Diss. ETH No. 12493

COMPUTING ECONOMIC EQUILIBRIA AND ITS APPLICATION TO INTERNATIONAL TRADE OF CO_2 PERMITS: AN AGENT-BASED APPROACH

Dissertation submitted to the

SWISS FEDERAL INSTITUTE OF TECHNOLOGY
ZURICH, SWITZERLAND

for the degree of

Doktor der Mathematik

presented by

BENNO PAUL BÜELER

Dipl. Math. ETH, Dipl. Ing.-Agr. ETH
born December 13, 1961
citizen of Schwyz (SZ)

accepted on the recommendation of
Prof. Dr. H.-J. Lüthi, examiner
Prof. Dr. B. Schips, co-examiner
Prof. Dr. J.-P. Vial, co-examiner

1997

To Anna, Joel
and my Parents

Leer - Vide - Empty

Acknowledgment

I am very grateful to Prof. D. Onigkeit for both supporting my mathematical studies and encouraging my environmental interests.

Prof. H.-J. Lüthi has my fullest gratitude for the freedom he gave me to do this work and his generous support in all kind of ways. His critical and fundamental discussions have decisively improved the work.

I am indebted to Prof. J.-P. Vial and Prof. B. Schips for their thorough and constructive review of the work.

S. Kypreos deserves the honor of having suggested the core subject of this work; furthermore, he contributed various regional models and data, and carefully reviewed the results. I am also obliged to researchers in the Netherlands and Sweden who supplied me with data.

Finally I thank all the people in the Institute and outside who supported me in many ways. Specifically, I appreciate the discussions relating to economy I have had with Jörg Wild, Stefan Felder and Werner Hediger.

Leer - Vide - Empty

Contents

List of Tables

List of Figures

Abbreviations

12–RT	12–Region Trade model
ACCPM	Analytic Center Cutting Plane Method (cf. Algorithm 2 page 23)
aeeifac	autonomous energy efficiency improvement factor (see page 118)
BNL	Bruckhaven National Laboratory (USA)
CES	Constant Elasticity of Substitution
CGE	Computable General Equilibrium
CH	Confoederatio Helvetica, i.e. Switzerland
CoGCPM	Center of Gravity Cutting Plane Method (cf. Algorithm 2 page 23)
CPM	Cutting Plane Method (Algorithm 2 page 23)
CPU	Central Processing Unit
EEP	Economic Equilibrium Problem (Definition 1.4 page 4)
EP	(Nash) Equilibrium Problem (Definition 2.4 page 17)
ETA	Energy Technology Assessment model
ETSAP	Energy Technology Systems Analysis Project
FPP	Fixed Point Problem (see (FPP) page 7)
GAMS	General Algebraic Modeling System
GDP	Gross Domestic Product
GHG	Greenhouse Gas
GNP	Gross National Product
IEA	International Energy Agency
IIASA	International Institute for Applied System Analysis (Austria)
KFA	Kernforschungsanstalt Jülich (Germany)
KKT	Karush-Kuhn-Tucker conditions (see Section A.1)
LP	Linear Programming
Macro	Macroeconomic growth model
Markal	Market Allocation
MESAP	Microcomputer based Energy Sector Analysis and Planning System
MINOS	Model Incore Nonlinear Optimization Solver
MM	Markal-Macro model
MM^{mr}	Markal-Macro multi region (see Section 7.1)
MUSS	Markal User's Support System
NCP	Nonlinear Complementarity Problem
NL	The NetherLands
NLP	NonLinear Programming
PSI	Paul Scherrer Institut (Switzerland)

QVIP	**Q**uasi**V**ariational **I**nequality **P**roblem (cf. Definition B.1 page 105)
RES	**R**eference **E**nergy **S**ystem
SMEDE	**S**imulation **M**odel for **E**nergy **DE**mand
SW	**SW**eden
VIP	**V**ariational **I**nequality **P**roblem (Definition 3.1 page 22)
WARP	**W**eak **A**xiom of **R**evealed **P**references (Definition 3.13 page 39)

Zusammenfassung

Seit der Postulierung der berühmten 'unsichtbaren Hand' von Adam Smith vor 200 Jahren haben Ökonomen eine ambivalente Haltung gegenüber Wettbewerbsgleichgewichten. Einerseits ist es *das* grundlegende Konzept der Marktwirtschaft und intuitiv einfach zugänglich, andererseits stellt dessen formale Handhabung grosse Probleme. So gelang z.B. erst in den dreissiger Jahren dieses Jahrhunderts ein erster Existenzbeweis von Gleichgewichten für bestimmte Modelle. Aber auch die algorithmische Handhabung selbst einfacher Modelle erweist sich vielfach als schwierig und erfordert im allgemeinen ein genaues Verständnis der Modell-Strukturen.

Von den zahlreichen Möglichkeiten, Wettbewerbsgleichgewichte formal zu behandeln, wird in dieser Arbeit der Fokus auf Agenten und deren Reaktion auf Preissignale gelegt. Dabei kann ein Agent je nach Situation verschiedenes repräsentieren: ein Konsument, ein Produzent, ein ganzer Wirtschaftssektor, eine geographische Einheit (Land), usw. Bei gegebenem Preis ist die Reaktion der Agenten definiert als Netto-Verkauf (Angebot minus Nachfrage, oder Export minus Import), was summiert über alle Agenten als Exzessfunktion bezeichnet wird.

Ein Wettbewerbsgleichgewicht mit einem nichtnegativen Gleichgewichtspreis ist dann gefunden, wenn entweder das Angebot und die Nachfrage übereinstimmen, oder aber das Angebot grösser als die Nachfrage und zugleich der zugehörige Preis Null ist.

Motiviert wird die Wahl, Wettbewerbsgleichgewichte auf der Ebene von Exzessfunktionen zu betrachten, durch eine Reihe von spezifischen Vorteilen: dazu zählen die breite Anwendbarkeit auf verschiedenste Gleichgewichtsprobleme, die einfache Integrierbarkeit bestehender beliebig heterogener Agenten in ein übergeordnetes Gleichgewichtsmodell, oder die offensichtliche Parallelisierungsmöglichkeit in der Behandlung der einzelnen Agenten. Dabei stellten sich die zwei letzten Punkte als entscheidend für das in dieser Arbeit konkret betrachtete Energie-Ökonomie-Modell MM^{mr} (Markal-Macro multi-region) dar. Diese Vorteile dürften auch für viele andere Modelle relevant sein. Allerdings erweist sich als einer der gravierendsten Nachteile dieser Sichtweise die für einen Konvergenzbeweis der angewandten Algorithmen im allgemeinen nicht gegebenen Struktur-Voraussetzungen der Exzessfunktion.

Aufgrund der entscheidenden Vorteile werden in dieser Arbeit zwei Heuristiken

entwickelt, die das Gleichgewichtsproblem basierend auf der Exzessfunktion lösen. Zum einen ist das ein Schnittebenenverfahren, welches im Rahmen von Variationsproblemen diskutiert wird, und zum anderen ein Fixpunktverfahren. Ein Beitrag dieser Arbeit findet sich dabei in der Diskussion der Monotonie der Exzessfunktion, welche sich als zentral für die Konvergenz des ersten Verfahrens heraustellt. Weiter wird die Behandlung der Unbeschränktheit der Lagrange-Funktion in gewissen Fällen untersucht. Diese Unbeschränktheit tritt bei der Dekomposition des Optimierungsproblemes auf, welches dem Fixpunktverfahren zugrundeliegt. Als erstes interessantes empirisches Resultat erscheint dabei die Robustheit des Fixpunktverfahrens, sofern eine spezifische primal-duale Beziehung zwischen den zwei Verfahren ausgenützt wird.

Um Algorithmen sinnvollerweise einzusetzen muss sichergestellt sein, dass eine Gleichgewichtslösung überhaupt existiert. In einer vergleichenden Diskussion werden verschiedene Beweisstrategien mit einigen Verallgemeinerungen und Ergänzungen für die Existenz eines Gleichgewichtes einander gegenübergestellt. Eine davon wird schliesslich an MM^{mr} angewandt.

Der mehr ökonomisch ausgerichtete Teil der Arbeit beginnt mit einem Exkurs zu Energie-Ökonomie Modellen unter dem Gesichtspunkt von CO_2 Emissions-Beschränkungen. Diskussionsbeiträge finden sich hier im Bereich des 'burden-sharing' und der Implementation von CO_2 Emissions-Zertifikaten.

