien au-dessous de la diagonale. On remplace donc f par K qui est une fonction étagée, à deux valeurs ; et la discontinuité est alors bien visible.





2. Passons au cas de deux dimensions et prenons un fragment E du plan R x R . Ce pourra être un disque, un carré, un simplexe, etc...



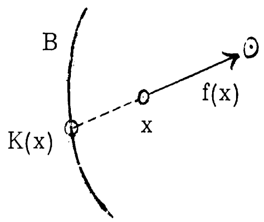
L'important est l'existence, ici encore, d'un Bord, qu'on nommera B.

Si l'on a une fonction f :

$$E \xrightarrow{f} E$$

qui à tout x $\in$ E fait correspondre f(x) $\in$ E, on peut ici encore construire une "classification" K qui rattache chaque point x de E à point K(x) du Bord.

Par exemple dans un espace affine, on cherche le point K(x) du bord tel que x est situé entre K(x) et f(x) . C'est déjà ce que nous avons fait pour le cas où E était un segment et B l'ensemble de ses deux points extrêmes. Dans le cas affine, la



convexité de E, par exemple assurera l'unicité de K(x) . Dans d'autres cas on pourra trouver d'autres façons de faire .

Il n'est pas inutile de noter que la fonction K(x) n'est pas seulement un artifice logique : dans de nombreuses applications, où la fonction f possède une signification concrète, la fonction K elle-même voundra dire quelque chose : c'est en somme un indicateur de la Tendance ou la Direction qui oriente le mouvement de x vers f(x). Dans le cas du segment il n'y avait que deux sortes de déplacement (gauche ou droite) . Ici il y en a une infinité . En empruntant à la topographie, on pourrait dire : le gisement, l'azimut, le cap, etc...

On notera que pour les points fixes x et f(x) étant confondus, le cap K n'est plus défini. Ainsi la fonction K déduite de f, n'est défini que sur E^* = l'espace E sauf les points fixes s'il y en a.

3. On comprend sans peine que la construction de l'application K peut être étendue à un nombre quelconque de dimensions. A titre d'exercice, signalons le cas d'un simplexe :

$$x = (x_1\, x_2 \ldots x_n)\,, \qquad x_i = 0,\;\; x_i \geq 0$$

est une fonction linéaire y = F x (F = matrice, telle que celles qu'on a vu dans nos premiers exemples) .

Le Bord d'un simplexe est constitué par les points $(x_1\, x_2 \ldots x_n)$ qui ont au moins une coordonnée nulle. Etant donné deux points x de Fx , on peut calculer un point Kx du bord tel que x soit une moyenne (donc intermédiaire) entre Fx et Kx . On pourra aussi penser au cas où E est une boule et B une sphère.

4. Revenons au cas de deux dimensions . Nous avons :

$$E^{*} \xrightarrow{K} B \qquad \text{(fonction Cap)}$$

Remarquons que si l'on prend un point du bord, s'il n'est pas fixe par f, il est fixe par K .

Nous devons donc étudier les relations (fonctionnelles) entre l'espace E et son bord B. Et si, comme c'est le cas, ce sont les fonctions continues qui nous interessent, nous nous limiterons aux applications continues.

L'idée fondamentale (elle remonte à l'Analysis Situs de Poincaré) est d'étudier aussi les applications : $B \longrightarrow E$

Intuitivement, en effet, on comprend qu'il ne puisse y avoir d'application continue

$$E \longrightarrow B$$

c'est à dire de "rétraction" d'un disque sur son bord (et qui laisse le bord invariant) . Mais pour le prouver il faut avoir quelques donnée sur la connexité du disque E .
Or cette structure (typiquement topologique) de E est décrite par l'ensemble des applications continues :

$$B \longrightarrow E$$

Dans le cas du disque, cela signifie l'ensemble de tous les circuits fermés (images continues de B) qu'on peut décrire dans E. Il n'est pas étonnant de retrouver la notion de cycle, déjà présente dans tous nos modèles (en particulier ceux de Cournot, c'est à dire d'action réciproque) .

Toutefois l'ensemble de tous les circuits fermés est un ensemble trop peu maniable. Il est préférable d'en faire une classification. Et c'est l'homotopie. Deux fonctions

$$B \xrightarrow{0} E \quad \text{et} \quad B \xrightarrow{1} E$$

sont homotopes (deux circuits appartiennent à la même classe) si l'on peut interpoler par

$$B \xrightarrow{\theta} E$$

avec : $0 \leqslant \theta \leqslant 1$. C'est, en somme, parler de segments (simplexes) dans notre ensemble de fonctions.

Or s'il n'y a pas d'obstacles (de trous dans le disque) il est clair que tous les circuits sont homotopes : il n'y a qu'une seule classe. Par contre s'il y a un trou (fut-il ponctuel) les chemins fermés se distinguent selon qu'ils entourent ou non le trou, et selon le nombre de "tours" qu'ils font autour de ce trou (et dans quel sens) .

5. En langage technique on dira que le groupe (car c'est un groupe suivant la loi de composition évidente) - le groupe fondamental de Poincaré

$$\pi[B \longrightarrow E]$$

est réduit à un seul élément tandis que

$$\pi[B \longrightarrow E^{*}]$$

sera le groupe des entiers Z s'il y a un seul trou, (et le groupe Z^n s'il y a n trous) .

Mais quand on a un circuit fermé

$$B \xrightarrow{\gamma} E$$

en composant avec

$$E^* \xrightarrow{K} B$$

on aura une application (continue) :

$$B \xrightarrow{\gamma K} B$$

Or il est facile de se faire une idee de :

$$\pi[B \longrightarrow B]$$

c'est à dire d'une classification (par homotopie) de toutes les applications de B en lui-même.

Dans le cas où B est un cercle , on retrouve le groupe Z des entiers (nombre des tours) ..

6. Finalement à partir de la fonction

$$E^* \xrightarrow{K} B$$

on a construit une autre fonction :

$$\pi[B \longrightarrow E^*] \longrightarrow \pi[B \longrightarrow B]$$

qui établit une correspondance entre deux groupes. On aboutit à un problème algébrique.

G. Th. Guilbaud

9 CATEGORIES ORDINALES

En ce qui concerne la théorie générale du calcul économique (et des équilibres), on doit à Cournot une première libération : l'idée de fonction quelconque (dont le statut mathématique n'a été défini qu'au début du XIX siècle) permet de se libérer de l'étude du calcul numérique : et Cournot lui-même souligne l'importance de cette libération. Toutefois, comme il était normal, les fonctions employées sont encore des applications de l'ensemble R des nombres réels (ou bien de R^n) en lui-même. L'écriture y = f(x) signifie (sans qu'on le dise toujours) ce que nous écririons aujourd'hui :

$$R \xrightarrow{f} R$$

Avec Pareto commence une seconde libération. Pareto en effet s'intéresse non seulement à l'action des hommes, mais au moteur de cette action, à savoir leurs intérêts, leurs "préférences". Et la notion de préférence renvoie à la notion mathématique d'ordre (ou plus précisément de préordre).

L'Utilité s'est spontanément présentée sous forme numérique ; soit que, avec Daniel Bernoulli on s'occupe de la décision qui est commandée par un calcul monétaire ; soit que, avec Bentham, on justifie les analogies d'un calcul monétaire et de cette maturation de la décision qu'on appelle encore "calcul" par le fait qu'il s'agit de comparer ce qui est susceptible de "plus" et de "moins". De même encore pour Dupuit, ou Grossen, ou Walras.

Mais Pareto a bien vu qu'un mouvement de repli était possible, qu'il suffisait, revenant à une position aristotélicienne, d'apercevoir une forme commune à toutes les délibérations d'un agent, et que cette forme commune aurait la préférence (mettre ceci avant cela) : tout intérêt (individuel ou autre) sera donc censé être représenté par un système cohérent de relations (binaires) de préférence. La cohérence requise étant celle d'une ordination (qui organise le superlatif à partir du comparatif) c'est à dire pour le logicien, une transitivité. Ainsi la notion d'ordre fournit une image mathématique de l'intérêt , moteur de tout choix ; si l'on appelle ordre toute relation binaire transitive.

Mais la notion d'ordre n'implique pas celle de nombre . C'est le contraire : les divers ensembles numériques, produits par l'activité mathématicienne (naturels, entiers, rationnels, réels) sont ordonnés, mais possèdent en plus des structures algébriques particulières (groupe abélien, anneau, corps) . On peut donc souhaiter une description des phénomènes indépendante de toute représentation numérique. Pareto n'y a jamais réussi complètement (comparer le "Manuale" et le "Corso" avec l'article de l' "Encyclopédie mathématique" , pour saisir le développement de la pensée parétienne) .

Aujourd'hui la chose est devenu beaucoup plus claire. On part de l'ordre total : s'il n'est pas strict (alors, avec Bourbaki, on le nommera pré-ordre) on peut définir des classes d'équivalence (ou d'indifférence) qui, elles, sont totalement et strictement ordonnés. C'est à dire constituent ce qu'on peut appeler une échelle. Mais on sait bien, depuis Cantor, qu'il existe une grande diversité des échelles. Rappelons sommairement quelques notions essentielles.
L'échelle la plus simple est constituée par un échelle finie qui "range des objets dans un certain ordre" :

$$a < b < c < \dots < z$$

Il y a des "intervalles" ouverts et vides : rien entre a et b . C'est le caractère "discret" de l'ordre. Il existe des échelles infinies discrètes : leur prototype est celle que Cantor nomme $\mathcal{Z}$ (initiale de Zahlen)parce qu'elle est représentée par le système des entiers:

$$\ldots < -2 < -1 < 0 < +1 < +2 < \ldots$$

Les problèmes concernant le superlatif et le comparatif sont d'une grande simplicité avec de telles échelles : il n'y a presque rien à en dire qui dépasse l'intuition et le sens communs. Mais , dans bien des cas, s'impose la nécessité d'interpoler, d'avoir des nuances. On considèrera donc des échelles possédant la propriété contraire de celle d'être "discretes" ("dichte" disait Cantor; en français on pourrait parler d'un ordre "dru", "fitto" en italien) : entre deux éléments quelconques il en existe toujours un troisième distinct de ces deux-là . Si $a < b$, l'ensemble $(x \mid a < x < b)$ n'est jamais vide.

Il est possible de construire une échelle d'ordre possédant cette propriété en utilisant une écriture de type numérique , c'est à dire un alphabet finit et par conséquent un ensemble dénombrable. Cantor a même démontré l'unicité de la solution, qui n'est autre que le type d'ordre η que possède l'échelle des nombres rationnels.
Mais cet ordre possède aussi une curieuse propriété, parfois gênante dans les applications (chaque fois qu'il faut faire des approximations) , et qu'on peut nommer selon la topologie (tout ordre total induit une topologie construire à partir des intervalles ouverts) on peut dire que l'échelle des rationnels n'est pas <u>compacte</u> . Cela signifie qu'il existe des sections qui ne sont pas des <u>intervalles</u> : traditionnelle-

ment on donnera l'exemple des rationnels dont le carré est inférieur à 2; non seulement aucun de ces nombres n'est supérieur à tous les autres; mais dans l'ensemble complémentaire aucun n'est supérieur à tous les autres ; mais dans l'ensemble complémentaire aucun n'est inférieur à tous les autres.

En d'autres termes, le superlatif ne fonctionne plus du tout comme dans les échelles finies.

Il existe par contre (mais non plus dénombrables) des échelles qui sont compactes ; alors pour toute partie d'un tel ensemble, il existe un supremum (le plus petit majorant) et un Infimum. Un segment réel est une telle échelle ; mais il y en a d'autres. Et l'on devra choisir, en connaissance de cause, le type d'ordre dont on pense qu'il est un modèle correct (ou le moins mauvais possible) des préférences considérées.

Chez Pareto, Edgeworth et plusieurs émules, ces exigences axiomatiques sont absentes : et le modèle des nombres réels semble s'imposer comme une loi de nature.

10 EQUILIBRES A LA PARETO

Revenons à l'idée d'équilibre chez Pareto : c'est essentiellement une confrontation de deux ou plusieurs ordres de preférence. Le prototype semble bien avoir été l'équilibre du consommateur.

L'état du consommateur étant le plus souvent figuré par un vecteur

$x = (x_1\ x_2 \ldots x_n)$ dans un espace convenablement choisi, la délibération du sujet porte sur deux aspects des choses :

1°) les couts : c'est d'abord un calcul linéaire qui s'écrit

$$p(x) = \sum p_i\ x_i$$

d'où résulte un ordre : $p(x) < p(y)$, qu'on lit :

l'état x est moins cher que y .

2°) les goûts : ici au contraire c'est un ordre qui est donné d'abord (et la question se pose de savoir si, et comment, on peut le ramener à un calcul) .

On dit x est plus agréable (ou utile) que y et on écrira: (utilité ou ophélimité)

$$u(x) > u(y)$$

Dans un cas fini et discret le problème est simple ; Soient quatre états a, b, c, d que le consommateur considère possibles et qu'il veut comparer. Supposons que les gouts soient :

a moins utile que b qui est moins utile que c; lui-même enfin moins utile que d

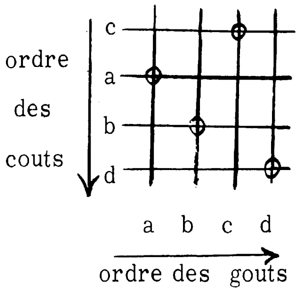
utilités : $a < b < c < d$

et les couts :

c moins cher que a moins que b moins que d

couts : $c < a < b < d$

Alors la comparaison se fait par le produit logique des deux



ordres de préférence.

Il est clair qu'on obtient un ordre partiel

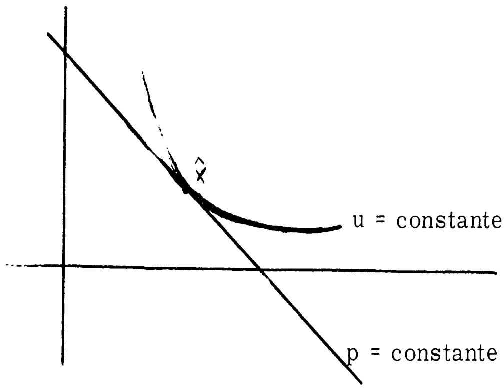
C'est plus utile et moins cher que a, donc préférable à tous points de vue.

par contre pour comparer c à d :

c est moins cher que d mais moins utile.

il faudra un arbitrage.

On sait que dans le schéma paretien, une limitation supérieure des couts est donnée (sous forme de revenu disponible) et le problème est alors de trouver l'optimum (selon l'utilité) lorsque le cout est fixé . On connait aussi la forme traditionnelle des variétés d'indifférence et des variétés linéaires de dépense, symbolisées par la figure classique :



Il est alors intéressant de se demander quelles sont les hypothèses introduites (parfois, chez les anciens auteurs, implicitement) pour

assurer la régularité des conclusions dont on aura besoin.
(Renvoyons par exemple à Debreu, Théorie de la valeur, Paris, Dunod, 1966 , pages 59 et suiv.)

Un seul exercice :

assurer l'optimum à cout constant, c'est trouver un équilibre tel que :

(1) pas plus cher que $\hat{x}$ $\Longrightarrow$ pas plus utile que $\hat{x}$

c'est à dire

$$p(x) \leqslant p(\hat{x}) \Longrightarrow u(x) \leq u(\hat{x})$$

Mais il est intéressant de se demander alors si l'on obtient le même résultat (dualité) en cherchant l'optimum (moindre cout) à l'utilité donnée. C'est à dire savoir si l'on a :

(2) pas moins utile $\Longrightarrow$ pas moins cher

c'est à dire :

$$u(x') \geqslant u(\hat{x}) \Longrightarrow p(x') \geqslant p(\hat{x})$$

Il est facile de voir que (1) et (2) ne sont pas équivalents en général. Parmi les hypothèses qui assurent que l'on a :
(1) $\Longrightarrow$(2) , on retient tout spécialement celle de continuité :

> L'espace est topologisé et l'on pose que :
> au voisinage de x il existe toujours des y qui sont plus utiles (et aussi des y qui sont moins chers : mais cela est assuré par la linéarité) .
> (voir Debreu, loc. cit.)

La confrontation d'ordres de préférence provenant de points de vue différents, va servir à Pareto dans sa construction d'un équilibre social.

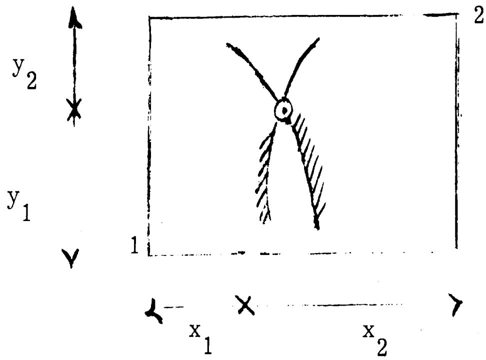
Commençons par un schéma, plus proche d'Edgworht que de Pareto, mais en fait universel. C'est le problème du partage.

Deux biens, deux individus.

Un programme de partage c'est un ensemble de quatre nombres positifs :

$$x_1 + x_2 = \text{donné}$$
$$y_1 + y_2 = \text{donné}$$

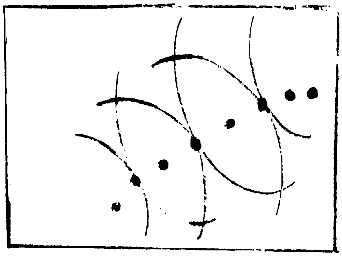
qui disent les parts respectives de Primus (1) et Secondus (2) dans le bien X et les parts en Y. On peut figurer par un point dans un rectangle (plus généralement ce serait un produit cartésien de simplexes).



Etant donné un point quelconque (x_1, x_2, y_1, y_2) il faut le comparer à tous les autres du domaine, et en particulier rechercher s'il n'existe pas un autre partage qui serait préféré à celui là à la fois par les deux partenaires.

Seuls seront déclarés "admissibles" ceux pour lesquels il est impossible de trouver une modification avantageuse à la fois pour les deux. Ce sont les " équilibres " au sens de Pareto.

Edgeworth les construisait comme l'ensemble des points de contact de deux lignes d'indifférence ,



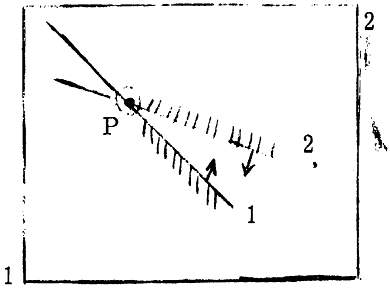
et cet ensemble est la Contract Curve de notre auteur. Bien entendu pour affirmer que c'est une "courbe", il faudra quelques hypothèses sur les lignes d'indifférence.

Notons un cas singulier dans lequel les indifférences seraient des variétés linéaires, c'est à dire que les utilités seraient calculées par :

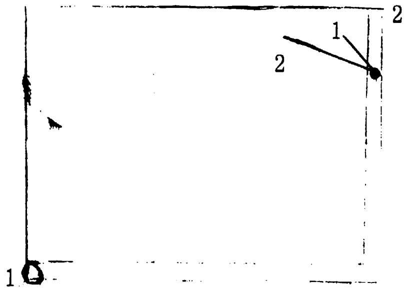
$$u(x) = u_1 x_1 + u_2 x_2$$

(pour chacun des partenaires, mais avec des coefficients différents pour l'un et l'autre) .

Alors il est immédiat de prouver que les points admissibles se trouven sur la frontière du domaine



partages jugés meilleurs que P par les deux



ensemble des partages admissibles.

On retrouve une "progammation linéaire" rudimentaire.
Ce modèle a été utilisé par Ricardo dans sa célèbre théorie du commerce international : les seules situations admissibles conduisent à la spècialisation (au moins l'une des quantités x_i y_i est nulle) .

11 LA CONTESTATION

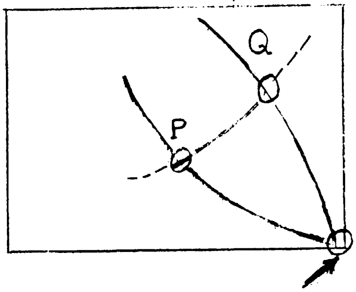
Cette propriété à laquelle Pareto attache tant d'importance, nous pourrons la qualifier d'optimum paretien , tout en notant que l'unicité d'un tel point n'est pas la règle, bien au contraire, Ainsi dans le problème de partage, selon Edgeworth , c'est tout un ensemble, formant la courbe des contrats qu'on peut raisonnablement débattre.
Mais on devra souvent introduire quelques conditions suplèmentaires que nous rangerons sous le vocable de Contestation. Ainsi dans le partage, il y a une situation initiale Primus "possède"

$x_1 = a$ $\qquad y_1 = 0$

et Secundus

$y_1 = 0$ $\qquad y_2 = b$

Alors on accordera à chacun d'eux droit de contester toute proposition qui n'améliore pas son état initial.



situation initiale

Sera dans ce cas admissible, seulement la partie PQ de la courbe des contrats située entre les deux indifférences passant par la situation initiale.

Mais il reste encore une indétermination dont la discussion, interminable, a constitué ce qu'on a appelé naguère la théorie du monopole bilatéral - théorie oubliée, non sans raisons.

Von Neumann et Morgenstern ont proposé dans leur Theory of Games, un mode de contestation assez particulier, en introduisant de surcroit l'idée féconde de coalition.
La f orme générale en est la suivante. (Je me permets de renvoyer, pour, les détails à Stratégie et Décisions économiques, Paris, C.N.R.S., 1954, pages V-1 à V-14).

Pour préférer une situation A à une situation B, un certain nombre d'agents économiques peuvent etre d'accord : ils trouvent tous avantage à passer de B en A.
Selon Pareto doivent seulement être prises en compte les préférences unanimes. Mais en fait, on sait bien que dans une société, les choses ne se passent pas toujours ainsi, et que les pouvoirs de contestation ne sont pas uniformément répartis.
Soient alors un ensemble N d'individus (en nombre n), et une partie K de cet ensemble.

$$K \subset N$$

nous applellerons K une coalition.

Pour Pareto la seule coalition intéressante est K = N. Edgeworth ajoute qu'il faut tenir compte, en sus, des opinions individuelles (c'est à dire des "coalitions" si l'on peut dire formées d'un seul individu) mais jusqu'à un certain point seulement : un individu ne peut imposer sa préférence de A à B que si B est inférieure à sa situation initiale.
Von Neumann et Morgenstern généraliseront cette idée en introduisant les préférences <u>efficaces.</u>

Une préférence unanime (coalition N) est toujours prise en compte. Mais les préférences d'une coalition K "A est préféré par K à B" peuvent l'être aussi certaines conditions: cette préférence est efficace si B appartient à un domaine donné (dépendant de K).

Les auteurs cités n'ont étudié les conséquences de cette idée que dans le cas du partage d'une somme d'argent (jeux). Et l'on constate, en ce cas, la diversité considérable des équilibres.
Au point qu'on a désespéré depuis de donner des conclusions générales assez précises. Mais on peut prendre le problème d'une façon legèrement différente et aboutir au moins à une conclusion intéressante.

Représentons les situations initiales de n agents économiques par n vecteurs choisis dans un espace convenable.
Soient $\vec{a}_1\ \vec{a}_2 \ldots \vec{a}_n$ ces situations initiales.
Proposons une redistribution définie elle aussi par n vecteurs:

$$\vec{x}_1\ \vec{x}_2 \ldots \vec{x}_n$$

Une telle redistribution est <u>possible</u> en anglais (feasible) sans apport extérieur, et par simple échange si :

(1) $$\sum x_i = \sum a_i$$

Peut-elle être contestée ?

1°) Soit un sous ensemble d'individus, que nous appellerons coalition K.

$$K \subset \{1, 2, \ldots, n\}$$

Supposons qu'ils puissent se mettre d'accord pour présenter un contre-projet de partage, une suite de vecteurs y_i, telle que chacun d'eux préfère y_i à x_i $(i \in K)$ et telle aussi que :

$$\sum_{i \in K} y_i = \sum_{i \in K} a_i$$

c'est à dire que la modification soit realisable par K seule (sans toucher aux autres) .
Alors nous pouvons, c'est notre postulat, ou si l'on préfère une définition de l'efficacité, déclarer que nous avons une contestation efficace du projet (x) par la coalition K .

Il suffit de traiter quelques exemples simples pour voir que l'ensemble des projets incontestables peut être fort compliqué à calculer, quand il existe. Mais on va faire un pas de plus.

2°) Il n'est pas déraisonnable de contester un projet, non seulement par la critique de ce qu'il réaliserait, mais encore par les principes qu'il suppose. Un individu, par exemple, peut plaider sa cause d'abord en s'associant avec ceux dont la situation est voisine de la sienne, mais aussi bien en élevant son cas particulier au niveau d'une règle générale. Il dira par exemple : si je n'étais pas seul dans la situation où je suis, (ou bien si nous n'étions pas si peu nombreux) nous pourrions faire entendre notre point de vue.

Essayons de traduire cela en termes mathématiques.
Prenons une coalition K et imaginons une coalition hypothétique K^* formée d'un nombre quelconque de répliques de chacun des membres. Soit une situation y jugée par chacun des membres de K meilleure que la situation x. Nous admettrons la contestation si l'on a :

$$\sum_{i \in K} n_i \vec{y}_i = \sum_{i \in K} n_i \vec{a}_i$$

quels que soient les entiers positifs n_i .

On peut écrire la même chose en parlant de moyennes pondérées:

$$\frac{\sum n\, y}{\sum n} = \frac{\sum n\, a}{\sum n}$$

Qu'arrive-t-il alors pour un projet incontestable c'est à dire qui ne peut être efficacement contesté par aucune coalition ?

Pour pouvoir repondre il faudra faire encore quelques hypothèses supplémentaires.

Ici se présente assez naturellement celle-ci :

L'ensemble des $\vec{y}_i$ préférés à $\vec{x}_i$ est un ensemble convexe.

On y adjoindra la continuité des préférences (voir leçon précédente) .

Si l'on considère alors divers domaines

$$\vec{y}_i \text{ préféré (par i) à } \vec{x}_i$$

aucune moyenne de plusieurs $\vec{y}_i$ ne peut être égale à la moyenne correspondante des $\vec{a}_i$ (sinon il y aurait contestation efficace de x) . Il en résulte alors une Séparation entre la situation initiale (les $\vec{a}_i$) et la fermeture convexe des partages préférés (les $\vec{y}_i$) c'est à dire un opérateur linéaire p tel que :

$$y_i \text{ préféré à } x_i \Rightarrow p(y_i) \geq p(a_i)$$

On en déduit $p(x_i) \geq p(a_i)$ et comme $\sum x_i = \sum a_i$, on a même

$$p(x_i) = p(a_i)$$

On retrouve un équilibre par un système de prix.

Inversement on peut montrer qu'un équilibre par les prix, c'est à dire un choix, pour chaque individu i , effectué à la manière du consommateur paretien d'une situation nouvelle $\hat{x}_i$ la meilleure parmi celles qui vérifient

$$p(x_i) = p(a_i)$$

un tel équilibre est incontestable.
En effet tout ce qui est plus utile (préférable) est trop cher .

$$y_i \text{ préféré par } i \text{ à } x_i \Rightarrow p(y_i) > p(x_i) = p(a_i)$$

Donc

$$\sum p(y_i) > \sum p(a_i)$$

ou bien :

$$p(\sum y_i) > p(\sum a_i)$$

(linéarité)

et il est impossible d'avoir la condition

$$\sum y_i = \sum a_i$$

qui permettrait la contestation.

On peut compliquer le modèle sans difficulté (voir Debreu, Intern. Eco. Rev., sept. 63, p. 235) en introduisant ce que la théorie appellera des producteurs. Mais il convient surtout de réfléchir à la signification du raisonnement .

On retrouve l'équilibre par les prix, chaque fois que l'on peut ériger en règle générale une situation particulière. C'est le sens même du calcul linéaire : fixer le prix d'une transaction, ce n'est pas seulement dire que tant de tonnes de blé seront échangées contre tant de monnaie, c'est donner en même temps tout un ensemble de transactions proportionnelles entre une quantité q et une somme pq (q varie, p reste fixe).

Bien entendu cela n'épuise pas la question. La contestation envisagée s'appuie sur la situation initiale, considérée comme une sorte de "droit de propriété", elle est donc très conservatrice. A l'égal du schéma paretien qui rejette à plus tard le problème de la répartition du revenu. Et comme on ne le sait que trop, cela ne résout rien si l'on fait du travail une marchandise "comme les autres".
Mais, comme on dit, c'est là une autre histoire.

CENTRO INTERNAZIONALE MATEMATICO ESTIVO
(C.I.M.E)

Harold W. KUHN

LOCATIONAL PROBLEMS AND MATHEMATICAL PROGRAMMING

Corso tenuto all'Aquila dal 30 agosto al 7 settembre
1965

LOCATIONAL PROBLEMS AND MATHEMATICAL PROGRAMMING

by

Harold W. Kuhn

1. Introduction. The most simple locational problem has its mathematical origin in classical geometry, where it is known as Steiner's Problem [1]. It appeared, in a slightly generalized form, in the pioneering work Über den Standort der Industrien of Alfred Weber [2]. This form of the problem, which we shall call the Steiner-Weber Problem, asks for a point in the plane that will minimize the weighted sum of distances to n given points in the plane. In spite of the simple and explicit form of the problem, relatively little is known about its solution, either analytically or computationally. The purpose of this paper is to discuss the problem from the point of view of mathematical programming. In Section 2, certain general properties and a set of necessary and sufficient conditions for a solution are derived. In Section 3, a problem dual to the Steiner-Weber Problem is formulated. This problem has a linear objective function and quadratic constraints; it possesses all of the desirable properties of the dual in linear programming and its solution yields a solution of the Steiner-Weber Problem trivially. In Section 4, some preliminary conclusions concerning computation are presented. Economic applications are the subject of a joint paper with R.E. Kuenne [3], to be published shortly; detailed computational methods will be treated in a later paper.

2. Statement of the Steiner-Weber Problem. Although the simpler properties of the problem have been amply discussed in the literature, for the sake of completeness and to establish notation, we shall resta-

te the problem and derive some basic results here.

Let there be given n distinct points $X_i = (x_i, y_i)$ in the plane and n positive weights w_i, where $i = 1, \ldots, n$. Furthermore, for $P = (x, y)$, let

$$d_i(P) = \sqrt{(x_i - x)^2 + (y_i - y)^2},$$

the Euclidean distance from P to X_i, for $i = 1, \ldots, n$. Then the Steiner-Weber Problem asks for a point P that minimizes

$$F(P) = \sum_i w_i d_i(P) \quad .$$

First note that, from the geometric interpretation of $d_i(P)$ as a Euclidean distance, it is obvious that $w_i d_i(P)$ is a convex function of P. Indeed, it is a strictly convex function of P except on the halflines ending at X_i ; on each of these halflines it is a linear function. Since the sum of (strictly) convex functions is (strictly) convex, the following result follows.

THEOREM 1 . If the n distinct points X_i are not colinear, then $F(P)$ is a strictly convex function of P. If the points are colinear, then $F(P)$ is piecewise linear on the line through them and strictly convex elsewhere.

COROLLARY. If the n distinct points X_i are not colinear, then $F(P)$ has a unique minimum.

For the rest of this section, attention will be restricted to the noncolinear case; the problems excluded by this restriction are clearly trivial. At this point, reasons both mathematical and physical (deriving from the string and weight model of the problem [4]) suggest consideration of the gradient of F (i.e., the resultant of the forces in the strings). To this end, let

$$R(P) = \sum_i \frac{w_i}{d_i(P)} (X_i - P)$$

if $P \neq X_i$ for all i. Obviously, the gradient is not defined at any vertex X_i. However, by physical analogy, set

$$R_j = \sum_{i \neq j} \frac{w_i}{d_i(P)} (X_i - P)$$

and extend the definition of $R(P)$ by letting

$$R(X_j) = \max(|R_j| - w_j, 0) \frac{R_j}{|R_j|} \quad \text{for} \quad j = 1, \ldots, n$$

(Here, as elsewhere in this paper, $|A|$ denotes the length of a vector A.) In this expression, the length of R_j is compared with the weight w_j. If $w_j > R_j$, then $R(X_j) = 0$; otherwise, a "gradient" of magnitude $R_j - w_j$ has been defined in the direction of R_j.

THEOREM 2. The point P minimizes $F(P)$ if and only if $R(P) = 0$.

PROOF. If P is not a vertex, then the convexity and differentiability of F imply that the first order condition $R(P) = 0$ is both necessary and sufficient for a minimum.

If P is the vertex X_j then consider a change from X_j to $X_j + tD$ where $|D| = 1$. Then, direct calculation yields:

$$\frac{d}{dt} F(X_j + tD) = w_j - R_j \cdot D \qquad \text{for} \quad t = 0,$$

and hence the direction of greatest decrease is $D = R_j / |R_j|$. (Here, as elsewhere in this paper, $A \cdot B$ denotes the inner product of the vectors A and B. Also $A^2 = A \cdot A$) Clearly, X_j is a local minimum if and only if

$$w_j - \frac{R_j^2}{|R_j|} \geqq 0$$

which is equivalent to $R(X_j) = 0$. Again , the convexity of F implies that $R(X_j) = 0$ is both necessary and sufficient for a global minimum.

THEOREM 3. If the point P minimizes F(P) then P is in the convex hull of the points X_i .

PROOF. If P is a vertex X_i then it is trivially in the convex hull. Otherwise , the condition $R(P) = 0$ yields P as a weighted sum of the points X_i with positive weights that add to one as follows :

$$P = \sum_i \frac{w_i}{d_i(P)} X_i \Big/ \sum_i \frac{w_i}{d_i(P)} .$$

COROLLARY. If the point P minimizes F(P) and is not a vertex, then P is not on the boundary of the convex hull.

PROOF. This is simply a consequence of the fact that the condition $R(P) = 0$ yields an expression for P as a positive weighted sum of the points X_i .

It is remarkable that Theorem 3 can be sharpened by a results which asserts that every point Q outside the convex hull is further from all of the X_j than some point P (which will depend on Q) in the convex hull . This theorem provides what appears to be a new characterization of the convex hull of n points.

DEFINITION. A point P _dominates_ Q with respect to the points X_i if $P \neq Q$ and $d_i(P) \leqq d_i(Q)$ for $i = 1, \ldots, n$. A point is said to be _admissible_ if it is not dominated by any other point.

THEOREM 4. A point P is admissible if and only if it is in the convex hull of the points X_i.

PROOF. We first prove that every point Q not in the convex hull is dominated by some point P in the convex hull. If this is not the case, then for all P in the convex hull,

$$d_i(P) > d_i(Q) \quad \text{for some} \quad i.$$

That is, the functions $f_i(P) = d_i(P) - d_i(Q)$ form a family of n continuous convex functions defined on a compact convex set which is such that

$$\max_i f_i(P) > 0 \quad \text{for all} \quad P.$$

Hence, by Ville's Lemma [5], there exist $w_i \geqq 0$ with $\sum w_i = 1$ such that

$$\sum_i w_i d_i(P) \geqq \sum_i w_i d_i(Q)$$

for all P in the convex hull. However, this means that Q solves the Steiner-Weber Problem with weights w_i, which contradicts Theorem 3. (Note that the fact that some w_i may be zero does not invalidate the argument. Merely restrict the problem to those X_i for which $w_i > 0$; these points have a convex hull which is a subset of the previous convex hull).

We must now show that every point in the convex hull is admissible. Clearly all of the vertices X_i are admissible. Let P and Q be two points in the convex hull of the X_i. Let A be an inadmissible point on the segment PQ joining them, i.e. A is dominated by some point B. Then all of the vertices X_i lie on the same side of the perpendicular bisector of AB as does B. But then so do P and Q since they are in the convex hull. However, this implies that A does also, which is a contradiction. To recapitulate, we have proved that the set of admissible points contains the vertices X_i and contains the segment joining any two points in the convex hull of the X_i. This implies that the set of admissible points contains the convex hull and completes the proof.

3. A Dual Problem. Several theories of duality of nonlinear programming are available. However, in most instances the dual program is difficult to give explicitly. Much more serious, in almost every case, its mere statement requires a solution to the original or primal program. Therefore, it is remarkable that the Steiner-Weber Problem possesses a dual with almost all of the useful properties of the duality of linear programming. In this section, we shall state the dual and verify these properties. (In a private comunication, C. Witzgall and R. T. Rockafellar have informed me that they have discovered the same dual.)

Let there be given n distinct points $X_i = (x_i, y_i)$ in the plane and n positive weights w_i, where $i = 1, \dots, n$. Let $U_i = (u_i, v_i)$ denote n two-dimensional vectors. Then the dual to the Steiner-Weber Problem asks for vectors U_i which maximize

(1) $$G(U_1, \dots, U_n) = \sum_i U_i \cdot X_i$$

subject to

$$\sum_i U_i = 0 \tag{2}$$

and

$$|U_i| \leqq w_i \qquad \text{for } i = 1, \ldots, n. \tag{3}$$

It should be noted immediately that this dual program is constructed entirely from the data of the primal program and does not require its solution.

PROPERTY 1. For any P and any U_i satisfying (2) and (3) (i.e., feasible U_i),

$$G(U_1, \ldots, U_n) \leqq F(P).$$

PROOF.

$$\begin{aligned} G(U_1, \ldots, U_n) &= \sum_i U_i \cdot X_i = U_i \cdot X_i - (\sum_i U_i) \cdot P \\ &= \sum_i U_i \cdot (X_i - P) \\ &\leqq \sum_i |U_i| \, |X_i - P| \\ &\leqq \sum_i w_i d_i(P) = F(P). \end{aligned}$$

Property 1 implies that equality of $F(P)$ and $G(U_1, \ldots, U_n)$ for any feasible U_i is a condition sufficient to establish optimality.

PROPERTY 2. Given any P that solves the Steiner-Weber Problem, there exist feasible U_i such that

$$G(U_1, \ldots, U_n) = F(P).$$

PROOF . We distinguish two cases : (A) P is interior to the convex hull of the X_i ; (B) $P = X_i$ for some i.