Aufbauend auf nationalen Energie-Ökonomie Modellen (Markal-Macro) werden unterschiedliche Konzepte zur Modellierung von CO_2 Emissions-Zertifikaten vorgestellt. Diese werden einerseits zur Integration der nationalen Markal-Macro Modelle im Mehrländermodell MM^{mr} benutzt. Andererseits werden die Konsequenzen der unterschiedlichen Zertifikats-Modellierung in diesem Kontext auch analysiert.

Beruhend auf Daten von Schweden, den Niederlanden und der Schweiz wurden die zwei entwickelten Heuristiken schliesslich an MM^{mr} erfolgreich getestet. Vorbehaltlich der bei Modellrechnungen zu machenden Relativierung der numerischen Resultate ergeben sich doch einige interessante ökonomische Einsichten. So erscheint der auf das Jahr 2000 diskontierte Preis für solche Zertikate umgerechnet bei etwa 20 Rappen pro Liter Treibstoff zu liegen, wenn eine 40%-ige Abnahme der Emissionen bis ins Jahr 2040 vorgegeben wird. Für dieses Szenario liegen die Verluste des BNP (Brutto-Nationalprodukt) im Bereich von 2% gegenüber einem Referenz-Szenario ohne Emissionsbeschränkung. Diese Verluste können um etwa einen Fünftel verringert werden, wenn statt fixen länderweisen Emissions-Beschränkungen handelbare Zertifikate eingeführt werden. Bemerkenswert ist auch die länderweise unterschiedliche Verteilung der Verluste gemessen am BNP. Da diese Verteilung der Verluste durch die Erstausstattung mit Zertifikaten direkt steuerbar ist, können solche Modelle bei der Aushandlung der Erstausstattung sowie möglicher Transferleistungen eine wichtige Entscheidungshilfe leisten.

Summary

Since Adam Smith postulated the 'invisible hand' 200 years ago, economists have had an ambivalent position towards competitive economic equilibria. On the one hand it is *the* fundamental paradigm of the market economy system and intuitively easy to understand. On the other hand its formal treatment poses considerable difficulties. The first proof of existence for certain models was possible only in the 1930's; but the algorithmic treatment of even simple models has often proved to be hard due to the need of an accurate insight into the concrete model-structure which can be hard to obtain.

There are various ways to formalize equilibria; in this work equilibria are formalized through the reaction of economic agents to price signals. Here 'agent' is used to denote different things, depending on the context: a consumer, a producer, a whole economic sector, or a geographic unit like a country, etcetera. The 'reaction' of an agent is defined as the net selling (supply minus demand, export minus import) which is determined by price. The summing of the reaction of all agents is called (market) excess.

An equilibrium with non-negative price is found when either supply equals demand or supply exceeds demand and the corresponding price is zero.

The choice to study equilibria on the level of the excess-function was motivated by a number of specific advantages including its broad applicability to different economic equilibrium problems, its simplicity of integrating existing and arbitrarily heterogeneous agents in an overall equilibrium model, and its possibility to treat agents in parallel. For MM^{mr} (Markal-Macro multi-region), the energy-economy model studied in this work, the last two advantages are of decisive value. A serious disadvantage of this excess-based view is the possible lacking of structural properties of the excess-function which are required for proving convergence of related algorithms to equilibria.

The above mentioned advantages, however, necessitated the development of two main heuristics to solve the equilibrium problem based on the excess-function approach. The first, the Cutting Plane Method (CPM), is derived from a formulation of the equilibrium problem as a Variational Inequality Problem (VIP). The second heuristic is a fixed point method.

Contributions to the solution of equilibrium problems include the mathematical analysis of monotonicity of the excess-function, the clarification of the central role of monotonicity when applying CPM, and the treatment of unboundedness of the Lagrangian-function in some cases. This unboundedness appears in the decomposition of the optimization problem which underlies the fixed point problem. Another contribution is the discussion and extension of different strategies to prove the existence of an equilibrium. One of these strategies is finally applied to MM^{mr}.

One of the study's empirical result is the robustness and convergence of the fixed point method when a specific dual relationship to the VIP is utilized.

The economically oriented part of this study starts with a discourse upon energy-economy models from the perspective of CO_2 emission bounds. Specific attention is given to burden sharing and the implementation of CO_2 emission-permits.

Based on national energy-economy models (Markal-Macro), different possibilities to model CO_2-permits are developed. First, the permits are used to integrate the national models in the international model MM^{mr}. Second, the consequences of the different permit strategies are analyzed.

Using data from Sweden, the Netherlands and Switzerland, the two heuristics are finally successfully tested. Even though the resulting numbers must be interpreted cautiously, some interesting economic trends can be observed. Assuming a CO_2 emission scenario which reduces linearly the emission by 40 % from 2000 to 2040, the average permit price is calculated to be 14 US cents per liter of fuel if discounted back to the year 2000. Furthermore, the GNP-losses are around 2 % compared to a reference case without emission bounds. In our model these losses can be reduced by one fifth if tradable permits instead of fixed national emission bounds are introduced. Significant economic differences were observed between nations. Because the distribution of gains and losses can be influenced directly by the initial endowment with permits, models like MM^{mr} can be useful as a decision support tool when initial endowments or transfer payments are negotiated.

Introduction

This study is motivated by an ecological concern: The rise of global mean temperature due to Carbon Dioxide emission. It investigates a possible strategy which could be used to reduce the level of this harmful emission.

CO_2 is widely recognized today as the single most influential greenhouse gas (GHG) emitted by human activities, and is therefor considered to be the main culprit in the observed rise of global mean temperature. Large sudden changes in the global climate seem to have happened regularly in prehistoric times and are in that sense part of the ecological system 'earth'. However, the great complexity of human society today makes mankind more vulnerable and sensitive economically and socially to this climatic change, and unfortunately, the less developed the country is the more it stands to suffer.

Under these circumstances politicians and decision makers must grapple with a number of dilemmas; for example, what abatement or mitigation strategies should be implemented if costs occur today but 'revenues' (avoidance of damage) are uncertain and might occur in a later date? And further, which of these strategies are politically viable, cost efficient and effective? What effects may have the implementation of such strategies on international equity and burden-sharing?

One of the strategies which has recently grown in popularity both economically and politically is tradable CO_2 emission permits. As a CO_2 abatement instrument emission permits are effective, cost efficient, and allow direct negotiation of burden-sharing by means of initial permit endowments.

The *focus of this study is to solve a competitive economic equilibrium problem* (EEP) resulting from international trade of CO_2 permits. The equilibrium problem is formalized using models representing the national economies called agents. Schemes which integrate various agents in an overall equilibrium framework are therefor investigated. Mathematically, agents are treated as oracles, which, given a price signal, return the resulting excess of supply minus demand of the goods traded.

While such an oracle-based perspective is attractive for model-builders who are free to design any kind of agents in any kind of modeling environment, its mathematical treatment presents difficulties. What mathematical structures can be exploited to solve such an agent-based equilibrium problem?

To date, exact methods can be roughly classified into two groups. On the one hand the fixed point based methods (see [33, 99]) and on the other hand a variety of methods which are usually based on a reformulation of the equilibrium problem as a variational inequality problem (VIP) or nonlinear complementarity problem (NCP), see [47, 81].

While fixed point methods require little mathematical structure to hold, and thus can in principle solve an equilibrium problem given by a set of arbitrary agents, their theoretical and practical performance is not convincing. These weaknesses are considerably worsened when the evaluation of an agent is costly. The second group of methods utilize the problem's structure to a larger extend and therefore exhibit, as a rule, a superior performance. But, being tailored for specific agent structures they can not be applied when agents are only given by an excess-oracle; or, in Harker's words [46], 'CGE-like models[1], or any numerical approach to equilibrium computation, suffers from the curse of specificity.'

Taking into account those difficulties, heuristic methods can be attractive. Two heuristic methods are presented here, a fixed point based and a VIP-based method, which share a certain dual relationship. Both give convincing performance in practice for our specific model and are also applicable to a wide range of different models. Importantly, both concepts strongly support the integration of arbitrary agents into an overall equilibrium framework without the requirement of a single modeling environment or a reformulation of the agents. In that sense both concepts can be useful tools in a decision support system where varying agents can be easily integrated and thereby a sensitivity analysis on the level of agent-models can be performed. However, while agents' integration for the VIP-based approach presents no difficulties, integration for the fixed point approach requires decomposition. Decomposition is used here in reverse, for integration and not for subdivision, and is crucial for the integration of agents, without their reformulation, in the fixed point based approach.

One important virtue of the VIP-based heuristic method is that if the excess-function is monotone, i.e. fulfills a structural assumption, then the method provably yields a solution. And furthermore, as stated by the theory of economics, monotonicity is likely to hold for economically reasonable agents.

The study starts with a general, theoretical exposition and moves in the subsequent chapters into a more specific discussion of the concrete equilibrium model 'Markal-Macro multi-region' MM^{mr}.

Chapter 1 introduces the basic definitions related to economic equilibrium problems and formally links them to MM^{mr}. Next it formulates a specific VIP as the first approach for treating equilibrium problems.
Chapter 2 discusses strategies of proving the existence of an equilibrium. One strategy will be used later to actually prove the existence of an equilibrium solution of MM^{mr}. Another strategy underlies the second approach for treating EEPs

[1] Computable General Equilibrium

called 'conceptual Negishi algorithm'.
Taking up the VIP from Chapter 1, *Chapter 3* addresses the notion of monotonicity, and introduces thereby two algorithms for solving the VIP.
Similarly *Chapter 4* builds on Chapter 2; it presents two Negishi-algorithms and discusses the resulting decomposition problem.
As a conclusion to the previous two chapters a qualitative comparison between the algorithms is given in *Chapter 5*. The comparison is extended by two representative advanced algorithms from the literature for solving equilibrium problems, thereby clarifying the advantages and disadvantages of the different equilibrium solution methods.
This mathematical focus is dropped in *Chapter 6* where some economic background to emission permits and related energy-economy models is given.
Chapter 7 is devoted to the construction of MM^{mr} and to the analysis of different aspects of introducing emission permits.
Finally, the results of applying the MM^{mr}-model to data from Sweden, the Netherlands, and Switzerland are presented in *Chapter 8*.

The appendix discusses some technical background. It starts in *Appendix A* with a brief compilation of the Karush-Kuhn-Tucker (KKT) theory and an introduction to VIPs.
Appendix B presents an alternative approach for the proof of the existence of a solution for EEPs based on an up-to-date view of VIP.
Appendix C introduces the models, Markal and Macro, which appear in the regional agents of MM^{mr}.
The concrete proof of an equilibrium for MM^{mr} is given in *Appendix D*.
Implementation details of the algorithms are presented in *Appendix E*.
An empirical comparison of the algorithms discussed in this work appears in *Appendix F*.

For the mathematically inclined reader Chapters 1–5 together with Appendix B are of more relevance. Economists may find Chapters 6–8 more rewarding, and politically oriented readers may want to focus solely on Chapter 8.

A final word on the burden of notation is in order. In principle the notation is designed to meet the specific needs of the different sections. The more economically oriented sections use more the specific economic notation, whereas the mathematical sections obey the notation of corresponding mathematical fields. As a consequence, the same abbreviation can designate different objects in different sections. Notational differences, however, are always made explicit. Among the notational conventions in this work the following should be observed:

(i) Let x be a vector with components x_i; if such a component is itself a vector, a scalar component is denoted by x_{ik}; if x is build up by the components $x_1, \ldots, x_n$ we write for the corresponding vector $x = (x_1, \ldots, x_n)$ and make no notational difference whether the parts x_i are vectors or scalars. Particularly, the usual compound, $x = (x_1^T, \ldots, x_n^T)^T$ for vectors x_i, is simplified to $x = (x_1, \ldots, x_n)$.

(ii) Vector relations are always meant component-wise, consequently addition and subtraction of vectors happen also component-wise.

(iii) Sequences $\{x^n\}$ are both symbolized by curly brackets and a superscript, e.g. n, omitting the subscript '$n \in \mathbb{N}$'; convergence to a point x is symbolized by a simple arrow, $x^n \to x$, dropping any '$n \to \infty$'; there are, however, rare instances where a superscript has a different meaning, so as in the initial endowment x_r^0 of region r, or in the trivial case of a power, e.g. x^2, but all these differences are clear from the context and do not appear mixed.

(iv) Avoiding the usual mathematical sloppiness, we differentiate between objects like x and $x(p)$; while the former is any quantity, the latter is a map with argument p. Such seemingly confusing notation is used where a semantic relation between x and $x(p)$ is emphasized; for example, while x_r denotes a variable of region r, $x_r(p)$ denotes the set of x_r where the utility is maximal for a given price p.

(v) Other basic notations include the following. $\mathbb{R}_+^m$ denotes the set of m-dimensional real vectors with non-negative components (≥ 0); $[a, b]$ is the closed interval from a to b; if round brackets are used the corresponding boundary point does not belong to the interval set, e.g. $[a, b) = \{x \mid a \leq x < b\}$; set designators stand for the whole set, e.g. T in $t \in T$, as well as for the last element in the set, e.g. $t \in \{1, \ldots, T-1\}$ where t gets all values of T but the last; the cardinality (number of elements) of a set T is written $|T|$.

(vi) ∇^k denotes the differentiation operator applied k times where, as usual, the exponent '1' is dropped. The map to which ∇^k is applied may be single or vector-valued. If the map depends on a vector-variable x but the differentiation is done only with respect to a subset $\bar{x}$ of components of x, we write $\nabla_{\bar{x}}^k$.

Vor dem Gesetze steht ein Türhüter. Zu diesem Türhüter kommt ein Mann vom Lande und bittet um Eintritt in das Gesetz. Aber der Türhüter sagt, dass er ihm jetzt den Eintritt nicht gewähren könne. Der Mann überlegt und fragt dann, ob er also später werde eintreten dürfen. "Es ist möglich", sagt der Türhüter, "jetzt aber nicht". Da das Tor zum Gesetz offensteht wie immer und der Türhüter beiseitetritt, bückt sich der Mann, um durch das Tor in das Innere zu sehen. *F. K. [58]*

Chapter 1

The Name of the Game: Economic Equilibrium Problems (EEP)

The chapter is organized as follows. In Section 1.1 a simple formalized economy, including a finite set of producers and consumers, is introduced; it follows essentially Negishi's [82] exposition, relaxing to some extend its assumptions, and defines the notion of welfare, Pareto optimality and economic equilibrium. While those definitions are used throughout this work, the Assumptions 1.1 will be needed in Section 2.2 when proving the existence of an economic equilibrium. Here a reference to the excellent monograph *Theory of Value* by Debreu [18] is apposite, where the concepts and assumptions used by Negishi are discussed more in depth.

Section 1.2 brings the general notation into formal correspondence with the basic structure of MM^{mr}, the concrete energy-economy model studied in this work. Different to the above abstract economy the production is here part of the 'consumer'.

We continue by a brief outlook on MM^{mr} in Section 1.3.

In Section 1.4, finally, we present possible equivalent formulations of the rather *abstract equilibrium conditions as complementarity* problem (NCP), variational inequality problem (VIP) and fixed point problem (FPP). They will be used in following parts of the work. Namely, (VIP) is the problem solved by the cutting plane methods discussed in Chapter 3.

Contributions of this chapter include the Sections 1.2 and 1.3.

A final notational remark is in order. The classical economic theory considers usually consumers and producers. We stick to this habit when presenting fundamental economic concepts. However, from a more general point of view they can be simply called (economic) 'agents'. In the concrete application MM^{mr}, finally, we talk of regions. Following these different standpoints we will use the appropriate notion in the different parts of this work.

1.1 A World of Producers and Consumers

A simplified economy can be thought of as a set of consumers (e.g. people) maximizing their utility and producers (e.g. firms) maximizing their profit. The profit in turn goes back to the consumers depending on the share of ownership.