CASE A. In this case, set

$$U_i = \frac{w_i}{d_i(P)}(X_i - P) \qquad \text{for} \quad i = 1, \ldots, n\,.$$

Then (see Theorem 2 and the definition of $R(P)$),

$$\sum_i U_i = R(P) = 0$$

and

$$|U_i| = \frac{w_i}{d_i(P)}\,|X_i - P| = w_i \quad \text{for} \quad i = 1, \ldots, n.$$

Furthermore, both of the inequalities in the proof of Property 1 become equations, the first because $U_i \cdot (X_i - P) = |U_i|\,|X_i - P|$ and the second because $|U_i| = w_i$ for $i = 1, \ldots, n$.

CASE B . Suppose $P = X_1$; then set

$$U_i = \frac{w_i}{d_i(P)}(X_i - P) \qquad \text{for } i = 2, \ldots, n\,,$$

and define

$$U_1 = -\sum_{i=1}^{n} U_i\,.$$

Then $\sum_i U_i = 0$ by definition, while $|U_i| = w_i$ for $i = 2, \ldots, n$ as before. Finally, $U_1 = -R_1$ and $R(X_1) = 0$ implies $w_j \geq |R_1| = |U_1|$. Therefore the U_i are feasible.

Again , both the inequalities in the proof of Property 1 become equations. The only change from the previous argument is in the first term. Here,

$$U_1 \cdot (X_1 - P) = |U_1|\;|X_1 - P| = w_1 d_1(P)$$

since all three expressions are zero .

Property 2 implies that any solution to the Steiner-Weber Problem provides a solution to the dual program in a trivial manner.

PROPERTY. 3 . Given any feasible U_i which solve the dual program, there exists a P such that

$$G(U_1, \ldots, U_n) = F(P) \ .$$

PROOF. Given feasible U_i which solve the dual program, we may apply the Kuhn-Tucker Theorem [6] to formulate necessary conditions for this maximum problem. The Lagrangian function to be used is :

$$\sum_i U_i \cdot X_i - P. \sum_i U_i - \frac{1}{2} \sum_i t_i(w_i^2 - U_i^2)$$

where $P = (x, y)$ and t_i are Lagrange multipliers. The necessary conditions for a maximum assert the existence of multipliers such that

$$X_i - P - t_i U_i = 0$$

$$\sum_i U_i = 0$$

$$w_i^2 \geqq U_i^2$$

$$t_i \geqq 0$$

$$t_i(w_i^2 - U_i^2) = 0$$

for $i = 1, \ldots, n$. Again we must distinguish two cases.

CASE A. All $t_i > 0$.

Then $|U_i| = w_i$ for all and

$$t_i = \frac{d_i(P)}{w_i} \qquad \text{for } i = 1, \ldots, n \ .$$

Since $U_i = \frac{w_i}{d_i(P)}(X_i - P)$, we again have equality of the objective functions and P is optimal.

CASE B. Some $t_i = 0$, say , for $i = 1$. Then $P = X_1$ and, since the points X_i are distinct, we must have $t_i > 0$ for $i = 2, \ldots, n$. Hence $|U_i| = w_i$ and $U_i = \frac{w_i}{d_i(P)}(X_i - P)$ for $i = 2, \ldots, n$ as before. Equality between the objective functions follows exactly as in

Case B of Property 2.

Property 3 implies that any solution to the dual program provides a solution to the Steiner-Weber Problem in the following trivial manner. If all $|U_i| = w_i$, the lines through X_i with direction U_i all go through P. To find P, two lines will suffice. Otherwise, at most one U_i can satisfy $|U_i| < w_i$, say for $i = 1$. Then $P = X_1$.

This completes the proofs of the properties which parallel in a remarkable manner the duality of linear programming.

4. Computational Considerations. The situation with regard to computation of solutions of a Steiner-Wever Problem has been aptly put in a recent note of Eisemann [7]:

> "While this is a somewhat celebrated problem and appears simple, there are not, apparently, very good methods for its solution, especially for large n. Since this problem becomes increasingly important in an economic context, better methods are needed."

The discussion of the preceding sections makes it clear that at least two methods of mathematical programming are applicable, in theory, to the problem or its dual. The first method is due to Zoutendijk [8] and has been named by him the method of "feasible directions." The second method is due to Rosen [9], who has called it the "gradient projection method." However, due to the choice between various variants of these procedures and the alternatives of the primal or dual problems, no definitive recommendation can be made at this time.

One practical suggestion may be helpful in applying almost any algorithm, however. This is the fact that initial feasible solutions for both problems are readily available. Namely, if we let

$$C = \sum_i w_i X_i / \sum_i w_i$$

denote the weighted center of gravity of the points X_i, C is generally close to the solution of the Steiner-Weber Problem and is clearly feasible. Moreover, setting

$$k = \min_i \frac{1}{|X_i - C|} \quad \text{and} \quad U_i = kw_i(X_i - C)$$

for $i = 1, \ldots, n$ one has an initial feasible solution for the dual problem.

The feeling is very strong that a special algorithm is called for in a problem of such simple and explicit form. One such algorithm will now be proposed although its convergence has not yet been established. This is an iterative procedure which transforms a trial point P in the convex hull into a new trial point $P' = T(P)$ in the convex hull. The definition of T is as follows :

$$P' = T(P) = \sum_i \frac{w_i}{d_i(P)} X_i \Big/ \sum_i \frac{w_i}{d_i(P)} .$$

For the sake of continuity, set $T(X_i) = X_i$ for $i = 1, \ldots, n$. The motivation of this transformation is given by the following theorem :

THEOREM 5 . If the point P minimizes $F(P)$, then P is a fixed point of T. If P is a fixed point of T that is not a vertex X_i, then P minimizes $F(P)$.

PROOF. This theorem is an immediate corollary of Theorem 2 above.

In effect , the procedure proposed is merely a naive attempt to solve the first order conditions $R(P) = 0$ iteratively. Perhaps more remarkable is the fact that this algorithm is a "long-step gradient method". Indeed, recalling that $R(P)$ is the gradient of $F(P)$ whenever it exists, direct calculation yields

$$P' = P + h(P)R(P)$$

where

$$h(P) = \prod_i d_i(P) \Big/ \sum_j \Big(w_j \prod_{i \neq j} d_i(P)\Big)$$

for all points P in the convex hull of the X_i . Thus, the method reduces to a gradient method with precalculated length of step.

Properties of this iterative scheme seem difficult to establish. In particular, although no counterexample is known, no proof has yet been established of the crucial <u>conjecture</u>

$$F(P') \leq F(P) \qquad \text{for all} \quad P.$$

This inequality has been satisfied in all examples which have been calculated to date. For an example of such calculation, consider the 24 cities in the Russian Ukraine which rank among the 100 most populous cities in U.S.S.R. The cities are listed in the following table with their weights w_i equal to the proportion each has of the population of 100 largest Russian cities. The x_i and y_i are degrees of north latitude and east longitude respectively, correct to the nearest full degree.

Localitions and Weights for Major Ukraine Cities

City	x_i	y_i	w_i	City	x_i	y_i	w_i
Minsk	54	28	.012	Zhdanov	47	38	.007
Lvov	50	24	.010	Stalino	48	37	.016
Kishinev	47	28	.005	Makeyevka	48	38	.008
Odessa	46	31	.014	Gorlovka	48	38	.007
Nikolsev	47	32	.005	Taganrog	47	39	.005
Kherson	47	33	.004	Krasnodar	45	39	.007
Sevastopol	45	34	.004	Rostov	47	40	.014
Simferopol	45	34	.004	Shakhty	48	40	.004
Krivoi Rog	46	34	.009	Kadiyevka	49	39	.004
Dneprodzerhinsk	48	35	.004	Kharkhov	41	37	.021
Dnepropetrovsk	48	35	.015	Kiev	50	31	.026
Zaporozhe	48	36	.010	Gomel	52	31	.004

Taking as P^o, the center of gravity of these cities, the results of four iterations are tabulated below.

	x_i	y_i
P^o	47.48	34.36
P^1	47.51	35.74
P^2	47.64	35.51
P^3	47.62	35.35
P^4	47.60	35.32

The degree of optimality of P^4 may be gauged by $R(P^4) = (-0.00082, 0.01516) \sim 0$.

BIBLIOGRAPHY

1 R. Courant and H. Robbins, What Is Mathematics?, New York, 1941, pp. 354-358.

2 Alfred Weber, Über den Standort der Industrien, Tübingen, 1909; Ch. III. The work has been translated as Alfred Weber's Theory of the Location of Industries, Chicago, 1929, by C. J. Friedrich.

3 H. Steinhaus, Mathematical Snapshots, New York, 1960, p. 119. The use of this physical model was suggested by Georg Pick in his Mathematical Appendix to Weber's book, and the essential idea can be found in more modern treatments of force systems. See, for example, G. Polya, Induction and Analogy in Mathematics, Princeton 1954, pp. 147-148, and W. Miehle, "Link-Length Minimization in Networks, " Operations Research, 6 (1958), pp. 232-243.

5 Jean A. Ville, "Sur théorie géneral des jeux ou intervient l'habilité des jouers " Applications aux Jeux de Hasard by Émilie Borel and Jean Ville, Tome IV, Fasicule II, in the Traité du Calcul des Probabilités et de ses Applications by Emile Borel (1938), 105-113.

6 Harold W. Kuhn and A. W. Tucker, "Nonlinear Programming" in Proc. Second Berkeley Symposium on Mathematical Statistics and Probability, Univ. of Calif. Press, Berkeley, 1951, pp. 481-492.

7 Kurt Eisenmann, " The Optimum Location of a Center", SIAM Review, 4(1962), 394-5.

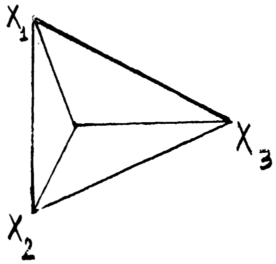
8 G. Zoutendijk, Method of Feasible Directions, Elsevier Publ. Co., Amsterdam, 1960.

9 J. B. Rosen, "The Gradient Projection Method for Nonlinear Programming, Part II Nonlinear Constrains, " SIAM Journal, 9 (1961), pp. 514-532.

ADDENDUM

In this short appendix, we shall give several details which were presented in the lecture but which are not contained in the preceding text.

First, the earliest formulation of the problem known to the author was given by Fermat at the end of his famous essay on maximum and minimum problems (ca 1635) . The problem seems to have been carried to Italy by Mersenne and solved by Torricelli (ca 1645?) . Fermats' problem was the special case n=3 and $w_1=w_2=w_3=1$ and the Torricelli solution is illustrated below.



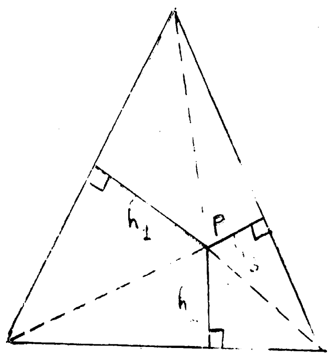
$$\angle X_1PX_2 = \angle X_2PX_3 = \angle X_3PX_1 = 120^\circ$$

The point P so determined is known in classical geometry as the <u>Fermat</u> or the <u>Torricelli point</u> determined by X_1, X_2, and X_3 . This result will appear as a consequence of an argument to be given shortly.

The general problem with weights w_1, w_2, and w_3 appears as an exercise in Simpsons' book on "fluxions", perhaps the first text book on the calculus. Even more surprizing is the fact that the duality given in the text was discoverd for the simple Fermat case by Fasbender (ca 1850) . We shall now present his theorem with an elementary proof.

LEMMA. Let P be any point in an equilateral triangle with altitude h and side S. Let h_1, h_2, h_3 be the lengths of the perpendicolars dropped to the three sides from P. Then : $h_1 + h_2 + h_3 = h$.

PROOF.



$$\frac{1}{2} Sh = \text{Area} = \frac{1}{2} Sh_1 + \frac{1}{2} Sh_2 + \frac{1}{2} S h_3 .$$

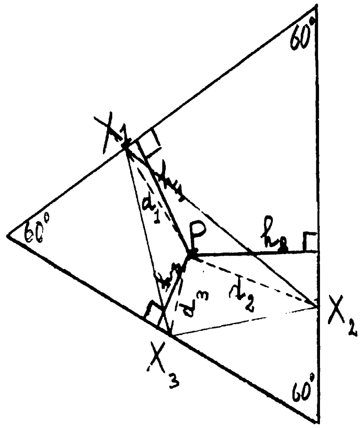
THEOREM. (Fasbender) Given any three points X_1, X_2, X_3 in the plane, let h be the altitude of any equilateral triangle circumscribing the triangle $X_1X_2X_3$. Let d_1, d_2, d_3 be the distances from any point P interior to the triangle $X_1 X_2 X_3$ to its three vertices.

Then

$$h \leqq d_1 + d_2 + d_3 .$$

(Notice that this is exactly the content of Property 1 above specialized to the Fermat case. The vectors U_1, U_2, U_3 are unit vectors perpendicular to the sides of the equilateral triangle and $h_i = U_i \cdot X_i$ for $i = 1, 2, 3$. By the Lemma, $h = h_1 + h_2 + h_3$)

PROOF.



$$h_1 \leqq d_1$$
$$h_2 \leqq d_2$$
$$h_3 \leqq d_3$$

$$h_1 + h_2 + h_3 = h \leqq d_1 + d_2 + d_3$$

COROLLARY. Equality holds if and only if PX_1, PX_2, PX_3 are perpendicular to the sides of the equilateral triangle. In this case, P minimizes the sum of distances and is the Torricelli point.

Concerning the computational algorithm presented in the text, the following information can be added. The algorithm is not new but seems to have been discovered by Weizfeld (Tohoku Journal 1936). As be proved,

$$P \neq P' \Longrightarrow F(P') < F(P).$$

Although he overlooked certain exceptional cases, covered by the qualifying clause in the following theorem, he also proved the convergence of the algorithm.

THEOREM. Consider the sequence of points $P^0, P^1, \ldots, P^k,$ $\ldots$ defined by

$$P^{k+1} = T(P^k) = \sum_i \frac{w_i}{d_i(P^k)} X_i \Big/ \sum_i \frac{w_i}{d_i(P^k)} .$$

If $P^{k+1} \neq P^k$ for all k, then $\lim_{k \to \infty} P^k = P$ and P minimizes $F(P)$.

Finally, several possible generalizations were discussed in the lecture, including non-linear costs, non-Euclidean distances, and adding additional points P_j with the role of P with each X_i associated with one P_j. Little progress has been made on any of the generalizations which have the utmost importance in applications.

CENTRO INTERNAZIONALE MATEMATICO ESTIVO

(C.I.M.E.)

THE MULTI-SECTORAL THEORY OF ECONOMIC GROWTH

Michio MORISHIMA

Corso tenuto all'Aqula dal 30 agosto al 7 settembre

1965

THE MULTI-SECTORAL THEORY OF ECONOMIC GROWTH

by

Michio Morishima (Osaka Univ. Japan)

Lecture 1

Growth Equilibrium in a Walrasian Model.

I devote this lecture to presenting the classical Walrasian model in its smallest scale and to examining it for the existence of a growth equilibrium. We assume that the economy we are going to be concerned with consists of many firms that are classified into two industries: the consumption-good industry and the capital-good industry.

It is assumed that a finite number of manufacturing processes are available to each industry. Let α_ι and λ_ι be the capital- and labour- input coefficients of the ι-th process of the consumtion-good industry ($\iota = 1,\ldots\mu$), and a_i and l_i the corresponding coefficients of the i-th process of the capital-good industry (i=1 ..., m). (We are following Professor Hicks in denoting prices and quantities referring to the consumption-good sector by Greek letters and those to the capital-good sector by the corresponding Latin letters). For the sake of simplicity we assume that all α's, λ's, a's and l's are positive and that the capital-good does not suffer wear and tear.

Let π be the price of the consumption good, p the price of the capital good, q the price of the capital service, and w the wage rate. Each process is evaluated at q and w ; each industry chooses the cheapest among possible μ or m processes; competition among the firms will make them obtain no supernormal profits. In equilibrium, we have

(I. 1) $\pi \leq q\,\alpha_\iota + w\,\lambda_\iota \qquad \iota = 1,\ldots,\mu$

(I. 2) $p \leq q\,a_i + w\,\ell_i \qquad i = 1,\ldots,m.$

Next classify citizens into two classes: capitalists and workers. Write K and L for the stock of the capital good and the number of the workers available at the point of time under discussion. When there prevails an equilibrium fully utilizing capital and labour, worker expect their income of the amount wL, while the income from owning capital is qK. The expected total income is the sum of them and is divided into consumption and savings.

Denote by γ the demand for consumption good and by s the quantity of the capital good that can be bought by spending the whole amount of the current savings. The total consumption is written $\pi\gamma$, and the total savings ps ; we have a definitional relationship,

(I. 3) $$\pi\gamma + ps \equiv qK + wL ,$$

which is usually referred to as Walras' identity.

A number of the consumption and the saving function have been proposed. We, however, follow Harrod in assuming that the real consumption γ and the real savings ps/π are constant fractions of the total real income; we have

(I. 4) $$\gamma = \varepsilon (qK + wL)/\pi$$

(I. 5) $$ps/\pi = b\,(qK + wL)/\pi$$

where the average propensity to consume ε and to save b are positive constants. As no more goods can be consumed than are produced and all savings are invested in the state of equilibrium, we have

(I. 6) $$\sum_{\iota} \xi_{\iota} \geq \gamma ,$$

(I. 7) $$\sum_{i} x_i \geq s ,$$

where ξ_{ι} is the output of the consumtion good produced by the ι-th process and x_i the output of the capital good produced by the i-th process.

The final set of the equilibrium conditions consists of the inequalities,

(I. 8) $$K \geq \sum_{\iota} \alpha_{\iota} \xi_{\iota} + \sum_{i} a_i x_i ,$$

(I. 9) $$L \geq \sum_{\iota} \lambda_{\iota} \xi_{\iota} + \sum_{i} \ell_i x_i ,$$

stating that no more factors can be used than are available in the economy. They implicitly assume that the capital stocks as well as workers are freely transferable from one firm to another . It is evident that the perfect transferability (particularly that of the capital stocks) is an unwelcome assumption, but it immensely simplifies matters. We shall use it until it is victimized in the von Neumann revolution discussed later.

We now have μ + m + 8 variables to fulfil μ + m + 6 inequalities and one identity. One of the four prices, say the price of the consumption good π, may be fixed at unity by the usual procedure of normalization; we have $p \sum x_i \leq ps$ from (I.1)-(I.6) and (I.8)-(I.9). This , together with (I.7), gives $p \sum x_i = ps$; that is, investment equal savings.

Let us non fix the real wage rate at some level w^0 , and find those values of p, q , ξ_ι , x_i γ s, K, and L which satisfy (I.1)-(I.9) with that w^0 .

We begin by elucidating the factor-price frontier that gives the correspondence between the real-wage rate and the rate of returns of the capital good, r = q/p . Let us arbitrarily pick out processes ℓ and i from among those available to the two industries. Consider the equations :

(I.10) $$1 = r\, p\, \alpha_\iota + w\, \lambda_\iota ,$$

(I.10') $$p = r\, p\, a_i + w\, \ell_i .$$

Eliminating p , we get

(I.11) $$r = \frac{1 - w\lambda_\iota}{a_i + w D} ,$$

where $D = \alpha_\iota \ell_i - \lambda_\iota a_i$. D is positive if the capital-good industry is less capital intensive than the consumption-good industry, and negative in the opposite case. Measure the real-wage rate w along the horizontal axis and the rate of returns along the vertical axis ; draw the curve (I. 11) on the plane. It is a downward sloping curve starting from $1/a_i$. and terminating at $1/\lambda_\iota$, and is convex (or concave) to the origin if D is positive (or negative). For each pair of processes, we can draw a similar curve, so that we obtain Figure 1 for $\mu = m = 2$, where we assume that $\frac{a_1}{\ell_1} < \frac{\alpha_2}{\lambda_2} < \frac{a_2}{\ell_2} < \frac{\alpha_1}{\lambda_1}$, and the activities are arranged such that $\lambda_1 \lessgtr \lambda_2$ and $a_1 \lessgtr a_2$.

We have μm possible pairs of activities. To each of them there corresponds a rate of returns given by (I. 11) ; we have, therefore, μm possible rates of returns. Among them the maximal one gives the equilibrium rate of returns corresponding the given rate of real wages. The correlation between them is traced out by the heavy broken curve in Figure 1 that is called the factor- price frontier. From it we see that if, for example, $w^{c} < w^{*}$, the consumption-good industry select the second process (α_2, λ_2) , while the capital-good industry the first activity $(a_1, \ell_1.)$. We also see that at w^{*} the capital-good industry may choose a mixture of the first and the second processes. The consumption-good industry is in a similar situation at w^{**} .

It is now clear that, given the real-wage rate w^{0} , the industries choose processes such that they yield the maximum rate of returns r^{0} ; the equilibrium prices are obtained by solving (I. 10) and (I. 10') corresponding to the processes chosen. The remaining problem is to find the equilibrium activity levels (ξ_ι , x_i^0) as well as the required capital K^0 and labour L^0 such that

(I. 12) $$\sum_\iota \alpha_\iota \xi_\iota + \sum_i a_i x_i = K \ ,$$

(I. 13) $$\sum_\iota \quad {}_\iota\xi_\iota + \sum_i \ell_i x_i = L,$$

(I. 14) $$\sum_\iota \xi_\iota = \gamma = \varepsilon \, (q K + w L) / \pi \ ,$$

(I.15) $$\sum_i x_i = s = b(q K + w L) / p.$$

In view of the Walras identity, it is seen that (I. 15) follows from the rest ; so that we concentrate our attention on (I. 12)-(I. 14).

As π^0 is set at 1 and $q^0 K^0 + w^0 L^0$ equals $\sum \xi_\iota^0$ + + $p^0 \sum x_i^0$, we have from (I. 14).

(I. 14') $$\sum \xi_\iota^0 = \varepsilon \, (\sum \xi_\iota^0 + p^0 \sum x_i^0) \ .$$

Hence $$\frac{\sum \xi_\iota^0}{\sum x_i^0} = \frac{\varepsilon}{1 - \varepsilon} \, p^0 \ .$$

It describes (like Kahn's multiplier) how the ratio of the output of the consumption good to the output of the capital good depends on the price of the capital good in terms of the consumption good.

By dividing (I. 12), (I. 13) and (I. 14') by $\sum x_i^0$, we have

(I. 16) $$\sum \alpha_\iota \eta_\iota^0 + \sum a_i z_i^0 = K^0 / (\sum x_i^0)$$

(I..17) $$\sum \lambda_\iota \eta_\iota^0 + \sum \ell_i z_i^0 = L^0 /(\sum x_i^0) \ ,$$

(I. 18) $$\eta^0 = \varepsilon \, (\eta^0 + p^0) \ ,$$

where $z_i^0 = x_i^0 / \sum x_i^0$, $\eta_\iota^0 = \xi_\iota^0 / \sum x_i^0$, $\eta^0 = \sum \eta_\iota^0$.

When a unique process pair (ζ, i) is chosen, $z_i^o = 1$ and $\eta_i^o = \eta^o = \frac{\varepsilon}{1-\varepsilon} p^o$, so that the rate of growth of the capital stock $g = \sum x^o / K^o$ is uniquely determined .

In Figure 1 we observe that r depends on w continuosly; and p and q are also shown to be continuous functions of w. Accordingly, the η determined by (I. 18) depends on w in a continuous way. This implies that so long as there corresponds a unique process pair to the real-wage rate w, the relationship between g and w is continuous. We have a continuous curve g(w) in each interval of w in which the same unique process pair in chosen. At the end of the intervals where the cheapest activity set are not unique, multiple rates of growth of the stock of capital are associated with the critical value of w, and the consecutive curves are connected with vertical segment. We thus have a staircase-like curve showing the correspondence between g and w (see Figure 2) . Celebrating Harrod , the pioneer of the modern theory of economic growth, we may call g(w) the warranted- rate - of - growth curve.

Let us now give our attention to the other blade of the scissors. We assume that the working population grows at a rate depending on the real-wage rate : the rate of growth of the labour force ρ is negative for very low level of the real - wage rate, zero for the subsistence level and then increases with a rise in the real-wage rate until it reaches a certain value, after which ρ may decrease but remains positive. It is seen that if the subsistence wage rate is lower that the technologically attainable maximum $1/\lambda_1$ of the real-wage rate, the natural - rate-of-growth curve $\rho(w)$ and the warranted-rate-of-growth curve g(w) have at least one intersection, at which the stock of capital and the labour force grow at the same rate.

Suppose $\bar{w}$ is a real-wage rate such that it generates an increase in K and L in the same proportion (say $\bar{\rho}$). Let the equilibrium values of π, p, q, ξ_1, x_i, K and L associated with $\bar{w}$ be denoted by the corresponding letters with bar, so that $\bar{\pi}, \bar{p}, \bar{q}, \bar{w}, \bar{\xi}_1, \bar{x}_i, \bar{K}$ and $\bar{L}$ satisfy the equilibrium conditions (I.1)-(I.9). We can easily see that when time goes on, a new equilibrium is established at $(\bar{\pi}, \bar{p}, \bar{q}, \bar{w}, e^{\bar{\rho}t}\bar{\xi}_1, e^{\bar{\rho}t}\bar{x}_i, e^{\bar{\rho}t}\bar{K}, e^{\bar{\rho}t}\bar{L})$. That is to say, when the real-wage rate $\bar{w}$ prevails, the stock of capital, the labour force and the outputs of the two industries grow forever at the common rate $\bar{\rho}$, and the prices remain unchanged. Such a state is referred to as a state of the long-run or the Golden Equilibrium, and $\bar{w}$ as the long-run or the golden real-wage rate.

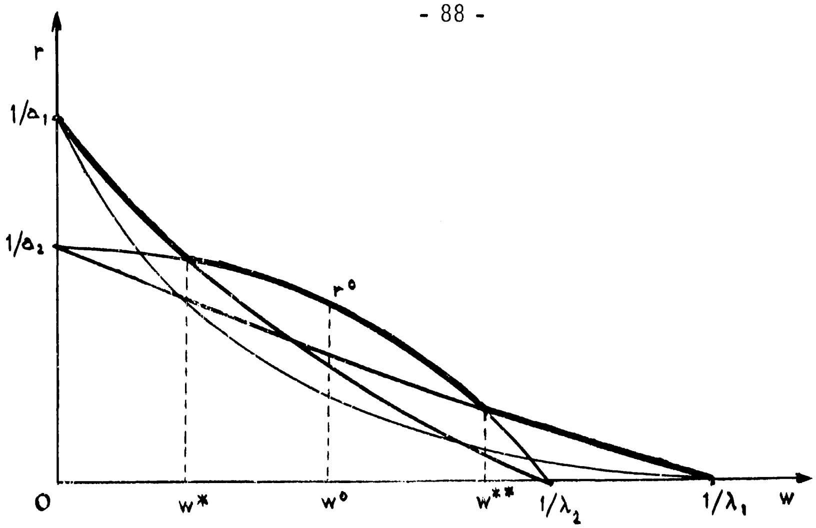


Figure 1

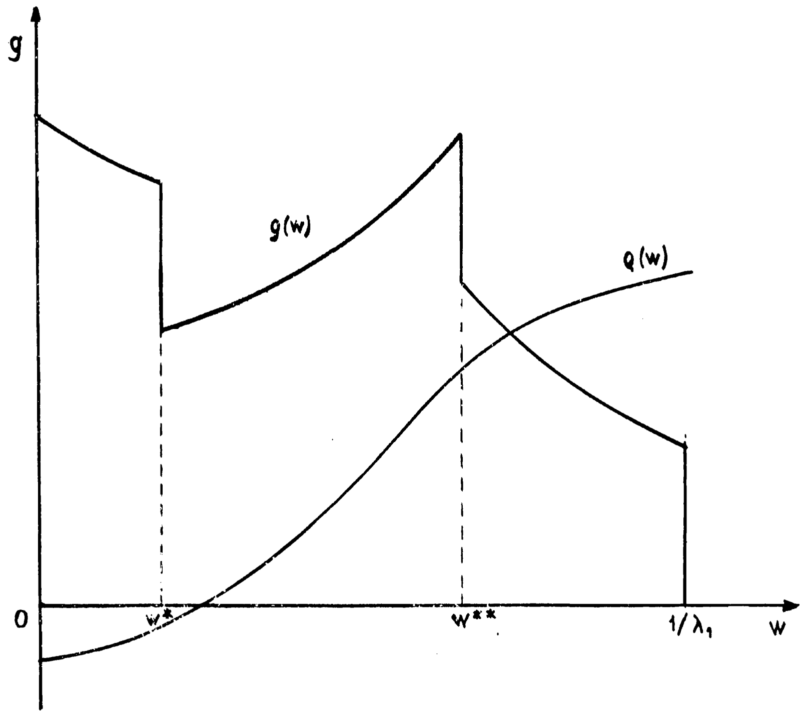


Figure 2

M. Morishima

Lecture 2

Stability of the Golden Equilibrium

We have seen that if the subsistence level of real wages is so low that it does not exceed its technologically attainable maximum, there is a solution to the system (I. 1) - (I..9). Alternatively, we can see that to any given values of K and L, say K° and L°, there corresponds a set of values $(\pi^{\circ}, p^{\circ}, q^{\circ}, w^{\circ}, \xi_i^{\circ}, x_i^{\circ})$ fulfilling inequalities (I. 1)-(I. 9). It is evident that with arbitrarily given K° and L° the solution may fulfill (I. 1)-(I. 9) with some strict inequalities; in such cases, however, the solution is required by the famous duality theorem of the activity analysis to fulfill the Profitability and Free Goods Rules. If the historically given endowments happen to equal the golden equilibrium endowments $\bar{K}$ and $\bar{L}$, we have a balanced growth of capital and labour; but they are so only by chance, so that we usually have a discrepancy between $g(w^0)$ and $\rho(w^0)$ which is to be removed by adjustments

We first deal with the normal case in which a unique set of processes and hence a unique rate of growth of the capital stock g° is associated with w°. The effects of increases in K and L at the rates g° and ρ° respectively can effectively be analysed by splitting them up into the effects of an increase (or a decrease) in K at the rate $g^{\circ} - \rho^{\circ}$, L being kept constant, and the effects of a proportional increase in K and L at the common rate ρ^{0}. It has been seen in Lecture 1 that the latter yields a proportional change in ξ_1, x_i, γ, s, but to make no effects on the relative prices. We, therefore, concentrate our attention on a solitary increase in the stock of capital, and show that it results in an increase in the real wage rate.

Suppose $g^0 = g(w^0)$ exceeds $\rho^0 = \rho(w^0)$. Suppose also that the relative prices remain constant in spite of a solitary increase in K. Constant relative prices imply that the same processes are selected as the cheapest before and after the increase in K. As each industry selects a unique process at w^0, all ξ's other that (say) ξ_γ and all x's other than (say) x_i are zero before and after the increase in K. When K increases with constant relative prices, γ and s, and hence ξ_γ and x_i that equal γ and s respectively, increase proportionately, so that a larger amount of labour is required for the production of goods. It is clear that this is incompatible with the constancy of L. Thus changes in the relative prices are necessarily induced.

It has been seen in Lecture 1 that the attainable range of the real wages $(0, 1/\lambda_1)$ may be divided into several sub-ranges, say $(0, w^*)$, (w^*, w^{**}) and $(w^{**}, 1/\lambda_1)$ in the case of Figure 2; the same processes are chosen as long as the wage rate changes within a sub-range. As the real-wage rate w^0 is taken so as to belong to the interior of one of the sub-ranges at the beginning of the course, it will still be in that interior after a (sufficiently) small increase in K. We may, therefore, assume that there is, for the time being, no change in the processes adopted.

Let (γ, i) be the process pair which is chosen. Let a small increase in K (say, $K^0 \to K^1$) give rise to changes in prices and levels of production, say $p^0 \to p^1$, $q^0 \to q^1$, $w^0 \to w^1$, $\xi_\gamma^0 \to \xi_\gamma^1$ and $x_i^0 \to x_i^1$. Note that L is taken as given; π is pegged on the level of unity by the normalization procedure.

We have two sets of equilibrium conditions before and after the increase in K :

$$\text{(II. 1)} \quad 1 = q^0 \alpha_l + w^0 \lambda_l \qquad \text{(II. 1')} \quad 1 = q^1 \alpha_l + w^1 \lambda_l$$

$$\text{(II. 2)} \quad p^0 = q^0 a_i + w^0 \ell_i \qquad \text{(II. 2')} \quad p^1 = q^1 a_i + w^1 \ell_i$$

$$\text{(II. 3)} \quad \xi_l^0 = \gamma^0 \qquad \text{(II. 3')} \quad \xi_l^1 = \gamma^1$$

$$\text{(II. 4)} \quad x_i^0 = s^0 \qquad \text{(II. 4')} \quad x_i^1 = s^1$$

$$\text{(II. 5)} \quad K^0 = \alpha_l \xi_l^0 + a_i x_i^0 \qquad \text{(II. 5')} \quad K^1 = \alpha_l \xi_l^1 + a_i x_i^1$$

$$\text{(II, 6)} \quad L^0 = \lambda_l \xi_l^0 + \ell_i x_i^0 \qquad \text{(II. 6')} \quad L^0 = \lambda_l \xi_l^1 + \ell_i x_i^1 .$$

(Note : We have $0 < w^0 < 1/\lambda_1$ = the technologically attainable maximum of w . It follows from $w^0 > 0$ that labour is not free. It also follows from $w^0 < 1/\lambda_1$ that q^0 is positive ; hence the capital stock is not free. Similarly , the consumption good and the capital good cannot be free, for $\pi^0 = 1$ and $p^0 > 0$. We thus find that (I. 6)-(I. 9) hold with equality.) Multiplying (II. 1) and (II. 2) by ξ_l^1 and x_i^1 respectively, and viewing (II. 6')-(II. 9'), we obtain

$$\text{(II.7)} \qquad \gamma^1 + p^0 s^1 = q^0 K^1 + w^0 L^0.$$

Similarly,

$$\text{(II. 8)} \qquad \gamma^1 + p^1 s^0 = q^1 K^0 + w^1 L^0$$

On the other hand, Walras' identity implies

$$\text{(II. 9)} \qquad \gamma^0 + p^0 s^0 = q^1 K^0 + w^1 L^0$$

$$\text{(II. 10)} \qquad \gamma^1 + p^1 s^1 = q^1 K^1 + w^1 L^0$$

Substracting (II. 7) from (II;9) , and (II. 8) from (II. 10), and adding, we finally obtain

(II.11) $$(q^1 - q^0)(K^1 - K^0) = (p^1 - p^0)(s^1 - s^0).$$

Let us now consider a third situation where capital and labour are available in the old equilibrium but prices are fixed at their new equilibrium values. Write

$$\gamma^2 = \varepsilon(q^1 K^0 + w^1 L^0), \quad s^2 = b\,\frac{q^1 K^0 + w^1 L^0}{p^1}.$$

In view of (II.8), we find that

(II.12) $$\gamma^0 + p^1 s^0 = \gamma^2 + p^1 s^2 .$$

Let us now show that if (II.12) holds, then

(II.13) $$\gamma^0 + p^0 s^0 < \gamma^2 + p^0 s^2 .$$

As $K^1 \neq K^0$, it is seen from (II.3 - 6) and (II. 3' - 6') That if $\gamma^0 > \gamma^1$ then $s^0 < s^1$ and vice versa ; and this implies $p^0 \neq p^1$. We can now show that the inequality

(II.14) $$\gamma^0 + p^0 s^0 \geqq \gamma^2 + p^0 s^2$$

contradicts (II.12) and hence we obtain (II.13). This is because (II.12) and (II.14) may be put in the following respective forms :

(II.12') $$\left(\varepsilon + \frac{p^1}{p^0} b\right)(q^0 K^0 + w^0 L^0) = (\varepsilon + b)(q^1 K^0 + w^1 L^0)$$

(II.14') $$(\varepsilon + b)(q^0 K^0 + w^0 L^0) \geqq \left(\varepsilon + \frac{p^0}{p^1} b\right)(q^1 K^0 + w^1 L^0),$$

from which we get

$$\left(\varepsilon + \frac{p^1}{p^0} b\right)\left(\varepsilon + \frac{p^0}{p^1} b\right) \leqq (\varepsilon + b)^2 ;$$

but the inequality sign is found to be in the wrong direction as $p^1 \neq p^0$.

This fact implies that the consumption and savings functions of the Harrodian type satisfy Samuelson's axiom of revealed preference. Subtracting (II.13) from (II.12), we get the familiar inequality

$$(p^1 - p^0)(s^2 - s^0) < 0 .$$

On the other hand, we at once have

$$s^1 - s^2 = b\, r^1 (K^1 - K^0),$$

where $r^1 = q^1/p^1$. Hence we may put (II.11) in the form

(II.15) $(q^1 - q^0)(K^1 - K^0) = b\, r^1 (p^1 - p^0)(K^1 - K^0) + (p^1 - p^0)(s^2 - s^0)$.

Since

$$q^1 - q^0 = -(w^1 - w^0)\,\lambda_1/\sigma_1 \quad \text{from (II.1) and (II.1')},$$

$$p^1 - p^0 = (w^1 - w^0)(\ell_i - a_i\lambda_1/\sigma_1) \quad \text{from (II.2) and (II.2')},$$

(II.15) can further be rewritten as follows :

$$\left[b\, r^1 \ell_i + (1 - b\, r^1 a_i)\,\lambda_1/\sigma_1 \right](w - w^1)(K^1 - K^0)$$
$$= (p^1 - p^0)(s^2 - s^0) < 0.$$

As $0<b<1$ and $0<r^1 a_i<1$, we have $1 - br^1 a_i > 0$. Hence we finally find that if $K^1 > K^0$, then $w^1 > w^0$.

We have so far assumed that a unique set of processes is associated with the real-wage rate w^0. What happens when multiple sets are associated with it, that is , when w^0 is a singular rate, say w^*, w^{**}, or $1/\lambda_1$ in Figure 2 ?

As before, changes in K and L at the rates $g(w^*)$ and $\rho(w^*)$ are examined by splitting them up into a proportional change in K and L at the rate $\rho(w^*)$ and a solitary change in K at the rate $g(w^*) - \rho(w^*)$. Since w^* is singular, either the consumption-good industry or the capital-good industry or both have at least two processes which they select as the cheapest among those available to them when the prices, p^*, q^*, w^*, prevail. These processes are different in the capital intensity. When there is a solitary increase in K, the industries will shift their production schedules so as to use more capital-intensive processes at greater rates that before. It can be shown that a solitary increase in K with the constant relative prices results in a decrease in the growth rate of the capital stock, because

$$g = s/K = b(q^*K + w^*L^*)/p^*K .$$

decreases when K increases.

If the rate of growth of the capital stock is still, after its fall, larger than $\rho(w^*)$, the whole adjustment process is repeated. The rate of growth g will finally rest at the level $\rho(w^*)$ if the staircase-like curve $g(w)$ and the natural-rate-of-growth curve $\rho(w)$ have an intersection in the vertical segment (corresponding to w^*) of the former; or the growth rate finally reaches the minimum rate of growth g_* associated with w^*. At $g = g_*$, a further solitary increase in K necessarily affects the relative prices; and the real wage will be increased. The economy departs from the singular point w^*, and gets into a normal range.

A similar argument can be made for the converse case where the actual rate of growth $g(w^*)$ is less than both the maximum rate of growth g^* associated with w^* and the rate of growth of the labour fource. The economy will climb up the vertical segment of the staircase from $g(w^*)$ to g^* (or to $\rho(w^*)$ if $\rho(w)$ and $g(w)$ have an intersection at w^*).

We are now in a position to be able to say definitely about the stability of the growth equilibrium.

We can easily observe from Figure 2 that even if we start from any rate of real wages, the economy finally settles on the path of the equilibrium growth at the rate $\rho(\bar{w})$ associated with the long-run equilibrium rate of real wages. Thus, if the long-run or golden equilibrium is unique, it has the global stability. When there are a number of the golden equilibria, a series of the short-run equilibria starting from any initial point eventually approaches one of the golden equilibria. (See Figure 3.) We find that the long-run equilibria corresponding to the lowest and the highest real-wage rates have the local stability. We also see that our system satisfies Samuelson's separation theorem stating that the consecutive equilibrium points $A_1, A_2, A_3, \ldots$ are alternatively stable and unstable.

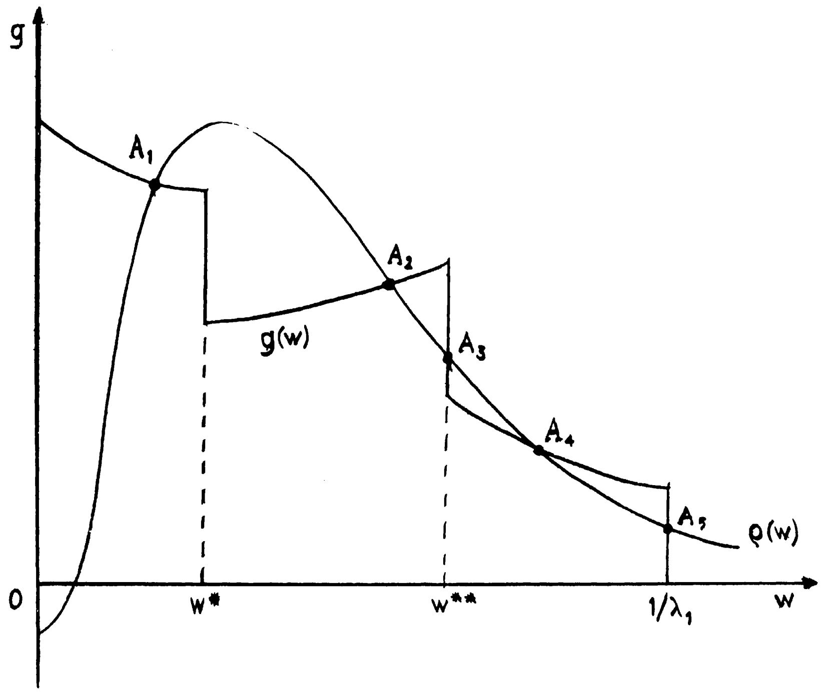


Figure 3.

Lecture 3

The von Neumann Revolution

In the previous model it is assumed that the consumption-good industry only produces the consumption good and the capital-good industry only the capital goods. At first sight, it looks very natural and admissible ; much simplicity is acquired by the assumption of no joint products, but a decisively important advancement in the theory of capital cannot be achieved so long as one refrains from confronting the "joint production" trouble.

Lecture 1 and 2 are based on an unrealistic assumption that all capital goods do not suffer wear and tear. The formal characters of the model will, however, be preserved if we may replace that assumption by the following more admissible one : The capital goods of old vintages (i.e. the capital goods produced several years ago, and hence now at a later stage of wear and tear) are physically equivalent to some less amounts of the brand new capital goods of the same kind. This is a useful assumption , but it still over-simplifies the age construction of the available endowments, so that the mortality of the capital goods cannot be treated well.

Only by treating capital goods of different vintages (hence at different stages of wear and tear) as qualitatively different goods, can we deal well with the age construction of capital equipment ; and those "different" capital goods appearing simultaneously with the products should be treated as the by-products of the manufacturing process in question.

On the von Neumann convention a process of production that uses an outfit of capital equipment is described as an operation that converts one bundle of goods -- the bundle of "inputs" including capital equipment left over from the preceeding period -- into another bundle of goods -- the bundle of "outputs" including qualitatively different capital equipment left over the following period. As long as we use some capital goods which have some years left to live, there must appear at least two goods in the list of "outputs" : ordinary products and "deformed" capital goods ; the process of production should, therefore, inevitably be multi-product, even if the output in the ordinary sense is produced without any by-product.

This treatment of capital goods enables us to get rid of the assumption of perfect transferability of capital goods, an unwelcome parasite in the previous "no joint production" model. One of the most important properties in which a brand-new fixed capital good differs from an old capital good of the same kind is, apart from the difference in productivity, its transferability. A new lathe, for example, will be sold to any factory which demands it, while a lathe which has already been set up in a shipyard will not usually be transferred to an aircraft factory, even though the former is over-equipped and has many lathes not used, while the latter operates at its full capacity. This asymmetry may be paraphrased the "<u>ex-ante</u> mobility <u>ex-post</u> immobility" of fixed capital goods, [1/] and any capital theory neglecting that important fact is far from being credited with ability to explain the reality.

Once we have been prepared to permit joint production and to treat fixed capital goods of different vintages as different goods, we shall also get ready to accept the idea of treating fixed capital goods installed in different factories as different goods. We can then rule out the transference of used fixed capital goods between factories by specifying the technology such that to no other factories than Factory A are production processes available to use those capital goods fixed in A. Each good is still transferable, but its transferability is limited within the "sector" to which it belongs. Machines installed in different factories are "different" goods belonging to different sectors. The same second-hand machines ("same" in the ordinary sense) may have different prices ; in particular, they become free goods in those sectors which are over-equipped with them. It should be noticed that such a "maldistribution" of capital goods may happen even in the Golden Growth Equilibrium discussed below.

Let us consider an economy where are a finite number m of different processes, each converting a bundle of n commodities into a different bundle. Let a_{ij} be the quantity of good j technically required per unit of process i, b_{ij} the quantity of good j produced per unit of process i, and ℓ_i the number of workers employed per unit of process i. Define matrices A and B and a column vector L as :[2/]

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

$$B = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \cdot & \cdot & \dots & \cdot \\ b_{m1} & b_{m2} & \dots & b_{mn} \end{pmatrix},$$

$$L = \{\ell_1, \ell_2, \dots, \ell_m\} .$$

We assume that

(IIIa) for each $j = 1, \dots, n$ there is at least one positive b_{ij}

(IIIb) all ℓ_i are positive .

The former implies that there is no good which cannot be produced, while the latter implies that labour is an indispensable factor of production for each process.

Let us now turn from technology to consumer's choice. As has been pointed out by Champernowne, the original von Neumann model of economic expansion is a slave-economy whose object is mere enlargement of production. The real-wage rate, on one hand, is fixed at the subsistence level such that workers can only but the minimum amounts of goods biologically required to subsist ; the whole capitalists' income, on the other hand, is accumulated for expansion. There is no room for consumer's choice.

A proper attention to consumer's choice has first been made in Chapter V of my earlier book. I have assumed that all workers have identical tastes as all capitalists do. There are only two different consumers in the economy, one of whom (the Worker) does not save at all, but the other (the Capitalist) saves a constant proportion of his income, the rest being spent for consumption 1/. Let d_j be the Capitalist's consumption of good j. Let U be his utility function, which is assumed to be strictly quasi-concave and quasi-homogeneous in its arguments. The quasi-concavity is the usual property implying that the marginal rates of substitution between goods decrease in all directions; while the quasi-homogeneity is rather unconventional and implies that a proportional change in the quantities of goods does not give rise to any change in the order of preferences. It is a restrictive assumption, for it makes unity of the Engel-elasticities of all goods (i.e., of the elasticities of demand with respect to income). But the proportionality between consumption and income is one of the necessary conditions for a balanced growth of the economy.

A Capitalist consumes only a constant proportion of his income; his budget constraint is written

$$\sum_j P_j d_j = cE ,$$

where c is his propensity to consume (which is a positive constant), E his income (profits) and P_j the price of good j. Maximizing his utility

$$u = U(d_1, d_2, \ldots, d_n)$$

subject to the budget constraint, we obtain the familiar conditions of a maximum that the marginal rates of substitution between goods should equal the rates of their prices, from which we can derive the demand functions

$$d_j = D_j(P_1, P_2, \ldots, P_n, E) \quad j = 1, \ldots, n .$$

We can show that they are homogeneous of degree zero in their arguments $P_1, \ldots, P_n$, and E, so that the multiplication of them by any positive number, say $1/(\sum_j P_j)$, makes the quantities demanded unaffected. Furthermore the elasticity of d_j with respect to E is unity; hence we may write

$$(*) \qquad d_j = f_j(y_1, y_2, \ldots, y_n) \frac{E}{\sum_j P_j} \qquad j = 1, \ldots, n,$$

where y_h is the normalized price $P_h/(\sum_j P_j)$ of good h. As these demand functions must fulfill the budget constraint, we have

$$\text{(III c)} \qquad \sum_j f_j(y_1, y_2, \ldots, y_n) y_j = c \qquad \text{for all } y_1, \ldots, y_n.$$

We assume that the Worker does not save at all ; but this does not necessarily imply that his wage is fixed at the subsistence level; in fact , it implies that he devotes all his income to consumption , however large it may be. Once the wage rate becomes greater than the minimum level for subsistence, there opens a possibility for the Worker to choose among consumption goods.

Let e_j be Worker's demand for good j, and W his wages. We make for the Worker, as we did for the Capitalist, the assumption of quasi-homogeneity (in addition to the assumption of strict quasi-concavity) to the effect that the Engel-elasticity of the Worker's consumption is unity. We may, therefore, write

$$(**) \qquad e_j = g_j(y_1, y_2, \ldots, y_n) \frac{W}{\sum_j P_j}, \qquad j = 1, \ldots, n,$$

(IIId) $\sum_j g_j(y_1, y_2, \dots, y_n)\, y_j = 1$ for all $y_1, \dots, y_n$;

because $e_1, \dots, e_n$ must satisfy the budget constraint

$$\sum_j e_j \; P_j = W .$$

For the sake of mathematical simplicity we shall also use the following assumption :

(IIIe) All Engel-coefficients $f_j(y_1, y_2, \dots, y_n)$ and $g_j(y_1, y_2, \dots, y_n)$ ($j = 1, \dots, n$) are non-negative, finite and continuous at every non-negative set of normalized prices $(y_1, y_2, \dots, y_n)$.

Let $r(t)$ be the rate of interest in period t, $P_j(t)$ the price of good j in period t, and $w(t)$ the money rate of good j in period t, and $w(t)$ the money rate of wages in period t . We define $\beta(t)$ as $1 + r(t)$ and call it the interest factor. As the period of production is assumed to be one unit of time, and wages are paid in advance , we have, in equilibrium,

(III. 1) $\sum_j b_{ij} \; P_j(t + 1) \leqq \beta(t)\left(\sum_j a_{ij} \; P_j(t) \; + \ell_i \, w(t)\right)$ $i = 1, \dots, m.$

or

$$BP(t + 1) \leqq \beta(t)\,(AP(t) \; + Lw(t)) \; ,$$

where $P(t)$ is an n-dimensional column vector $\left\{P_1(t), \dots, \; P_n(t)\right\}$
(1) means that :

Total cost (including interest on working and fixed capital) of any process $\geqq$ Total receipts from that process.

Next let $q_i(t)$ be the intensity at which process i operates in period t. As an unprofitable process i is not used, the intensity of that process i which fulfills (III. 1) with strict inequality should be zero. It is obvious that for the other processes, $q_i(t)$'s are non-negative; we have , therefore,

$$\text{(III. 2)} \quad \sum_i{}' \sum_j b_{ij} P_j(t+1) q_i(t) = \beta(t) \Big[\sum_i \sum_j a_{ij} P_j(t) \quad q_i(t) + \big(\sum_i \ell_i q_i(t)\big)\, w(t) \Big];$$

in other words,

$$q(t)BP(t+1) = \beta(t)(q(t)\,AP(t) + q(t)Lw(t)) \ ,$$

where $q(t)$ is an m-dimensional row vector $\left[q_1(t), \ \ldots, \ q_m(t) \right]$.

Let us now turn from prices to "demand-supply" (or "input-output") relations . It is evident that no more goods can be consumed in any period than are available in that period. The amounts of goods that are available and "supplied" in period t are "outputs" produced in the previous period t - 1; while the "demand" consists of the "inputs" in the production process and the personal consumption by the Worker and the Capitalist.

We have :

$$\text{(III. 3)} \qquad \sum_i b_{ij} q_i(t-1) \geqq \sum_i{}' a_{ij} q_i(t) + d_j(t) + e_j(t) \qquad j = 1, \ldots, n,$$

or

$$q(t-1)\, B \geqq q(t)A + d(t) + e(t) \ ,$$

where $d(t)$ and $e(t)$ are n-dimensional row vectors $\left[d_1(t), \ \ldots, d_n(t) \right]$ and $\left[e_1(t), \ \ldots, \ e_n(t) \right]$. And the "free-good rule" discussed above ensures that at equilibrium prices $P_1(t), \ldots, P_n(t)$ will hold the equation

(III. 4) $\sum_i \sum_j b_{ij} q_i(t-1)P_j(t) = \sum_i \sum_j a_{ij} q_i(t)P_j(t) + \sum_j d_j(t)P_j(t) + \sum_j e_j(t)P_j(t),$

which may more conveniently be put in the matrix form,

$$q(t-1)BP(t) = q(t)AP(t) + d(t)P(t) + e(t)P(t) .$$

It is evident that there are no outputs, no inputs, no wages and no profits without productive activities ; the conditions (III. 3) and (III. 4) are thus met when $q_i(t-1) = q_i(t) = 0$ for all $i = 1, \ldots, m$. But such an obviously meaningless state should be put outside our scope. We make, therefore, an additional condition that the total value of all goods produced be positive :

(III. 5) $$\sum_i \sum_j b_{ij} q_i(t-1)P_j(t) > 0,$$

or

$$q(t-1)BP(t) > 0.$$

The final condition describes the equilibrium in the labour market. In the original von Neumann model it is assumed that the supply of labour can be expanded indefinitely at the subsistence level of real wages so that it completely ignores the problem of deficiency of labour, one of the most serious obstacles to a rapid growth. In fact, it is a defect of von Neumann's theory of growth that no attention is paid to Harrod's observation that the natural rate of growth sets a limit to the maximum average value of the actual rate of growth over a long period.

Let ρ_t be the growth factor (1 + the rate of growth) of the labour force in period t, and N the number of workers in period 0 ; then the supply of labour in period t will be $\rho_{t-1} \cdots \rho_1 \rho_0 N$. Since the demand for labour in period t is $\sum_i \ell_i q_i(t)$, the demand-supply balance is described by

$$\sum_i \ell_i \; q_i(t) = \rho_{t-1} \cdots \rho_1 \rho_0 \; N .$$

It would be realistic to assume that the "natural rate of growth" $\rho - 1$ depends on the real-wage rate that to assume its constancy. But it will help the reader's comprehension to begin with examining the simpler case.

When ρ is constant, we have

$$\sum_i \ell_i \; q_i(t) = \rho^t N \qquad \text{or} \qquad q(t)L = \rho^t N.$$

Let us concentrate our attention on a state of balanced growth, where prices, wage rate and interest rate are constant over time, and the intensities of production grow at a constant geometric rate. We have

$$\alpha q_i(t-1) = q_i(t), \; P_j(t-1) = P_j(t), \qquad w(t-1) = w(t), \; \beta(t-1) = \beta(t),$$

$$(i = 1, \ldots, m; \; j = 1, \ldots, n) \text{ for all } t ,$$

where α is $1 +$ the rate of balanced growth.

In the state of balanced growth we may delete letters $t - 1$, t, etc. denoting dates. Dividing (III. 1) and (III. 2) by $\sum_j P_j(t)$ and $\sum_i q_i(t)$, we have

$$\text{(III. 1')} \qquad \sum_j b_{ij} y_j \leqq \beta \Big(\sum_j a_{ij} y_j + \ell_i \Omega \Big) \qquad i = 1, \ldots, m,$$

$$\text{(III. 2')} \qquad \sum_i \sum_j b_{ij} x_i y_j = \beta \Big(\sum_i \sum_j a_{ij} x_i y_j + \big(\sum_i \ell_i x_i \big) \Omega \Big) ,$$

where x_i is the normalized intensity, y_j the normalized price, and Ω the real-wage rate , $w / \sum_j P_j$. Similarly, divide

(III. 3) , (III, 4), (III.5) by $\sum_i q_i(t-1)$ and $\sum_j P_j(t)$.

In view of ($*$) and ($**$) we may put (III. 3)-(III. 5) into the following respective forms,

$$\text{(II. 3')} \quad \sum_i b_{ij} x_i \geqq \alpha \sum_i a_{ij} x_i + (\beta - 1)\Big[\sum_i \sum_k a_{ik} x_i y_k + \big(\sum_i \ell_i x_i\big)\Omega\Big] f_j(y)$$
$$+ \alpha\big(\sum_i \ell_i x_i\big)\Omega g_j(y) \quad j=1, \ldots, n,$$

$$\text{(III. 4')} \quad \sum_i \sum_j b_{ij} x_i y_j = \alpha \sum_i \sum_j a_{ij} x_i y_j + (\beta - 1)\Big[\sum_i \sum_j a_{ij} x_i y_j + \big(\sum_i \ell_i x_i\big)\Omega\Big] c$$
$$+ \alpha\big(\sum_i \ell_i x_i\big)\Omega ,$$

$$\text{(III. 5')} \quad \sum_i \sum_j b_{ij} x_i y_j > 0 ,$$

because $E(t) = (\beta(t-1)-1)\Big[\sum_i \sum_j a_{ij} q_i(t-1) P_j(t-1) + \sum \ell_i q_i(t-1)\, w(t-1)\Big]$,

$$W(t) = \sum_i \ell_i q_i(t))\ w(t) .$$

Finally (III. 6) is equivalent to the following two equations :

$$\text{(III. 6')} \quad \alpha = \rho ,$$

and

$$\big(\sum_i \ell_i x_i\big)\xi = N ,$$

where ξ (a scalar) is the absolute level of the intensities in period 0 ; once x_i's are determined, ξ is trivially obtained. In the following, therefore, our attention is focused on the inequalities (III. 1')-(III. 6'), which may more simply be written in terms of matrices and vectors as :

(III. 1') $$By \leqq \beta(Ay + \Omega L) ,$$

(III. 2') $$xBy = \beta x(Ay + \Omega L) ,$$

(III. 3') $$xB \geqq \alpha((xA + x\Omega Lg(y)) + (\beta - 1)(xAy + x\Omega L)f(y) ,$$

(III. 4') $$xB_y = (\alpha + (\beta - 1)c)(xAy + x\Omega L) ,$$

(III.5') $$xBy > 0 ,$$

(III. 6') $$\alpha = \rho .$$

It is evident that if the labour force grows at a very high rate, it is impossible to provide all workers in the economy with opportunities to work ; unemployment of labour is inevitable, so that the condition (III. 6') for full employment cannot be satisfied. In order to establish the existence of a solution $(x_1, \dots x_m, y_1, \dots, y_n, \alpha, \beta, \Omega)$ to (III. 1')-(III. 6'), we need not only the conditions (III a, b) on technology, and the conditions (III c, d, e) on Engel-coefficients, but also some condition restricting the value of the rate of growth of the labour force. We assume :

(III .f) the natural rate of growth is so low that there is a non-negative set $(x_1, x_2, \dots, x_m)$ such that for all goods $j=1, \dots, n$

$$\sum_i b_{ij} x_i > ((\rho - c)/s)(\sum_i a_{ij} x_i) .$$

FOOTNOTES

1. In my <u>Equilibrium, Stability, and Growth</u>, it has been assumed that when capitalists' incomes are non-positive, their consumption is zero and when it is positive, it is proportional to their income . In this lecture, however, we assume that the Capitalist always consumes a constant proportion of his income even if it is negative. By doing so, formulas are greatly simplified. But when it is necessary,

proper attention is paid to, switching the Capitalist's consumption pattern at the zero income level.

Lecture 4

Existence of Golden Equilibrium

In order to prove the existence of solutions to (III. 1') - (III. 6'), given Assumptions(III a)-(III f), we use the following notation :

$$M(y) = \frac{1}{\rho} B - A - \frac{1}{\rho} (\beta-1) Ay f(y),$$

$$N(y) = Lg(y) + \frac{1}{\rho} (\beta-1) Lf(y) .$$

As I have shown in my earlier book , and many authors have also observed, there is a relationship between the rate of interest (β -1) and the rate of growth. Let s be the Capitalist's average propensity to save ; then s=1-c . At a state of fulfilling (III. 1') - (III. 6') (if it exists) , we must have

$$\rho - 1 = (\beta - 1)s ,$$

(that is, the equilibrium rate of interest must equal the rate of growth divided by the Capitalist's propensity to save). (This immediately follows from (III. 2'), (III. 4') and (III. 6')) . As s and ρ are constants exogenously given, the rate of interest at the state of balanced growth is uniquely determined, independent of the intensities, prices and wages. Hence M(y) and N(y) may be considered as depending on y only.

Taking the relationship between ρ , β and s into account, and in view of Assumptions (III c, d), we can rewrite (III. 1') and (III. 3') in the following simple forms :

(IV. 1) $$M(y)y \leqq \Omega N(y)y ,$$

(IV. 3) $$xM(y) \geqq \Omega xN(y) .$$

These inequalities can be examined from the game-theoretic point of view. Let y^* be an arbitrary set of normalized prices (i.e. a non-negative vector with unit sum). Consider the following set of inequalities :

(IV. 1*) $$M(y^*)y \leqq \Omega N(y^*)y \ ,$$

(IV. 3*) $$xM(y^*) \geqq \Omega xN(y^*) \ .$$

The matrices M and N now assume constant values, and we can interpret (IV. 1*) (IV. 3*) as a game between two persons : the "entrepreneur" (the maximizer) and the "market" (the minimizer).

Given prices and the real-wage rate, the pay-off matrix $M(y^*) - \Omega N(y^*)$ is determined.

It is obvious that if the wage rate Ω is set too low, the game is in favour of the entrepreneur, and vice versa. Accordingly there is a finite wage rate at which the game is fair ; that is, neither player can expect positive gains, so that the value of the game is zero. (IV. 1*) and (IV. 3*) imply that the wage rate should be "fair".

The existence of a fair wage-rate and the fairness of the wage-rate determined by (IV. 1*) and (IV. 3*) are proved as follows. As the natural rate of growth is positive ($\rho > 1$) and $c + s = 1$ by definition, it is at once seen that $(\rho - c)/s$ is greater than ρ. In view of Assumption (III f), we find that there is a non-negative intensity set $(x_1, x_2, \ldots, x_m)$ such that

$$\sum_i (b_{ij} - \rho a_{ij})\, x_i > 0 \qquad j = 1, \ldots, n.$$

It also follows from (III. b) and (IIIc, d) that every row of $N(y^*)$ has at least one positive entry for all y^*. Therefore, we may choose x's so that for all $j = 1, \ldots, n$,

$$\sum_i [m_{ij}(y^*) - \Omega n_{ij}(y^*)]\, x_i > 0 \quad \text{if } \Omega \text{ is negative and very small,}$$

and choose y's so that for all $i = 1, \ldots, m$,

$$\sum_j [m_{ij}(y^*) - \Omega n_{ij}(y^*)]\, y_j < 0 \quad \text{if } \Omega \text{ is positive and very large.}$$

Thus, for very small negative Ω the entrepreneur may always expect a positive gain irrespective of the choice of prices of the market, while for very large positive Ω the latter may always expect a negative gain irrespective of the former's choice of activities, the value of the game (the expected pay-off) is in favour of the entrepreneur or the market respectively, according as Ω is negative and sufficiently small or positive and sufficiently large. Hence there is a fair wage-rate Ω at which the value of the game is zero.

Next suppose the wage rate were unfair, so that the value of the game were not zero; we would have

$$v = \max_x \min_y \sum_i \sum_j [m_{ij}(y^*) - \Omega n_{ij}(y^*)]\, x_i y_j \neq 0,$$

where $x_i \geqq 0$, $\sum_i x_i = 1$, $y_j \geqq 0$, and $\sum_j y_j = 1$. First, assume $v>0$.

Then there is an $x^o = (x_1^o, \ldots, x_m^o)$ such that

$$\sum_i [m_{ij}(y^*) - \Omega n_{ij}(y^*)]\, x_i^o > 0 \quad (j = 1, \ldots, n) .$$

These strict inequalities imply that there is no $y = (y_1, \ldots, y_n)$ fulfilling (IV. 1*) . To show this, suppose $y^o = (y_1^o, \ldots, y_n^o)$ satisfies (IV. 1*) ; as $y_j^o \geqq 0$ and $\sum_j y_j^o = 1$, we have from the above inequalities

$$\sum_i \sum_j [m_{ij}(y^*) - \Omega n_{ij}(y^*)]\, x_i^o y_i^o > 0 .$$

On the other hand, as $x_i^o \geqq 0$ and $\sum_i x_i^o = 1$, we see from (IV. 1*) that the bilinear form takes on a non-positive value ; this is a contradiction . Similarly for the case with $v < 0$. Hence the wage rate fulfilling (IV. 1*) and (IV. 3*) should be fair .

Let us now define $T(y^*)$ as a set of all n-dimensional non-negative vectors y with unit sums which satisfy (IV. 1*) at a fair wage rate Ω . To any preassigned y^* there corresponds a set $T(y^*)$. We can show that this correspondence fulfills the requirements of Kakutani's fixed-point theorem :

Let S be a non-empty , bounded, closed, convex subset in a Euclidean space. If a multi-valued correspondence $z \to R(z)$ from S to S is upper semicontinuous and $R(z)$ is non-empty and convex for all $z \in S$, then there is a point $\bar{z}$ such that $\bar{z} \in R(\bar{z})$.

Now let S be

$$\{y \mid y_j \geqq 0, \sum_j y_j = 1\} .$$

To any preassigned y^* there corresponds a fair wate-rate Ω , with which is always associated a non-empty set $T(y^*)$; by construction $y^* \to T(y^*)$ is a multi-valued (possibily a single-valued) correspondence from S to S, ($T(y^*)$ is a subset of S).

The upper semicontinuity of the corrispondence is verified as follows. Let y^{*k} and y^k be any sequences of points in S that converge to y^* and y respectively. Corresponding to each y^{*k} , we have the fair wage rate Ω^k and the "optimum-price" set $T(y^{*k})$. Suppose $y^k \in T(y^{*k})$; then

$$\sum_j m_{ij} (y^{*k}) y_j^k \leqq \Omega^k \sum_j n_{ij}(y^{*k}) y_j^k \qquad (i = 1, \dots, m)$$

for every k. The fairness of Ω^k implies that with each Ω^k

is associated a non-negative intensity vector x^k (with unit sum) such that

$$\sum m_{ij}(y^{*k})x_i^k \geqq \Omega^k \sum n_{ij}(y^{*k})x_i^k \qquad (j=1, \ldots, n).$$

We can show that Ω^k is bounded and does not tend to infinity when k does. Let Ω and x be limit points of $\{\Omega^k\}$ and $\{x^k\}$ respectively, when k tends to infinity. It is obvious that the above inequalities holding for every k must hold in the limit. As we have from (III e)

$$m_{ij}(y^*) = \lim_{k\to\infty} m_{ij}(y^{*k}), \quad \text{and} \quad n_{ij}(y^*) = \lim_{k\to\infty} n_{ij}(y^{*k}),$$

we get

(IV, 1*) $$\sum_j m_{ij}(y^*)y_j \leqq \Omega \sum_j n_{ij}(y^*)y_j \qquad (i = 1, \ldots, m),$$

(IV. 3*) $$\sum_i m_{ij}(y^*)x_i \geqq \Omega \sum_i n_{ij}(y^*)x_i \qquad (j = 1, \ldots, n).$$

These show the fairness of the limiting wage rate Ω. Thus y satisfies (IV. 1*) at a fair wage rate ; hence $y \in T(y^*)$. This proves the upper semicontinuity of the correspondence $y^* \to T(y^*)$.

Next suppose there are two price-systems y^0 and y^1 that belong to $T(y^*)$. Let the fair wage rate associated with them be Ω^0 and Ω^1 respectively. Without loss of generality we may assume

$$\Omega^0 \geqq \Omega^1 .$$

It is clear that we have the following three sets of inequalities :

$$\sum_j m_{ij}(y^*)y_j^0 \leqq \Omega^0 \sum_j n_{ij}(y^*)y_j^0 \qquad (i=1, \ldots, m),$$

$$\sum_j m_{ij}(y^*)y_j^1 \leqq \Omega^1 \sum_j n_{ij}(y^*)y_j^1 \qquad (i=1, \ldots, m),$$

$$\sum_i m_{ij}(y^*)x_i^0 \geqq \Omega^0 \sum_i n_{ij}(y^*)x_i^0 \qquad (j=1, \ldots, n) .$$

As Ω^0 does not fall short of Ω^1, we have, a fortiori,

$$\sum_j m_{ij}(y^*)y_j^1 \leqq \Omega^0 \sum_j n_{ij}(y^*)y_j^1 \qquad (i=1, \ldots, m) .$$

Hence, for all μ such that $0 \leqq \mu \leqq 1$, we have

$$\sum_j m_{ij}(y^*)((1-\mu)y_j^0 + \mu y_j^1) \leqq \Omega^0 \sum_j n_{ij}(y^*)((1-\mu)y_j^0 + \mu y_j^1) \quad (i= = 1, \ldots, m) .$$

These m inequalities for y , together with the above n inequalities for x , show that all convex combinations of y^0 and y^1 belong to the optimum set $T(y^*)$. Hence it is a convex set .

We now find that the existence of a point such that $\bar{y} \in T(\bar{y})$ is ensured by the Kakutani theorem , so that

$$\sum_j m_{ij}(\bar{y})\bar{y}_j \leqq \bar{\Omega} \sum_j n_{ij}(\bar{y})\bar{y}_j \qquad (i=1, \ldots, m) ,$$

where $\bar{\Omega}$ is a fair wage rate associated with $\bar{y}$. The fairness of $\bar{\Omega}$ implies the existence of a non-negative, non-zero intensity vector $\bar{x}$ such that

$$\sum_j m_{ij}(\bar{y})\bar{x}_i \geqq \bar{\Omega} \sum_j n_{ij}(\bar{y})\bar{x}_i \qquad j = 1, \ldots, n .$$

We have thus shown that inequalities (IV. 1) and (IV. 3) (hence (III. 1') and (III. 3')) are fulfilled at the fixed point ($\bar{\Omega}$, $\bar{x}$, $\bar{y}$) . Once they are established, it is easy to show that all other equilibrium conditions are also satisfied at that point. As $\bar{x}_i \geqq 0$ and $\sum_i \bar{x}_i = 1$, we have from (IV. 1)

$$\sum_i \sum_j m_{ij}(\bar{y})\bar{x}_i\bar{y}_j \leqq \bar{\Omega} \sum_i \sum_j n_{ij}(\bar{y})\bar{x}_i\bar{y}_j .$$

Similarly, we have from (IV. 3) and the non-negativity of $\bar{y}$

$$\sum_i \sum_j m_{ij}(\bar{y})\bar{x}_i\bar{y}_j \geqq \Omega \sum_i \sum_j n_{ij}(\bar{y})\bar{x}_i\bar{y}_j \quad .$$

It is evident that these expressions can simultaneously hold only with equality ; hence

$$\text{(IV. 2)} \qquad \sum_i \sum_j m_{ij}(\bar{y})\bar{x}_i\bar{y}_j = \bar{\Omega} \sum_i \sum_j n_{ij}(\bar{y})\bar{x}_i\bar{y}_j$$

Remembering the relations between the equilibrium rate of interest β - 1, the warranted rate of growth α - 1 and the natural rate of growth ρ - 1, we can convert the equation (IV. 2) into the alternative forms (III. 2') and (III. 4') .

We have seen that in equilibrium the rate of growth equals the rate of interest times the Capitalist's propensity to save. Accordingly, we have $\beta = (\rho - c)/s$. By virtue of Assumption (III f) we find that there is an activity vector x such that $x(B - \beta A) > 0$. Hence, $x(B - \beta A)\bar{y} > 0$.

On the other hand, it is clear from (III. 1') that he cannot choose activities so as to make the wage payments less than the value of net outputs. Hence the wage payments are positive at the x; it is therefore seen that the equilibrium real-wage rate $\bar{\Omega}$ is positive.

It is now evident that in the state of equilibrium growth the total value of inputs (including the wages) is positive ; to that it follows from (III. 2') that the value of the equilibrium outputs is also positive. Hence (III. 5') is fulfilled at $(\bar{\Omega}, \bar{x}, \bar{y})$. We thus find a "golden equilibrium" solution to the model .

Lecture 5.

Flexible Population Growth

We have so far been concerned with a case in which the labour force increases at a constant rate $\rho - 1$, independent of the real wage rate and other factors. We now, assume that :

(Va) The rate of growth (ρ-1) of the working population is a continuous function of the real wage rate, Ω, such that it is negative (but ρ is still positive) for very low levels of Ω, and zero for the subsistence level of Ω, and positive (but less than some finite number) for Ω exceeding the subsistence level.

We also assume, in place of (III f), that :

(V b) For $\hat{\rho}$ = min ρ such that $\rho \geq \rho(\Omega)$ for all Ω, there is a non-negative intensity vector $(x_1, x_2, \ldots, x_m)$ such that for all goods j=1, . . . , n,

$$\sum_i b_{ij} x_i > \hat{\rho} \left(\sum_i a_{ij} x_i \right) .$$

The only change in the system (III. 1')-(III.6'), which is caused by this shifting from the rigid to the flexible population growth, is a replacement of (III. 6') by

(III. 6") $$\alpha = \rho(\Omega) ;$$

that is to say, we must find an Ω such that both the warranted and the natural rate of growth corresponding to it are equal to each other. A natural and familiar way of finding it would be to trace out on a (ρ, Ω) plane the natural growth rate curve $\rho(\Omega)$ and the golden wage rate graph, and to find their intersection.

An alternative way is our old procedure to reduce the system consisting of (III-1')-(III.5') and (III.6") to a "game" between the "entrepreneur" and the "market". Let us choose an arbitrary set y^* of normalized prices, and a non-negative real wage rate Ω^*. The former determines the Engel-coefficients, and the latter the natural rate of growth, which in turn determines the equilibrium rate of interest [1/]. Throughout this section we assume that the workers operating process i (say stevedores) may have different tastes from those operating different process i' (say pilots) ; the Engel-coefficients of workers are, therefore, denoted by $g_{ij}(y^*)$. We may write

$$m_{ij}(y^*, \Omega^*) \equiv \frac{1}{\rho(\Omega^*)} b_{ij} - a_{ij} - \frac{1}{\rho(\Omega^*)} \left[(\beta(\Omega^*)-1) \sum_k a_{ik} y_k^* \right] f_j(y^*)$$

and

$$n_{ij}(y^*, \Omega^*) \equiv \ell_i g_{ij}(y^*) + \frac{1}{\rho(\Omega^*)} (\beta(\Omega^*)-1) \ell_i f_j(y^*) ,$$

respectively. We obtain the following two inequalities which define the "game" :

$$(V.1^*) \qquad \sum_j m_{ij}(y^*, \Omega^*) y_j \leqq \Omega \sum_j n_{ij}(y^*, \Omega^*) y_j \quad , \quad i = 1, \ldots, m ,$$

$$(V.3^*) \qquad \sum_i m_{ij}(y^*, \Omega^*) x_i \geqq \Omega \sum_i n_{ij}(y^*, \Omega^*) x_i \quad , \quad j = 1, \ldots, n .$$

As we have seen in the last lecture, these inequalities give the fair wage rates Ω and the optimum strategies y of the market, and the optimum strategies x of the entrepreneur. Let $T(y^*, \Omega^*)$ be the set of all possible pairs of the fair wage rate Ω and the optimum price set y, corresponding to (y^*, Ω^*). It is obvious that any y belonging to $T(y^*, \Omega^*)$ is an n-dimensional non-negative vector with unit sum, but some Ω in $T(y^*, \Omega^*)$ may be negative. However, since we have $\rho(\Omega^*) \geqq 1$ for any Ω^* exceeding the subsistence level, and $\rho(\Omega^*) > 0$ for any non-negative Ω^*, and since by (V a) $\rho(\Omega^*)$ is bounded from above, $m_{ij}(y^*, \Omega^*)$ and $n_{ij}(y^*, \Omega^*)$ are bounded from below as well as from above. In view of the fact that every row of the matrix $N(y^*, \Omega^*)$ has at least one positive entry[2], the boundedness of the coefficients $m_{ij}(y^*, \Omega^*)$ and $n_{ij}(y^*, \Omega^*)$ implies the boundedness of the fair wage rate ; that is, there are two finite numbers, Ω_0, and Ω_1, such that for any (y^*, Ω^*), any Ω belonging to $T(y^*, \Omega^*)$ is not less that Ω_0, nor is it greater than Ω_1.

Ω_0 may be negative, while Ω_1 is definitely positive. If some fair wage rates are negative, the correspondence $(y^*, \Omega^*) \rightarrow T(y^*, \Omega^*)$ is not a transformation from the non-negative set

$$S = \left\{ (y, \Omega) \mid y_j \geqq 0, \sum_j y_j = 1, \ \Omega \geqq 0 \right\}$$

into itself; that is , some (y, Ω) in $T(y^*, \Omega^*)$ does not belong to S, although the point (y^*, Ω^*) does. Accordingly, we cannot directly apply to that correspondence any kind of fixed-point theorems because no correspondence other that those from some set into itself is regarded by the fixed-point theorems now available to us. We must set up a scaffolding before we begin to build our house .