Definition 1.1 *(cf. [82])*

I	*(finite) set of consumers;*
J	*(finite) set of producers;*
m	*dimension of the space of goods;*
$x_i \in \mathbb{R}^m_+$	*consumption of consumer i;*
$x_i^0 \in \mathbb{R}^m_+ \setminus \{0\}$	*initial endowment of consumer i;*
$y_j \in \mathbb{R}^m$	*production vector of firm j;*
$F_j(y_j) : \mathbb{R}^m \to \mathbb{R}$	*production function of firm j;*
$\lambda_{ij} \geq 0$	*profit share of firm j distributed to consumer i;*
$U_i(x_i) : \mathbb{R}^m_+ \to \mathbb{R}$	*utility of consumer i with consumption vector x_i;*
$p \in \mathbb{R}^m_+$	*price vector.*

Note that by definition x and p are non-negative. The production function characterizes the set of possible (feasible) production vectors, that is, y_j is a feasible production of firm j if and only if $F_j(y_j) \geq 0$. The profit share fulfills $\sum_i \lambda_{ij} = 1$.

As mentioned in the introduction we abbreviate $x := (x_1, \ldots, x_I)$ or $U(x) := (U_1(x_1), \ldots, U_I(x_I))$; furthermore, vector-relations are meant component-wise, e.g. $F(y) \geq 0$ means $F_1(y_1) \geq 0, \ldots, F_J(y_J) \geq 0$. The quantity $e := e(x, x^0, y) := \sum_i (x_i^0 - x_i) + \sum_j y_j$ is called excess (supply minus demand) and is defined for any feasible (x, y). Later, based on optimal vectors $(x(p), y(p))$, a different excess definition $e(p)$ will be given which is a function of the price p.

We make in this chapter the following assumptions:

Assumption 1.1 *(cf. [82])*

1. *$U(x)$ is once continuously differentiable, non-decreasing, strictly increasing in at least one good, and concave;*

2. *$F(y)$ is once continuously differentiable and quasi-concave;*

3. *$\exists y^* : F(y^*) > 0$ and $0 < \sum_i x_i^0 + \sum_j y_j^*$;*

4. *With $Y_j := \{y_j \mid F_j(y_j) \geq 0\}$ and $Y := \sum_j Y_j = \{y \mid y = \sum_j y_j, y_j \in Y_j\}$ it must hold*

(a) $0 \in Y_j$: *possibility of no production;*

(b) $Y \cap \mathbb{R}^m_+ = 0$: *'negation of the land of Cockaigne' (no free lunch);*

(c) $Y \cap (-Y) = 0$: *irreversibility of the production process.*

Let us briefly comment these assumptions. Non-decreasing means $x' \geq x$ implies $U(x') \geq U(x)$; U is strictly increasing in at least one component, if for each $x \geq 0$ and for all $i \in I$ there is k such that $U_i(x'_i) > U_i(x_i)$ if $x'_i \geq x_i$ and $x'_{ik} > x_{ik}$. The latter requirement is very modest, because otherwise we may have constant utilities, for which trivial equilibria and welfare maxima can be given. Concavity of the utility functions implies convexity of all level sets and can be interpreted economically as non-increasing marginal utility.

The second assumption—quasi-concavity of $F(y)$—is equivalent with convex production sets.

The third condition, sometimes denoted as Slater condition, is motivated by mathematical reasons permitting to apply the theorem of Karush-Kuhn-Tucker (KKT). At the same time it is one of the easier regularity conditions to be verified by economic arguments.

By the fourth condition we impose a reasonable economic behavior of the producers.

As a first step to approximate the behavior of a real economy, it makes sense to require that all consumers are simultaneously 'optimal', e.g. are in a state of maximal utility. Such a multiobjective maximization can be explicited by assigning each consumer a weight $\alpha_i \geq 0$ and maximizing the sum of the weighted consumer utility calling it *welfare maximum.*[1]

Definition 1.2 *(Welfare maximum, [82]) Given a normalized weight vector* $\alpha_i \geq 0$, $\sum_{i \in I} \alpha_i = 1$, *an allocation* (x^*, y^*) *is called a welfare maximum if it solves*

$$\left.\begin{array}{llll} \max \sum_{i\in I} \alpha_i U_i(x_i) & & & \\ s.t. \sum_{i\in I} x_i \leq \sum_{i\in I} x_i^0 + \sum_{j\in J} y_j & & \textit{(no excess of demand)} & (1.1) \\ F_j(y_j) \geq 0 \ \forall j \in J & & \textit{(condition of production)} & (1.2) \end{array}\right\} \quad (1.3)$$

Note that the consumers are not restricted by an individual monetary budget; only the overall demand is restricted by the total supply (1.1) and the feasibility of production (1.2). Giving weights to the consumers implies a distribution, that is, the share of the overall wealth given to a consumer is implicitly determined

[1] In the economic literature the notion of (social) welfare is more generally defined as a monotone function $W(U(x)) : \mathbb{R}^{m\cdot|I|} \to \mathbb{R}$, that is, if $U(x') \geq U(x)$ we have $W(U(x')) \geq W(U(x))$.

by his weight. To overcome this restriction, a more general concept called *Pareto optimality* can be used to characterize the optimality of the state of an economy. The idea is, if a consumer can be made better off and everybody else does not loose, then the former state can be improved, i.e. is not optimal. Reformulated we have the following

Definition 1.3 *(Pareto optimal, [82]) An allocation (x, y) is called Pareto optimal, if it satisfies (1.1) and (1.2), and if there is no allocation (x', y') satisfying also (1.1) and (1.2) and for which $U(x') \geq U(x)$, $U(x') \neq U(x)$.*

Assumption 1.1 implies that the set of welfare optima with $\alpha > 0$ and the set of Pareto optimal points coincide, cf. Theorem 2.4. Both the concept of welfare and Pareto optimality do not consider explicitly prices, or, to state it differently, both the wealth of consumers expressed by their initial endowment x^0 and the scarcity of different goods expressed by pricing are not taken into account. Those lacks are overcome in a state called competitive equilibrium. Here—given a price p—all consumers maximize their utility subject to a budget, and all firms maximize their profit. The key concept lies in the balance of supply and demand, by which the set of possible equilibria is restricted to an extend making even the existence of one equilibrium questionable.

Definition 1.4 *(Competitive economic equilibrium problem, EEP, [82]) A vector (x^*, y^*, p^*) is called a competitive equilibrium if*

(a) y_j^* solves for each firm $j \in J$

$$\left.\begin{array}{ll} \max & p^{*T} y_j \\ \text{s.t.} & F_j(y_j) \geq 0; \end{array}\right\} \tag{1.4}$$

(b) given the budget $M_i := p^{*T} x_i^0 + \sum_j \lambda_{ij} p^{*T} y_j^*$, x_i^* solves for each consumer $i \in I$

$$\left.\begin{array}{ll} \max & U_i(x_i) \\ \text{s.t.} & p^{*T} x_i = M_i; \end{array}\right\} \tag{1.5}$$

(c) $p^* \geq 0$, $e(p^*) = e(x^*, x^0, y^*) \geq 0$ (no excess of demand over supply) and $p_i^* e_i^* = 0$, $i = 1, \dots, m$ (equality of demand and supply for non-free goods).

The complementarity condition in (c) expresses that only scarce goods ($e_i(p^*) = 0$) can have a positive price $p_i^* > 0$.

Welfare and economic equilibria have a close relationship; first note that both are 'economically efficient' in the sense of equal marginal utility relations. Next, on a formal level we can observe a sort of duality: On the one hand welfare maxima always fulfill the (overall) excess constraint (1.1) but not necessarily the (individual) budget constraints in (1.5); on the other hand, given an arbitrary (not necessarily equilibrium) price p the maximization of the firms (1.4) and consumers (1.5) fulfills the budget constraint but not necessarily the excess constraint. The latter is exactly the condition for an equilibrium price. The relation between economic equilibria and welfare will be discussed more in depth in Section 2.2.

1.2 Simplification for Utility Maximizing Agents

The formalism presented in the preceding section can be simplified if every firm belongs exactly to one consumer, i.e. $\forall j \in J \; \exists i \in I : \lambda_{ij} = 1$. Denote by J^i all firms owned by consumer i. Then the next proposition states that the utility maximization for all consumers subject to the production constraints of the corresponding firms is equivalent to the simultaneous profit maximization of the firms and the utility maximization of the consumers. Note that in the following proposition the price p is an exogenously fixed parameter; to have a proper (finite) solution requires positiveness of certain components of p.