Suppose Ω_0 is negative, and define the set S_0 as

$$S_0 = \left\{ (y, \Omega) \mid y_j \geqq 0 , \sum y_j = 1, \ \Omega_0 \leqq \Omega \leqq \Omega_1 \right\}.$$

It is clear that $T(y^*, \Omega^*)$ belongs to S if (y^*, Ω^*) belongs to S. Next, we define the natural growth rate function $\rho(\Omega)$ such that $\rho(\Omega) = \rho(0)$ for all negative Ω not less than Ω_0. It is easily seen that if $\Omega_0 \leqq \Omega^* \leqq 0$, the sets $T(y^*, \Omega^*)$ and $T(y^*, 0)$ coincide; hence, $T(y^*, \Omega^*)$ belongs to S_0 for negative Ω also. We thus find that

$(y^*, \Omega^*) \to T(y^*, \Omega^*)$ is a transformation from S_0 into itself.

It would lighten the reader's burden to tell him at this moment the tool which plays a most significant role in a later argument. It is the fixed-point theorem due to S. Eilenberg and D. Montgomery (a generalization of Kakutani's fixed-point theorem), which may be stated in a general form we shall use, as follows :

Let Q be a non-empty, closed, bounded, convex subset of a Euclidean space. (i) If a multi-valued correspondence $z \to R(z)$ from Q to Q is upper semicontinuous, and (ii) if the set $R(z)$ is contractible for all z in Q, then R has a fixed point, i.e., there is a point $\bar{z}$ which is in $R(\bar{z})$. In our later application, $Q = S_0$, $z = (y, \Omega)$, and $R(z) = T(y, \Omega)$. Clearly, S_0 (the product of the simplex $\left\{ y_j \geqq 0, \sum_j y_j = 1 \right\}$ and the line segment, $\Omega_0 \leqq \Omega \leqq \Omega_1$) is a non-empty , closed bounded, convex subset of Euclidean (n+1)-space. Moreover, as we have just seen, $(y, \Omega) \to T(y, \Omega)$ is a correspondence from S_0 to S_0. Hence , if we succeed in establishing the contractibility of the set $T(y, \Omega)$ for all (y, Ω), as well as the upper semicontinuity of that correspondence, then we can immediately apply the theorem to find a fixed point $(\bar{y}, \bar{\Omega})$.

Let us first be concerned with the upper semicontinuity. Let (y^{*k}, Ω^{*k}) and (y^k, Ω^k) (k = 1, 2, . . . , ad infinitum) be any sequence of points in S_0 that converge to (y^*, Ω^*) and (y, Ω), respectively Suppose each (y^k, Ω^k) is in $T(y^{*k}, \Omega^{*k})$; then

$$\sum_j m_{ij}(y^{*k}, \Omega^{*k}) y_j^k \leqq \Omega^k \sum_j n_{ij}(y^{*k}, \Omega^{*k}) y_j^k \quad , i = 1, \dots, m$$

for all k. As Ω^k is the fair wage rate corresponding to (y^{*k}, Ω^{*k}), there is a non-negative vector x^k (with unit sum)such that

$$\sum_i m_{ij}(y^{*k},\Omega^{*k})x_i^k \geqq \Omega^k \sum_i n_{ij}(y^{*k},\Omega^{*k})x_i^k, \quad j=1,\dots,n.$$

The continuity (III e) of Engel coefficients implies that $m_{ij}(y, \Omega)$ and $n_{ij}(y, \Omega)$ are continuous at every point (y, Ω); consequently, the above two sets of inequalities, which hold for every k, must hold in the limit also. We have

$$\sum_j m_{ij}(y^*, \Omega^*)y_j \leqq \Omega \sum_j n_{ij}(y^*, \Omega^*)y_j, \quad i = 1,\dots,m,$$

$$\sum_i m_{ij}(y^*, \Omega^*)x_i \geqq \Omega \sum_i n_{ij}(y^*, \Omega^*)x_i, \quad j = 1,\dots,n,$$

where x is a limit point of the sequence $\{x^k\}$ ($k = 1, 2, \dots$, ad infinitum). These inequalities imply the fairness of Ω, and y fulfills the requirements to be an element of $T(y^*, \Omega^*)$. Therefore, (y, Ω) belongs to $T(y^*, \Omega^*)$. This implies that the correspondence $(y^*, \Omega^*) \rightarrow T(y^*, \Omega^*)$ is upper semicontinuous.

The same argument as Lemma 1 in Chapter V of my Equilibrium, Stability and Growth leads to the contractibility of the set $T(y, \Omega)$. Let Ω' and Ω'' be the largest and the smallest fair real wage rate when prices and the real wage rate are fixed at y^* and Ω^*. Let y'' be an optimum price set associated with Ω'', so that $(y'', \Omega'') \in T(y^*, \Omega^*)$. We shall show that for any $(y, \Omega) \in T(y^*, \Omega^*)$ and for any non-negative μ not exceeding unity, convex combinations $((1-\mu)y + \mu y'', (1-\mu)\Omega + \mu\Omega')$ belong to $T(y^*, \Omega^*)$. Then it at once follows from the definition of contractibility that $T(y^*, \Omega^*)$ is contractible to a point (y'', Ω') is itself.

We first show that for any $(y, \Omega) \in T(y^*, \Omega^*)$,

(V..4) $\qquad (y, (1-\mu)\Omega + \mu\Omega') \in T(y^*, \Omega^*) \quad$ for all μ in $[0, 1]$.

As Ω' is fair, there is a non-negative vector x' (with unit sum) such that

$$\sum_i m_{ij}(y^*, \Omega^*)x'_i \geqq \Omega' \sum_i n_{ij}(y^*, \Omega^*)x'_i \quad , \quad j=1,\dots,n .$$

Therefore, the inequalities

$$\text{(V.5)} \qquad \sum_i m_{ij}(y^*, \Omega^*)x'_i \geqq \psi \sum_i n_{ij}(y^*, \Omega^*)x'_i \quad , \quad j=1,\dots,n$$

hold *a fortiori* for any ψ between Ω and Ω'. On the other hand, we have

$$\sum_j m_{ij}(y^*, \Omega^*)y_j \leqq \Omega \sum_j n_{ij}(y^*, \Omega^*) y_j \quad , \quad i = 1,\dots, m ,$$

because $(y, \Omega) \in T(y^*, \Omega^*)$. Hence, for any ψ between Ω and Ω', we have

$$\text{(V.6)} \qquad \sum_j m_{ij}(y^*, \Omega^*)y_j \leqq \psi \sum_j n_{ij}(y^*, \Omega^*)y_j \quad , \quad i=1,\dots,m$$

a fortiori. It is easily seen that (V.5) and (V.6) imply (V.4).

As a trivial corollary of the above argument, we get

$$(y'', (1-\mu)\Omega + \mu\Omega') \in T(y^*, \Omega^*) \quad \text{for all} \quad \mu \quad \text{in} \ [0,1] .$$

Hence,

$$\text{(V.7)} \qquad \sum_j m_{ij}(y^*, \Omega^*)y''_j \leqq \psi \sum_j n_{ij}(y^*, \Omega^*)y''_j \quad , \quad i=1,\dots,m$$

for all ψ between Ω and Ω'. It now follows from (V.6) and (V.7) that for any μ in $[0,1]$,

$$\text{(V.8)} \quad \sum_j m_{ij}(y^*, \Omega^*)\left[(1-\mu)y_j + \mu y''_j\right] \leqq \psi \sum_j n_{ij}(y^*, \Omega^*)\left[(1-\mu)y_j + \mu y''_j\right] \quad , \quad i=1,\dots,m .$$

Clearly, (V.5) and (V.8) imply that $((1-\mu)y + \mu y'', \psi) \in T(y^*, \Omega^*)$. As ψ may be any number between Ω and Ω', it may be $(1-\mu)\Omega + \mu\Omega'$. Hence, we have shown that for any non-negative

μ not exceeding unity ,

$$((1-\mu)y + \mu y'', (1-\mu)\Omega + \mu \Omega') \in T(y^*, \Omega^*).$$

Thus the contractibility of the set $T(y^*, \Omega^*)$ is verified . 3/

We are finally in a position to be able to apply the Eilenberg-Montgomery fixed-point theorem, ensuring the existence of an equilibrium point $(\bar{y}, \bar{\Omega})$, such that

$$\text{(V.9)} \qquad \sum_j m_{ij}(\bar{y}, \bar{\Omega})\bar{y}_j \leqq \bar{\Omega} \sum_j n_{ij}(\bar{y}, \bar{\Omega})\bar{y}_j \quad , \quad i = 1, \dots, m \ ,$$

$$\text{(V.10)} \qquad \sum_i m_{ij}(\bar{y}, \bar{\Omega})\bar{x}_i \geqq \bar{\Omega} \sum_i n_{ij}(\bar{y}, \bar{\Omega})\bar{x}_i \quad , \quad j = 1, \dots, n \ ,$$

where $\bar{x}$ and $\bar{y}$ are non-negative vectors with unit sums, and $\bar{\Omega}$ is in the closed interval $[\Omega_o, \Omega_1]$. These inequalities can be converted into the alternative forms (III. 1') and (III. 3') in terms of outputs, material inputs, labour inputs, Capitalist's and Worker's consumption. It is easily seen that (III. 6") holds at $\bar{\Omega}$; (III. 1'), together with (III. 3') implies both (III. 2') and (III. 4'). Thus all inequalities other that (III. 5') are seen to be satisfied at $(\bar{x}, \bar{y}, \bar{\Omega})$.

It remains to show that a positive value of output is produced at $(\bar{x}, \bar{y}, \bar{\Omega})$, as well as to show that the equilibrium real wage rate $\bar{\Omega}$ is strictly positive. These problems are closely related to each other ; in fact, the first problem is merely a direct consequence of the second problem.

Let us suppose the contrary : that $\bar{\Omega}$ is negative . As, for all negative $\Omega \geqq \Omega_o$, $\rho(\Omega)$ is set so as to equal $\rho(0) < \hat{\rho}$, it follows from (V. b) that $\bar{\Omega}$ has non-negative intensity vector x (with unit sum) such that for all goods j = 1, ..., n ,

$$\sum_i b_{ij} x_j > \rho(\bar{\Omega})(\sum_i a_{ij} x_i) .$$

Hence,

$$\sum_i \sum_j \left(\frac{1}{\rho(\bar{\Omega})} b_{ij} - a_{ij}\right) x_i \bar{y}_j > 0 . \tag{V.11}$$

As $\rho(\bar{\Omega}) = \rho(0) < 1$, the equilibrium rate of profits is negative,so that by virtue of our switching rule, the Capitalist does not consume at all. This implies that

$$m_{ij}(\bar{y}, \bar{\Omega}) = \frac{1}{\rho(\bar{\Omega})} b_{ij} - a_{ij} .$$

It is at once seen from (V.9) and (V.11) that

$$\bar{\Omega} \sum_i \sum_j n_{ij}(\bar{y}, \bar{\Omega}) x_i \bar{y}_j > 0 .$$

Considering the fact that the coefficients $n_{ij}(\bar{y}, \bar{\Omega})$ are non-negative, we find that the real wage rate $\bar{\Omega}$ is strictly positive. This is a contradiction; hence $\bar{\Omega}$ must not be non-positive.

Let us now turn to the first problem. The equation (III.2') which has already been established states that the value of the equilibrium outputs equals the total value of inputs (including the wages). As the real wage rate is positive, and labour is an indispensable factor of production, the wage payments should be positive. The value of outputs is therefore positive ; (III.5') is verified.

Foot notes

1 Because we have

$$\beta - 1 = (\rho - 1)/s$$

in equilibrium. It is shown, however, that when $\rho(\Omega^*) < 1$, the profits are negative, so that the Capitalist does not consume at all. Hence, the Capitalist's average propensity to save is unity ; we have

$$\beta - 1 = \rho - 1$$

in equilibrium. (See my book, pp. 144-5) . Corresponding to this, the last term of each $m_{ij}(y^*, \Omega^*)$ and that of each $n_{ij}(y^*, \Omega^*)$ disappear when $\rho(\Omega^*) < 1$. The switching occurs only at those points where $\rho - 1$ vanishes . Hence , it cannot be a cause of discontinuity.

2/ Note that $N(y^*, \Omega^*)$ is non-negative, because if $\beta(\Omega^*) < 1$, the second term of the formula of $n_{ij}(y^*, \Omega^*)$ vanishes, according to our switching rule.

3/ If all workers have identical tastes, i.e., $g_{1j}(y^*) = g_{2j}(y^*) = \dots = g_{mj}(y^*)$ for all $j = 1, \dots, n$, it is shown that the set $T(y^*, \Omega^*)$ is convex ; this dispenses with the Eilenberg-Montgomery theorem to find a fixed point, and we may effectively use the Kakutani theorem. But, when $g_{ij}(y^*) \neq g_{i'j}(y^*)$ for some j we must use the former, instead of the latter, as the following example shows.

Suppose, for simplicity, $c = 0$ and $s = 1$, so that $f_1(y^*) = f_2(y^*) = 0$. Suppose

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 2.2 & 2.2 \\ 2.2 & 0 \end{pmatrix}, \quad L = \begin{pmatrix} 4 \\ 1 \end{pmatrix},$$

$$\rho(\Omega^*) = 1.1, \quad \begin{pmatrix} g_{11}(y^*) & g_{12}(y^*) \\ g_{21}(y^*) & g_{22}(y^*) \end{pmatrix} = \begin{pmatrix} 0.5 & 0.5 \\ 1 & 0 \end{pmatrix} .$$

Then

$$M(y^*, \Omega^*) = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \quad N(y^*, \Omega^*) = \begin{pmatrix} 2 & 2 \\ 1 & 0 \end{pmatrix} .$$

We find the critical values of the fair wage rates Ω' and Ω'' are 1 and 1/2 respectively ; we also find that any

Ω between these two critical values is fair. Moreover, we can see that when $1/2 < \Omega < 1$, $(y_1, y_2) = (0, 1)$ is the only optimal price set associated with it, and when $\Omega = 1$, any price set (y_1, y_2) such that $y_1 + y_2 = 1$ is optimal. It can now easily be seen that the set $T(y^*, \Omega^*)$ is not convex but contractible.

Lecture 6

The Turnpike Theorem

It has been seen in Lecture 2 that the path of full employment growth of capital and labour will eventually approach the balanced growth (or Golden Equilibrium) path, even if it starts from any given endowments of capital and labour. But this result holds true only in two-sector models. The von Neumann procedure of treating capital goods of different stages of wear and tear as different goods requires that models should inevitably be multi-sectoral : and in such models consisting of more than two sectors, the stability of the Golden Equilibrium is not preserved. In fact, the stability depends, inter alia on the form of the saving function, the state of technology (the capital intensity of the consumption-goods industry versus that of the capital-goods industry), the production periods of various goods, the mechanism of price formation, and so on. Another kind of the stability concept (that is, that in the sense of programming) has, however, been proposed by Dorfman, Samuelson and Solow, to the effect that any long-run efficient path starting with historically given initial endowments and aiming at a capital structure specified by the planning authorities runs near the von Neumann balanced growth path (i.e., the Turnpike) for most of the programming period. The efficient growth cannot be achieved by a blind maintainance of the full utilization of capital and labour, but is attained under the dictatorship of the Ministry of Economic Planning that orders factories to discard some capital goods (but not all) if the situation necessitates.

In the following we make, for simplicity, several assumptions due to von Neumann, most of which will be removed in a more far-reaching discussion in a later lecture. (1) As soon as an an excess demand for labour is found in the economy, the supply of labour is expanded by natural growth or by immigration, to whatever extent is necessary . (2) The Worker only consumes and the Capitalist only saves. (3) The Worker buys consumption goods in fixed amounts. That is to say , the consuption of each good per worker is independent of prices as well as the real-wage rate ; we have

$$g_j(y)\Omega = c_j \qquad j = 1, \ldots, n,$$

where c's are constant. (4) The real-wage rate is adjusted so as to enable the Worker to buy goods of the amounts, $c_1, \ldots, c_n$; so that

$$\Omega = \sum_j c_j y_j \quad .$$

Substituting from these and considering $f_j(y) = 0$, we may put the basic inequalities (III. 1') - (III. 5') in the original von Neumann forms :

$$\text{(VI. 1)} \qquad \sum_j b_{ij} y_j \leqq \beta \sum_j c_{ij} y_j \quad , \qquad i = 1, \ldots, m,$$

$$\text{(VI. 2)} \qquad \sum_i \sum_j b_{ij} x_i y_j = \beta \sum_i \sum_j c_{ij} x_i y_j$$

$$\text{(VI. 3)} \qquad \sum_i b_{ij} x_i \geqq \alpha \sum_i c_{ij} x_i \quad , \quad j = 1, \ldots, n,$$

$$\text{(VI. 4)} \qquad \sum_i \sum_j b_{ij} x_i y_j = \alpha \sum_i \sum_j c_{ij} x_i y_j$$

$$\text{(VI. 5)} \qquad \sum_i \sum_j b_{ij} x_i y_j > 0 \; ,$$

where c_{ij} is the sum of a_{ij} and $\ell_i c_j$.

It is the result of Kemeny, Morgenstern and Thompson that the above system has non-negative solutions $(\bar{x}, \bar{y}, \bar{\alpha}, \bar{\beta})$ such that $\bar{\alpha} = \bar{\beta}$, if it is assumed, in addition to (III a) , that :

For each $i = 1, \dots, n$ there is at least one positive c_{ij}. These assumptions, however, do not ensure the uniqueness of the solution. We refer to those solutions which give the largest α (the maximum rate of growth) as the von Neumann solutions, and assume that:

(i) the von Neumann intensity vector $\bar{x}$ is unique,

(ii) the von Neumann price vector $\bar{y}$ is strictly positive.

The former implies some independence of processes, while the latter rules out the possibility of free goods, so that $\bar{x} = (\bar{x}_1, \dots, \bar{x}_m)$ fulfills each inequality of (VI. 3) with equality.

Let us now consider feasible paths of length T starting from the stocks of goods $(b_1^o, b_2^o, \dots, b_n^o)$ inherited from the past and leaving the stocks of the amounts $(b_1^T, b_2^T, \dots, b_n^T)$ at the end of period T for production in the future beyond T. They must be feasible throughout the whole period T; hence for all $j = 1, \dots, n$,

$$b_j^o \geqq \sum_i c_{ij} q_i(1)$$

(VI. 6) $$\sum_i b_{ij} q_i(t) \geqq \sum_i c_{ij} q_i(t+1) \qquad (t = 1, \dots, T-1)$$

$$\sum_i b_{ij} q_i(T) = b_j^T ,$$

where $q_i(t)$ is the level of activity i (or the non-normalized intensity of process i) in period t. No restriction is made on the initial stock vector b^o except that:

(iii) b^o is given such that the economy can move from b^o to a point on the von Neumann path in a finite number, say T_o, of periods.

Next, suppose the terminal composition of stocks, say $b^* = (b_1^*, b_2^*, \dots, b_n^*)$ at which the system aims, is specified by the planning authorities. The only restriction on b^* is that :

(iv) b^* is given such that the economy can move from a point on the von Neumann path to a state with the capital composition b^* in a finite number T_1 of periods.

We may normalize b^* such that

$$\bar{y}_1 b_1^* + \ldots + \bar{y}_n b_n^* = 1 .$$

Let us evaluate all the feasible paths in terms of the final composition b^*. Denote the minimum of the ratios, b_1^T/b_1^*, ..., b_n^T/b_n^* by u ; call a feasible path of length T an optimal path of length T, if it maximizes the index u of attainment of the aim among all feasible paths. We can now state the Turnpike theorem as follows :

<u>If the programming period T is sufficiently long, then any optimal path starting from b^0 will remain most of that period within a very small neighbourhood of the von Neumann path.</u>

Following Hicks, all available processes may be classified into two groups : the "top" and the "non-top" processes. All processes such that $\sum_j b_{ij}\bar{y}_j = \beta \sum_j c_{ij}\bar{y}_j$ are arranged in the top class, while all others in the non-top. Non-top processes are unprofitable when the equilibrium prices and interest rate prevail, so that they enter into the equilibrium vector $\bar{x}$ with zero intensity.

We also classify processes into the von Neumann and the non-von Neumann processes. The processes, if their $\bar{x}_i$ are positive, are called the von Neumann processes ; otherwise they are the non-von Neumann . There is a suitable von Neumann price set $\bar{y}$, in terms of which the top and the von Neumann classifications coincide . In the following we choose such a $\bar{y}$. It was shown by

Gale that the number of the von Neumann process m^* is at most as large as the number of goods n. We define a "top" state as a state of affairs where all the non-top processes are not used ; then the von Neumann equilibrium state is a particular "top" state, where top processes are used with the balanced intensities, $\bar{x}_1$, $\bar{x}_2, \ldots, \bar{x}_m$. Following Hicks again, we shall sometimes refer to it as the "top-balanced" state.

It will be seen that when the programming period T becomes very long, any optimal path will approach some sequence of the top states, which in turn converges to the von Neumann path. Let b_j^t denote the stock of good j available at the end of period t, i.e.:

$$b_j^t = \sum_i b_{ij} q_i(t) \qquad j = 1, \ldots, n .$$

Consider a stock index in which the stoks of various goods are weighted by their equilibrium prices ; we have

$$K_t = \sum_j \bar{y}_j b_j^t$$

The feasibility condition (VI.6) implies

$$\text{(VI. 7)} \qquad \sum_i \sum_j b_{ij} q_i(t) \bar{y}_j \geqq \sum_i \sum_j c_{ij} q_i(t+1) \bar{y}_j .$$

On the other hand, we have from the equilibrium condition (VI.1)

$$\text{(VI. 8)} \qquad \bar{\beta} \left[\sum_i \left(\sum_j c_{ij} \bar{y}_j \right) q_i(t+1) \right] \geqq \sum_i \left(\sum_j b_{ij} \bar{y}_j \right) q_i(t+1) .$$

In view of the definition of the index K we obtain from (VI. 7) and (VI, 8)

$$\bar{\beta} K_t \geqq K_{t+1} \quad \text{or} \quad K_{t+1}/K_t \leqq \bar{\alpha} ,$$

because $\bar{\alpha} = \bar{\beta}$; in words, the capital index cannot grow at a rate greater than the von Neumann rate of growth. If and only if all stocks are fully employed in period $t+1$, (VI. 7) holds with equality; and if and only if top processes are used in period $t+1$, (VI. 8) holds with equality. Under both these conditions, $K_{t+1}/K_t = \bar{\alpha}$, that is, the capital index increases at the "top" rate.

It follows from Assumptions (iii) and (iv) that among feasible paths there are paths that start from given b^o to reach a point on the von Neumann path at the end of period T_o and then run on that path until the end of period $T-T_1$ to get to a point with the aimed capital composition b^* at the end of period T. Take one of such paths, and call it the "bottom" path, because any optimum path must be at least as high as that path at the end of period T. Let the rate of growth of the index K in period t be η_t along the bottom path and θ_t along an optimum path; the former gives K_T of the value

$$\bar{K}_T = K_o(1+\eta_1)(1+\eta_2)\ldots(1+\eta_T) ,$$

while the latter gives

$$K_T = K_o(1+\theta_1)(1+\theta_2)\ldots(1+\theta_T) .$$

It is seen that the optimality requires $K_T \geqq \bar{K}_T$. 1/

Let us rearrange the series of the rates of growth $\{\theta_t\}$ $(t=1,2,\ldots, T)$ in order that

$$\theta_{t_1} \leqq \theta_{t_2} \leqq \ldots \leqq \theta_{t_T} .$$

Let θ_{t_N} be less than the von Neumann rate of growth by δ. Then

$$1 + \rho_{t_i} \leqq \bar{\alpha} - \delta \qquad i = 1, \dots, N.$$

$$1 + \rho_{t_j} \geqq \bar{\alpha} - \delta \qquad j = N+1, \dots, T\ ;$$

hence, we find that

$$K_T \leqq K_0 (\bar{\alpha} - \delta)^N \; \bar{\alpha}^{T-N} .$$

Since the botton path is so constructed that the index K increases at the von Neumann rate from period $T_0 + 1$ to $T - T_1$, we may write

$$\bar{K}_T = K_0 J \bar{\alpha}^{T-T_0-T_1} .$$

where

$$J = (1+\eta_1) \dots (1+\eta_{T_0})(1+\eta_{T-T_1+1}) \dots (1+\eta_T) .$$

As K_T cannot fall short of the bottom value $\bar{K}_T$, we have, a fortiori,

$$J \bar{\alpha}^{T-T_0-T_1} \leqq (\bar{\alpha} - \delta)^N \bar{\alpha}^{T-N} ;$$

hence we get

$$J \bar{\alpha}^{-(T_0+T_1)} \leqq \left(\frac{\bar{\alpha} - \delta}{\bar{\alpha}}\right)^N ,$$

where it is seen that $J\bar{\alpha}^{-(T_0+T_1)}$ is a constant independent of the length of the optimum path. The equation

$$\text{(VI. 9)} \qquad J \bar{\alpha}^{-(T_0+T_1)} = \left(\frac{\bar{\alpha} - \delta}{\bar{\alpha}}\right)^N$$

gives the maximum difference δ which the N-th smallest rate of growth ρ_{t_N} can take from the von Neumann rate; and for all $i > N$, $1+\rho_{t_i}$ should be between $\bar{\alpha}$ and $\bar{\alpha} - \delta$. It is at once seen from (VI. 9) that δ tends to zero as N tends to infinity.

Now let ε be a small positive number, and define N as the integer wich is nearest to ε/T . When T tends to infinity, N will also tend to infinity, but the ratio of N to T becomes smaller and smaller, finally approaching zero. Thus, not only the number of those periods in which the rate of growth of K (along the optimum path) is approximately equal to the von Neumann rate increases indefinitely, but also its ratio to the total number of periods approaches unity. Accordingly, if the programming period T is sufficiently long, we may take a very long period, throughout which the difference between the von Neumann growth rate and the optimum growth rate is negligible.

Let us arrange processes in order that the first m^* processes are the top processes and others the non-top. If the weights of the non-top processes were not negligible, the growth rate of K would significantly be less than the von Neumann rate. Hence, throughout the long period we have just found, non-top processes can safely be neglected ; we have for that period

$$\text{(VI.10)} \qquad \sum_i^* b_{ij}\, q_i(t) \geqq \sum_i^* c_{ij}\, q_i(t+1) \qquad j = 1, \dots, n,$$

where $\sum_i^*$ denotes the summation over the "top" range, i.e. from $i = 1$ to $i = m^*$. If (VI.10) holds with strict inequality for some good, some quantities of that good are not used and thrown away in period $t + 1$. As we are assuming that there is no free good in the economy, this means a discarding of a valuable stock. Therefore, the capital index K cannot increase at the maximum rate (because it can do so if and only if all the available stocks of goods are fully utilized by the top processes). Hence we must have

$$\text{(VI.11)} \qquad \sum_i^* b_{ij} q_i(t) \simeq \sum_i^* c_{ij} q_i(t+1) \qquad j = 1, \dots, n,$$

for most of the time as the period of programming becomes very long.

Noting that $m^* \leqq n$, let us consider the following equations;

$$\text{(VI.12)} \qquad \sum_i^* b_{ij} q_i(t) = \sum_i^* c_{ij} q_i(t+1) \qquad j = 1, \dots, m^*$$

$$\text{(VI.13)} \qquad \sum_i^* b_{ik} q_i(t) = \sum_i^* c_{ik} q_i(t+1) \qquad k = m^* + 1, \ldots, n .$$

Equations (VI.12) are solved by the usual method; that is, all possible particular solutions are obtained by finding latent roots of the characteristic equation :

$$\text{(VI.14)} \qquad \left| B^* - \lambda C^* \right| = 0 ,$$

where B^* stands for the $m^* \times m^*$ matrix (b_{ij}) and C^* for the $m^* \times m^*$ matrix (c_{ij}). If the rank of C^* is m^*, then (VI.14) gives m^* roots ; but no a priori information is given about the rank of C^*, so that we only know that (VI.14) gives at most m^* roots. Let $\bar{m}$ ($\leqq m^*$) be the total number of the latent roots, and denote them by $\lambda_1, \ldots, \lambda_{\bar{m}}$. It is well known that the general solution to (VI.12) can be written

$$\text{(VI.15)} \qquad q_i(t) = \sum_k^{\bar{m}} \nu_k q_{ik} \lambda_k^t \qquad i = 1, \ldots, m^* ,$$

where $Q_k = (q_{1k}, q_{2k}, \ldots, q_{m^* k})$ is the characteristic solution associated with λ_k; and $\nu_1, \ldots, \nu_{\bar{m}}$ depend on the initial conditions.

It is noticed that when $\bar{m} < m^*$, the initial values of $q_1(t)$, $q_2(t), \ldots, q_{m^*}(t)$ cannot arbitrarily be given ; we must start from a point in the non-negative part of the convex polyhedral cone spanned by $\bar{m}$ vectors $Q_1, Q_2, \ldots, Q_{\bar{m}}$. It is also noticed that further restrictions are imposed on the initial position in order for the solution (VI.15) to (VI.12) to satisfy $n - m^*$ equations (VI.13) too. In fact , if

$$\text{(VI.16)} \qquad \sum_i^* b_{ij} q_{ik} \neq \lambda_k \sum_i^* c_{ij} q_{ik} \qquad \text{for some } j = m^* + 1, \ldots, n ,$$

then the ν_k corresponding to the particular solution $q_{ik} \lambda_k^t$ must be zero. It is seen however, that there is a particular solution which satisfies the whole equations (VI. 12) and (VI. 13) for any value of t, that is, the von Neumann balanced growth solution ; $\bar{\alpha}$ is a latent root of (VI. 14) and $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_{m^*}$ are not only the characteristic solutions associated with $\bar{\alpha}$ but also fulfill

$$\sum_i^* b_{ij} \bar{x}_i = \bar{\alpha} \sum_i^* c_{ij} \bar{x}_i \qquad j = m^* + 1, \ldots, n .$$

Let us now assume that :(v) (VI. 16) holds for all negative and complex roots with modulus $\bar{\alpha}$.

This is the "joker" ruling out exceptions discussed later. If there is a latent root λ_k such that $|\lambda_k| > \bar{\alpha}$, it is seen that the initial point is to be situated in a more restricted zone in order that (VI. 12) and (VI. 13) should hold for a long period. Suppose λ_k is a positive number greater that $\bar{\alpha}$. Then some of $q_{1k}, q_{2k}, \ldots, q_{m^*k}$ must be negative ; because otherwise there would be a balanced growth solution at a rate λ_k greater than $\bar{\alpha}$. This contradicts the definition of the von Neumann balanced growth as a state of balanced growth at the maximum rate. If , say , q_{1k} is negative, then $q_1(t)$ will become negative sooner or later, because $q_{1k}\lambda_k^t$ will dominate other terms of $q_1(t)$ when t is large. In order that (VI. 12) and (VI. 13) should hold for a long period, the importance of the particular solution $q_{ik}\lambda_k^t$ in the general solution (VI. 15) must be negligible ; that is, ν_k must be a very small number. Similar observations are made for negative and complex roots which are greater than $\bar{\alpha}$ in modulus. All particular solutions corresponding to those λ_k with modulus $\bar{\alpha}$ do not enter the general solution because of (VI. 16) being assumed.

We have seen that if the programming period T is sufficiently long, any optimum path has a long stretch throughout which (VI.12) and (VI.13) are approximately fulfilled. This implies that at the beginning of that stretch the optimum path has already arrived in the restricted zone just discussed above. Some ν_k must be zero, and the ν_k corresponding to those λ_k which is greater than $\bar{\alpha}$ in modulus must be zero or negligible. Every particular solution entering the general solution with a ν_k of some "macroscopic" magnitude has a characteristic value λ_k which is less that $\bar{\alpha}$ in modulus, unless it is the von Neumann balanced growth solution. It is now clear that as time goes on, the von Neumann solution will become dominant in the general solution. This means that the optimum path remains most of the programming period within a very small neighbouring cone of the von Neumann path.

The Turnpike theorem is not an exception of the rule that every rule has its exception. There is no reason (economic or mathematical) why the "joker" (v) should be in our hand. If it is not, there is a path viable over an infinitely long period that does not converge to the equilibrium path of balanced growth and oscillates around it with a relatively constant amplitude.

1/ Since the bottom path is a path that realizes the composition b^* at the end of period T, the ratios $\bar{b}_1^T/b_1^*, \ldots, \bar{b}_n^T/b_n^*$, are equal to each other, where $\bar{b}_j^T$ is the "bottom" stock of good j at the end of period T. It follows from the definition of the optimality that any optimum path must have an index u of attainment of the aim which is at least as high as the ratio $\bar{b}_j^T/b_j^*$. In view of the definition of u, we have

$$u \leqq b_j^T / b_j^* \qquad j = 1, \ldots, n \quad .$$

Hence $\bar{b}_j^T \leqq b_j^T$ $(j = 1, \ldots, n)$. Thus we get $\bar{K}_T \leqq K_T$.

Lecture 7 .

Existence of Golden Equilibrium Once More

The arguments in the previous lectures were based on a tacit assumption that the production of goods by means of goods and labour was independent of the feeding of labour ; the former was described in terms of the matrices of input and output coefficients and the vector of the labour-input coefficients, all of which were considered as technologically given constants; as a matter of fact, however, it is difficult to obtain precise figures of those coefficients, especially the figures of the labour-input coefficients, without knowing how and on what level people are fed. The productivity of labour dependes not only on the technology in the narrowest sense but also on the workers' state of health , their living and working conditions, and so on. We remember that in a slave economy, with the wages fixed at a subsistence level, the productivity of labour was low, so that it was replaced by a more productive system, the capitalist economy. It is not surprising to see that outputs of goods will not decrease, even though the allocation of available goods among industries and families becomes favorable to the latter ; in fact, the positive indirect effect on outputs of a transfer of goods from industries to families though an improvement in the welfare of the workers is so strong as to overcome the negative direct effect on outputs of the decrease in industrial inputs. It is thus seen that the production of goods and the feeding of men should be treated as an inseparable process.

Let $x_i(t)$ be the amount of good i available at the begining of period t, and $m(t)$ the number of people living in the economy at the same point of time. $x_i(t)$ is allocated among industries and families to the production of goods and to the production of men, respectively; the former part is denoted by $x_i^I(t)$, the latter part by $x_i^F(t)$. Obviously,

$$x_i(t) = x_i^I(t) + x_i^F(t) .$$

There are n kinds of goods, and the vector $(x_1(t), x_2(t), \ldots, x_n(t))$ is denoted by $x(t)$; similarly, $x^I(t)$ and $X^F(t)$ represent $(x_1^I(t), x_2^I(t), \ldots, x_n^I(t))$ and $(x_1^F(t), x_2^F(t), \ldots, x_n^F(t))$, respectively.

The number of people at the end of period t, $\ell(t)$, depends on the feeding $x^F(t)$ in that period, as well as the number of people at the beginning. If good i is perishable, the families have no stock of that good at the end of the period, but if it is durable, they will enter the next period with certain amounts of the secondhand good i. We assume that goods are all perishable when they are delivered to households, but may be durable if they are used in the industries. The families have, therefore, no outfit of goods at the end of the period. We may write a biologically feasible transformation of goods and men as follows :

$$(x^F(t), m(t)) \longrightarrow \ell(t) .$$

It is evident that the amounts of goods 1, ..., n produced in period t are determined primarily by the industrial inputs $x^I(t)$ and number of workers employed m(t) ; so that we have a tecnologically feasible transformation :

$$(x^I(t), m(T)) \longrightarrow y(t) .$$

where y(t) is the vector of the quantities of goods which the industries will have at the end of period t ; a component of y(t), say

$y_i(t)$, stands for the sum of the quantity of good i newly produced during the period, and the quantity of the same good that the industries will have at the end of period t after using, for one period, $x_i^I(t)$ for production. But, as we have seen, there can be no definite input-output relationship that is purely technological; the amounts of goods produced, $y(t)$, depend not only on the industrial inputs, $x^I(t)$, and the employment of labour, $m(t)$, but also on how workers are fed in period t. It is thus seen that the set of all possible transformations of industrial inputs $x^I(t)$ and $m(t)$ into industrial outputs $y(t)$ will not be well defined, even though there is no change in the technology, unless the level of the welfare the workers are enjoying is specified. It is seen that the argument cannot conveniently run in terms of the set of the "purely technological possibilities". We consider instead, therefore, the set Π of all possible transformations of social inputs $(x(t),\ m(t))$ into social outputs $(y(t), \ell(t))$. It is remarked that $y(t)$ depends not only on $x^I(t)$ and $m(t)$ but also on $x^F(t)$. It is also remarked that many different final states, $(y(t), \ell(t))$, can of course correspond to the same initial states, $(x(t), m(t))$, but the set Π itself remains constant so long as there is no change in the biology and the technology.

Through this paper each process converting the social inputs $(x(t), m(t))$ into the social outputs $(y(t), \ell(t))$ will be referred to as an "aggregate" production process. Let $(x, m) \rightarrow (y, \ell)$ and $(x_*, m_*) \rightarrow (y_*, \ell_*)$ be any two aggregate processes. Each of them, say the former, is made up of labour feeding process, $(x^F, m) \rightarrow \ell$, and an industrial production process, $(x^I, m; x^F) \rightarrow y$ such that

$$x = x^I + x^F ;$$

similarly, the latter aggregate process is analyzed into $(x^F_*, m_*) \rightarrow \ell_*$ and $(x^I_*, m_*; x^F_*) \rightarrow y_*$ such that

$$x_* = x^I_* + x^F_* .$$

Consider now an economy in which an outfit of goods, $x + x_*$, is available to $m + m_*$ workers. We can feed the bundle of goods x^F to m workers, and x^F_* to m_* workers ; then they will grow at the rates ℓ/m and ℓ_*/m_*, respectively, so that there will be, at the end of the period, $\ell + \ell_*$ people in the economy. After having fed workers, goods are available for production in the amounts $x + x_* - x^F - x^F_*$ $(= x^I + x^I_*)$. They are allocated among workers, the m of whom, being fed with x^F, can use x^I to produce y, while the remaining m_* workers fed with x^F_* can convert x^I_* into y_*. It is, therefore, always possible for the economy to close the period with goods of the amounts $y + y_*$. This fact may be paraphrased as follows :

(i) If aggragate production processes $(x, m) \rightarrow (y, \ell)$ and $(x_*, m_*) \rightarrow (y_*, \ell_*)$ are feasible, then a third aggregate process $(x + x_*, m + m_*) \rightarrow (y + y_*, \ell + \ell_*)$, formed by operating the two simultaneously, is also feasible.