Proposition 1.1 *Assume $\lambda_{ij} = 1$ if $j \in J^i$ and zero otherwise, and let the sets J^i be a disjoint covering of J. Furthermore, let the prices for scarce goods be positive, i.e., for all k for which there exists $i \in I$ such that U_i is strictly increasing in the k^{th} component we have $p_k > 0$. Then (x^*, y^*) is a solution of*

$$\left.\begin{array}{rl} \max & U_i(x_i) \\ \text{s.t.} & p^T x_i = p^T x_i^0 + \sum_{j \in J^i} p^T y_j \\ & F_j(y_j) \geq 0 \quad \forall j \in J^i \end{array}\right\} \quad \forall i \in I \tag{1.6}$$

if and only if y_j^* is a solution of (1.4) $\forall j \in J$ and x_i^* is a solution of (1.5) $\forall i \in I$.

The proof uses the characterization of optima by Karush-Kuhn-Tucker, cf. Appendix A. Because $U_i(x_i)$ is strictly increasing in at least one good, we can replace the budget constraint $p^T x_i = p^T x_i^0 + \sum_{j \in J^i} p^T y_j$ in (1.5) and in (1.6) by $p^T x_i - p^T(x_i^0 + \sum_{j \in J^i} y_j) \leq 0$ and know that the corresponding multiplier in the KKT-condition is positive. Choose any $i \in I$; from Assumption 1.1 follows the equivalence of the KKT-points and the optima for each of the three maximization problems (1.4), (1.5) and (1.6). Thus it is sufficient to verify the equivalence of the KKT-conditions of (1.4) and (1.5) with those of (1.6) for any $i \in I$ and all $j \in J^i$, which are, respectively

$$\left.\begin{array}{rl} -p - \mu_j \nabla_{y_j} F_j(y_j) & = 0 \\ \mu_j & \geq 0 \end{array}\right\} \quad \forall j \in J^i \tag{1.7}$$

for (1.4),

$$\left.\begin{array}{rl} -\nabla_{x_i} U_i(x_i) + \mu p & = 0 \\ \mu & > 0 \end{array}\right\} \tag{1.8}$$

for (1.5) and

$$\left.\begin{array}{rl}
-\begin{bmatrix} \nabla_{x_i} U_i(x_i) \\ 0 \\ \vdots \\ 0 \end{bmatrix} - \sum_{j \in J^i} \lambda_j \begin{bmatrix} 0 \\ \vdots \\ \nabla_{y_j} F_j(y_j) \\ \vdots \\ 0 \end{bmatrix} + \lambda \begin{bmatrix} p \\ -p \\ \vdots \\ -p \end{bmatrix} & = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \\
\lambda_j & \geq 0 \;\; \forall j \in J^i \\
\lambda & > 0
\end{array}\right\} \quad (1.9)$$

for the aggregated problem (1.6). Setting $\lambda = \mu$ and $\lambda_j = \mu_j \lambda$ for all $j \in J^i$ yields the desired equivalence of (1.9) with (1.7) and (1.8). ∎

Note that the non-negativity of x does not affect the proven equivalence.

1.3 Formal Link to Markal-Macro Multi-Region (MM^{mr})

MM^{mr} (Markal-Macro multi-region) considered in this work and presented in detail in Chapter 7 has exactly the structure (1.6). The overall equilibrium model integrates a set R of regions connected by trade of CO_2 emission permits and other goods. Dropping for convenience the regional index r (respectively i in (1.6)) the utility of one region is defined as discounted sum of logarithms of consumption

$$U(C) = \sum_{t=1}^{T} d_t \cdot \log C_t$$

over a set T of time periods. The consumption is determined by the gross domestic product (GDP) minus investment costs, which are subsequently specified by a set of constraints. The constraints—represented by $F(y) \geq 0$ in (1.6)—can be divided into two parts: A small nonlinear part responsible for the aggregated macro-economic structure, and a large linear part describing in detail the energy related sector.

1.4 EEP as Variational Inequality Problem (VIP) and Other Formulations

An economic equilibrium problem (EEP) defined in definition 1.4 can be characterized in a more transparent way by using the notion of the excess map:

Definition 1.5 *Let $y_j(p)$ and $x_i(p)$ be the set of solutions to (1.4) and (1.5) respectively; then the excess map $e(p)$ is defined by*

$$e(p) := \sum_{i \in I} \left(x_i^0 - x_i(p)\right) + \sum_{j \in J} y_j(p). \tag{1.10}$$

Note that in general $e(p)$ may be set-valued. Based on $e(p)$ EEP is subsequently formulated as generalized nonlinear complementarity problem (GNCP), generalized variational inequality problem (GVIP), and as generalized fixed point problem (GFPP). Here 'generalized' designates the set-valuedness of $e(p)$; to simplify the notation, we drop the 'G' for generalization in the rest of this exposition and mean by e.g. VIP both GVIP and VIP, where the valuedness of $e(p)$ is given by the context.

As a direct consequence of Definition 1.4 we find an equilibrium to be characterized by the following NCP:

$$\text{find } p^* \in \mathbb{R}_+^m \text{ such that } \exists \xi \in e(p^*) \text{ with } \left.\begin{array}{rcl} \xi & \geq & 0, \\ \xi^T p^* & = & 0. \end{array}\right\} \tag{NCP}$$

This (NCP) can also be formulated as VIP, which is the basement of the VIP-solution approach discussed in Chapter 3:

$$\text{find } p^* \in \mathbb{R}_+^m \text{ such that } \exists \xi \in e(p^*) \text{ with } \xi^T(p - p^*) \geq 0 \;\forall p \in \mathbb{R}_+^m. \tag{VIP}$$

Here (NCP) and (VIP) are tags for the specific problems from above, while NCP and VIP denote the corresponding class of problems. The equivalence of (NCP) and (VIP) is proven in the following proposition.

Proposition 1.2 *(cf. [61]) p^* solves (NCP) if and only if it solves (VIP).*

⌈ Assume p^* solves (NCP), then for the corresponding ξ we have $\xi^T(p - p^*) = \xi^T p \geq 0 \;\forall p \in \mathbb{R}_+^m$, i.e. p^* solves (VIP).

Assume now p^* solves (VIP); obviously we must have $\xi \geq 0$ due to $\xi^T(p-p^*) \geq 0$, because for any $\xi_i < 0$ we could otherwise choose a sufficiently large $p_i \gg 0$ such that $\xi^T(p - p^*) < 0$. Let us check now complementarity; from the last consideration we know $\xi_i p_i^* \geq 0$, by setting $p := p^*$ except for $p_i = 0$, we get $\xi^T(p - p^*) = -\xi_i p_i^* \geq 0$ or equivalently $\xi_i p_i^* \leq 0$, thus $\xi_i p_i^* = 0$ for any i. ⌋

As a third possibility we formulate EEP as a fixed point problem:

$$\text{find } p^* \in \mathbb{R}_+^m \text{ such that } p^* \in P_{\mathbb{R}_+^m} \circ (\mathbf{1} - e)(p^*). \tag{FPP}$$

By $P_{\mathbb{R}_+^m}$ we mean the orthogonal projection onto $\mathbb{R}_+^m$, and by $\mathbf{1}$ the identity map. The structure of this fixed point problem will be discussed more in depth in Section 2.1. Based on the characterization of projections given in Lemma A.7, we have equivalence of (VIP) and (FPP).

Proposition 1.3 *(cf. [23]) p^* solves (VIP) if and only if it solves (FPP).*

The definition of a solution p^* to (VIP) is $\exists \xi \in e(p^*)$ such that $\xi^T(p - p^*) \geq 0 \; \forall p \in \mathbb{R}^m_+$. By multiplying with -1 and adding $p^{*T}(p - p^*)$ on both sides we have the equivalent relation

$$p^{*T}(p - p^*) \geq (p^* - \xi)^T(p - p^*) \qquad \forall p \in \mathbb{R}^m_+.$$

Applying Lemma A.7 yields equivalently $p^* = P_{\mathbb{R}^m_+}(p^* - \xi)$, and with $\xi \in e(p^*)$ we get finally equivalence to $p^* \in P_{\mathbb{R}^m_+} \circ (\mathbb{1} - e)(p^*)$.

In our practical problem MM^{mr}, described in Chapter 7, $e(p)$ is single valued because the underlying problem has the structure (1.6) with a strictly concave utility function. For this and because it simplifies both intuition and proofs, we will restrict ourselves in the following discussion mostly to single-valued operators.

Note that for economic problems with a scalar budget constraint like (1.5) or (1.6) the excess map is homogeneous of degree 0, that is, $e(\lambda p) = e(p) \; \forall \lambda > 0$. This allows to restrict the feasible price set $\mathbb{R}^m_+$ to the unit-simplex Δ which is compact and defined by

$$\Delta := \{p \in \mathbb{R}^m_+ \; : \; \sum_{i=1}^{m} p_i = 1\}, \tag{1.11}$$

where the dimension m of the embedding space is chosen accordingly to the problem.

Finally we should point to the aggregation level in the problem formulations above. Instead of using the very aggregated excess map which hides the structure of the underlying optimization problems, one might as well formulate the $|I| + |J|$ simultaneous maximization problems together with the no-excess condition (c) in Definition 1.4 directly as complementarity problem, and based on that also as VIP and fixed point problem. The basic idea is to catch the maximization problems in their respective Karush-Kuhn-Tucker systems of equations, cf. Garcia and Zangwill [33], and then solve those systems of equations simultaneously.

Whereas from a mathematical and algorithmic point of view it is in general advantageous to work on a disaggregated level where more information is available, the situation might be different in practice like in case of our model, where such a simultaneous formulation as set of equations or complementarity problem is not available.

Als der Türhüter das merkt, lacht er und sagt: "Wenn es dich so lockt, versuche es doch, trotz meines Verbotes hineinzugehen. Merke aber: Ich bin mächtig. Und ich bin nur der unterste Türhüter. Von Saal zu Saal stehen aber Türhüter, einer mächtiger als der andere. Schon den Anblick des dritten kann nicht einmal ich mehr ertragen."
F. K. [58]

Chapter 2

Is There a Solution to EEP?

In 1874 Leon Walras presented in *Éléments d'Économie politique pure ou Théorie de la Richesse sociale* [102] a formalization of general equilibrium theory; there he argued that an economic equilibrium exists by stating that there is an equal number of variables and equations in the underlying set of equations. Indeed, it is easy to construct economic equilibrium problems which, having an equal number of variables and equations, do not possess an equilibrium solution. It took almost 60 years until 1935, when Abraham Wald [101] gave a first mathematically satisfactory answer (for a thorough discussion see John [55]).

Later, in the early fifties, the proof of existence was given in the totally different setting of fixed point theory; this allowed the relaxation of some of the conditions required by Wald, and at the same time made the proofs considerably easier. Since then most strategies for proving the existence of equilibria use finally a fixed point argument. The adverse side of this elegant mathematics is its non-constructive nature, i.e. it can not be used directly to actually find an equilibrium.

Ten years later Lemke [69] resolved this question partially by suggesting an algorithm for solving bimatrix games. A more general approach was developed by Scarf [95] another ten years later, which today is seen as a variant of 'path-following', cf. Zangwill and Garcia [104]. This constructive view in turn admitted new variants to prove the existence of equilibria, and it is exactly such an equation-based approach which will be used in the case of the MM^{mr}-model. The reason why an abstract fixed point theorem can not be directly applied is the lack of structure in the excess function $e(p)$.

The chapter is structured as follows. In Section 2.1 Kakutani's Fixed Point Theorem is discussed and the obstacles in applying it to our excess-based problem are clarified.

Based on the foundations of Chapter 1, Section 2.2 presents Negishi's approach to the problem of proving the existence of a competitive economic equilibrium. From Negishi's theory a general fixed point heuristic called 'conceptual Negishi algorithm' is derived, which will be detailed in Chapter 4.

Section 2.3 discusses a path-following approach following Garcia and Zangwill [33], which will be used in Appendix D to actually prove the existence of an equilibrium of MM^{mr}.

Let us also point to Appendix B, where an up-to-date VIP-based approach following Yao [103] is discussed. This yields on the one hand the Negishi-based proof in Section 2.2 again, and makes on the other hand the proof of existence applicable to a wider range of structural assumptions. Similar to the path-following approach, the VIP under consideration does not rely on the aggregated excess function $e(p)$, but on a direct formulation of the optimality conditions of the underlying maximization problems.

The contributions in this chapter are as follows. In Section 2.2 the original assumptions used by Negishi [82] are relaxed, and based thereupon the existence proof is given. In Section 2.3 the agents have a different structure compared to the discussion in Garcia and Zangwill [33], and furthermore the constraint qualifications are changed in order to make the concepts applicable to our model MM^{mr}.

2.1 Kakutani's Fixed Point Theorem

Consider a convex set $C \subset \mathbb{R}^n$, denote by 2^C the set of subsets of C and consequently by a point-to-set map $f : C \to 2^C$ a map relating to each $x \in C$ a set $f(x) \subset C$. Then we call x^* a *fixed point* of f if $x^* \in f(x^*)$. Furthermore, f is called *convex* if $f(x)$ is convex for all $x \in C$. Besides convexity we need a second property called closedness to assure the existence of a fixed point.

Definition 2.1 *(Closedness, [30]) Let $\{x^k\} \subset C$ be any convergent sequence with limit point $x \in C$, $x^k \to x \in C$, and choose for all $k \in \mathbb{N}$ a $y^k \in f(x^k)$ such that $\{y^k\} \subset C$ is convergent in C, $y^k \to y \in C$. If for all such sequences $\{x^k\}$ and $\{y^k\}$ we have $y \in f(x)$, then f is called closed.*

Furthermore, we also introduce the notion *open*:

Definition 2.2 *(Openness, [30]) Let $\{x^k\} \subset C$ be any convergent sequence with limit point $x \in C$, $x^k \to x$. If for any $y \in f(x)$ we can choose for all $k \in \mathbb{N}$ a $y^k \in f(x^k)$ such that $\{y^k\} \subset C$ converges towards $y \in C$, then f is called open.*

A set-valued map which is both open and closed is called *continuous*. Interpreting these properties, closedness inhibits a sudden 'contraction' of the sets $f(x)$ if x varies slightly, whereas openness inhibits a sudden expansion.

A proof of the following fixed point theorem due to Kakutani (1941) can be found e.g. in Heuser [49, p. 614], or in Garcia and Zangwill [33]. Its relevance for mathematical economics can not be overestimated, and one of its applications is presented in the next section.

Theorem 2.1 *(Kakutani's Fixed Point Theorem) Let $C \in \mathbb{R}^n$ be a non-empty, compact and convex set. If the map $f : C \to 2^C$ is nonempty, closed and convex, then f has a fixed point in C.*

To clarify what is essential for the existence of a fixed point, let us give the following simpler but nevertheless equivalent

Theorem 2.2 *(Brouwer's Fixed Point Theorem)*[1] *Assume $C \subset \mathbb{R}^n$ is a nonempty, convex and compact set. Then every continuous point-to-point map $f : C \to C$ has at least one fixed point.*

Hence it is continuity of f together with convexity and compactness of C which guarantees the existence of a fixed point. Let us briefly outline the connection of Kakutani's Fixed Point Theorem to the problem of existence of an equilibrium. As stated in Proposition 1.3, the economic equilibrium problem EEP can be formulated equivalently as fixed point problem using the map $f(p) := P_\Delta \circ (\mathbb{1} - e)(p)$, where P_Δ denotes the orthogonal projection map onto the unit-simplex Δ, $\circ$ abbreviates the concatenation of maps, and $(\mathbb{1} - e)(p) := p - e(p)$. If $e(p)$ is single-valued we have equivalence of continuity of the fixed point map $f(p)$ and $e(p)$ due to the continuity of the projection P_Δ. In view of (FPP), Proposition 2.2 together with continuity of the excess map implies the existence of an equilibrium. In case of a set-valued excess map we must have both convex values and closedness, cf. Theorem 2.1. However, a projection of a convex set onto another convex set does in general not produce a convex set, and hence the fixed point map $f(p)$ may be non-convex even if $e(p)$ is convex for all $p \in \Delta$. Furthermore, even if convexity could be assured closedness is a demanding property.