By taking $x = x_*$, $m = m_*$, $y = y_*$, and $\ell = \ell_*$, it clearly follows from (i) that if $(x, m) \rightarrow (y, \ell)$ is a feasible process, then $(\lambda x, \lambda m) \rightarrow (\lambda y, \lambda \ell)$ is also feasible for any <u>integer</u> λ. But (i) does not imply that the same relationship holds for <u>any non-negative number</u> λ. We must have an independent assumption, that is, that of the divisibility of aggregate processes, in order to obtain the law of constant returns, not only for discrete changes but also for continuous changes in scale :

(ii) If an aggregate process $(x, m) \rightarrow (y, \ell)$ is feasible, $(\lambda x, \lambda m) \rightarrow (\lambda y, \lambda \ell)$ is also feasible for any non-negative λ.

Although it would be an extremely difficult empirical question to ask whether the aggregate production possibility set π contains all its boundary points (or whether it is closed or not), it is true that much analytical advantage is gained by assuming it. Moreover, the closedness of the set π, if it is put in the following equivalent form (iii), is not seen to be implausible and unreasonable. There would be no reason to reject it ; we have just accepted a similar mathematical condition - that is, the divisibility of the aggregate production process for the purpose of analytical convenience. We thus have :

(iii) If a transformation $(x, m) \rightarrow (y, \ell)$ is not feasible, then all transformations differing from it only infinitesimally are also infeasible.

Together with (i) and (ii), the condition (iii) implies that π is a closed convex cone in $2(n+1)$-dimensional Euclidean space .

By putting $\lambda = 0$ in (ii), we find that nothing can produce nothing; that is, an aggregate process of transforming (0, 0) into (0, 0) is feasible. This, however, does not mean that the converse is also true. Without an independent assumption, the above three conditions do not logically rule out the possibility that something is produced from nothing. But as a matter of fact it is universally true that when $x = 0$, workers have nothing to eat and nothing to use for production, and hence we would have $y = 0$ and $\ell = 0$. On the other hand, it is also obviously true that dogs can only be borne by dogs, so that if $m = 0$, then $\ell = 0$. We may thus make the following postulate :

(iv) If an aggregate process $(x, m) \rightarrow (y, \ell)$ is feasible with $x=0$, then $y = 0$ and $\ell = 0$; and if $m = 0$, then $\ell = 0$.

An additional assumption we impose is that which is usually referred to as the assumption of free disposal of goods. It asserts that any goods can be disposed of without any additional inputs. Under this assumption, by disposing of $x_* - x$ we can transform (x_*, m) into (y, ℓ), which can in turn be transformed into (y_*, ℓ) by disposing of $y - y_*$ without additional inputs. We thus find that :

(v) If an aggregate process $(x, m) \rightarrow (y, \ell)$ is feasible, then any other process $(x_*, m) \rightarrow (y_*, \ell)$ such that $x_* \geqq x$ and $y_* \leqq y$ is also feasible.

On the other hand, the free disposal of labour is possible only in a shameless world like a slave society, but never in any modern society, of course. I am very sorry to confess that in our hypothetical economy, however, persons of no use can be eliminated in the following way : Suppose there are m_* workers living in an economy where goods are available in the amounts x. Suppose also that m workers out of m_* can transform x into (y, ℓ). In such a situation, we may group the m_* workers into two classes consisting of m and $m_* - m$ members, respectively. We may then employ m workers so as to transform x into (y, ℓ), and leave $m_* - m$ workers without being fed. It follows from (i) that by operating two processes $(x, m) \rightarrow (y, \ell)$ and $(0, m_* - m) \rightarrow (0.0)$ simultaneously, the initial state ((x, m_*) can be transformed into (y, ℓ).

Thus , if) $(x, m) \rightarrow (y, \ell)$ is feasible, then $(x, m_*) \rightarrow (y, \ell)$ is feasible, for any m_* such that $m_* \geq m$; workers can be freely disposed of since they will die of starvation. (People have to work in order to live.) It is, of course, an extremely severe and immoral rule, but a logical consequence of the assumptions (i) and (iv) .

Let us now make two further assumptions on the aggregate production processes. One of them means the possibility of growth, and may be stated as follows :

(vi) Among feasible aggregate processes there is at least one such that $y > x$ and $\ell > m$.

The other is that which ensures the indecomposability of the system. Suppose $m > 0$, and arrange goods in two mutually exclusive groups, G_1 and G_2 . A set of goods G_1 is said to be independent of G_2 if those goods can be produced from themselves, that is, without consuming any goods in G_2. We assume that :

(vii) The system is indecomposable (or irreducible) in the sense that there are no proper independent subgroups of goods.

This hypothesis implies that if $m > 0$ and $y_i > 0$ for all goods in the group G_1, then there are some goods j in G_2 with $x_j > 0$. It is a restrictive and even unrealistic assumption, for it is not difficult to take exceptions to it. From any actual "Input-Output" tables so far obtained, however, we may observe that nation-wide economies are likely to be indecomposable if individual goods are aggregated into "composite" goods whose number is of a manageable size.

The final assumption we require has some relation to the famous (or notorious) Malthusian law of population; Let $(x, m) \rightarrow (y, \ell)$

be a feasible production process ; define the rate of growth of the "gross" production as g such that

$$g + 1 = \min(y_1/x_1, \ldots, y_n/x_n) .$$

It is, by definition, the smallest among the n rates of growth of outputs. With a free, but recapitulating interpretation, the Malthusian law may be formulated as a "law" requiring that for each feasible process the rate of growth of the "gross" production is less than the corresponding rate of growth of population , $\ell/m - 1$. This is a strong assumption ; but if it is weakened into the form of the statement (viii) below, it will be accepted even in a contemporary, highly advanced society with a low actual (but not maximum) rate of growth of population :

(viii) Being compared at their highest feasible values, the rate of growth of population is greater than that of the "gross" production.

We make (viii) because, since we assume that there are no technical improvements, it will not be far from reality to assume that human beings can grow at a rate faster than the maximum growth rate of the gross production, although it is true that they cannot grow as fast as mice.

As for the capitalists' demand for consumption goods, we accept all the assumptions in Lecture 3. First, all capitalists have identical tastes, and the "Capitalist" consumes a constant fraction of his income. Secondly, his demand for consumption remains unchanged when we have a proportional change in all prices and income. Moreover, it is assumed that the Capitalist's Engel-coefficients are independent of the level of his income ; that is to say, his consumption demand for each good has the income elasticity which is unity. We thus have

$$d_j = f_j(p_1, p_2, \ldots, p_n)e \qquad j=1, \ldots, n,$$

where d_j is the Capitalist's demand for the consumption good j, p_i's are normalized prices, and e represents the Capitalist's normalized income (profits). Since at any prices a constant proportion of e is devoted to the consumption, $\sum_j p_j d_j$, we have,

$$\sum_j f_j(p_1, p_2, \ldots p_n)p_j = c \qquad \text{for all } p_i\text{'s},$$

as before where c is the Capitalist's average propensity to consume.

The state of the Golden Equilibrium (or the von Neumann equilibrium) can now be defined in terms of the following five sets of inequalities: Let α_j be the growth factor (i.e., 1+the rate of growth) of the total (industrial and labour-feeding) inputs of good j. As the amount $y_j - f_j(p)e$ of good j, which is left after deducting the Capitalist's consumption $f_j(p)e$ from the output y_j is available for the production in the next period, the total input of j can expand at a rate of $\alpha_j - 1$ such that

$$y_j - f_j(p)e \geqq \alpha_j x_j \quad .$$

As for labour, the total number of workers available at the beginning of the next period is ℓ, which should not fall short of α_{n+1} (the growth factor of the employment of labour) times the total number of employed workers m in the present period; thus,

$$\ell \geqq \alpha_{n+1} m \ .$$

At the state of balanced growth, α_j's are obviously equal to each other so that we get

(VII. 1) $$\bar{y}_j - f_j(\bar{p})\bar{e} \geqq \bar{\alpha}\bar{x}_j \qquad j=1,\ldots, n,$$

(VII. 1') $$\bar{\ell} \geqq \bar{\alpha}\bar{m},$$

where $\bar{\alpha}$ is the common rate of growth, $\bar{p}$ the equilibrium price vector, $(\bar{x}, \bar{m}) \rightarrow (\bar{y}, \bar{\ell})$ the Golden Equilibrium process, and $\bar{e}$ the profits that accrue.

The prices of goods are determined by the rule of competition. If there is an excess demand for a good, the price of that good is increased, while if there is an excess supply, it will be decreased. It follows from this rule that the prices of those goods that are overproduced in the state of the Golden Equilibrium cannot remain stationary unless they have already fallen to the bottom. Thus, the price equilibrium is only established when zero prices are charged for those goods that fulfill the inequalities (VII. 1) with strict inequality. We have , therefore,

(VII. 2) $$\sum_j \bar{p}_j\bar{y}_j - \sum_j(\bar{p}_j f_j(p))\,\bar{e} = \bar{\alpha} \sum_j \bar{p}_j\bar{x}_j \;.$$

A similar rule is often applied to the labour market. But we assume that the wage rate is fixed at a level such that the workers can just buy the necessaries of life by spending the whole amount of their wages :

(VII. 2') $$\bar{w}\,\bar{m} = \sum_j \bar{p}_j \; \bar{x}_j^F \quad ,$$

where $\bar{w}$ is the wage rate per worker, and $\bar{x}_j^F$ the amount of good j consumed by the workers in the state of the Golden Equilibrium.

The third set of inequalities insists that the principle of the maximum rate of profit hold good in the Golden Equilibrium. It may be put into two different (but equivalent) forms , according to how we define the costs of production. Let us choose one from among feasible aggregate production processes $(x, m) \rightarrow (y, \ell)$ such that $\ell \rightarrow \bar{\alpha}m$. From the viewpoint of the society as a whole, the total costs of that process and the profits accruing from it are evaluated at

$$\sum_j \bar{p}_j(x_j^I + x_j^F) \quad \text{and} \quad \sum_j \bar{p}_j y_j - \sum_j \bar{p}_j(x_j^I + x_j^F)$$

respectively, if the equilibrium prices prevail. Its rate of profit (i.e., the ratio of the latter to the former) will not exceed the golden rate of profit, $\bar{\beta} - 1$, realized along the the path of balanced growth by virtue of the principle of the maximum rate of profit; in fact, it requires that no feasible aggregate process can yield profits at a rate greater than the Golden Rate unless the reproduction of the labour force is sacrificed. Hence, for all feasible processes $(x, m) \rightarrow (y, \ell)$ such that $\ell \geqq \bar{\alpha}m$, we

$$\text{(VII. 3)} \qquad \sum_j \bar{p}_j y_j \leqq \bar{\beta}\left[\sum_j \bar{p}_j(x_j^I + x_j^F)\right] .$$

On the other hand, from the point of view of the firms it is clear that the costs of production should not be defined as $\sum_j p_j(x_j^I + x_j^F)$ but $\sum_j p_j x_j^I + wm$. It is, of course, true that goods are consumed in the amounts $x_1^I + x_1^F, \ldots, x_n^I + x_n^F$ in the process of the aggregate production. But the excess of $\sum_j p_j x_j^F$ over wm is charged to the workers, while the excess of the latter over the former makes an addition to their net worth. The firms are unconcerned about the family finances of the workers, and would choose a process that yields profits at the maximum rate on the costs in the sense of the second definition. As the equality of the workers' income and their expenditure does not necessarily follow, the two definitions of the costs of production are not identical. Thus, the principle of the maximum rate of profit may not hold in one sense, even though it holds in the other sense. But under the assumption that the workers are always paid such wages as are just enough for purchasing the necessities of life, we have the equality, $w\,m = \sum_j p_j x_j^F$, not only for $(\bar{x}^F, \bar{m})$, but also for all other feasible feeding processes (x^F, m). When the equilibrium prices prevail, we have

$$\text{(VII. 2")} \qquad w\,m = \sum_j \bar{p}_j x_j^F \quad ;$$

the wage rate is flexible so as to establish this equality. Then no discrepancy is found between the two definitions of the production costs. The principle of the maximum rate of profit requires that

$$\text{(VII. 3')} \qquad \sum_j \bar{p}_j y_j \leqq \bar{\beta}\left(\sum_j \bar{p}_j x^I + w\,m\right)$$

for all feasible processes subject to the qualification, $\ell \geqq \alpha m$; and it immediately follows from (VII. 2") and (VII. 3) .

Since the rate of profit must be maximized in the state of the Golden Equilibrium, the above inequalities, (VII. 3) and (VII. 3') hold with equality at $(\bar{x}^I, \bar{x}^F, \bar{m}, \bar{y})$. We thus have

$$\text{(VII. 4)} \qquad \sum_j \bar{p}_j \bar{y}_j = \bar{\beta} \sum_j \bar{p}_j (\bar{x}_j^I + \bar{x}_j^F) = \bar{\beta}\left(\sum_j \bar{p}_j \bar{x}_j^I + \bar{w}\,\bar{m}\right) ,$$

which, in view of the definition of $\bar{e}$ as the (normalized) total profits, may be written in the following alternative form :

$$\bar{e} = (\bar{\beta} - 1) \sum_j \bar{p}_j (\bar{x}_k + \bar{x}_j^F) = (\bar{\beta} - 1)\left(\sum_j \bar{p}_j \bar{x}_j^I + \bar{w}\,\bar{m}\right) .$$

The final inequality is required so as to make the golden e-quilibrium economically non-trivial. It means that all goods produced at that state should not be free; i.e., the equilibrium otuputs $y_1, \ldots, y_n$ evaluated at the prices, $p_1, \ldots, p_n$, are summed up to be positive :

(VII.5) $$\sum_j \bar{p}_j \bar{y}_j > 0 \quad .$$

The whole argument of establishing the existence of a Golden Equilibrium is higlhy mathematical, but is similar, in its essential character, to the proofs given in Lectures 4 and 5 .

Lecture 8

Two turnpike theorems : Finite state and consumption en route

In this lecture we are concerned with an economy, where, as in a communitarian economy, there are no capitalists and no bourgeoisie at all . We shall arrive at two different turnpike theorems ; one of them is based on the usual definition of the efficiency of growth paths in terms of the final state which the economy aims at reaching at the end of the programming period. This criterion for choosing paths will be justified if it is our object to bequeath properties to the "future" as much as we can . But if children and grand children are not to be sacrificed in the interest of great grand children, their consumption should also be the desiderata in making the long-run growth programme. Not only the final state but also intermediate states or consumptions en route must be evaluated according to some principle of welfare judgment. The unhappiest woman would be the wife of a man who stakes his whole life on happiness and glory in his last moments .

Let us assume that capitalists consume, in the state of the Golden Equilibrium of a capitalist economy, some goods which are valuable in the communist economy. We can then observe that the golden equilibrium rate of growth of a communist economy with no bourgeoisie is ceteris paribus greater than that of a capitalist economy where the capitalists for industrial investment the remainder of their income after deducting from it the part devoted for their private consumption. Thus the communist turnpike is the true turnpike that gives the fastest balanced growth.

Let us now assume the uniqueness and positivity of the communist turnpike ; that is, there is only one Golden Equilibrium aggregate process $(\bar{x}, \bar{m}) \rightarrow (\bar{y}, \bar{\ell})$, and $(\bar{x}, \bar{m})$ is a strictly positive vector. Consider a sequence of feasible aggregate processes $(x(t, m(t)) \rightarrow (y(t), \ell(t)$ such that $y(t-1) \geqq x(t)$ and $\ell(t-1) \geqq m(t)$ $(t=1, 2, \ldots, T)$, and call it, as before, a feasible path of order T. All feasible paths of the same order T may be compared with each other with respect to their final outcomes $(y(T), \ell(T))$, but in disregard of all the intermediate outcomes $(y(t), \ell(t))$ $(t=1, \ldots, T-1)$ that may be considered as goods in process developing into products at the final date. We call a feasible path of order T, $(x(t), m(t)) \rightarrow (y(t), \ell(t))$, $t=1, \ldots, T$, an efficient path of order T if there is no other feasible path of the same order, $(x'(t), m'(t)) \rightarrow (y'(t), \ell'(t))$, $t=1, \ldots, T$ such that inequalities $y_j'(T) \geqq y_j(T)$ and $\ell'(T) \geqq \ell(T)$ hold with at least one strict inequality, while $x_j'(1) \leqq x_j(1)$ and $m'(1) \leqq m(1)$ for all $j = 1, \ldots, n$. We shall show that all efficient paths thus defined in terms of the final states converge on the turnpike in a broad (or average) sense that the average output per man (discounted by the turnpike, or von Neumann growth factor $\bar{\alpha}$) over T periods of each good j,

$$\text{(VIII. 1)} \qquad \frac{\dfrac{y_j(1)}{\bar{\alpha}} + \dfrac{y_j(2)}{\bar{\alpha}^2} + \ldots + \dfrac{y_j(T)}{\bar{\alpha}^T}}{\dfrac{\ell(1)}{\bar{\alpha}} + \dfrac{\ell(2)}{\bar{\alpha}^2} + \ldots + \dfrac{\ell(T)}{\bar{\alpha}^T}}, \quad j = 1, \ldots, n,$$

approximates the turnpike output per man $\bar{y}_j/\bar{\ell}$ when T tends to infinity. It is, of course, true that this convergence does not necessarily imply the convergence of $y_j(t)/\ell(t)$ on $\bar{y}_j/\bar{\ell}$ that is asserted by the usual turnpike theorems ; but it still implies that all efficient paths cannot depart from the turnpike in the sense that if they are averaged, the averages are close to $\bar{y}_j/\bar{\ell}$'s, although they may regularly or irregularly fluctuate around the turnpike.

The following argument will give a proof of the weak turnpike theorem. It is assumed that the amounts of goods $j = 1, \ldots, n$ available in the economy at the beginning of period 1, $y_1(0), y_2(0), \ldots, y_n(0)$, and the number of workers at the same point of time $\ell(0)$, are given so as to fulfill

$$y_j(0) \geqq \mu \bar{y}_j, \qquad j = 1, \ldots, n \text{ and } \ell(0) \geqq \mu \bar{\ell}$$

with a positive number μ. It then follows from the costless disposal of labour and goods implied by the assumption (i), (iv) and (v) in Lecture 7, that the following path of any order t is feasible :

$$(y(0), \ell(0)) \rightarrow (\mu\bar{\alpha}\,\bar{y}, \mu\bar{\alpha}\bar{\ell}),$$

$$(\mu\bar{\alpha}\,\bar{y}, \mu\bar{\alpha}\bar{\ell}) \rightarrow (\mu\bar{\alpha}^2\,\bar{y}, \mu\bar{\alpha}^2\bar{\ell}),$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$(\mu\bar{\alpha}^{t-1}\,\bar{y}, \mu\bar{\alpha}^{t-1}\bar{\ell}) \rightarrow (\mu\bar{\alpha}^t\,\bar{y}, \mu\bar{\alpha}^t\bar{\ell}),$$

because we may derive on the turnpike for t periods by entering it by disposing of goods and labour in the amounts $y_j(0) - \mu\bar{y}_j$ $(j = 1, \ldots, n)$ and $\ell(0) - \mu\bar{\ell}$ at the beginning of period 1. If the turnpike social outputs $\mu\bar{\alpha}^t\,\bar{y}_j$ $(j = 1, \ldots, n)$ and $\mu\bar{\alpha}^t\bar{\ell}$ in some period t are greater than the corresponding social outputs $y_j(t)$ $(j=1, \ldots, n)$ and $\ell(t)$ of an efficient path ; i.e., if

$$\text{(VIII. 2)} \qquad \mu\bar{\alpha}^t\,\bar{y}_j > y_j(t), \quad j = 1, \ldots, n, \quad \text{and } \mu\bar{\alpha}^t\bar{\ell} > \ell(t),$$

then we may find a positive number ν such that

$$\mu \bar{\alpha}^t \bar{y}_j \geqq x_j(t+1) + \nu \bar{\alpha}^t \bar{y}_j, \qquad j = 1, \ldots, n,$$
$$\mu \bar{\alpha}^t \bar{\ell} \geqq m(t+1) + \nu \bar{\alpha}^t \bar{\ell}.$$

It is evident that costless disposal of goods and labour in appropriate amounts at the beginning of period t+1 enables us to get out from the turnpike into a hybrid path obtained by super-adding the original efficient path and the turnpike. We have

$$(x(t+1)+\nu \bar{\alpha}^t \bar{y},\ m(t+1)+\nu \bar{\alpha}^t \bar{\ell}) \rightarrow (y(t+1)+\nu \bar{\alpha}^{t+1} \bar{y},\ \ell(t+1)+\nu \bar{\alpha}^{t+1} \bar{\ell}),$$
$$(x(t+2)+\nu \bar{\alpha}^{t+1} \bar{y},\ m(t+2)+\nu \bar{\alpha}^{t+1} \bar{\ell}) \rightarrow (y(t+2)+\nu \bar{\alpha}^{t+2} \bar{y},\ \bar{\ell}(t+2)+\nu \bar{\alpha}^{t+2} \bar{\ell}),$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$(x(T)+\nu \bar{\alpha}^{T-1} \bar{y},\ m(T)+\nu \bar{\alpha}^{T-1} \bar{\ell}) \rightarrow (t(T)+\nu \bar{\alpha}^{T} \bar{y},\ \ell(T)+\nu \bar{\alpha}^{T} \bar{\ell}),$$

where $y(s) \geq x(s+1)$ and $\ell(s) \geqq m(s+1)$ for $s = t+1, t+2), \ldots, T-1$. We thus find a feasible path of order T that starts from the same initial position $(y(0), \ell(0))$ as that of the original efficient path, and arrives at a better terminus $(y(T) + \nu \bar{\alpha}^T \bar{y}, \ell(T) + \nu \bar{\alpha}^T \bar{\ell})$ than that of the latter. This is an apparent contradiction, for the efficiency of a path requires that there is no feasible path superior to it. Hence, we do not have (VIII. 2) in any period t, so that for any t there is at least one j_t such that

(VIII. 3) $$y_{j_t}(t) \geqq \mu \bar{\alpha}^t \bar{y}_{j_t}$$

or

(VIII. 4) $$\ell(t) \geqq \mu \bar{\alpha}^t \bar{\ell}$$

Let us now write $k = \mu \min(\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_n, \bar{\ell})$ which is positive by assumption. It follows from (VIII. 3) and (VIII. 4) that either the discounted output of some good j, $y_j(t)/\bar{\alpha}^t$, or the discounted number of workers, $\ell(t)/\bar{\alpha}^t$, is as great as k. In viewing that $y_j(t)$ and $\ell(t)$ are non-negative, we may a fortiori say that the contemporary sum, $\sum_j y_j(t)/\bar{\alpha}^t + \ell(t)/\bar{\alpha}^t$, can never be less than k. Taking the total sum of the sums over T periods, and writing

$$w(T) = \sum_1^T \sum_j \frac{y_j(t)}{\bar{\alpha}^t} + \sum_1^T \frac{\ell(t)}{\bar{\alpha}^t} ,$$

we get $w(T) \geqq Tk$. It can now immediately be seen that $w(T)$ becomes larger and larger when T tends to infinity, and this result will play a most important role in the following discussion of the weak convergence of efficient paths.

As aggregate processes are subject to constant returns to scale, it follows from the feasibility of the original undiscounted processes that the following discounted processes are also feasible :

$$\left(\frac{x(t)}{\bar{\alpha}^t}, \frac{m(t)}{\bar{\alpha}^t}\right) \rightarrow \left(\frac{y(t)}{\bar{\alpha}^t}, \frac{\ell(t)}{\bar{\alpha}^t}\right) \qquad t = 1, \dots, T .$$

In view of the super-additivity of processes ensured by the assumption (i), we have a feasible process,

$$(X(T), M(T)) \rightarrow (Y(T), L(T)) ,$$

which is the sum of the above T feasible processes, i.e., $X(T) = \sum_1^T \frac{x(t)}{\bar{\alpha}^t}$, $M(T) = \sum_1^T \frac{m(t)}{\bar{\alpha}^t}$, $Y(T) = \sum_1^T \frac{y(t)}{\bar{\alpha}^t}$, and $L(T) = \sum_1^T \frac{\ell(t)}{\bar{\alpha}^t}$. Taking $y(t-1) \geqq x(t)$ and $\ell(t-1) \geqq m(t)$ $(t = 1, \dots, T)$ into account, we have

$$\text{(VIII.5)} \qquad X(T) \leqq \frac{x(1)}{\bar{\alpha}} + \frac{Y(T)}{\bar{\alpha}} - \frac{y(T)}{\bar{\alpha}^{T+1}} , \text{ and } M(T) \leqq \frac{m(1)}{\bar{\alpha}} + \frac{L(T)}{\bar{\alpha}} - \frac{(T)}{\bar{\alpha}^{T+1}} .$$

Moreover, since goods and labour are freely disposable, we obtain a feasible aggregate process,

$$\left(\frac{x(1)}{\bar{\alpha}} + \frac{Y(T)}{\bar{\alpha}}, \frac{m(1)}{\bar{\alpha}} + \frac{L(T)}{\bar{\alpha}}\right) \rightarrow (Y(T), L(T)),$$

which is normalized into

(VIII. 6) $$\left(\frac{x(1)}{\bar{\alpha}w(T)} + \frac{Y(T)}{\bar{\alpha}w(T)}, \frac{m(1)}{\bar{\alpha}w(T)} + \frac{L(T)}{\bar{\alpha}w(T)}\right) \rightarrow \left(\frac{Y(T)}{w(T)}, \frac{L(T)}{w(T)}\right)$$

by dividing all the components of the process by the sum $w(T) = \sum_j Y_j(T) + L(T)$. It is evident that $\left(\frac{Y(T)}{w(T)}, \frac{L(T)}{w(T)}\right)$ belongs to a bounded, closed region $\left\{(y, \ell) \mid y_j \geqq 0, \sum_j y_j + \ell = 1\right\}$, so that it has a limit point (y^*, ℓ^*) as T tends to infinity. Furthermore, as has already been seen, w(T), which is at least as large as Tk, becomes very large when T is large enough; we may therefore safely neglect $x(1)/(\bar{\alpha}\ w(T))$ and $m(1)/(\bar{\alpha}w(T))$. We thus have from (VIII. 6) a limiting feasible process ,

$$\left(\frac{y^*}{\bar{\alpha}}, \frac{\ell^*}{\bar{\alpha}}\right) \rightarrow (y^*, \ell^*) \quad \text{or} \quad (y^*, \ell^*) \rightarrow (\bar{\alpha} y^*, \bar{\alpha} \ell^*),$$

which gives a balanced growth at the Golden Equilibrium rate . We get $y^* = \bar{y}$, and $\ell^* = \bar{\ell}$ from the assumed uniqueness of the communist Golden Equilibrium. We have thus shown that there is a long-run tendency for the average discounted output per man (VIII. 1) to approximate the turnpike output per man; the weak convergence of efficient paths is established.

It must be admitted, however, that the above argument does not ensure the strong or proper convergence (i.e., convergence of each $y_j(t)/\ell(t)$ on $\bar{y}_j/\bar{\ell}$) which is usually asserted by the turnpike theorem ; to get it we need some additional assumption, say the assumption of strong super-additivity.

Consider any two feasible processes. By super-adding them - i.e., by operating them jointly - we can form a third process in which the output for any commodity is not less than the sum of the corresponding outputs for that commodity in the two given processes, and a similar relationship holds for labour as well. If the output of some good (or the number of workers as an output) in the third process is greater than the sum of the corresponding outputs (or the sum of the corresponding numbers of workers) in the two (i.e., if the composition of the two processes makes a positive contribution to productivity), then we say that strong super-additivity prevails between the two feasible processes. This is a sufficient condition for the strong convergence, but is not fulfilled by the von Neumann polyhedral technology. Its empirical foundation would be flimsier than those of the other assumptions. Thus we may conclude that the strong turnpike theorem is not so generally accepted as the weak one.

So far the workers' consumption of goods has been discussed only from the viewpoint of the aggregate process of producing goods and manpower, but not from the viewpoint of the utility it brings to them. It is nevertheless true that even a meal for a dog may be criticized by a veterinary surgeon (and by the dog himself), so we may introduce, with stronger justification, a utility (or welfare) function which describes the preference ordering (by the workers or planning authorities) of various labour-feeding activities. Let $y_j(t-1)$ be the amount of good j available at the beginning of period t, and $x_j^I(t)$ the amount of good j used as industrial input in the same period. It is evident that the workers' consumption $x_j^F(t)$ in period t cannot exceed the remainder of the stock $y_j(t-1)$ after subtracting the input $x_j^I(t)$:

$$y_j(t-1) \geqq x_j^I(t) + x_j^F(t) \quad , \qquad j = 1, \ldots, n.$$

For each good we have a stream of the workers' consumption ,

$$\left\{x_j^F(1),\ x_j^F(2), \ldots,\ x_j^F(T),\ y_j(T)\right\}, \qquad j = 1, \ldots, n,$$

of order T+1, ending with $y_j(T)$; $x_j^F(t)$ is consumed by $m(t)$ or by $\ell(t-1)$ workers, while the disposal of $y_j(T)$ is in the hands of the $\ell(T)$ "future" people . By supposing society has a subjective time-preference factor γ such that $\bar{\alpha} \geqq \gamma \geqq 1$, we may aggregate the consumption of good j on different dates into a discounted average :

$$C_j(T) = \frac{x_j^F(1) + \dfrac{x_j^F(t)}{\gamma} + \ldots + \dfrac{x_j^F(T)}{\gamma^{T-1}} + \dfrac{y_j(T)}{\gamma^T}}{T+1} \qquad j=1, \ldots, n.$$

Similarly, we have the discounted average of the time stream of the number of workers :

$$N(T) = \frac{m(1) + \dfrac{m(2)}{\gamma} + \ldots + \dfrac{m(T)}{\gamma^{T-1}} + \dfrac{(T)}{\gamma^T}}{T+1} \quad .$$

When a Golden Equilibrium path starts from a point $(\bar{x}, \bar{m})$ in balanced proportions at the beginning of period 1, the amounts of goods consumed by the workers and the size of the labour force grow like

$$x^F(t) = \bar{\alpha}^{t-1}\bar{x}^F, \text{ and } m(t) = \bar{\alpha}^{t-1}\bar{m} \qquad j=1, \ldots, n,$$

respectively , and we have, at the end of period T, goods and labour in the amounts

$$\bar{\alpha}^{T-1}\bar{y}_j \text{ and } \bar{\alpha}^{T-1}\bar{\ell}, \qquad j=1, \ldots, n.$$

They are discounted to give

$$\bar{C}_j(T) = \frac{\bar{x}_j^F + \dfrac{\bar{\alpha}\bar{x}_j^F}{\gamma} + \dots + \dfrac{\bar{\alpha}^{T-1}\bar{x}_j^F}{\gamma^{T-1}} + \dfrac{\bar{\alpha}^{T-1}\bar{y}_j}{\gamma^T}}{T+1} \qquad j=1,\dots,n, \quad j=1,\dots,n,$$

$$\bar{N}(T) = \frac{\bar{m} + \dfrac{\bar{\alpha}\bar{m}}{\gamma} + \dots + \dfrac{\bar{\alpha}^{T-1}\bar{m}}{\gamma^{T-1}} + \dfrac{\bar{\alpha}^{T-1}\bar{\ell}}{\gamma^T}}{T+1}$$

The preference ordering (of the workers of the authorities) between the Golden average state $(\bar{C}(T), \bar{N}(T))$ and any feasible $(C(T), N(T))$, is described by a utility function $U(C(T), N(T))$, which has, in addition to the usual properties of U such as continuity, quasi-concavity, and so on, the following two :

(a) the quasi-homogeneity in the sense that

$$U(C(T), N(T)) \geqq U(\bar{C}(T), \bar{N}(T)) \quad \text{if and only if}$$

$$U(\lambda C(T), \lambda N(T)) \geqq U(\lambda \bar{C}(T), \lambda \bar{N}(T)) \text{ for any } \lambda > 0,$$

and (b) the positive utility of the Golden Equilibrium in the sense that

$$U(\bar{C}(T), \bar{N}(T)) > U(0, 0) .$$

It follows from the former that if $N(T) = \bar{N}(T)$, then

$$U(C(T), N(T)) \geqq U(\bar{C}(T), \bar{N}(T)) \text{ implies } U(C(T)/N(T), 1) \geqq U(\bar{C}(T)/\bar{N}(T), 1)$$

and vice versa ; that is to say no matter how large the population is, a path with the same discounted average number of persons as that of the Golden Equilibrium can consistently be compared with it in terms of the average consumption per man, $C(T)/N(T)$ and $\bar{C}(T)/\bar{N}(T)$;

while the property (b) implies that the Golden Equilibrium is preferred to the "desert," with nothing - no fruits, no water, ..., and no inhabitants. It is seen that (a) is rather restrictive, but (b) is very natural, in spite of the fact that it will not be accepted by Zen Buddhists who find an ecstacy in the state of nothing .

Let us assume, as in the previous section, the uniqueness of the communist turnpike (or the turnpike which is obtained when capitalists devote all their income to accumulation) ; that is to say that there are no two different communist Golden Equilibrium processes that remain different after normalization ; they are always normalized into the same process $(\bar{x}, \bar{m}) \rightarrow (\bar{y}, \bar{\ell})$. We also assume that goods and labour are available in positive amounts at the starting point, i.e., $(y(0), \ell(0)) > 0$. Being armed with these assumptions, we can establish the Consumption Turnpike Theorem which states that when the programming period T is taken long, any optimal path to maximize the utility U(C(T), N(T)) converges to the turnpike in the weak or average sense that

$$\lim_{T \rightarrow \infty} \frac{Y_j(T)}{L(T)} = \frac{\bar{y}_j}{\bar{\ell}} , \qquad j = 1, \ldots, n ,$$

where , as before ,

$$Y_j(T) = \sum_1^T \frac{y_j(t)}{\bar{\alpha}^t} , \text{ and } L(T) = \sum_1^T \frac{\ell(t)}{\bar{\alpha}^t} .$$

The criterion of choosing growth path is different from that of the Final State Turnpike Theorem, but the proof is not different in its essence .

Like the Final State Turnpike Theorem, the Consumption Turnpike Theorem does not assert anything about the convergence to the turnpike of time components $x_j^F(t)/m(t)$ and $y_j(t)/\ell(t)$, $j = 1, \ldots, n$

of the discounted averages $C_j(T)/N(T)$ and $Y_j(T)/L(T)$, $j = 1, \ldots, n$, if it is given in the weak form. What additional condition (or set of additional conditions) is enough for ensuring the strong convergence.? The one which at once comes into our mind is the strong super-additivity that has been seen in the case of the Final State Theorem to be a sufficient condition for strong convergence. It is confirmed that an exactly parallel argument can be made in the present case, but, as has been mentioned, the strong super-additivity will lack firm empirical foundations.

As for the Consumption Turnpike Theorem, however, we have an interesting case worth discussing. So far, society has been supposed to maximize the utility that depends upon the discounted averages of consumption streams and the population stream, $C_j(T)$ and $N(T)$, $j = 1, \ldots, n$. This implies that no preference is present between two paths which give the same averages; it is highly probable, however, that society will be more or less fluctuation-averting in the sense that a more stabilized stream of consumption of goods is preferred to those with fluctuations.

Let S_j be the average of deviations of the per capita consumption of good j from its equilibrium value ; i.e. ,

$$S_j(T) = \left(\left| \frac{x_j^F(1)}{m(1)} - \frac{\bar{x}_j^F}{\bar{m}} \right| + \left| \frac{x_j^F(2)}{m(2)} - \frac{\bar{x}_j^F}{\bar{m}} \right| + \ldots + \left| \frac{x_j^F(T)}{m(T)} - \frac{\bar{x}_j^F}{\bar{m}} \right| + \left| \frac{y_j(T)}{\ell(T)} - \frac{\bar{y}_j}{\bar{\ell}} \right| \right) \Big/ (T+1) .$$

It is clear that the discounted average consumption per man $C(T)/N(T)$ strongly converges to the Golden value if and only if $\lim_{T \to \infty} S_j(T) = 0$. now assume that the planning auyhorities maximize the utility:

$$U = U(C(T),\ N(T),\ S(T)), \quad \text{where} \quad S = (S_1, \dots, S_n),$$

which preserves the properties (a) and (b) with respect to C(T) and N(T) ; i.e.,

(a') $U(C(T), N(T), S(T))) \geqq U(\bar{C}(T), \bar{N}(T), 0)$ if and only if

$U(\lambda C(T), \lambda N(T), S(T))) \geqq U(\lambda \bar{C}(T), \lambda \bar{N}(T), 0)$ for any $\lambda > 0$,

(b') $U(\bar{C}(T), \bar{N}(T), 0) > U(0, 0, 0)$.

An additional property that is used to establish the strong convergence is the assumption of fluctuation aversion :

(c) U is a decreasing function of S .

It follows from the free disposal of goods and labour, and the uniqueness of the turnpike, that goods and labour are not free at the state of Golden Equilibrium ; this, in turn, implies that $y_j(t)$, $x_j^F(t)$, m(t) and $\mathcal{L}(t)$ cannot grow at a long-run rate greater that $\bar{\alpha} - 1$; hence, $C(T)/\bar{N}(T)$ and $N(T)/\bar{N}(T)$ are bounded from above . By putting $\lambda = 1/\bar{N}(T)$ in (a'), we have

$$(VIII.7) \qquad U\left(\frac{C(T)}{\bar{N}(T)}, \frac{N(T)}{\bar{N}(T)}, S(T)\right) \geqq U\left(\frac{\bar{C}(T)}{\bar{N}(T)}, 1, 0\right)$$

for an optimal path. From the negativity of the effect of S(T) on U and the boundedness of $C(T)/\bar{N}(T)$ and $N(T)/\bar{N}(T)$, it is seen that there shold be an S^* such that $S^* \geqq S(T)$; otherwise the inequality sign of (VIII.7) will be reserved. An when the negative effect of S(T) on U is very strong , i.e. the very strong fluctuation - aversion prevails, S^* is nearly equal to zero ; then $S_i(T) \cong 0$ for $i = 1, \dots, n$.