To enlighten the difficulties with closedness, following Flippo [30], let us look at a single agent represented by a mathematical programming problem $\max U(x)$ subject to $G(x) \leq b$, where x, $G(x)$ and b are vectors of appropriate dimension, and where p is part of b. Denote by $\phi(b)$ the feasible set map and by $\omega(b)$ the corresponding optimal set map. Our excess map can be understood as (part of) $\omega(b)$, and hence closedness of $e(p)$ follows from closedness of $\omega(b)$. The latter is essentially given if both the objective $U(x)$ and $\phi(b)$ are continuous [30, Theorem 2.1 and Corollary 2.1]. As mentioned above, continuity of the set-valued map $\phi(b)$ requires closedness and openness. While closedness of $\phi(b)$ is basically implied by a continuous $G(x)$ [30, Theorem 2.3], openness requires a constraint qualification like the one of Mangasarian-Fromovitz [30, Theorem 3.4].[2] As discussed in [30], there is little hope to weaken those requirements, because they

[1] In Brouwer's original formulation C is the closed unit ball in $\mathbb{R}^n$.

[2] Flippo [30] discusses only the case where solely the right hand side may vary. In order to translate the consumers problem containing a budget constraint into such a framework, both x and p must be treated as variable. In a second step the 'variable' p is then equalized to components of the right hand side b. Thereby the convexity of the budget constraint is destroyed, and consequently [30, Theorem 3.1–3.3] can not be applied.

are too strongly connected to a 'well-behaved' optimization problem. This may explain why proving existence of an equilibrium is usually not done by examining the overall excess, but by studying some underlying mathematical structures.

In the next section, Negishi's successful application of Kakutani's Fixed Point Theorem, based on an exploration of the structure of the agents, is demonstrated.

2.2 Negishi's Approach

This section is based on Negishi's proof for the existence of an equilibrium [82] and requires the definitions and assumptions of Section 1.1. Compared to Negishi we weaken some assumptions and modify the proofs accordingly. We start by a direct consequence of Assumption 1.1 which guarantees the existence of a finite solution to both the consumer and producer problem. Moreover, we can conclude that the set of solutions must be convex.

Lemma 2.3 *([82], Lemma 1) If Assumption 1.1 is fulfilled and if there is no excess of demand over supply, i.e.* $e \geq 0$*, then the domain of* x *and* y *is nonempty, convex and compact.*

Applying this prerequisite we have under Assumption 1.1 almost equivalence of Pareto optimal states and welfare maxima:

Theorem 2.4 *([82], Theorem 2) If Assumption 1.1 holds, then for any weighting vector* $\alpha^* \geq 0$ *there is a welfare maximum represented by the utility vector* U^**. Furthermore, for* $\alpha > 0$ *an allocation is a welfare maximum if and only if it is Pareto optimal.*

Note that if either the feasibility sets are non-convex or U is non-concave, the equivalence of welfare maxima and Pareto optimality is violated.

The following theorem relates equilibria to a subset of welfare maxima characterized by the weighting vector α. Based on this relation a fixed point map is set up and Kakutani's theorem can finally be applied to prove the existence of an equilibrium.

Theorem 2.5 *([82], Theorem 4) Let* p *correspond to the Lagrange-multiplier of (1.1) in the welfare problem. Then*

1. *at any welfare maximum the conditions (a) and (c) of a competitive equilibrium (Definition 1.4) are fulfilled;*
2. *condition (b) in Definition 1.4 is satisfied if and only if* $0 < \alpha_i = 1/\delta_i$*, where* δ_i *is the marginal utility of income of consumers, i.e.* δ_i *is the Lagrange-multiplier of the budget constraint in the utility maximization problem (1.5) of consumer* i*.*

Both claims can be shown by comparing the KKT-conditions of the underlying optimization problems.

The equation $\alpha_i = 1/\delta_i$ is only given in this canonical setting, where the equilibrium price equals the dual multiplier p of (1.1). Obviously the welfare maximization problem produces the same *primal* result for any scaling $\lambda\alpha$ with $\lambda > 0$, and also the utility maximization problem is *primal* invariant with respect to a scaling λp for $\lambda > 0$. This is relevant for restricting the feasible set of α when applying Kakutani's theorem below, but also in the practical implementation where the relation $\alpha_i = 1/\delta_i$ is satisfied only up to scaling.

In order to prove the existence of an equilibrium we have to sharpen Assumption 1.1 slightly. This is necessary for two reasons. First we need a *convex* valued and *closed* fixed point map to apply Theorem 2.1, which is achieved by assuming concavity of F. Secondly we want each consumer to have a positive amount of 'money' to spend even if some (but not all) prices are zero. This is needed in the proof of Theorem 2.6 to derive an equilibrium from a fixed point. Several conditions can be considered implying a positive monetary endowment. Negishi [82] presumes strictly increasing utilities (i.e. $(\nabla U_i(x_i))_k > 0$ $\forall x_i \geq 0$, $\forall i \in I$ and $\forall k = 1, \ldots, m$) together with $x_i^0 \geq 0$, $x_i^0 \neq 0$ and $F(0) \geq 0$, resulting in a positive monetary endowment. But these assumptions are rather strong, because they imply $p > 0$ in any welfare solution (where $\alpha \in \Delta$), which collapses the complementarity and variational inequality problem to finding a zero of the excess map $e(p)$. Here we impose the following relaxed assumptions:

Assumption 2.1

1. *$F_j(y_j)$ is concave for all $j \in J$.*
2. *There is a good k for which all utilities are strictly increasing, and for which all consumers have a positive endowment.*

The second condition is easily fulfilled if a good is introduced which represents a monetary numéraire and if each consumer is endowed with a (small) positive amount.

Under Assumptions 1.1 and 2.1 all welfare maxima are equivalently saddle-points of the Lagrange dual function

$$L(x, y, p, \mu; \alpha) := \alpha^T U(x) + p^T e(x, y) + \mu^T F(y). \tag{2.1}$$

The saddlepoint map ψ is defined as

$$\psi(\alpha) := \arg \min_{(p,\mu)\geq 0} \max_{x \geq 0, y} L(x, y, p, \mu; \alpha) \tag{2.2}$$

and is obviously a non-empty, *convex*-valued point-to-set map: given $(p, \mu) \geq 0$, maximization with respect to (x, y) appears in concave summands only, and given (x, y), minimization with respect to (p, μ) appears in linear, thus convex, summands only.

Theorem 2.6 *(cf. [82], Theorem 5) Under Assumptions 1.1 and 2.1 there exists a competitive equilibrium.*

The proof is based on Kakutani's fixed point theorem; the underlying fixed point map $\phi : (\,\cdot\,^k) \to \{(\,\cdot\,^{k+1})\}$ is constructed by a concatenation of three maps, where the brackets $\{(\,\cdot\,)\}$ indicate set-valuedness:

$$\begin{array}{l}
(\alpha^k, x^k, y^k, p^k) \\
\quad\Big\downarrow\ \overline{\psi}(\alpha^k)\text{: (extended) saddlepoint map} \\
(\alpha^k, \{(x^{k+1}, y^{k+1}, p^{k+1})\}) \\
\quad\Big\downarrow\ \nu\text{: normalization of } p \\
(\alpha^k, \{(x^{k+1}, y^{k+1}, p'^{k+1})\}) \\
\quad\Big\downarrow\ \upsilon\text{: upgrade of } \alpha^k \\
\{(\alpha^{k+1}, x^{k+1}, y^{k+1}, p'^{k+1})\}
\end{array}$$

The extended saddle point map $\overline{\psi}$ differs from ψ in that it contains α but drops μ. The normalization map $\nu(p)$ is defined by $p \mapsto p/\sum_i p_i$, and the upgrade map υ by

$$\alpha_i \;\mapsto\; \frac{\max(0, \alpha_i + p'^T(x_i^0 + \sum_{j\in J} \lambda_{ij} y_j \;-\; x_i))}{\sum_i \max(0, \alpha_i + p'^T(x_i^0 + \sum_{j\in J} \lambda_{ij} y_j \;-\; x_i))}.$$

In the course of the proof it is shown that the denominator in the map υ is positive, and also ν is well defined. Hence for all $k \in \mathbb{N}$ both α^k and p^k are in the unit-simplex of appropriate dimension.

In order to apply Kakutani's fixed point theorem we first have to verify that ϕ is closed. We do this by exploiting that continuous (point-to-point) maps preserve closedness, and the Cartesian product of two closed maps is closed (this is always on the background of $\mathbb{R}^n$-topology, where all metrics are equivalent).