We obtain the strong convergence of $x_j^F(t)/m(t)$ to $\bar{x}_j^F/\bar{m}$, j=1,...,n.

A sharp eye will not fail to find that strong convergence of the consumption per man, $x_j^F(t)/m(t)$, does not imply that of the output per man $y_j(t)/\ell(t)$ to $\bar{y}_j/\bar{\ell}$, and that it is the latter property that the strong turnpike theorem asserts. A simple device to get it would be to assume, instead of the aversion of fluctuations in per capita consumption of goods, that the people or the authorities prefer a more balanced growth of outputs to a growth at the same long-run rate but with fluctuations. Such an aversion of fluctuations in outputs would easily be incorporated into the present model only by re-defining $S_j(T)$ as

$$S_j(T) = \left(\left| \frac{y_j(1)}{\ell(1)} - \frac{\bar{y}_j}{\bar{\ell}} \right| + \left| \frac{y_j(2)}{\ell(2)} - \frac{\bar{y}_j}{\bar{\ell}} \right| + \dots + \left| \frac{y_j(T)}{\ell(T)} - \frac{\bar{y}_j}{\bar{\ell}} \right| \right) \Big/ \; T \; .$$

It may be noted, however, that no justification will be found for the aversion of output fluctuations ; in fact, people will be prepared to accept cycles of outputs if they are technically required for an optimal growth . We may , therefore, conclude by saying that the introduction of the device of very strong fluctuation-aversion will eliminate cycles of the per capita consumption of goods, but not those of the per capita outputs ; we still cannot get rid of cyclic exceptions from the turnpike theorem.

CENTRO INTERNAZIONALE MATEMATICO ESTIVO

(C.I.M.E.)

B. MARTOS - J. KORNAI

EXPERIMENTS IN HUNGARY WITH INDUSTRY-WIDE AND ECONOMY WIDE PROGRAMMING

Corso tenuto all'Aquila dal 30 agosto al 7 settembre 1965

EXPERIMENTS IN HUNGARY WITH INDUSTRY-WIDE AND ECONOMY-WIDE PROGRAMMING.

by

B. MARTOS - Institute of Economics of the Hung. Acad. Sci.

J. KORNAI - Computing Centre of the Hung. Acad. Sci.

1. Introduction

The aim of our lecture is to make known some experiments which were and are carried out by Hungarian economists and mathematicians in order to apply the mathematical programming methods to the preparation of economic plans aiming at more favourable proposals, than those based on traditional planning methods.

We do not mention in our lecture other mathematical planning methods, - e.g. the input-output analysis - which are also used in Hungary. We speak only about the application of mathematical programming methods. Furthermore we pass over the wide-spread, intrafirm results in operations research, which apply sometimes programming techniques as well. We deal only with more wide-ranging problems: planning a whole industry, more industries or the economy as a whole

Although our lecture deals with applications of mathematical programming, its character is non-mathematical, because in most cases the research workers were not compelled to invent new mathematical techniques, they only had to seek for ways and means, how to use in practice the theoretically wellknown procedures .

2. Sectorial programming.

The first experiments with mathematical programming methods in economic planning of industries started in our country about the end of the fifties. In our opinion there is a twofold reason for achieving the first results in sectoral programming :

1./ In the Hungarian economic system - then and now- a very considerable part of economic decisions is made by organizations spanning a whole industry, in the powers of a ministry or of a board of industrial directors. The firm itself possesses only a very limited power to choose its line of products, its buying and selling partners, etc. The bulk of investment is decided upon by higher industry-wide authorities. Thus the application of programming methods naturally turned to those economic units , where the essential decisions arise.

2./ Of course, the most important economic decisions are made not on sectoral , but on economy-wide level,by the Council of Ministers or by the Central Planning Board. The main reason now, why the research workers did not start with economy-wide planning, is that they discovered : the lack of theoretical knowledge, practical experience, data availability and suitable computing equipment all dictated for the beginning a moderate framework. They could turn to economy-wide planning only later, armed by due experience.

The very first sectoral programming models were of a short-run character, they served annual "operative" planning purposes. Accordingly they did not contain investment activities. A quadratic programming model for transporting and refining sugar-beet[1] and a model for production and foreign trade of paper and pulp [2] is worth mentioning

---

[1] Elaborated by G. Kondor [3]

[2] The first version of this model has been set up by a team headed by Dr. I. Aczél [1] and further developped by B. Martos [6] .

here. Even the first programming models, which are used now regularly, were very difficult to handle with our small computers (the latter had several hundred variables and more than a hundred constraints[3]). But the mathematical programming in the paper and pulp industry yielded, as compared to the plans set up by traditional methods, about 8 % savings in import costs, besides it kept the domestic paper consumption on the same level and did not change the domestic labour and material requirements.

Simultaneously begun the first experiments with sectoral models for a longer period containing investment-activities, too. The first of this kind was the model of the cotton-weaving industry[4], followed by the man-made fibres industry[5]. With the man-made fibre model several sets of computations were performed, as objectiv function also a nonlinear, concave cost function was applied and the risk, arising from future price fluctuations, was taken into account as well. Another sectoral long-range model was fitted to the bauxit-aluminium industry[6].

The main significance of all these sectoral models - though their results were apt to be employed in sectoral decisions - consists in preparing the soil for an economy-wide model. Two characteristic deficiencies of the sectoral models called the attention to the necessity of such an economy-wide model. First : the sectoral models inevitably assumed at the outset that certain resources of the economy had been distributed previously among the industries. But these preliminary allocations were clearly non-optimal.

---

3) The first version was computed in the Soviet Union.

4) 5) J. Kornai [4]

6) B. Martos-J. Kornai-A. Nagy [7]

Only the dual solutions of the sectoral models can give some orientation as to the efficiency of the original and assumed allocation. Second : however emphatically the modelbuilders endeavoured that the objective function of a sectoral model should not only express the particular interests of the sector in question but that of the whole economy, too , they could not entirely achieve the latter because of the artificial nature of the existing price-system and other reasons.

All these considerations naturally led to the requirement that the programs of single industries should be interconnected in the framework of an economy-wide model. The present state of the research work tending toward this end will be discussed in a more detailed manner in what follows.

3. / The purposes of the economy-wide programming.

A team of several hundred economists, mathematicians, engineers and practical planners is at present engaged in Hungary in the preparatory work of the first experimental economy-wide programming project.[7)] The research work is in direct connection with drawing up the national economic plan for the years 1966 - 1970 .

It is but natural that our work should show many analogies with the applications of mathematical programming in other fields. Its essential characteristic, on the other hand, lies in the fact that we have endeavoured to employ the method under socialist conditions and on the national level, as a tool in drawing up the five-year plan and as an integral part of actual practical planning work.

In the construction of our model we endeavoured to meet

[7)] Sponsored by the National Planning Board, the research work is carried out under the direction of J. Kornai.

the following requirements.

1/ We set ourselves the task to construct a computable model. Accordingly, we have employed a linear programming model. We are fully aware of the fact that by using non-continuous variables in our model beside the continuous ones, by representing certain relationships by non-linear equations, by carrying out stochastic programming etc., a truer reflecting of economic reality could be achieved. However, for the purposes of the first economy-wide programming project we thought it preferable to revert to the most elementary type of programming model easiest to manage - and even so we had to surmount extreme difficulties as regards computing techniques.

2/ Our model has been constructed in a way as to make it in several respects conform to the structure and the index system of traditional planning. In doing so we had two aims in mind. First, to be able to meet the greatest possible part of the project's immense data requirements by drawing on data compiled for the purposes of traditional planning. In addition, we should like our results to be, as far as possible, directly comparable with the plan targets determined by traditional methods. (The latter will in the sequel be called the official program) In its relation to traditional planning our research work shows thus a peculiarly dual character : on the one hand, it relies on the official program, on the other hand, it competes with it.

The second requiremente - that of conformity to the structure of traditional planning - means at the same time that we make our mathematical model to simulate to a certain extent the course and the patterns followed up to now in setting up the five-year plans. This may prove useful in working out a theory of socialist planning .

3/ Our calculations must be sufficiently detailed to yield utilizable information not only for the central work carried out at the National Planning Board but also for planning at the ministry level. Extreme aggregation in the sector breakdown of the model must be avoided and planning must be possible within the project framework for the most important investments, productive and foreign-trading activities.

Unfortunately, the latter requirement - together with the first one of computability - has compelled us to make certain concessions : we had to renounce to making the model express the time schedule of the activities. The main argument for the less aggregate but non-dynamical model as against the more aggregate but multi-period model was that traditional planning, too, belonged to the former category. Long-term planning for 15-20 years is still in the elementary stage in this country and the preparation of the five-years plans is not as yet connected in any organized form with working-out long-term plans to cover 15 to 20 years periods. Had we, therefore, constructed a programming model covering two or three five-year-plan periods, we would not have been able to meet the second requirement mentioned above ; the data compiled for the purposes of traditional planning would not have provided an adequate basis for our work, nor would it have been possible to compare the results obtained with the official program.

We are well aware that the single-phase character of the model constitutes the greatest weakness of our research work, compelling us as it does to make a number of highly simplifying assumptions. It is to be hoped that in a second experimental calculation it will be possible to remedy this deficiency.

The national model is composed of 40 sector models. In the initial stage of the calculations each sector model has an "individual life" - it lends itself to independent interpretation and constitutes a tool in the service of lower-level planning.

The individual sectors are generally responsible for several - six to ten in some cases fifteen to twenty - product groups, aggregates composed of a variety of concrete products. (E. g. bricks, or enamelware, or TV receiver sets etc.) In the following, both the product aggregates and the services will be called products.

With the concept of the product thus defined, our economy-wide model contains a total of some 400 products. The major part of the variables figuring in the model and representing economic activities is connected with some product (e.g. the establishment of a plant to produce the product in question, its production itself, or its import or export etc.) .

The variables representing the economic activities are divided into two main groups : capital transformation variables and operation variables .

The capital transformation variables represent the economic activities which result in transforming a certain part of the capital stock, of the production and turnover funds, from a definite state in 1966 into a definite state in 1970. Let us give a few examples.

- The 1966 state of a definite productive, transport or service capacity is preserved on the unchanged technological level until 1970 . The transformation will in such cases require inputs of a maintenance character.

- A definite industrial plant, power station, railway line etc. is in 1966 already in existence. This will be transformed in the course of the five-year plan period in a definite way and brought into a terminal state which is different from the initial one. (E. g. the plant's old machinery will be partially replaced by modern ones ; the railway line will be provided with an automatic safety equipment ; some branch of plant cultivation will be further mechanized etc.) Here, the activities aimed at the preservation and the transformation of the 1966 state will be closely interrelated.

- An entirely new plant, transport or service capacity will be brought into existence by 1970. In this case we are dealing with a transformation which creates a 1970 capacity in the place of a "zero capital stock" in 1966.

In these examples we spoke of production funds and the transformation of productive capital stocks. The treatment of the country's stock of foreign assets and liabilities is to a certain extent analogous. The credit-raising and debt-repayment activities figuring in the model will increase or decrease the terminal 1970 state of assets and liabilities in the money markets abroad as compared to the initial state in 1966.

The common purpose of all capital transformation variables is to create capacities and possibilities for the economic operations in 1970. Let us now give a few examples also of the operation variables.

- The production of a definite product in 1970. Under this heading are classed all productive activities, including those which will contribute only indirectly to the production of the 400 products of our model .(E. g. the variables representing the production of various semi-finished products which will not leave the plant in this form, as well as the compounding of such products in the same plant in some of the chemical sectors etc)

- The export and import of definite products in 1970 in definite market relations .

- The collection of interest accrued on credits granted abroad and the payment of interest on foreign debts in 1970.

From what has been said it will be clear that the program obtained by means of the model constitute a complex investment, technical development, production, international financial, export and import plan.

Our model contains a great number of constraints. The first main group of the constraints limits the capital transformation activities, particularly under three aspects :

- From the side of the initial state : e.g. the activities to preserve the initial state are limited by the stocks existing in 1966.

- From the side of the terminal state : e.g. the surplus capacity which can be created by means of the technical recomstruction of an old plant is technically given.

- From the side of the inputs required for the transformation: investment resources are considered limited. Within these, and in accordance with the conventions of traditional planning, separate limits are set to the quotas of domestic, socialist import, and capitalist import machinery available for investment purposes; to the quota of construction etc. Separate limits are also imposed on the utilization of certain specific types of machinery and construction activity.

Another main group of constraints limits the 1970 activities, primarily under the following aspects :

- Technological equations : the regulation of technical relations between the raw materials, semi-finished and finished products.

- The obligation to satisfy final domestic demand. The non-productive requirements of the population and public institutions are taken given.

- Foreign trade constraints. The upper bounds representing the marketing difficulties in export; export obligations undertaken under international agrèements as lower bounds; the representation of certain "tie-up"deals such as e.g. export obligations to set off purchases of scarce materials and commodities ; global balances of payments and trade ; and so on.

- The constraints of living labour available. First of all, a non-specified global constraint of the labour force is prescribed. (But this is meant only as the upper constraint ; the total utilization of the potential supply of labour will not be regarded as compulsory, because the non-utilized manpower might also be utilized to bring about an increase in leisure time.) In certain domains - such as the supply of engineers or skilled labour, or in research capacity etc. - the available "intellectual capital" is considered as specially limited. An upper bound is also set to the wage fund; this is needed to prevent spending capacity to exceed the supply of commodities and services envisaged when determining final domestic demand.

- The scarce natural resources. These include, first of all, the area of cultivable land, with the different grades of soil quality. Geological resources also come under this heading.

Finally, the third main group of constraints includes those regulating the relationship between capital transformation variables on the one hand and operation variables on the other. Whenever this relationship is unequivocally determined, capital transformation and the activity in 1970 are represented by a common variable. Let us assume e.g. that with a capacity already in operation in 1966 and to be preserved on an unchanged technological level, a single commodity can only be produced. Then, the same variable will represent in the model the preservation between 1966 and 1970 , and the operation in 1970. In other cases, however, it will be expedient to connect the two types of activity by means of constraints. E.g. the same productive capacity will produce a variety of products; then, the constraint will prescribe that the capacity requirements of the alternative productive activities for 1970 must not exceed the capacity brought into being by that deadline.

In the sector-level calculations various types of objective function are employed parallelly. (E.g. the minimization of costs in 1970 ; the maximization of the surplus of the 1970 balance of payments etc) . The sector models were used to carry out a number of sensitivity tests and parametric programming calculations, and the experiences thus gained have served as the basis for our proposals submitted to the management of the sector concerned.

4. The linking up of the individual sector models : the "two level planning"

In reality, the sector models are not autonomous but related to one another by numerous links. From this point of view the constraints of the sector models may be classed into two main groups, viz. intrasectoral and intersectoral constraints.

The intrasectoral constraints regulate the "internal affairs" of the sector. They include those of the technological equations which describe the flow of products within the sector ; the constraints of the sector's initial capital stock, initial capacities, etc. ; the individual export marketing constraints relating to the sector's products.

The intersectoral constraints, on the other hand, regulate the sector's "external affairs". They include all equations which describe the flow of products between the individual sectors. (E.g. electric energy constitutes the output of the electric-energy sector and an input for all other sectors - the balance of electric energy must therefore be considered an intersectoral constraint for all sectors.) The same applies to the allocation of the resources which are being drawn upon by several sectors (e.g. the gross investment quota, the wage fund etc.).

Let us introduce the following notation :

A_i = matrix of the coefficients figuring in the intersectoral constraints of the i-th sector. (In our model i=1, 2, ..., 40, but for the sake of a more general formulation we will speak of n sectors.)

B_i = matrix of the coefficients figuring in the intrasectoral constraints of the i-th sector

u_i = vector of the intersectoral constraint constants of the i-th sector

b_o = vector of the macroeconomic bounds of the intersectoral constraints

b_i = vector of the intrasectoral constraint constants of the i-th sector

c_i = vector of the objective function coefficients of the i-th sector

x_i = the program of the i-th sector

$x = \left[x_1, x_2, \ldots, x_n \right]$ = the national program

x_i^{off} = the official program of the i-th sector as worked out on the basis of traditional methods

x^{off} = the official national program

As pointed out before, the calculations are carried out in two phases. In the first phase, each individual sector carries out its own programing project separately ; in the second phase the sector models are linked up and combined into a single large-scale economy-wide model. The separation of the two phases is motivated solely by practical considerations. The sectors will not be ready with their respective models at the same time and it is our wish to use the period of waiting for one another to carry out useful calculations. Besides, experience has shown that in the early stages the models will contain many errors the elimination of which requires repeated checking calculations and practical tests. It is considerably easier to carry out these operations of "running in" on the sector models than on the national model with its very large dimensions.

Let us now consider the <u>first phase</u>. First of all, we will determine the intersectoral constraint vector u_i so as to conform to the official sector program x_i^{off} :

(1) $$u_i = A_i \; x_i^{off}$$

When constructing the model it should be ensured that the official program satisfies also the intrasectoral constraints :

(2) $$B_i x_i^{off} = b_i$$

Should, in exceptional cases, this condition not be fulfilled, we will proceed to correct the official program and will in the sequel consider this corrected program, which satisfies condition (2), as x_i^{off}.

Now we will proceed, by means of the electronic computer, to determine program x_i^*, which means the solution of the following linear programing problem

(3a) $$A_i x_i = u_i$$

(3b) $$B_i x_i = b_i$$

(3c) $$x_i \geqq 0$$

(3d) $$c_i' x_i \longrightarrow \max!$$

We would like to avoid the term "optimal" to distinguish program x_i obtained as solution of problem (3), as the optimality of this program is rather relative. (It will depend to a considerable extent, among others, on intersectoral constraint vector u_i.) We prefer to call it the <u>dominant</u> sector program, because it generally dominates the official sector program x_i^{off}; both x_i^* and x_i^{off} satisfy the condition (3a) to (3c), and at the same time the dominant program is considerably better than the official one from the point of view of objective function (3d)[8].

---

8) For the sake of simplicity, we will disregard the theoretically not entirely impossible case where $x_i^* = x_i^{off}$.
In our actual practice up to now this has not yet occurred.

Let us now describe the <u>second phase</u>. Here, the sector models are combined into a single large-scale national model. We are faced with the following linear programing problem :

(4a) $$A_1 x_1 + A_2 x_2 + \dots + A_n x_n = b_o$$

$$B_1 x_1 = b_1$$

(4b) $$B_2 x_2 = b_2$$

$$\dots$$

$$B_n x_n = b_n$$

(4c) $$x_1 \geqq 0, \quad x_2 \geqq 0, \dots, x_n \geqq 0$$

(4d) $$c'_1 x_1 + c'_2 x_2 + \dots c'_n x_n \rightarrow \max !$$

In relation to our computing possibilities the dimensions of problem (4) are enormous ; it contains, in fact, several thousand variables. We will, therefore, have to content ourselves with an <u>approximation</u> of the macroeconomic program which constitutes the exact solution of problem (4) . Before dealing with the method of approximation, it should be pointed out that by the end of the first phase there will be already at least two national programs available, viz.

$$x^{off} = \left[x_1^{off}, \ x_2^{off}, \ \dots, \ x_n^{off} \right],$$ the official

program of the national economy ; and

$$\bar{x} = \left[x_1^*, \ x_2^*, \ \dots, \ x_n^* \right]$$, the aggregate of the dominant sector

programs obtained as a result of the calculations carried out in the first phase, - as the solution of problems (3) . Let $\bar{x}$ be called in the following the national program of the first approximation.

The macroeconomic constraint vector of the intersectoral constraints - b_o - can now be determined, in a manner analogous to calculation (1), as follows :

(5) $$b_o = \sum_{i=1}^{n} A_i x_i^{off} = \sum_{i=1}^{n} u_i$$

The following statements can now be made about the two national programs already known :

1) Both the official national program x^{off} and the national program of the first approximation $\bar{x}$ are feasible, i.e. they satisfy the conditions (4a) to (4c) .

2) The national program of the first approximation $\bar{x}$ dominates the official national program : according to objective function (4d) it is more advantageous than the latter.

In the second phase of the calculations the aim is to find a program which is more advantageous, according to objective function (4d) than the one of the first approximation, i.e. which dominates the official program to an even greater extent.

From what has been said it becomes clear that in the two phases of the calculations we are gradually "drifting away" from the official program. In the first phase vector u_i, the official intersectoral distribution of the intersectoral constraints as derived from the official sector program in accordance with equation (1) , was still considered binding. In the second phase this restriction is already removed. (I.e. $A_i x_i$ may be greater or smaller than u_i, as the case may be.) It is now only the constraint b_o relating to the national economy as a whole that is derived from the official program in accordance with equation (5), while for the distribution of the constraints b_o among the sectors our mathematical model is left free scope.

It is an obvious precondition to constructing problem (4), to linking up the individual sector models, that each intersectoral constraint should be strictly analogously interpreted in each sector model. In some cases this is easy to ensure, in others, however, it involves great difficulties. Let us mention but one typical example. The output sector will generally strive to plan its output or outlet in a considerably more detailed breakdown than the user sector will be able to state its requirements. (E. g. the output of the man-made fibres industry is given in more detailed breakdown in the man-made fibres industrial sector than the man-made fibres input of the textile industry in the textile-industrial sector.) In the economy-wide problem linked up according to (4) it will, therefore, be necessary to insert in the model's corresponding places certain desaggregating constraints and variables. The latter will serve to break up the aggregated requirements of the user sectors for the output sector concerned.

In the course of linking up the sector models, a whole range of other problems will also emerge (thus e.g. in connection with the constraints regulating the output and distribution of the sectors which are producing capital goods etc.) - these can, however, not be dealt with here because of our restricted space.

The given state of computing techniques in this country does not allow a direct solution of the largescale programming problem (4) by means of some usually employed algorithm (as, for instance, the simplex method) . Instead, we will have to employ one of the so-called decomposition methods, , making use of the block-diagonal arrangement of the matrices B_i .

After theoretical investigations in several directions and practical calculatory tests we have come to the decision to carry out our first experimental calculations on the basis of the Dantzig-Wolfe method [2]. Similarly to the other decomposition methods, this is a rather lengthy procedure. It has, however, the great advantage that it brings about a monotonic improvement in the value of the objective function. Thus, we will obtain a workable result even in the case that we are compelled to stop the iteration before attaining an optimum.[9)] Moreover, the Dantzig-Wolfe method affords a possibility to profit at the beginning of the economy-wide calculations from the programs worked out in the first phase of sector-level programing, and to improve the value of the objective function as against $\bar{x}$, the national program of the first approximation,

---

9) Another method had been worked out originally from the purpose of the economy-wide planning problem by J. Kornai and the mathematician T. Lipták [5] . The method in question is based on the game-theoretical interpretation of the problem and utilizes the so-called method of fictitious play. The principal advantage of this latter method of fictitious play consists in the fact that here the dimensions of the model will practically not be limited by the computer's storing capacity. The intersectoral part of the problem will, as a matter of fact, require no "regular" linear programing but only the carrying out of a set of considerably simpler operations (such as the calculation of arithmetic means etc.). However, the method is not monotonic : while it approaches the optimum the value of the objective function is strongly fluctuating. It is exactly the monotonic character of the convergence that we consider the principal advantage of the Dantzig-Wolfe method.

from the first step of the iteration already.

The economy-wide planning carried out on the basis of decomposition methods was termed two-level planning. (The term had been originally introduced to designate the algorithm using fictitious play: but in our opinion its generalization is wholly justified.) The work of planning is being carried out on two "levels" : partly within the sectors themselves and partly at the National Planning Board, the central government agency responsible for and directing the sectors. On both levels , a certain amount of initial information will be available. Moreover, in the course of the planning process, information will flow between the two levels. The information "output" of the Planning Board's calculations will constitute the information "input" of the sector-level calculations and conversely. It is exactly here that the various decomposition methods differ from one another : in what to consider the initial information at the two levels : in the information that is flowing between the two levels ;and in the character of the calculations employed in their processing .

The practice of traditional planning also shows a procedure of a similar character. Previous to drawing up the individual five-year plans, the government - or rather , on the government's behalf the Planning Board- would lay down the so-called "planning methodology", prescribing the various phases of national economic planning ; determining the extent to which the central plan figures must be broken up and "planned back" by the various ministries; and so forth. The decomposition methods will determine the algorithms of the complete planning process, algorithms that will ensure the approximation of the program with the maximum objective function value (or, e.g. in the case of the Dantzig-Wolfe method, its attainment in a finite number of steps).

In the course of traditional planning, the so-called method of plan coordination is used to ensure the equilibrium of the plan. What this means in actual practice is that the Planning Board and the staff of the planning departments in the ministries try to coordinate their self-established plan figures. The coordination of a long-term plan entails hundreds and thousands of proposals and counter-proposals, memoranda and counter-memoranda, minor and major conferences, debates and telephone calls. The plan which will finally appear to most participants in the coordination more or less acceptable is shaped in bargaining and in collective discussions between many hundred planners. This intricate process may be regarded as an "exploration", by means of trial and error, of the solution of an immense equation system consisting of several ten-thousand plan figures as unknowns and of the equations which express their mutual relationships. The mathematical programing of the national plan - and especially the method of two-level planning - provides a mathematical formulation of the plan coordination process, of the general coordination of the plan figures, and mechanizes it by means of the electronic computer.

On the basis of the traditional planning methods it will take 2 to 3 years to draw up a five-year plan. During that period sevelar complete plan proposals may be drawn up one after the other, based always on the latest informations and on the latest instruction received from the governing political bodies. But never so far have several complete plan variants been submitted at the same time and parallely to the decision-makers in Hungary. And it is here that the significance of mathemetical programing lies, in the fact that by means of sensitivity tests and parametric programming it lends itself for drawing up parallelly e whole range of complete national plan variants.

What is more, these plan variants will not be simply feasible (i.e. complying with the model's constraint constants) and realistic plans in the state of equilibrium but will be efficient plans at the same time. (It will be remembered that the criterion of an efficient plan is that there exists no other plan superior to it in every respect, only plans which may be more advantageous in one respect and less advantageous in another respect than the efficient plan.)

It will be clear from the above that the task of our investigations is not to define a single national program which would then be unequivocally recommended for execution. Our task will be successfully fulfilled on the day when the two-level mathematical model of the national economy is completed and available for economic administration to carry out experiments on all levels : to work out plan variants and to study the possible consequences of their own decisions. The data of the model will be continually replaced in accordance with the latest informations : the activities which are no longer timely will be eliminated and the newly emerging possibilities included. The mathematical programming model should thus become a permanent tool of continuous planning.

5. Results of the work so far.

At the date of submitting the manuscript (July 1965) we already have achieved some results in some parts of the work on two-level planning. The suggestions as to the common structure of the sectoral models have been fully detailed. Now we are essentially in the "first phase" of the whole process, that is, in the period of constructing and solving the sectoral models. The interconnection of the sectors, the first solution for an economy-wide program will ensue presumably in the first half of the next year.

The following table shows the present state of the sect oral programs.

Industries	Number of the sector - models		
	projected	set up	solved
Fuel, energy, chemistry	9	7	2
Metallurgy, engineering	11	4	-
Textille, paper	4	4	4
Food	5	5	5
Building, b. material	3	2	-
Agriculture	6	3	-
Transport	2	1	-
Sum	40	26	11

Let us see some results from the area of foodstuffs, one of the most advanced from our point of view[10)]

One model includes the sugar, the swetmeat and the alcohol industry, the number of structural variables is 34, that of the constraints is 50. The computation was accomplished in two versions. The difference between them consisted partly in the export and import prices,

[10)] The sectoral models of the food-industry have been elaborated under the direction of P. Bod.

partly in the second version the available amount of sugar-beet, manpower and investment found being 10 % higher. Both programs are differing considerably from the "official plan" of the sector. For instance the first computation showed that in contrast with the official plan a part of the produced molasses is to be exported and the arising lack of alcohol should be covered by import. This program vielded 8 % more foreign exchange income than the official one. A second computation showed profitable to prolong the producing period of the sugar refineries and instead of the reconstruction of these refineries it brought the production of fodder-yeast in prominence.

On the other hand another sectoral program, which included the vegetable oil, detergents and cosmetic products, showed no structural difference from the official program, the surplus yield was 5,5, %.

The third model consisted of the mills, the breweries and the wineries . We solved here again two versions, departing from different assumptions. Both versions left the official plan of the wineries unchanged. But they differed one from another and both from the official plan considerably as regards the mills and breweries. They showed improfitable to renew some obsolete equipments.

These few examples suffice to prove that already in the first phase of the work we got some non-trivial results in economic sense, helping to improve our five-year plan.

We expect the full development of the "two-level programing" to yield further valuable contribution to this task.

Furthermore we expect that our experiments will contribute in assessing the domains of the national economy, where application of mathematical programming is appropriate, and separate those, where other means, (traditional planning methods, or traditional market-forces), would serve with higher efficiency.

REFERENCES

[1] I. ACZÉL : Egy iparág optimális programjának meghatározása korszeru módszerekkel (Establishment of the optimum program for an industry by up-to-date methods). Ipargazdaság, 1961. N. 1.

[2] G. B. DANTZIG - Ph. WOLFE : The decomposition algorithm for linear programs. Econometrica, 29/1961/ 767-778.

[3] G. KONDOR : Combined transportation and processing program for sugar-beet. Economics of Planning, (forthcoming).

[4] J. KORNAI : Mathematical planning of economic structure. Publ, House of the Hung . Acad. Budapest. - North-Holland Publ. Co. Amsterdam. (forthcoming) .

[5] J. KORNAI :-T. LIPTÁK: Two-level planning. Econometrica, 33 (1965) 141-169.

[6] B. MARTOS - A. GELEI : gyástton a papiripar? Egy programozás eredményei. (What to produce in the paper industry? Results of a programming process) Ipargazdaság, forcoming)

[7] B. MARTOS - J. KORNAI - A. NAGY : Lineáris programozási modell a magyar aluminiumipar távlati tervezésében. (Linear programming model for perspective planning of the Hungarian aluminium industry.) Comp. Center. Hung. Ac. Sci. (Mimeographed) .

DISCUSSION

M. MORISHIMA : In the data which are obtainable from the traditional planning and in the official programm itself there may be certain errors. These errors are transferred into the coefficients of the model.

B. MARTOS : Yes, you are right. But as I pointed out in my lecture the enormous data requirement of the model could not be satisfied from any other source. On the other hand if we used other data or corrected those used by the planning authorities this would mean that our assumptions are different and we would be unable to compare the result of our programming with the official plan.

M. MORISHIMA : What are the criteria used when comparing the official plan with the result of your linear programming model ?

B. MARTOS : We do not work with just one objective function. A variety of objective functions will be considered and a corresponding variety of plan proposals will be presented to the authority. They may then see what are the consequences of the different objectives and may choose one among these programs, to accept it as the official plan . We do not suppose that the authority has a predetermined objective function. Having a variety of plan proposals it is, however, possible to decide on them by considering their content and not just the objective functions.

CENTRO INTERNAZIONALE MATEMATICO ESTIVO

(C.I.M.E.)

PROBABILITY DISTRIBUTION PROBLEMS CONCERNING STOCHASTIC PROGRAMMING PROBLEMS

András PRÉKOPA

Corso tenuto all'Aquila dal 30 agosto al 7 settembre

1965

PROBABILITY DISTRIBUTION PROBLEMS CONCERNING STOCHASTIC PROGRAMMING PROBLEMS.

by

András PRÉKOPA

Tulane-University - New Orleans

1. Introduction .

Different kinds of stochastic programming models are formulated in the present mathematical programming literature. Their solutions lead to linear or non-linear deterministic programming problems. There are, however, a number of problems, mainly probability distribution problems , which remained unsolved which are nevertheless important and necessary to solve in order to be able to handle effectively these stochastic optimization problems.

The main types of stochastic programming problems are the following.

a) Chance constrained programming. The problem is to minimize the expectation of a functional $z(c, x)$ under the condition that

$$P(g(A, x) \geq b) \geq \alpha \tag{1.1}$$

where g is a certain function of the elements of the matrix A and of the elements of the unknown vector x . α is a prescribed probability usually near 1. In many practical cases the above problem reduces to the simpler form :

$$\text{minimize } E(c' x)$$

subject to the condition

$$P(Ax \geq b) \geq \alpha . \tag{1.2}$$

The matrix A and the vector b are partly or entirely random. Among several papers where this problem is investigated we mention [20] and [21] .

One practical problem belonging to this class is the nutrition problem. We have several different kinds of meals each of which contains a certain amount of the different nutrients (the number of which varies from problem to problem). We wish to program the different amounts of meals for a period say two weeks for a population containing a certain number of individuals. Our problem is formulated for one individual whose demand in the different nutrients is therefore a random vector b . If $\alpha = 0,95$ then our solution will satisfy roughly speaking the nutrient demand of at least the 95 % of the polulation with the lowest expected price. . If c is constant then $E(c'x) = c'x$.

b) <u>Two-stage stochastic programming problem.</u> This problem was formulated by Dantzig and Madansky [22] and is the following . We have the following constraints

(1.3) $$A x + B y = b , \quad x \geqq 0 , \quad y \geqq 0$$

where b is a random vector. For a given $x \geqq 0$ and for a fixed realization of b we have an optimal solution $y(b, x)$ to the problem

(1.4) $$B y = b - A x , \quad y \geqq 0$$

provided the problem is solvable i.e. the set of y's satisfying the above equation is non-empty. We wish to minimize with respect to x the functional

(1.5) $$c'x + E\left[f'y(b, x)\right]$$

where c and f are fixed vectors of the same dimension as x and y, respectively .

Such problems arise in case of a random demand where e.g. x may represent certain stocks of the considered goods, raw materials on which we decide previously in time but not mathematically and after that perform an optimization problem having observed the actual demand b.

c) Paying penalty for infeasibility. Several models are formulated in which we have to decide on a non-random x such that it minimizes the total expected cost which consists of two parts, an original cost of the activities and the expected cost what one has to pay for the eventual non-feasibility of the solution x .; In connection with this problem we refer to the paper [23] .

d) Minimum expected deviation model. A penality what one would pay for the infeasibility cannot, however, be formulated in all cases. In other words in certain problems it would be hard to say what is that sum what one pays for infeasibility. We may use instead of c) a model of the following type. If we originally had the problem

$$\text{minimize} \quad c'x$$
$$\text{subject to} \quad Ax \geqq b$$

but A, b, c are partly or entirely random then we may formulate instead the following :

(1.6) $$\text{minimize} \quad E\left[\max_{1 \leqq i \leqq m}\left(b_i - \sum_{k=1}^{n} a_{ik} x_k\right)\right]$$

subject to the condition

(1.7) $$E(c'x) \leqq d$$

where d is a given price level and the a_{ik}'s, b_i's, x_k's are elements of A, b, c, respectively. . If c does not depend on x then $E(c'x) = (E(c))x$.

e) Programming by expectations. We replace each random variable in our programming problem by its expectation and carry out the solution with these.

In problems a) b) c) d) we may obviously replace the expectation by some other functional defined on the probability distributions of which the expectations were taken.

Due to the practical meaning of the expectation, problems a) b) c) d) have a good background just in the cases where the same problem arises great many times under similar circumstances. In unique cases a minimum expectation principle does not realize the gain what we expect, completely similarly to games governed by chance. We may carry out our program by using the expectations but this will be legitimate just under special circumstance as we shall see later.

One sees from this presentation that all stochastic programming models require the solution of some probabilistic problems.

The chance constrained programming problem needs the expression of the probability in (1.1) or (1.2) what we do not have (except very special cases) even in the case if our random variables have a joint normal distribution which is perhaps the most important from the point of view of the practical application. Concerning the two stage programming problem where only b is random the problem to be solved

is to find the expectation of $f'y(b, x)$ where y is the random optimal solution of (1.4) for a given x. Concerning problem d) the expectation of the maximum of the random variables $b_i - \sum_{k=1}^{n} a_{ik} x_k$, $i=1, \dots, m$ is needed. The author of the present paper has some quite recent results concerning these problems assuming normal distributions of the random variables appearing in the problems and using a

Hermite-polynomial development of the probabilities in question but we do not enter into details in this connection.

Some other unsolved problems concerning stochastic optimization problems are mentioned in [24] .

Our principal aim here is to discuss the probability distribution of the optimum value of a random linear programming problem a question which refers to d) , the programming by expectations. We consider the situation in such a way that the true values of the parameters are the random values, and the right way to carry out the solution would be the programming by these actual random data. This is, however, impossible sometimes for some reason e.g. we do not know these values in advance but the predicted values which are usually the expectations. We therefore program by the expectations and are interested in the probability distribution of the true random optimum μ around the optimum value belonging to the expectations. We solve this problem under certain special circumstances.

2. Remarks on the Probability Distribution of the Optimum of a Random Linear Program -

We shall consider linear programming problems

$$\mu = \max\ c'x \tag{2.1}$$

$$Ax = b,\ x \geqq 0\ ,$$

where A is an m x n matrix, c and x are n-, and b is an m-dimensional vector. We shall suppose A, b, c have random elements and components, respectively. As μ is a function of the variables in A, b and c,

$$\mu = \mu\,(A,\ b,\ c\ ) \tag{2.2}$$

is also a random variable and its probability distribution is that what we are interested in. This problem is of a basic importance and is conceivable as a stochastic sensitivity analysis of a linear programming model. The question how the transformation A, b, c $\rightarrow \mu$ operates under the presence of random influences in A, b and c does not play just the role of a sensitivity analysis, however. In fact in A, b, c we may have not just random disturbances but random variables of significant variation.

The mentioned problem in its general form has been considered by Tintner, 2 , 3 , and Babbar 1 . In these papers it is supposed that the random variation does not change the optimal basis in the sense that the subscripts of the optimal basis vectors remain the same for all possible values of A, b, c . Thus finding the probability distribution of μ is equivalent to finding the probability distribution of a (also random) linear functional defined over the random solution of a set of linear equations. In this respect it is also possible to proceed in two different wys. Either to develop μ into a finite Taylor series and use the leading, linear terms as an approximation to μ and obtain its probability distribution, or to consider the components of the solution as fractions of random determinants, approximate their distributions by the normal law and again approximate by the normal law the fraction of two normally distributed variables. This method has the handicap that produces sophisticated approximative formulae.

In sections 2., 3., and 4. we consider systems of linear equations, the probability distribution of a random linear functional defined over the solutions and apply this theory for our original problem concerning random linear programs. Our approximative formulae for the characteristics of μ especially for the dispersion will be particularly simple, as simple as possible in this general formulation of the problem from the point of view of practical application, involving the primal and dual optimal solutions of the linear programming problem carried out with the expectations and the covariances of the present random variables. We express our statements in limit theorems and list carefully all mathematical suppositions. Our results are generally formulated containing the essential features of the problem and allowing the possibility of specialization when facing with a particular problem.

The results of the present paper, however, apply to the case when the random elements in the technology matrix keep the optimal basis (basis subscripts), obtained by computing with the expectations, with a high probability. To more general questions we return in subsequent papers.

As in the present approach the principal aim is to reduce μ to a sum of random variables, the asymtotic normality of this sum will besupposed. In the particular cases where the sum in question contains an increasing number of independent random variables e.g. A has independent elements or it is enough if its rows or columns are independent, the limit distribution theory of sums of double sequences of independent random variables can be applied . In this respect the reader is referred to the book [19] . If independence does not occur then we may suppose the joint normality of the random variables in A, b, c which is satisfying or suppose simply the sum in question to be normally distributed but there is no detailed general theory of the limit distributions of sums of double sequences of dependent random variables. On the other hand any particular problem contains some specific way how the random elements intervene from the knowledge of which we may assume the normality of the approximative sum.

3. System of random linear equations.

Consider the following system of linear equations.

$$\sum_{k=1}^{m} a_{ik} x_k = b_i, \quad i = 1, \dots, m. \tag{3.1}$$

Let us denote by B the matrix of the equations , by b the vector consisting of the b_i's as components and introduce a further m-dimensional vector c. If B is non-singular then (3.1) has a unique solution B^{-1} b. Suppose now that B, c, b are all random and all elements and components have finite variances. We are interested in the probability distribution of the functional

(3.2) $$\mu = c'R\,b\,, \quad \text{where } R = B^{-1}\,.$$

In order to avoid complications in the notations we shall suppose that B is independent of the couple c, b and c, b are also independent of each other. This supposition does not play any significant role here. We shall denote. by $a_1, \dots, a_m$ the columns of B and by D_{ik} the cross covariance matrix of a_i and a_k, i.e.,

(3.3) $$D_{ik} = E\left[(a_i - a_i^{(o)})(a_k - a_k^{(o)})'\right],$$

wher $a_i^{(o)} = E(a_i)$, $i = 1, \dots, m$, the prime means the transpose and E the expectation. The expectations of B, b, c, a_{ik}, b_i, c_j will be denoted by B_o, b_o, c_o, $a_{ik}^{(o)}$, $b_i^{(o)}$, $c_j^{(o)}$, respectively. R_o will denote B_o^{-1}. The covariance matrices of γ and β will be denoted by C and F, respectively.

Disregarding for a while of the random nature of our quantities we shall give the finite Taylor development of μ around the expectations closing by the second order terms. It can be done by using a formula well-known in matrix theory saying that if the inverse of a non-singular square matrix is R and we modify the original matrix by adding to the element standing in the ith row and kth column then the inverse of the modified matrix will be

(3.4) $$R - \frac{\xi}{1+r_{ki}} \begin{pmatrix} r_{1i} \\ \vdots \\ r_{mi} \end{pmatrix} (r_{k1} \dots r_{km})$$

Hence if we change B in the manner described above and consider the change in the functional then we obtain

(3.5) $$\mu(a_{ik}+\xi) - \mu(a_{ik}) = \frac{\xi}{1+r_{ki}\xi}\, y_i x_k$$

where

$$y_i = \sum_{j=1}^{m} C_j r_{ji}$$

From this it follows that

(3.6) $$\frac{\partial \mu}{\partial a_{ik}} = y_i x_k \ , \quad i,k = 1,\dots, m.$$

By a double application of the formula (3.4) we get

(3.7) $$\mu(a_{ik}+\xi) + \mu(a_{ik}-\xi) - 2\mu(a_{ik}) = = \xi^2 \, y_i x_k \, \frac{r_{ki}}{1-\xi^2 r_{ki}^2}$$

hence

(3.8) $$\frac{\partial^2}{\partial a_{ik}^2} = \lim_{\xi\to 0} \frac{(a_{ik}+\xi) + \mu(a_{ik}-\xi) - 2\mu(a_{ik})}{\xi^2} = y_i x_k r_{ki}\, , \ i,k=1\dots, m.$$

We can determine similarly the mixed partial derivatives. The result is the following

(3.9) $$\frac{\partial^2 \mu}{\partial a_{ik}\, a_{pg}} =$$

$$= \lim_{\substack{\xi\to 0\\ \eta\to 0}} \frac{\mu(a_{ik}+\xi, a_{pg}+\eta) - \mu(a_{ik}+\xi, a_{pg}) - \mu(a_{ik}, a_{pg}+\eta) + \mu(a_{ik}, a_{pg})}{} =$$

$$= x_k y_p r_{gi} + x_g y_i r_{kp} \qquad \text{for } |i-p| + |k-g| > 0 .$$

Let us finally mention the derivatives where c_i and b_i are involved :

(3.10) $$\frac{\partial \mu}{\partial c_j} = x_j \; , \quad \frac{\partial \mu}{\partial b_j} = y_j \; , \quad \frac{\partial^2}{\partial c_i \partial b_j} = r_{ij} \; ,$$

$$\frac{\partial^2}{\partial a_{ik} \partial c_j} = x_k r_{ji} , \quad \frac{\partial^2}{\partial a_{ik} \partial b_j} = y_i \, r_{kj} \; , \quad i, j, k = 1, \ldots, m \; .$$

Let us introduce the notations.

(3.11) $$y' = c'R \quad , \quad y' = c_0' R_0 \, ,$$
$$x = R\, b \quad , \quad x_0 = R_0 b$$

and denote by $x_i^{(o)}$, $y_j^{(o)}$ the components of x_o, y_o, respectively. Furthermore

(3.12) $$a_{ik} - a_{ik}^{(o)} = \xi_{ik} \, , \quad B - B_0 = \Xi \, ,$$
$$c_i - c_i^{(o)} = \gamma_i \quad , \quad c - c_0 = \gamma \, ,$$
$$b_i - b_i^{(o)} = \beta_i \quad , \quad b - b_o = \beta \quad , \quad i, \; k = 1, \ldots, m.$$

With the aid of these notations the mentioned Taylor-development is the following

$$\mu - \mu_0 = - \sum_{i,k=1}^{m} y_i^{(o)} \xi_{ik} x_k^{(o)} + \sum_{i=1}^{m} y_i^{(o)} \gamma_i + \sum_{k=1}^{m} x_k^{(o)} \beta_k \; +$$

$$+ \frac{1}{2} \sum_{i,k=1}^{m} y_i \xi_{ik}^2 \, r_{ki} \, x_r \; +$$

$$+ \sum_{i,k,j=1}^{m} x_k \, r_{ji} \, \xi_{ik} \gamma_j \; + \sum_{i,k,j=1}^{m} y_i r_{kj} \xi_{ik} \beta_j \; +$$

$$+ \sum (x_k y_p r_{gi} + x_g u_i r_{kp}) \, \xi_{ik} \xi_{pg} ,$$

$$|i-p| + |k-g| \; > \; 0$$

or in a concise form

$$(3.14)\qquad \mu - \mu_0 = -y_0' \Xi x_0 + y_0' \gamma + x_0' \beta + \rho ,$$

where the error term ρ is equal to

$$(3.15)\qquad \rho = -\frac{3}{2} x' \Xi_2 y + 2 y' \Xi R x + + \gamma' R \Xi' x + y' \Xi R \beta ,$$

and Ξ_2 is the matrix consisting of the entries $\xi_{ik}^2 r_{ki}$. In the above development in the error term x, y and R, which are functions of b, c, B, are taken at a point $b_i^{(o)} + \vartheta \beta_i$, $c_i^{(o)} + \vartheta \gamma_i$, $a_{ij}^{(o)} + \vartheta \xi_{ij}$, where $0 < \vartheta < 1$. The leading term in (3.14) ha 0 expectation and variance

$$(3.16)\qquad \sigma^2 = \sum_{k=1}^{m} x_i^{(o)} y_0' D_{ik} y_0 x_k^{(o)} + y_0' C y_0 + x_0' F x_0 .$$

If the column of B are independent random vectors as it can be supposed in some practical cases, then σ^2 reduces to

$$(3.17)\qquad \sigma^2 = \sum_{i,k=1}^{m} (x_k^{(o)})^2 y_0' D_{kk} y_0 + y_0' C y_0 + x_0' F x_0 ,$$

which further reduces if all elements of B are independent . If also c and b have independent components then we have

$$(3.18)\qquad \sigma^2 = \sum_{i,k=1}^{m} (y_i^{(o)})^2 \sigma_{ik}^2 (x_k^{(o)})^2 + \sum_{i=1}^{m} (y_i^{(o)})^2 s_i^2 + \sum_{k=1}^{m} (x_k^{(o)})^2 t_k^2 ,$$

where

$$\sigma_{ik}^2 = E(\xi_{ik}^2) , \quad s_i^2 = E(\gamma_i^2) , \quad t_k^2 = E(\beta_k^2) .$$

In the next sections we shall give sufficient conditions under which $(\mu - \mu_0)/\sigma$ has an asymptotic normal distributions. This will mean from the practical point of view that μ has an asymptotic normal distribution with the expectation μ_0 i.e. the value of the functional belonging to the expectations of all values involved and the variance given by (3.16) which may specialize.

4. Limit distribution theorems for random linear equations .

First we prove a lemma.

Lemma . Let $H_1, H_2, \dots$ be a sequence of events with the property that

$$\lim_{N \to \infty} P(H_N) = 1.$$

Let further η_N and ζ_N be two sequences of random variables where η_N has a limit distribution i.e.

$$\lim_{N \to \infty} P(\eta_N < x) = G(x)$$

at every point of continuitu of $G(x)$ and ζ_N tends stochastically to 0 (in symbols $\zeta_N \Rightarrow 0$) i.e.

$$\lim_{N \to \infty} P(|\zeta_N| > \varepsilon) = 0 \text{ for every } \varepsilon > 0.$$

(ζ_N has a degenerated limit distribution) . Under these conditions

$$\lim_{N \to \infty} P(\eta_N + \zeta_N < x \mid H_N) = G(x)$$

at every point of continuity of $G(X)$.

This lemma is essentially Cramer's lemma (see [18] p. 254) expressed in a slightly modified form therefore we omit the proof

In order to obtain limit theorems we can proceed in two directions. We may keep m; the size of the system of equations, fixed while the random disturbances have a slowing down tendency. This is the case e.g. when the random disturbances are due to some inaccuracy in the measurements of the data which show up a decreasing tendency using more data or in other terms a larger sample. The other possibility is to increase m. In this case we shall also suppose implicitly that the random elements are small as compared to the expectations but for the convergence to the normal distribution the increasing size of the matrix B contribute also. First we formulate two theorems in general forms.

Theorem 1 . It a function $f(z_1, z_2, \ldots, z_k)$ has continuous second order derivatives in some convex neighbourhood of the point $z_1^{(o)}, z_2^{(o)}, \ldots, z_k^{(o)}$ where k is fixed and for a sequence of random vectors

$$(\xi_1^{(N)}, \xi_2^{(N)}, \ldots, \xi_k^{(N)}) \text{ with } E(\xi_i^{(N)}) = 0,\ i=1, \ldots, k,$$
$$N=1, 2, \ldots,$$

the following conditions are satisfied :

1./ $\xi_i^{(N)} \Rightarrow 0$ if $N \to \infty$, $i = 1, \ldots, k$

2./ $\lim\limits_{N \to \infty} P(\frac{1}{\sigma_N} \sum_{i=1}^{k} \frac{\partial f^{(o)}}{\partial z_i} \xi_i^{(N)} < x) = G(x)$

at every point of continuity of $G(x)$ where $\frac{\partial f^{(o)}}{\partial z_i}$ means the derivative $\frac{\partial f}{\partial z_i}$ taken at $z_1^{(o)}, z_2^{(o)}, \ldots, z_k^{(o)}$ and σ_N is the dispersion of

$$\sum_{i=1}^{k} \frac{\partial f^{(o)}}{\partial z_i} \xi_i^{(N)} \quad ,$$

then we have 3./ $\frac{1}{\sigma_N} \sum_{i,j=1}^{k} \frac{\partial^2 f^{(1)}}{\partial z_i \partial z_j} \xi_i^{(N)} \xi_j^{(N)} \to 0$, if $N \to \infty$,

$$\lim_{N \to \infty} P\left\{ \frac{1}{\sigma_N} \left[f(z_1^{(o)} + \xi_1^{(N)}, \ldots, z_k^{(o)} + \xi_k^{(N)}) - f(z_1^{(o)}, \ldots, z_k^{(o)}) < x \right\} = \right.$$
$$= G(x) \; .$$

<u>Proof</u>. Let H_N denote the even that $(z_1^{(o)} + \xi_1^{(N)}, \ldots, z_k^{(o)} + \xi_k^{(N)})$ is in that neighborhood of $(z_1^{(o)}, \ldots, z_k^{(o)})$ where f has continuous second order derivatives . In this case

$$(4.1) \qquad f(x_1^{(o)} + \xi_1^{(N)}, \ldots, z_k^{(o)} + \xi_k^{(N)}) - f(z_1^{(o)}, \ldots, z_1^{(o)}, \ldots, z_k^{(o)}) =$$
$$= \sum_{i=1}^{k} \frac{\partial f^{(o)}}{\partial z_i} \xi_i^{(N)} + \frac{1}{2} \sum_{i,j=1}^{k} \frac{\partial^2 f^{(1)}}{\partial z_i \partial z_j} \xi_i^{(N)} \xi_j^{(N)} ,$$

where $\frac{\partial^2 f^{(1)}}{\partial z_i \partial z_j}$ is the second order derivative taken at

$$z_1^{(o)} + \vartheta \xi_1^{(N)}, \ldots, z_k^{(o)} + \vartheta \xi_k^{(N)} \text{ and } 0 < \vartheta < 1 \; .$$

According to Condition 1.) $\lim_{N \to \infty} P(H_N) = 1$. Let us divide by σ_N on both sides in (3.1) . Then the second term on the right hand side tends stochastically to 0 according to Condition 3.) . Let us denote this term by ξ_N while the first term by η_N . Then a direct application of the Lemma completes the proof.

Condition 3.) is clearly fulfilled if

$$\frac{1}{\sigma_N} \xi_i^{(N)} \xi_j^{(N)} \Rightarrow 0, \quad i, j = 1, \dots, k.$$

Before Theorem 2. we mention the notion of a star shape domain. An open neighbourhood K around a point $(z_1, \dots, z_k)$ in the k-dimensional space is called that if the intersection of K whit any ray

$$(z_1 + t\,\xi_1, \dots, z_k + t\,\xi_k), \qquad t > 0$$

is an open interval. This may contain in particular every point of the ray. The point $(z_1, \dots, z_k)$ is called the seed of the domain. This notion will be important to extend the possibility of the Taylor-series development around the given point as large as possible.

For later purpose we introduce a notion, the notion of a maximal domain of a star shape around a non-singular matrix B_0 which by definition consists of alla matrices of the form

$$B_0 + t\,\Xi$$

where for any given Ξ , t runs continuously from 0 until the above sum becomes singular. That singular matrix is excluded, however.

Theorem 2. Suppose that we have a sequence of functions of increasing number of variables $f_N(z_1, z_2, \dots, z_{Nk_N})$, where $k_N \to \infty$ if $N \to \infty$, and each f_N has a neighbourhood of a star shape around a point $(z_{N1}^{(o)}, \dots, z_{Nk_N}^{(o)})$ where its second order derivatives exist and are continuous. Suppose furthermore that we have a double sequence of random variables $\xi_1^{(N)}, \xi_2^{(N)}, \dots, \xi_{k_N}^{(N)}$ with 0 expectations and finite variances, satisfying the following conditions

$$1./ \quad \lim_{N \to \infty} P\left\{ (z_{N1}^{(o)} + \xi_1^{(N)}, \dots, z_{Nk_N}^{(o)} + \xi_{k_N}^{(N)}) \in K_N \right\} = 1 ,$$

where K_N is the above mentioned neighborhood

At every point of continuity of $G(x)$ $\left(\frac{\partial f_N^{(0)}}{\partial z_i}, \frac{\partial f_N^{(1)}}{\partial z_i z_j}\right.$ have the same meaning as in Theorem 1) .

$$3./ \quad \frac{1}{\sigma_N} \sum_{i,j=1}^{k_N} \frac{\partial^2 f_N}{\partial z_i \partial z_j} \xi_i^{(N)} \xi_j^{(N)} \Rightarrow 0 \quad \text{if} \quad N \to \infty \; .$$

In this case we have

$$\lim_{N \to \infty} P\left\{ \frac{1}{\sigma_N} f_N(z_{N1}^{(o)} + \xi_1^{(N)}, \ldots, z_{Nk_N}^{(o)} + \xi_{k_N}^{(N)}) - f(z_{N1}^{(o)}, \ldots, z_{Nk_N}^{(o)}) < x \right\} =$$

$$= G(x)$$

at every pont of continuity of $G(X)$.

The proof is completely similar to that of Theorem 1.

It is worth mentioning that the fulfillment of Condition 1 in Theorem 2 may be the cause of the slowing down tendency of random elements as well as the increase of K_N or both .

In both theoremes we used the same idea Cramer used when proving the asymptotic normality of functions of moments (see [19] pp. 366--367, in that case the numer of variables was fixed). We can apply these theorems for random linear equations . In the following theorem we shall omit the subscript N which would refer to the fact that we have a sequence of random elements. Thus all our previous notations concerning random equations are applicable .

Theorem 3. Suppose that m, B_o, c_o, b_o are fixed, B_o is non-singular and introduce the following conditions :

1.) $\sigma_{ik} \to 0$, $i, k = 1, \ldots, m$,

2.) $P\left\{\frac{1}{\sigma}(-y' \Xi x_o + y'\gamma + x_o' \beta) < x\right\} \to \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}} du,$

for every x, $-\infty < x < \infty$;

3.) $\frac{\rho}{\sigma} \to 0$.

In this case we have for every x

(4.2) $$P\left(\frac{\mu - c_0' R_0 b_0}{\sigma} < x\right) \to \Phi(x) .$$

Proof. Theorem 3 is an immediate consequence of Theorem 1 applied for the function $\mu = \mu(A, b, c)$ of $m^2 + 2m$ variables. A_0, b_0, c_0 as the point around which the Taylor-series development is taken and Ξ, γ, β as the sequence of random $m^2 + 2m$ - dimensional vectors. We just have to mention that Condition 1. in Theorem 1. is ensured by Condition 1., of Theorem 3. Various consequences of this theorem can be derived. Among them we mention the simplest.

Corollary. Suppose that the $m^2 + 2m$ random variables in Ξ, γ and β have a normal joint distribution and

$$\sigma_{max} \to 0, \quad \frac{\sigma^2_{max}}{\sigma} \to 0 \text{ where } \sigma_{max} = \max (\sigma_{ik}, s_i, t_k) .$$

Then (3.2) holds

Proof. All what we have to verify is the fulfillment of Condition 3., in Theorem 3. If we look at the detailed expression of ρ given by the last terms in (3.13), we see that separately each term of that sum divided by σ converges stochastically to 0. In fact considering the quadratic terms ξ^2_{ik}/σ we see by the Markov-inequality that

$$P\left(\frac{\xi^2_{ik}}{\sigma} > \varepsilon\right) \leq \frac{E(\xi^2_{ik})}{\sigma} \leq \frac{\sigma^2_{max}}{\sigma} \to 0 .$$

For all other terms the Chebischev-inequality can be applied.

Theorem 4. Consider a sequence of matrices and vectors B_0,

c_o, b_o and a corresponding random sequence Ξ, γ, β (the subscripts are omitted), where m, the size of the matrices (equal to the dimensions of the vectors) tends to infinity. Suppose that all B_o matrices are non-singular. To every B_o in the sequence there corresponds a maximalneighbouhood of a star shape K where $B_o + \Xi$ is non-singular and the Taylor development around B_o applies. Suppose that

1.) $P(B_0 + \Xi \in K) \longrightarrow 1$,

2.) $P\left\{\frac{1}{\sigma}(-y_o' \Xi x_o + y_o' \gamma + x_o' \beta) < x\right\} \longrightarrow \Phi(x), \; -\infty<x<\infty,$

3.) $\frac{\varrho}{\sigma} \Rightarrow 0$.

Under these conditions (*)

$$P\left\{\frac{1}{\sigma}(\mu - c_o' R_o b_o) < x\right\} \to \Phi(x), \qquad -\infty < x < \infty .$$

The proof of this theorem is similar to that of Theorem 3

Analyzing the conditions here, 1.) and 2.) are realistic as the size of the matrix increases The crucial point is Condition 3.) which may very easily fail. In fact first of all the 4th term containing the squares ξ_{ik}^2 may not be negligible as compared to σ. It does not have in general 0 expectation even in the case when all random variables are independent. It seems therefore to be advisable to attach the sum

$$\frac{1}{2} \sum_{i,k=1}^{m} y_i^{(o)} \xi_{ik}^2 r_{ki}^{(o)} x_k^{(o)}$$

to the leading term changing it into

(*) Instead of $\Phi(x)$ we may suppose other distribution function too.

$$(4.3) \qquad - \sum_{i,k=1}^{m} y_i^{(o)} \zeta_{ik} \left(1 - \frac{1}{2} \zeta_{ik} r_{ki}\right) x_k^{(o)} + y_o' \gamma + x_o' \beta .$$

We may then assert that μ has an asymptotic normal distribution with the expectation

$$(4.4) \qquad c_o' R_o b_o + \frac{1}{2} \sum_{i,k=1}^{m} y_i^{(o)} \sigma_{ik}^2 r_{ki}^{(o)} x_k^{(o)}$$

and variance (2.16) where D_{ik} has to be replaced by

$$(4.5) \qquad D_{ik} - T_i D_{ik} T_k - T_i D_{ik} - D_{ik} T_k$$

and T_i is a diagonal matrix consisting of elements $r_{1i}^{(o)}, \ldots, r_{mi}^{(o)}$ in the diagonal. The same sum what we added to the leading term has to be subtracted from the remainder and it is more realistic to say that the new remainder divided by the dispersion of the new leading term tends stochastically to 0 .

5. Application to random linear programs.

Consider the linear programming problem

$$(5.1) \qquad \mu_0 = \max c_o' x$$

subject to the conditions

$$(5.2) \qquad A_o x = b_o , \quad x \geqq 0 ,$$

and suppose that it has a unique optimal basis B_o which for the sake of simplicity we suppose to be the set of vectors $a_1^{(o)}, \ldots, a_m^{(o)}$./
We suppose that A_o has rank m .
Consider also the problem

$$(5\ 3) \qquad \mu = \max c' x$$

subject to the conditions

(5.4) $$Ax = b, \quad x \geqq 0 \quad ,$$

where A, b, c have random elements, components, respectively. Concerning these problems we apply the same notations as concerning random equations in sections 2. and 3. but we observe that A has mn elements and c has n components.

We suppose also that B_0 is non-degenerated. There is then a neighbourhood of A_0, b_0, c_0 in which the problem (5.3)-(5.4) will preserve the subscripts of the optimal basis. Keeping fixed m and n, consider a sequence of random matrices, vectors A, b, c, respectively. If we suppose that

(5.5) $$_{max} \to 0 ,$$

where $$_{max} = \max\ (\sigma_{ik}, t_i, s_k, i=1,\dots,m; k=1,\dots, n) ,$$

the probability that $B = (a_1, \dots, a_m)$ will be the optimal basis to problem (5.3)-(5.4), tends to 1. Hence according to our Lemma $(\mu - \mu_0)/\sigma$ / will have the same asymptotic probability distribution unconditionally or conditionally, given that B is optimal. If further Conditions 2., and 3., are also satisfied in Theorem 3. where all quantities vectors, matrices are taken from the random equation Bx = b and c means the vector consisting of the first m components of that used in (5.3) then we may state that

<u>Theorem 5.</u> The optimum value μ of the random programming problem (5.3)-(5.4) has asympyotic normal distribution with the expectation μ. the optimum of the program taken with the expectations in each place and variance (3.16) where x_0, y_0 are the primal and dual optimal solutions of the first problem, more exactly x_0 is a part of

the primal optimal solution consisting of the basic components. The meaning of D_{ik}, C, F remains unchanged. Asympotic normality means that the probability that $(\mu - \mu_i)/\sigma < x$ tends to $\Phi(x)$.

It is seen from these that the present approach gives a particularly simple result and very advantageous from the practical point of view because in the characteristics of the random variable such vectors and matrices appear as the primal and dual optimal solutions x_0, y_0 and D_{ik}, C and F, the covariance matrices of the random variables involved.

We may also apply Theorem 4. by considering a sequence of programming problems, $m \to \infty$, $n \to \infty$. There we suppose that at each problem with A_0, b_0, c_0 there is a unique finite, non-degenerated optimum and the probability that the optimal basis has the same column subscripts in problem (5.1) - (5.2) and in (5.3) - (5.4), tends to 1. Then if we take into account our Lemma, the results of Theorem 4 are applicable, where x_0 and y_0 have the same meating as before

One practical conclusion of these results is the following : if for some reason we solve the linear programming problem with the expectations e.g. with predicted prices and predicted technology coefficients but we have information about their random variation then we may set up confidence limits for the optimum value which would have been the result if we had programmed with the particular realization of the random data in A, b, and c.

BIBLIOGRAPHY

1. M.M. Babbar, Distributions of Solutions of a Set of Linear Equations (With an application to Linear Programming), J. Amer. Stat. Assoc. 50 (1955), pp. 854-869.

2 G. Tintner, Stochastic Linear Programming with Applications to Agricultural Economics, Second Symposium on Linear Programming National Bureau of Standards, Washington, D.C., Vol. 1 (1955). pp. 197-227.

3 G. Tintner, Les Programmes Linéaires Stochastiques, Revue de Economie Politique, 67 (1957).

4 G. Tintner, A Note on Stochastic Linear Programming, Econometrica, 28 (1960) pp. 26-34.

5 H. M. Wagner, On the Distribution of Solutions in Linear Programming, Problems, J. Amer. Stat. Assoc. 53 (1958), pp. 161-163.

6 J. V. Talacko, On Stochastic Linear Inequalities, Trabajos Estadisticos, 10 (1959), pp. 89-112.

7 A. Madansky, Inequalities for Stochastic Linear Programming Problems, Management Science, 6 (1960), pp. 197-204.

8 S. Vajda, Inequalities in Stochastic Linear Programming, Bull. International Stat. I st., 36 (1958), pp. 357-363.

9 H. W. Kuhn and R. E. Quandt, An Experimental Study of the Simplex Method, Proc, Symp. Appl. Math. 15 (1963), pp. 107-124.

10 A.T. Lonseth, Systems of Linear Equations with Coefficients Subject to Error, Annals of Math. Stat. 13 (1942), pp. 332-337.

11 A. T. Lonseth, On Relative Errors in Systems of Linear Equations, Annals of Math. Stat. 15 (1944) , pp. 323-325.

12 A. T. Lonseth, The Propagation of Errors in Linear Problems, Trans. Amer. Math. Soc. 62 (1947) pp. 193-212.

13 G. E. P. Box and J. S. Hunter, A Confidence Region for the Solution of a Set of Simultaneous Equations with an Application to Experimental Design, Biometrika, 41 (1954) , pp. 190-199.

14 R. E. Quandt, Probabilistic Errors, in the Leontief System, Naval Res. Log. Quart. 5 (1958) , pp. 155-170.

15 H. D. Mills, Marginal Values of Matrix Games and Linear Programmes, Linear Inequalities and Related Systems, Princeton Univ. Press. 1956, pp. 183-193.

16 A. C. Williams, Marginal Values in Linear Programming, J. Soc. Ind . and Appl. Math. 11 (1963) , pp. 82-94.

17 E. Bodewig, Matrix Calculus, North-Holland Publ. Comp. Amsterdam, 1956.

18 H. Cramer, Mathematical Methods of Statistics, Princeton Univ. Press. 1946.

19 B.V. Gnedenko and A. N. Kolmogorov. Limit Distributions for Sums of Independent Random Variables, translated by K. L. Chung, with an appendix by J. L. Doob, Addison-Wesley Publ. Comp. Inc ., Reading, Mass. , 1954.

20. A. Charnes and W.W. Cooper, Chance-Contrained Programming, Management Science 6(1959) , pp. 73-79.

21 A. Charnes and W.W. Cooper, Deterministic Equivalents for Optimizing and Satisficing under Chance Constraints, Operations Research 11 (1963), pp. 18-39

22 G. B. Dantzig and A. Madansky, Two Stage Linear Programming under Uncertainty, Proceedings of the Fourth Berkeley Symposium edited by J, Neyman, Berkeley, 1961.

23 H. Theil, Some Reflection on Static Programming under Uncertainty, Weltwirschaftliches Archiv, 87 (1961), pp. 129-138.

24 R. Wets, Chance Constrained Programming, BoeingScientific Research Laboratories 1965 (unpublished) .

CENTRO INTERNAZIONALE MATEMATICO ESTIVO

(C.I.M.E.)

RAGNAR FRISCH

GENERAL OUTLOOK ON A METHOD OF ADVANCED AND DEMOCRATIC MACROECONOMIC PLANNING

Corso tenuto all'Aquila dal 30 agosto al 7 settembre 1965

GENERAL OUTLOOK ON A METHOD OF ADVANCED AND DEMOCRATIC MACROECONOMIC PLANNING

by

Ragnar Frisch

Lecture No. .1

I shall start by giving you a bird's-eye view of the whole problem, adopting a general and philosophical attitude with absolutely no mathematics. Later you will have mathematics and just as much as you want. But today I will speak in general terms.

Rational and democratic planning.

What is rational and democratic planning ? I will give you my personal views on this matter. It will be my personal views because I think that much of what is currently called economic planning does not deserve that name.

It is lacking in two respects: it is not rational enough and it is not democratic enough.

I stress both these aspects because the problem I am driving at is much more ambitious than just to increase the long term average growth rate of gross national product, say , from 4 to 6 percent. My purpose is to make economic planning at high aspiration level one of the pillars of a living democracy.

I want a society which is a living democracy, not only a formal one with free elections, so-called freedom of speech, a so-called free press, and so on, but a democracy that is living in the sense of actually engaging as many as possible of the citizens to take an active part in the affairs of the small community where they are living, and even an active part in the affairs of the nation as a whole.

I will give an example of what I mean by living democracy. Some years ago I undertook, together with Mrs. Frisch, a lecture trip to Iceland visiting amongst others also some of the small communities in the northern parts of the country. I was invited on this trip in order to help saving Iceland from becoming a member of the Western European Common Market. In the small agricultural communities in the north the population depends nearly exclusively on grass farming. This was in the middle of the haying season. In one place there had gathered an audience of 60 people. Think of what a gathering of 60 people means in this very sparsely populated country and in the middle of the haying season. Some of them had travelled 60 kilometers to come to the meeting. They brought along with them long papers to present and discuss after my lecture. Now, this is living democracy.

I want to establish advanced national planning methods as a pillar for developing our modern society into a truly living democracy. This is my purpose.

<u>Analytical and deciding bodies</u>.

When we start discussing this problem, the first thing we have to do is to distinguish between <u>analytical</u> bodies and <u>deciding</u> bodies.

The causal connections in and the structure of modern economic society are very complicated. So, therefore, we need experts to study it. We need economists and statisticians and medical men, sociologists and technical engineers and so on. All these groups and many others are <u>needed</u> as experts in order to lay bare the nature of the economic steering problem and to suggest solutions, suggest alternatives.

But this being the situation, there is a great danger that we may be pushed into an expert dictatorship, an expert tyranny. There is a great danger in modern society of getting into a situation where we are dominated by the experts. The danger is that the experts might presenter their findings in such a way that they put the political bodies into a forced situation where these bodies have no free choice. This is what happens too frequently. The experts, instead of working out a technique of studying alternatives, present just one alternative. They say to the political body: "Now , you politicians, answer yes or no. But you must decide quickly, time is pressing". This happens too frequently and this is what we must avoid.

It is necessary that the political bodies, the bodies that are established through a political machinery with the clear purpose of bearing the responsibility of decision, come into the picture early in the analysis and have an opportunity of following what the experts are doing. Not following the details, of course, but following the general aspects, the framework of the investigations. This is vital from the political and social point of view.

Perhaps the experts or the government representatives will say that these matters are too complicated and too difficult to understand for the common man who is elected as a member of the deciding body. This was precisely the attitude taken by Mr. Adolf Hitler. He said: "This is too complicated for you people to understand, I and my experts will think for you, you only have to obey" . Now, we know what that led to. And it is a similar type of deadly danger which we are facing when we engage in a system that leads to expert dictatorship.

So, we must, already in an early stage of an investigation, get the deciding organs into the picture and make them understand as much as possible of the analytical machinery.

It is possible to proceed in this way. Common people of average intelligence, elected by some democratic machinery have a great power to understnad what is essential from the social and broad political view point. I will give you an example. There is a question in Norway, at this moment, of introducing a big pension scheme for the population. The elected people in Parliament, most of them, will not be to able to follow the details of the actuarial mathematics involved in this problem, certainly not. But they are surely in a position to understand the great social and economic implications of the scheme and to make their decisions accordingly.

The fact that the great social and economic implications can be understood by the democratic body of the people, if the findings of the experts are presented in an appropriate form, is certainly a very fortunate situation. Indeed, this fact is the solid foundation on which we must build our democratic society. This is the basis of our work when we want to make rational and advanced planning a pillar of living democracy.

Continuous cooperation.

We need a continuous cooperation between the experts and the politicians.

Both parties have a lot to learn from each other. The experts will learn what it means to carry a tremendous responsibility for the future of the nation, what it means to carry the responsibility of making wise decisions for steering the economy. The experts will have a lot to learn by becoming acquainted with this type of thinking. And, on the other hand, the politicians, those who are to decide, also have a lot to learn. They have to learn about the enormous analytical power that is present in a systematic study based on macroeconomic models and mathematical programming. As an example take the great project of building a new big aluminium factory at Husnes in western Norway. In this case the Norwegian Storting (Parliament) had to make important decisions. A great number of different considerations entered into the problem. Indeed, there were so many different considerations involved that it was next to impossible to keep track of all of them simultaneously just by verbal reasoning. The appended chart indicates the complexicity of the problem. And it was a question of deciding quickly. An enormous advantage could have been gained if this whole thing had been formalized in a practical way and with continuous cooperation between the two sides : On one hand the analytical experts working with programming models, and on the other hand the politicians who had to make the final decisions.Unfortunately this was not done in this case.

Even more necessary are similar studies of the overall problems of social, economic and regional development of the country .

Misuses of the work "planning".

In what does it consist, this system of rational economic planning?

I do not think that there is any word in economic or social life in general that has been misused to the extent that the phrase "economic planning" has.

In its crudest form "planning" is something like the child's exercise when it writes out a list of all the Christmas gifts it would like to have, even without adding up the items to see what the total amounts to. At a somewhat more advanced stage one does make a summation of the list but does not go further than that. Investment "planning" at its crudest stage means just adding up investment plans which the private firms have announced. Such a procedure , of course, cannot really be called investment planning.

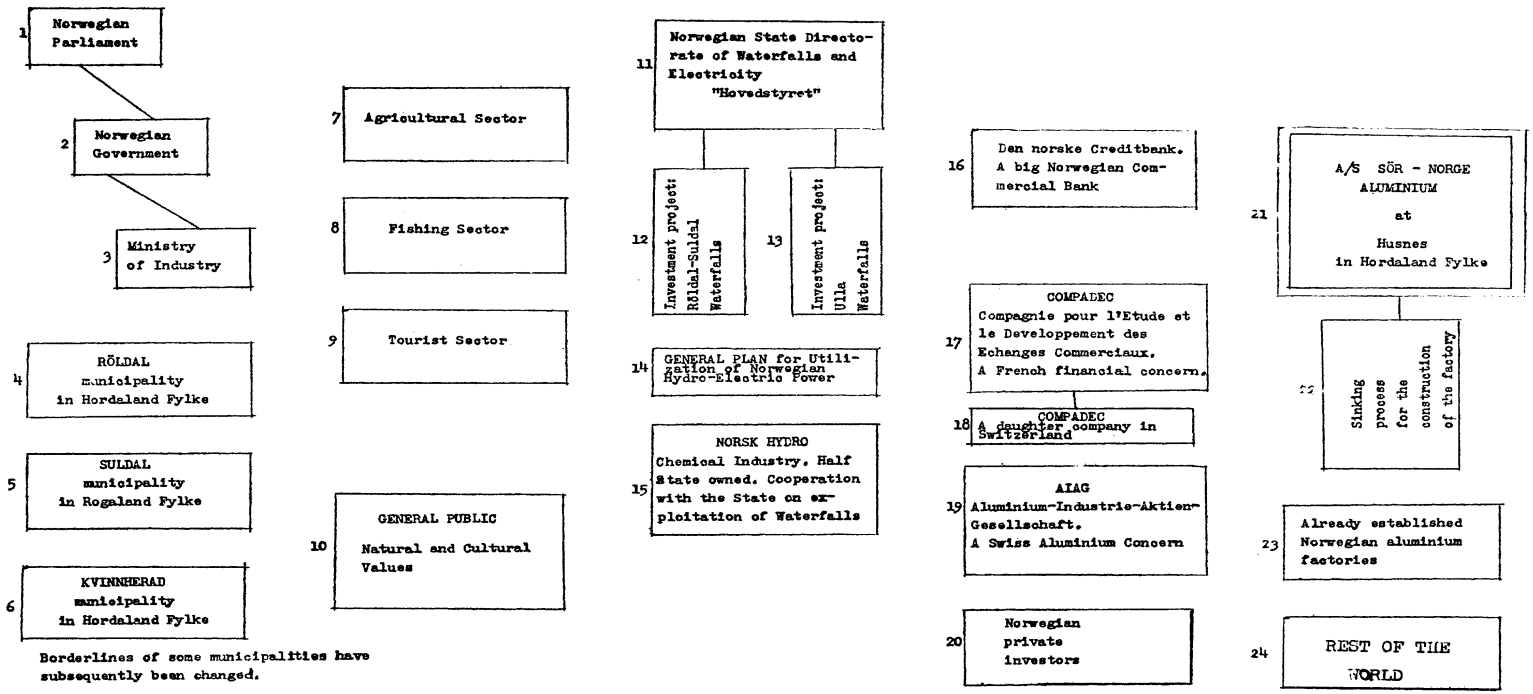
One step further will be when the experts investigate whether the sum total in this list is feasible, in view of the conditions that are imposed on the problem in the form of available resouces, for instance available raw material and labour and foreign exchange, etc. Then , if the sum total is not feasible, the experts themselves may sit down and juggle a little bit with the figures so as to make them fit into a picture which they think presents a feasible solution, and then put this list up to the politicians and ask for a decision of "yes or no".

The poor politicians, what can they do if they are presented with a list that is made feasible in this way ? They have absolutely no practical possibility of seeing to it that this feasible list has the best composition, the optimum composition amongst all feasible lists, the optimum being define by the preferences of the politicians regarding the goals to be attained.

The basic default of a list presented in this way will exist, even if the list has been worked out by mobilizing an army of highly competent people, all these types of experts which I mentioned a little while ago. Even if all these experts have been at work, the list might have nothing to do with a plan in the sense that I have defi-

CHART INDICATING THE COMPLEX NETWORK

to be considered in connection with the big investment project: Establishment of a new Aluminium Factory at Husnes on the west coast of Norway, some 70 km south of Bergen. Relations amongst the units indicated ought to have been studied by mathematical programing.



Borderlines of some municipalities have subsequently been changed.

ned it : a plan that is rational and democratic.

Three stages of planning.

A rational and democratic plan must be worked out in three stages : In the first stage, one has to analyze and find out and describe in as much detail as is needed the structure of the economy. What sorts of relations exist between the variables one has to consider? Secondly, one has to proceed to an analyses of the preferences. What do the politicians want to obtain ? And what would be a feasible development path that is optimal according to these preferences ? Thirdly, we have the implementation problem. What means of implementation are at disposal and what means will the politicians allow when they try to steer the economy in such a way that it will come as near as possible to the optimal path?

These are the three crucial stages of economic planning. As a common catchword for the first two stages : study of the structure and of the preferences, we may use the word "selection". This is a coined word, but convenient . By selection I mean structure and preference. In addition to this we have the implementation problem.

So, let us attack the whole problem from the view point of this sub-division.

It is clear that at all three stages we need the help of experts. The expert's work is a necessary but certainly not a sufficient condition for finally finding a happy solution.

In the case of the structure the experts are sovereign. They don't need much help from the politicians in order to study the structure. The study of the structure is in all essentials their sole responsibility.

When we come to a study of the preferences, the politicians must, of course, play a major role . I say major role, not the only role, because even in the study of this problem it is necessary to rely on the work of the experts. It really takes quite a bit of analytical work to find out what the preferences of the responsible politicians are. And not only does it take a considerable amount of analytical work to find out what the preferences actually are, but it also takes analytical work to express these preferences in a language which our electronic computers can understand. So, while the work of the politicians is absolutely essential and forms the substance matter at this second stage, the work of the experts is by no means superfluous. I will, of course, explain in much more detail later what I mean by this. This is only to give you a birds -eye view.

Thirdly, with regard to the means of implementations. This is an equally important part of the complete analysis. What sort of means of implementations i.e. means of steering the economy, will the politicians allow ? This is essentially a political question . If you restrict very heavily the list of the means of implementation which you will allow politically, you restrict, of course, heavily your opportunity of manoeuvring. By heavily restricting the manoeuvrability you may not be able to steer the economy in the direction you want. You will not have a sufficient number of degrees of freedom in order to produce the desired solution.

Degrees of freedom. Degrees of manoeuvrability.

Let me give a very simple illustration of what is meant by degrees of freedom. Consider the gadget they use on board a ship in order to take the goods from the ship's hull and put them down on the pier. Here you must have two degrees of freedom. You must be able to move the goods vertically up and down, and you must be able to swing them to the side. Suppose you had a gadget with only one degree of freedom : only making it possible to raise and lower a container with goods, but not having the possibility of swinging it to the side . Now what will happen ? You would start by raising it, but would find out that this did not bring the container down on the pier , and you would say "What will I do? Let me try to lower it and see what happens". This being done you find that you have just put the container down in the ship's hull again. Then you may perhaps say, "Let me try once more, but this time not lowering it quite so much as the first time". You would finally discover that you are facing an impossible situation . You must have two degrees of manoeuvrability in order to bring the container from the ship's hull down on the pier.

This examplifies something that comes in at an enormously magnified scale when we are trying to steer an economy in such a way as to obtain the large number of things which the politicians want to obtain . We must have many degrees of freedom in our implementation system. Rapid economic growth and social justice cannot be achieved at the same time just by raising or lowering the interest rate.

And if we do have several means of implementation we must use them in such a way as to take full account of, and be in conformity with , the structural aspects of the economy. The study of the means of implementation is again an example of a situation where there is need for constant cooperation between the experts of various types and the politicians.

Aggregate vs. detailed plans. Global vs. partial plans.

Proceeding now to discussing this set-up a little more in detail, we have to distinguish between aggregate plans and detailed plans. An aggregate plan is a plan where values and quantities are lumped together so that we get only a small number of variables to end up with. For instance, total consumption in society, total investment , total needs for foreign loans, etc. Just ending up with very few variables. This is aggregate planning. It does not mean that a great number of things have been left out. In principle everything may be included. The plan may be global in the sense that all investments should be included, all investments should be included and so on. Only we do not specify these items in detail.

On the other hand consider a very detailed plan - a plan which contains a great number of specifications. For instance so much is needed of the various types of consumer goods in each of the various regions of the country. Even perhaps at the various seasons of the year. And so on. This would be a detailed plan. Obviously it is physically and materially impossible to construct a plan that is to be global and detailed at the same time. It is completely unrealistic to think of a plan of this character. I am sure that today no respectable planner would maintain that his intention is to construct a plan which is at the same time global and detailed.

So we are facing the enormous problem of what you may call compartmentalization. The problem of how to break the problem up into parts. You can have a global plan that is much aggregated. You can have a number of detailed plans but they must, by necessity, be partial, such as partial plans concernig the consumption structure within a specific part of the population or a specific geographical region.

How should we compartmentalize ? This question involves two aspects : the breakdown into types of investigations and the problem of distributing decisional power. Here too the continuous cooperation between the politicians and the analytical experts is indispensible.

Soviet experiences in compartmentalization.

Let me give you an example of the enormous practical importance of the compartementalization problem. In the Soviet Union they started in with the ministerial system of planning. This was a highly centralized system. There was one ministry for each category of goods . And each such ministry was responsible for union so far as this category of goods was concerned. Then after some time they found out that this system entailed great difficulties of various sorts. For instance, one ministry may not receive in due time the goods it needed from other ministries, and this may induce it to set up its own factories for spare parts and other things needed for its own production. Of course, this was very much contrary to the principle that the national plan should be global. Finally the ministerial system broke down and in 1957 they passed from this system to a more regionally oriented system with planning within geographical regions. But this was only shifting from one kind of difficulties to another. For instance : How should the interests or the regions be reconciled ?

Coordination amongst plannings at different levels.

The problem of how to compartmentalize, that is, how to split the problem up, is important both from the analytical viewpoint and from the decisional view point. But this is not all. Having decided on a specific system of breakdowns, an even more important problem emerges : How should one ensure a perfect coordination between these various parts that have been defined through the compartmentalization?

I would like to state that the problem of compartmentalization and how to ensure complete coordination between the compartmentalized fields, is the essence of rational and democratic planning.

We may also call this the pyramidation problem. There must be a pyramid of plans. At the highest level we must have a heavily aggregated model, global and aggregated, and worked out with the specific purposes of serving the top level decisional organ . In a Western democracy this top level decisional organ would be an elected parliament. In other parts of the World it might be a junta of powerful men . Or it may be a single dictator. But in all cases, under all political systems, there is one organ which is the top level deciding organ. And for the use of this top level decisional organ there must be constructed a global and, therefore, by necessity aggregated model. The top level authority will retain its final decisional power, but as an indispensable tool for making wise decisions it must have at the tip of its fingers an aggregated and global plan constructed specially for its own use. Without this it will be in the same situation as the master of a ship who has been deprived of his compass.

No sensible planning analyst will, of couse, maintain that his findings represent a universal truth which one can rely on just like one relies on the telephone hook to find the number one needs. He knows full well that on the top of his findings one must always rely on one's sense of smell. What a plan and a planning model can do, is to push forward by leaps and bounds the line of demarcation where one has to fall back onone's sense of smell .

The top level authority will decide on how to compartmentalize with respect to the next lower level in the pyramid. This decision will primarily be concerned with the compartmentalization of decisional power. The analytical machinery needed in the next lower compartments must be worked out through a system of consultations between the top level organ and the next lower bodies.

The system of the next lower bodies will be compartmentalized according to a cross classification with respect to types of goods and with respect to geographical regions. This means that the problem of coordination between bodies on the same level in the pyramid is a two dimensional one.

On the lower level and inside of any of the organs on this lower level we have precisely the same type of problems as that which the top level body was facing. I mean the same from the formal view point. Analytically and with respect to further distribution of decisional power we have formally the same type of problem. Therefore , in what I shall have to say later in these lectures I shall concentrate mainly on the type of problems which the top level organ is facing. This type is representative of the kinds of problems we find also on the lower levels of the pyramid.

The dynamics of planning.

All these aspects of planning will have to be carried out in a truly dynamic way . The essence of the problem is to lay bare how each situation grows out of the foregoing situation with respect to a certain number of its aspects. Not necessarily in the sense that we have a completely deterministic growing out of the preceeding situation. This would leave us with absolutely no manoeuvrability, no possibility of steering the economy, at some subsequent moment. At any given moment something is fixed and something is left open for new decisions. Some aspects will have to be taken as data because of decisions made previously. These are the already-committed-to aspects of the economy. How these aspects grow out of the preceding situations must be brought out explicitly in the model, and the remaining degrees of freedom must be studied. What degrees of manoeuvrability are still left after taking account of the decisions previously made ? This is the essence of the dynamic aspect. It has, of course, both an analytical and a decisional side.

Different time horizons.

How far should we look into the future when we work out a plan? This raises again two problems : the time horizon problem and the truncation problem.

First with regard to the time horizon problem . It has no meaning to work out a plan to last for a thousand years, because in the course of such a long span of time nearly everything loses itself in the haze of uncertainty. We must concentrate most of our efforts on a much shorter run . But, of course, we must not be too short-sighted. We will need at least three kinds of plans. We will need an annual plan, and a medium term plan with anything between 4 and 7 years perspectives, and we will need a long term plan with a perspective of 20 to 50 years. Some sort of breaking up with respect to time length there must be. This breaking up is only another aspect of the dynamic feature of the planning process.

And here again we face the problem of coordination. There must be a perfect coordination between the plans of different time lengths. The one year plan must be a part of, a specification of, a detailing of, the items that are included in the medium term plan. And the medium term plan must again be conceived of as a part of the long term plan. This coordination between plans of different time spans has been a very weak point in planning up till now.

The truncation problem

A few words regarding the truncation problem. This is a special aspect of the distinction between different time lengths. A long term plan can not contain as many details as a short run plan. The longer the perspective the greater will be the number of items that lose themselves in the haze of the future. How should we handle this transition from a stage where a certain item is handled seriously and explicitly, to a stage where it is more or less neglected or lumped together with other items ? This is the truncation problem.

Moving planning.

When we work out a medium term plan, say a five year plan, the meaning is of course not that we will so to speak put the whole economy in a Procustean bed defined by this five year plan and not make any departure from this plan in the next five years , regardless of what happens in the world around us and regardless of what new experiences we may make. We must be prepared to accept changes and to introduce revisions in the plan.

This leads to what I call moving planning. A five year plan must be worked out on a moving basis. Every single year we must work out a new five year plan . Each such new five year plan must take account of the irrevocable decisions which have been made in the past. These irrevocable decisions represent the already-committed-to aspects of the new five year plan.

For more than ten years I have in writing and speaking advocated the use of this moving principle in planning and have indicated how the moving should be handled. In my mathematical models I have introduced a systematic distinction between these two kinds of magnitudes : the variables that are decisional at the time of planning and those that are already-committed-to at this time . A specific system of notation is introduced to distinguish between these two categories of magnitude s. Thi s is the methodological essence of moving planning. The planning point of time is introduced as a second time variable besides the happening point of time , buth otherwise each new plan in the series of moving plans will be worked out according to the standard procedure for establishing a plan with the time horizon in question.

Lecture No. 2.

I will continue today to present the bird's-eye view of the various aspects of planning.

Structural and administrative aspects of planning.

The first thing I want to discuss is the distinction between the structural aspects and the administrative aspects of planning.

By structural aspects I mean all those things in the economy which can not be changed directly by a Parliament decision, or more generally by a human action. As an example of this consider the number of tons of bauxite needed in order to produce one ton of aluminium. This coefficient is an important question for Norway because we are heavy aluminium producers. This technical coefficient can not be changed directly by Parliament decision. Of course, indirectly and in a very small way it might be changed by Parliament granting credits for technical research which may, perhaps, in the end change a little bit this technical coefficient. But at least in the short run or in the medium run this is a technical coefficient that can not be changed by Parliament decision. Another example is the way in which human need for and craving for food decrease as the supply of food increases. This biological and physiological law is something that can not be changed by Parliament decision. All such things pertain to the structure of the economy. They are structural properties of the economy.

All other things, on the contrary, are what I call administrative aspects. In this vast field of all other things come such things as taxation rules, subsidy rules, rules for the operation of the banking system, fixation of bank rates, systems of social measures, such as the pension system, further import and export rules, etc., all these things can be changed by Parliament decision or by a decision of a subordinate body to which Parliament has delegated its power. These things are administrative properties of the economy, taking the word administrative in a very broad sense.

It is obvious that this distinction between structural and administrative aspects of the economy must be fundamental in any analysis of planning. A structural aspect is something which we have to accept whether we like it or not. Whether I am the bluest sort of a conservative politician or the reddest sort of a communist politician, I have to accept the gravitational force. The other things - the administrative things - we can discuss and make the object of a political fight. But the structural aspects we must accept whether we like it or not.

Imposing conditions means restricting manoeuvrability.

From the logical view point the structural aspects of the economy constitute a set of conditions which must be imposed on the solution we are looking for. And this set of conditions limits, of course, our degree of manoeuvrability in the economy.

The acceptance of a specific system of administrative rules and fixations constitutes an additional set of conditions. Therefore, if we pool the two sets together, the structural conditions and the administrative conditions, we have a larger and more exacting set of conditions. As long as we only impose the structural conditions we have a fairly high degree of manoeuvrability. But after having imposed the additional set of conditions which are expressed by the administrative rules and fixations, our degree of manoeuvrability is much lower. Therefore, if we require that a specific system of administrative rules and conditions is to be maintained unchanged i.e. if we say that we have to accept a specific administrative system that is defined in considerable detail, we will be left with practically no manoeuvrability. Then we will have to take the evolution of the economy as it comes. Then we have adopted the on-looker's view point, and have given up any initiative of steering the economy.

If we do want to steer the economy, we must consider the possibility of making changes in the administrative set of rules and fixations. When we speak about "measures of implementation" in steering the economy, it is precisely changes in the administrative rules and fixations we have in mind. In other words, admitting a large list of administrative means of implementation is the same thing as to admit the possibility of a large number of changes in the administrative rules and fixations. To sum up: Means of implementation consist in possible changes of administrative rules and fixations.

However liberal we may be in allowing a large list of possible changes in administrative rules and fixations, there are certain conditions which are unavoidable and which we must accept whether we like it or not, namely, the structural conditions. They are always there and limit our manoeuvrability. They constitute so to speak a rock bottom set of manoeuvrability-limiting conditions. This is why it is particularly interesting to study this set of conditions.

Preference analysis.

I must now go a little deeper into the problems of the preference analysis.

No rational action is possible before we have made clear to ourselves what we really want to obtain. Anybody will understand how absurd it would be if a person would walk up to the information desk in a bus terminal and ask the girl at the desk: "What bus line should I take? ", without making any further explanation. No answer to such a question is possible before one has made clear to what particular part of the country one wants to go. Similarly, in the case of a sharp-shooting competition it has no meaning to say : "Did I hit the mark? " when the shot was fired before the target had been put up.

It has been a source of constant wonder to me to observe that when it comes to the steering of the economy it seems that the public and the debating politicians, without any symptom of uneasiness, act and talk as if it should be possible to answer questions about whether one type of measure is an "effective" means or an "ineffective" means, without having specified what they want to obtain by using these means.

It is as if the public and the debating politicians think that it is possible to pick out the effective means of implementation as distinct from the ineffective ones, a "good" course to follow or a "bad" course to follow, to pick out these things without having specified what they want to obtain through the steering of the economy. It has been a constant wonder to me that people seem to think this way.

In economic and political life we hear daily a host of questions ; "Should the bank rate on loans be raised?" "Should the wage rate of industrial workers be increased?" "Should we build a road between points A and B in the country? " "Should the dried cod fish industry of Norway be monopolized?" "Should we promote investments that will give employment to many people , or should we on the contrary promote such investments which will save labour?" Or to quote some questions of a more general sort: "Should we aim at a high rate of increase in the gross national product, or should we put more emphasis on a socially justified distribution of it ?" . "Should we aim , above all, at keeping the price level under control?". " Or should we sacrifice the stability of the price level and put more emphasis on the increase of the gross national product?" . "Should we sacrifice a part of the increase of the total gross national product in order to be able to increase the living standard of one specific social group, say the fishermen or the industrial workers?"

When we ask such questions as these, it is impossible to give a meaningful answer without having made a quite explicit and quite precise statement regarding what we would like to see achieved in the economy.

And not only it if necessary to be quite explicit about what we want to obtain, but we must also be aware of the fact that such questions as these I mentioned, can't be answered separately. We must understand that they are knit together internally through the structure of the economy.

Direct and indirect effects.

The necessity of considering everything in conjunction springs from the fact that any measure we take in economic life has many effects. Some of them are direct and so obvious that even the superficial observer can see them. Others are indirect and work through the complex network of the whole economic structure and, therefore, can not be seen and understood by the casual observer and the general public, but need a profound analysis by economists and statisticians and a number of other experts in order to be discovered. To quote but a single example : Dr. Petter Jakob Bjerve, director of The Central Bureau of Statistics in Norway and his associates have recently found that an outlay of, say 100 mill. N. Kr. for housebuilding will entrail an import of 7,5 mill. if we only take account of the direct effects which the housebuilding will entail , but no less than 27.9 mill. imports if all effects, direct as well as indirect, are taken into consideration. Total import effect is nearly four times the direct import effect.

When the politicians are to base their decision on an economic analysis, they must not only take account of the direct effects but also of all the indirect ones, and we must take account of the fact that all these effects, whether direct or indirect , are tied together in a complex network that cannot be understood just by verbal reasoning. Sometimes each of the indirect effects may be small, perhaps even insignificant at first sight, but if they are pooled together, they may constitute a mighty stream that flows perhaps in the opposite direction of the one in which the direct effect flows.

This means that we should always be sceptical about what casual observers, say newspaper men write about "effects" of economic measures. Such writings are nearly always based only on the direct and very obvious effects.

An analysis of effects which is to be a real help to the politicians must take fully into account both the direct and the indirect effects, and the way in which all the effects are knit together in a complex network.

R. Frisch

The use of models

There is only one way in which these things can be studied in a meaningful way, namely through the construction of a well worked-out model . This is a conceptual scheme with variables and mathematical relations by which the economist and planner tries to keep track of all the direct and indirect effects and of the way in which they are tied together.

In particular, let us consider a structural model . This is a model where we only consider the structural aspects, i. e. those aspects of the economy which we have to take account of whether we like it or not. As a first - and fundamental - step in the analysis it is necessary to construct such a rock-bottom model as distinct from an administratively contaminated model.

The distinction between these two types of models resides in the type of relations which we pick as the logical basis for constructing the model. In the rock-bottom model we only take account of such relations which have the structural aspect : which are such that we have to accept them whether we like it or not. Or, in other words, such relations which are invariant under a change of the administrative rules and fixations. It is obvious that this type of rock-bottom model must be of tremendous importance when we try to find out what particular kind of changes in administrative rules and regulations that might be advantageous in steering the economy. If this is the purpose of the analysis, we must base it on something which itself is stable and fixed under all the possible changes whose effects we want to study.

I do not say that an administratively contaminated model can never be useful. It may be of use in certain partial and specific problems where we have stated a priori that there are a number of the administrative rules and fixations which we are to accept and which we do not want to change.

Such administratively contaminated models have, however, less practical applicability in development planning than one usually thinks.

Among Western economists there is all too much emphasis, sometimes perhaps unconsciously, put on building up models that are highly administratively contaminated and particularly contaminated by such types of rules and fixations and assumptions that are specific to the free market system.

Such models are particularly useless in an underdeveloped country. Here there is a great need to concentrate on the structural aspects of the problem and to be very free in admitting perhaps drastic means of implementations.

I had a very good proof of this when I lectured in the International Summer School in The Oslo University this summer. There was present a number of students from under-developed countries. One of them intervened and said in substance : "Here we call in Western economists and ask them to give us advice, and they present solutions which are derived from the well-known Western type of economy. They do not have the freedom of thought to see the problem in its structural aspect". He did not use the word "structural", but what he said amounted to that. And he asked me what they ought to do. I told him, "Send them back". That was my advice.

We must first of all base ou thinking on the rock-bottom structural analysis, and only afterwards, if at all, pay any attention to administratively contaminated models.

The problem of compromise and consistency.

This digression about the nature of models: a means of taking account of several things at the same time was necessary. We are now in a better position to revert to the question of preferences and discuss it in a more articulated way.

By way of examples I suggested a number of specific questions : should we raise the bank rate ? Should the wage rates of the industrial worker be raised ? Etc. A specific economic measure which may work positively for the obtainment of one of the things we like, may work negatively for many of the other things we like. So, we are facing a tremendous problem of compromise. This compromise is absolutely fundamental. It represents the essence of the political side of macroecønomic planning problem. No final answer to what is "optimal" can be given without an explicite and clear formulation of how the compromise is to be made. Before deciding about which bus to take, we must have cleared our minds about what part of the country we want to go to. For clarity of thought and honesty we should state publicly where we stand regarding preferences and prorities. And, equally important, we must have become consious of the fact that we cannot go to all parts of the country at the same time.

This is a nearly completely neglected question in the discussion amongst politicians. We are in Norway at this moment approaching a general election to be held on 12 and 13 September. And, of course, there have been posters and pamphlets from the various political parties . What do these posters and pamphlets say ? They simply list a number of good things which every body would like to have : No. 1, 2, 3, 4, etc. Therefore, vote with party XX. Then an other party will list the same things: No. 1, 2, 3, 4 , but now the conclusion will be that one should vote with party YY. But this is not all. The list of things No. 1, 2, 3, 4, is not only a list that is common to all parties, but the list itself is often inconsistent . It is often impossible to have all the items No. 1, 2, 3, 4, at the same time.

Ladies'shoes should be small and neat on the outside, and at the same time large and comfortable on the inside. Therefore, if I started a political party with the platform that I want to have ladies' shoes that are small and neat on the outside and big and comfortable on the inside, I would be sure of having a big following.

To summarize: Before we can hope to reach anything like a rational and democratic system of planning we must face the situation squarely. One measure will work positively in one direction and negatively in another direction. It is only through the construction of a well thought out decision model that we can succeed in keeping track of all effects, direct as well as indirect. And in addition, we must state in a very specific and explicite way what sort of compromise we like, i.e. what our preferences are. When the model and the preferences are available, a technique of mathematical programming will help us to find that combination of economic measures which will be optimal in the sense that, first this combination is feasible , and, second, amongst all feasible combinations it is that particular combination which is most condusive to the attainment of that particular compromise which we favour.

How to formulate macroeconomic prefereces.

Some people think that it is hopeless to arrive at anything like a formalization and definition of a coordinated system of preferences regarding the economic constellation. They think that the only thing we can do is to let the political parties continue their dog fighting and propaganda about individual economic measures which have to be decided upon one by one. This time honoured procedure is irrational in the extreme, and it is nothing but defeatism to think that we are doomed to stick to this irrational procedure forever. We must find a better procedure.

One of the popular arguments in favour of the defeatistic attitude is that there are so many different opinions about the preferences. Some people think that we ought to concentrate only on increasing the gross national product as quickly as possible, forgetting nearly everything else. Others think we have to concentrate more on ways and means of increasing the living standards of industrial workers, even at the risk of reducing somewhat the rate of growth of gross national product. Some will think most about the present, others will think more about the future. And so on. Such differences of opinion are by some taken as a proof that it is impossible to formalize the political preferences.

This is a big mistake. This is really one of the biggest pitfalls in the discussion of this matter. Differences of opinion, of course, there are . One social group may have one type of preference and another social group may have other preferences, and different persons may have different preferences, and even the same person may have different preferences at different points of time. All this is, of course, true. But there exists a machinery for settling differences of opinion. This machinery is simply the political system of the country. The system of the country - whatever this system may be - is just invented in order to settle differences of opinion. This is the purpose of the political system. What we have to do is to apply this very system for the formalization of a system of social preference , instead of applying it for deciding about concrete economic measures one by one.

It is not the task for us as economists and as social engineers to go into a complete and detailed discussion of the political system. We are allowed to hope for a constant evolution in a democratic direction. Democracy is not a stationary thing that can be established once and for all . It is a thing we must constantly work towards . But at any given moment we will realistically - and provisionally - have to accept an existing political system as a machinery for formalizing the system of social preferences which we need in order to arrive at rational basis for economic policy.

Before discussing a technique for reaching a compromise on a unified social system of preferences, let us consider the simpler problem of formalizing the preferences of a given political party.

The preference of a given political party.

What can the technical expert, particularly the economist working on the plan, do in order to find out what the preferences of a given political party are ? His work will proceed through three phases.

The first phase simply consists in his making a systematic use of his general knowledge of the political atmosphere in the country, in particular the political atmosphere in the particular party in question. The expert will have formed an opinion, a tentative opinion, about what the preferences of this party would look like if they were formalized in way that fits in with the expert's model and is expressed in a language that will be understandable to the electronic computer.

In the second phase the expert - on the basis of his tentative formalization - will work out a system of interview questions through which the system of preferences will be further quantified. These interview questions are of the following type : What would you , politicians, choose if you had the choice between two packages of economic results, for instance one package with, say, 3 percent unemployment and an annual inflation rate of 5 percent, and another package with, say, 5 percent unemployment and an inflation rate of 1 per cent . I have worked out a number of memoranda on the organization of such interview questions, most of them from the University of Oslo Institute of Economics, and one from the University of Pittsburg. Here let me just state that through such an interview system wisely constructed the expert planner will be able to form a fairly definite and precise idea about the preferences of the political authorities of the party in question.

In the third phase the expert will go back to his electronic computer in which he had already entered the data on the structure of the economy. To this he will now add the formalization of the preferences in the quantitative form as he now has it.
And from this will come out a solution, the structurally optimal development path for the economy. Optimality being defined through the preferences of this party.

Let me remind you that I am now thinking only of the structurally optimal solution, not of the administratively optimal solution. I am now thinking of the solution which is only subject to the structural conditions, i.e. those conditions which we have to accept whether we like it or not. The model to be used now has many degrees of freedom, many degrees of manoeuvrability, precisely because we have not added more conditions on the problem than are strictly necessary. Consequently the machine, having a large number of degrees of freedom at its disposal , will be able to produce a solution which is very high from the view point of the preferences as they have been formulated. The ensuing solution will be the structurally optimal solution under these preferences.

The expert will come back to the politicians with this solution. And the politicians, having now seen the complete solution, will, perhaps, say : "No, this was not really what we wanted... We have to change this particular aspect of the solution" .

The expert will understand more or less precisely what sort of changes are needed in the formulation of the preferences in order to produce a solution that comes closer to what the politicians now say they want. This leads to a discussion back and forth - a continuation of the cooperation between the political side and the expert side - in order to work step by step towards a preference formulation such that the politicians can say about the ensuing solution :"All right, this is really what we would like to see" .

I will revert once more to the ladies shoes being small on the outside and big on the inside. Perhaps the politicians in this cooperation will insist on having things that are inconsistent. Therefore they will always be dissatisfied with the solution which the expert brings back. When the expert has understood that such a situation is reached, he will politely say : "Your Excellencies, I am sorry but you can not have all these things on which you insist." And Their Excellencies, being intelligent persons, will understand the philosophy of the ladies'shoes, and, thereforeacquiesce with a solution which is not what they like completely but is at least something which is better than other alternative shapes of the development path which have emerged from the tentative solutions.

Even if we did not go any further with the formalization of the system of preferences than to work out such an analysis separately for each political party, an enormous gain would be obtained in the economic political discussions.

A compromise between the preferences of different political parties.

But, we sould not stop point. We should proceed to a discussion of what sort of political compromise that might be reached in the formulation of a unified system of social preferences . And then having reached this compromise formulation of the preferences, there would appear a compromise structurally optimal solution. Here too an iteration between politicians and experts would take place.

The top political authority - in a Western country it would be the elected Parliament - oughtto concentrate most of its efforts on a discussion of this compromise on the formulation of the system of preferences, instead of using practically all of its time on discussing one by one the specific economic measures that have been proposed, and for each of these measures deciding whether to accept it or not.

The latter procedure will be needed as a final check, but only as a final check. Before this final stage of the discussion is reached, one should settle on a formulation of the compromise system of preferences and discuss the ensuing structurally optimal solution. This is what ought to take up most of the time and energy of the top level authority. A suggestion of a voting system for this type of political discussion is given in my paper in the international journal "Economics of Planning " No. 1-2, 1965.

Only after this fundamental framework for the economic policy has been decided upon, and then in connection with the decision on what means of implementation to use, should individual measures be decided upon.

In my mind this procedure will be the most essential aspect of a reform that could foster a development in the direction of a rational and democratic form of planning. By following this the Parliament would concentrate its time and energy on that sort of things which a Parliament is able to digest and see the consequenses of, and on that sort of things for which we need a Parliament decision. When this is done, many details can safely be left to the experts. And the Parliament will get more time for discussing vital issues.

Impossibility of listing alternatives.

Sometimes we hear the suggestion that instead of going to the trouble of discussing preferences in the way I have indicated, one should leave it to the experts to put on the table of the politicians a number of alternatives for the development path of the nation's economy, and ask the politicians to choose amongst these alternatives.

This is an absolutely impossible way of proceeding. The reason for this is simply the following. In political life there is a nearly infinite number of specific economic issues that may come up : Should the interest rate be raised ? Should the road between points A and B be built ? Etc. I have previously mentioned a number of examples of such specific questions. If we should ask the experts to produce a liste of feasible alternatives for the development path of the economy, a list that would be comprehensible enough to cover alternative answers to all the various specific questions that are under debate, the list of possible development paths would have to contain millions and millions of alternatives. The number of alternatives would multiply by cross classifications.

Such a list is impossible for the simple reason that the experts would be physically unable to analyse and present all these millions and millions of alternatives. And even if these millions and millions of alternatives could be analysed and put on the table of the politicians, the politicians would be absolutely drowned in information. They would not know where to start or where to end in discussing which alternative to choose.

On the electronic computer one speaks of "information death" when one has made the mistake of coding the computer in such a way that it prints out too many of the itermediate results that it has reached in the course of its work. The poor politicians would suffer a similar "information death" if they found on their table the hypothecal list of the millions and millions of feasible economic development paths, amongst which they would have to choose.

In rational economic planning there is no other possibility than to start in patiently on a discussion of the compromise system of social preferences and in this way to lay the foundation for finding structurally optimal solution. This solution would contain as many details as the top level political authority is able to digest . If the power of digestion is small, the model used must be heavily aggregated.

The means of implemenation.

The administratively optimal path as distinct from the structurally optimal path.

When it comes to a choice of the means of implementation, i.e. a choice of the type of changes in administrative rules and fixations which are to be taken as politically permissible, we must remember that the more restricted we make the list of these politically permissible changes the more we will restrict the degree of manoeuvrability in steering the economy. Suppose we have started by determining the structurally optimal development path over the subsequent years. This path will represent the absolutely best thing we can possibly obtain, because here we have only imposed such conditions as are unavoidable. If we now assume a rather restricted list of means of implementation that are to be politically permissible, we have restricted our manoeuvrability so much that it will probably not be possible actually to reach, actually to implement, a development path that is as desirable (from the viewpoint of the compromise system of social preferences) as the structurally optimal development path. We have to acquiesce with something that is less desirable. This we have to do because we have restricted the list of means of implementation that are politically permissible .. This is the penalty for being conservative.

We will get a sound piece of information if we bring out by how much the administratively optimal solution which emerges from a restricted list of politically permissible means of implementation, lag behind the structurally optimal solution. This "lagging behind" being defined through the compromise system of social preferences.. By such a comparison we would visualize the penalty we have to pay for restricting the list of politically permissible means of implementation. And it would show that the more radical we are, in the sense of admitting as possible a great number of substantial changes in administrative rules and fixations, the larger is the possibility that we will be able to reach closely to the structurally optimal solution. To find something that is better than the structurally optimal solution is, as we know, impossible.

The structurally optimal solution was only an intermediate step on our road towards the final plan. The final plan - and only this - is concrete in the sense of being something which we will try to go through with in practice. The structurally optimal solution was an intermediate step which was needed in order to bring out fully the implications of the structure of the economy and our preferences. In the subsequent implementation analysis we lay bare how, and how far political prejudices may prevent us from reaching a really high level of those things which we consider as desirable from the viewpoint of the compromise system of social preferences.

Control and communication.

Working out a plan, and settling on a system of means of implementation is not enough. There must be current control of how the plan works in practice. There must be a standing watch. And this watch must be organized. It embraces three stages.

R. Frisch

I. Current reporting.

The first stage in the control work consists in organizing current reporting on what actually happens after the plan has been put into effect . Here we meet specific problems of communication : How can we get reliable information about the actual situation and about the existing tendencies towards a change in the situation? Information comes from different sources.

In the first place we get information throught current and quickly available statistics. In this connection sample surveys in various forms, including opinion polls taken anongst expert observers, play a particularly important role because such surveys can be carried out quickly.

In the second place information becomes available through a loyal political opposition. The opposition must be alert and pin down those features of the development which it thinks are unfortunate and contrary to the intention of the dynamic plan that has been adopted. Here we have one of the most important features that gives strenght to a free democracy. Under a dictatorship - whether it be a direct political dictatorship or indirectly a dictatorship through experts of the kind I have denounced - it frequently happens that the central authority does not become informed, or does not become informed early enough, when something goes wrong. Of course, not everything which a political opposition brings forward can be taken at its face value. Sometimes it must be taken with a big grain of salt. Here applies the old saying that one ought to listen to most of what is being said, but believe only a small part of it. But the voice of the political opposition must not be brushed aside just because it comes from the opposition. Under a dictorship - and even under a government with a strong absolute majority in Parliament - this happens too often.

In the third place information becomes available through a free press. Here is a second important feature that gives strenght to a free democracy. Also in the case of the free press much of the information must, of course, be taken with a big grain of salt, but one must not brush it aside just because it is only something that "has been in the newspaper."

II. Rapid processing of the incoming information.

The second stage in the planning control work is a rapid processing in the central planning authority of all currently incoming information, with a view to finding out if any of the means of implementation ought to be changed in order to bring the development path of the economy cleser up to the ideal that is represented by the structurally optimal solution. This means that one is fully aware of the possibility that errors might have been committed in the earlier analyses and political discussions on the effects to be expected from the application of specific means of implementation

In this study of possible changes in the means of implementation one must be fully aware of the stability problem. A given change in the means of implementation which at first sight seems to be susceptible of counteracting an undesirable feature of the economic development , may on further scrutiny prove to have precisely the opposite effect. If one is uncritically led by such first sight impressions, one may set up oscillations in the economic system, oscillations that may become wilder and wilder the more one tries to keep them down.

Let me use an illustration . Suppose we have a pendulum that has started to swing while we want it to be at rest. If in this situation we apply an outside force that presses the pendulum towards the right when it has swung to the left, and vice versa, we can be sure of only producing more violent oscillations, because the application of such an outside force is tantamount to increasing the intensity of the gravitational field. Just as two men placed one on each side of a swing with a child, will keep the child constantly swinging if each of them pushes the child towards the central position whenever the swing with the child comes to his side.

If damping of oscillations is what we want, we must procede in an entirely different way. We must, for instance, use the principle of the oil-brake or the air-brake. This will dampen the intensity of the swing because it is tantamount to increasing the friction.

Similar considerations, expressed in a refined mathematical model, must guide the central planning authority when it proposes changes in the means of implementation , on the basis of incoming information on happenings in the economy. This involves continuous work both analytically and with respect to political discussions.

III. The counteraction.

The third stage of the planning control work is the counteraction, that is to say : The carrying out of such changes in the means of implementation which through the scientific analysis and the political discussions , utilizing the incoming information, have been found desirable.

These three stages in the control work: Current reporting, rapid processing and counteraction, can be illustrated by the functionning of an ordinary room thermostat. The thermostat is set for a given temperature. This illustrates the structurally optimal plan. If the room temperature falls too much below the set temperature, the thermo-sensitive spring in the thermostat contracts. This illustrates the incoming information. This contraction closes an electric circuit which starts the oilburner. This illustrates the counteraction, i.e. the changes in the means of inplementation. When the temperature in the room rises above the set temperature, the spring in the thermostat relaxes. This illustrates a new round of incoming information. The electric circuit is cut off, which illustrates a new round of counteractions, i.e. of changes in the means of implementation.

Such series of operations in three stages : Current reporting, rapid processing and counteraction , is something that finds application not only in small technological processes and in the complicated system of causes and effects in economic and social like . It can be applied in nearly every conceivable field, for instance in medicine or in the biological processes that govern the development of living organism etc. The new science of cybernetics studies the principles that govern all processes that can be brought into the three-stage conceptual framework of current reporting, rapid processing and counteraction.

In the sphere of control and communication subsequent to the putting into effect of an economic and social plan , we much to learn from cybernetics. But the general cybernetic principles must not be swallowed uncritically. They must be digested, understood and sometimes reshaped with a view to the special application we want to make of them in economic and social planning .