To begin with let us prove closedness of the saddle-point map ψ; under Assumptions 1.1 and 2.1 we find that for any given α in the unit-simplex the set of welfare maxima $\{(x, y, p, \mu)\}$ equals the set of saddle-points in (2.2). Assume now a sequence $\{\alpha^l\} \subset \Delta$ converging to some $\alpha^* \in \Delta$, and choose for each $\alpha^l, l \in \mathbb{N}$, from the set of saddle-points a (x^l, y^l, p^l, μ^l) converging to some (x^*, y^*, p^*, μ^*). We have to show that (x^*, y^*, p^*, μ^*) is in the set of saddle-points of α^*. Looking at the following characterization of saddle-points

$$L(x^l, y^l, p, \mu; \alpha^l) \;\geq\; L(x^l, y^l, p^l, \mu^l; \alpha^l) \;\geq\; L(x, y, p^l, \mu^l; \alpha^l)$$

which holds for all $l \in \mathbb{N}$ and all feasible (x, y, p, μ), we derive from continuity of L in x ,y, p, μ and α

$$L(x^*, y^*, p, \mu; \alpha^*) \;\geq\; L(x^*, y^*, p^*, \mu^* \alpha^*) \;\geq\; L(x, y, p^*, \mu^* \alpha^*),$$

hence (x^*, y^*, p^*, μ^*) is in the set of saddle-points of α^* implying closedness of $\psi(\alpha)$. To accept closedness of the extended saddle-point map $\overline{\psi}$, note that it can be built by a projection $(\alpha^k, x^k, y^k, p^k) \mapsto \alpha^k$, followed by $\psi : \alpha^k \mapsto \{(x^{k+1}, y^{k+1}, p^{k+1}, \mu^{k+1})\}$, followed by a projection $\{(x^{k+1}, y^{k+1}, p^{k+1}, \mu^{k+1})\} \mapsto \{(x^{k+1}, y^{k+1}, p^{k+1})\}$ and finished by building the Cartesian product with α^k, that is, $(\alpha^k, \{(x^{k+1}, y^{k+1}, p^{k+1})\})$.

Next, from $\alpha^k \in \Delta$ and the second item in Assumption 2.1 we observe $p^{k+1} \geq 0$ and $p^{k+1} \neq 0$, implying continuity of ν.

Before showing continuity of v in the last step, we first verify that it is well-defined, i.e. $\sum_i \max(0, \alpha_i + (M_i - p'^T x_i)) > 0$ for any $\alpha \in \Delta$ and its resulting (x, y). Assume contrarily $\sum_i \max(0, \alpha_i + (M_i - p'^T x_i)) \leq 0$; this implies $M_i - p'^T x_i \leq 0$ for all i, and because $\alpha \in \Delta$ there is an i with $\alpha_i > 0$ forcing $M_i - p'^T x_i \leq -\alpha_i < 0$. But then $\sum_i (M_i - p'^T x_i) < 0$ contradicting $\sum_i (M_i - p'^T x_i) = p'^T e = 0$. Thus, $\sum_i \max(0, \alpha_i + (M_i - p'^T x_i)) > 0$ and the map is well defined.

Based on that the upgrade mapping v is obviously continuous leading to the verification of closedness of ϕ. Furthermore, non-emptiness and convex values of ϕ can be easily verified using the fact that ψ has these properties. The final condition to be verified is that ϕ has a compact and convex set of definition, which is a consequence of Lemma 2.3 and the scaling onto the unit-simplex of α and p. By Kakutani's Theorem we thus have the existence of a fixed point.

The verification that any fixed point is a competitive equilibrium is a consequence of the positive monetary endowment $M_i > 0$ for each consumer, induced by Assumption 2.1. Assume in a fixed point α there is an $i \in I$ where $\alpha_i = 0$. From the definition of the fixed point map we have then $M_i - p^T x_i \leq 0$ which is only possible if $p^T x_i > 0$. But this contradicts the assumption of a welfare maximum being maximal, because $p^T x_i > 0$ implies that x_i has positive components for goods with positive dual multiplier in the welfare problem. This implies in turn that by setting $x_i = 0$ we can strictly increase the overall welfare which, in view of $\alpha_i = 0$, contradicts optimality of the welfare solution. Thus $\alpha > 0$ and by Theorem 2.5 the only point left to prove is that the budget constraint holds for all $i \in I$, which, for $\alpha > 0$, follows directly from the fixed point property. ∎

Extending Theorem 2.6, Ginsburgh and Waelbroeck [36] show that for *every* competitive equilibrium there is an $\alpha > 0$, such that the solution $(x(\alpha), y(\alpha), p(\alpha))$ of the corresponding welfare problem equals the equilibrium solution. This proves that under the Assumptions 1.1 and 2.1 the set of equilibria is contained in the set of welfare states.

It is interesting to see that this set-relation can be reversed in a restricted sense: given $\alpha > 0$ with a welfare solution $(x(\alpha), y(\alpha), p(\alpha))$ fulfilling $x_i(\alpha) - \sum_{j \in J} \lambda_{ij} y_j(\alpha) \geq 0$, we endow each consumer with

$$x_i^{0e} := x_i(\alpha) - \sum_{j \in J} \lambda_{ij} y_j(\alpha) \geq 0;$$

then the thereby defined economic equilibrium problem has the same solution as the original welfare problem. To see this note that $(x(\alpha), y(\alpha))$ is feasible for the optimization problems in Definition 1.4(a) and (b), and that at $(x(\alpha), y(\alpha))$ the KKT-conditions for both the consumer and producer problems are also fulfilled.

In this restricted sense equilibria and welfare are equivalent, but it is the point of view which sets them apart: while equilibrium problems allow a direct control of the endowment and thereby of the resulting distribution, the welfare approach allows only to choose α which determines the distribution implicitly. This subtle difference forms the fundamental distinction between centrally planned economies and ('free') market economies. Superfluous to say that this distributional effect is also of central importance when a community of countries agrees on trading CO_2 permits. The price for the different modeling philosophy is the change from a (convex) optimization problem in case of welfare problems to equilibrium problems, formulated e.g. as variational inequality problem, complementarity problem or fixed point problem.

Finally, note that the view of a fixed point problem leads to solving approaches like path following, cf. Garcia and Zangwill [33], which can be computationally very demanding. In practice it turns out, however, that some heuristic methods, based on the following general concept, are in our case very fast.

Algorithm 1 Conceptual Negishi Algorithm

(i) Choose a set of initial weights α^0 and set $k = 0$.

(ii) Solve the Negishi welfare problem and compute thereby the regional excess e_r^k and the dual price p^k of the excess constraint.

(iii) Stop if all regional budget constraints ${p^k}^T e_r^k$ are (close to) zero. Otherwise set $k := k + 1$, update the weight vector α^k and return to (ii).

The crucial part in this concept is the update step $\alpha^k \to \alpha^{k+1}$; two strategies are discussed in Chapter 4 and numerically compared in Appendix F.5.

2.3 The Path Following Concept

This section is based on Garcia and Zangwill [33]. The concept and the relationship to our problem MM^{mr} is given in some detail, because the proof of existence of a solution to MM^{mr} is based on this idea. The reason why we can not simply verify the assumptions underlying the Negishi proof of existence has to do with the more complex structure involved in the consumers of MM^{mr}.

At its heart the path-following (homotopy) concept solves a system of (nonlinear) equations by following a path from a known solution of the somehow twisted problem to a real solution of the original problem.

To bridge the gap to our EEP we proceed in several steps; first the *economic* equilibrium problem EEP is transferred into an equivalent Nash equilibrium problem (EP) by introducing an artificial price-agent. EP consists of $|I|+|J|+1$ optimization problems which can be formulated equivalently as systems of equations by using the theory of Karush-Kuhn-Tucker, given some regularity conditions hold. A simultaneous solution to the $|I|+|J|+1$ systems of equations is an equilibrium solution of the original EEP and can, in principle, be found by applying a path-following approach. Taking the last step first the basic idea in the path-following concept can be described as follows:

1. Extend the original system of equations by introducing a scalar t, such that for $t = 0$ we have a unique known solution v^0, while at $t = 1$ it coincides with the original system.

2. Follow the path $v(t)$ of the system's solution from $t = 0$ to $t = 1$, thereby solving the original problem.

As a simple example consider the linear system $Av = b$ with a non-singular $A \in \mathbb{R}^{n \times n}$. Take any $v^0 \in \mathbb{R}^n$ and choose d according to $Av^0 = b + d =: \bar{b}$. The extended system is $Av = \bar{b} - td$ where, at $t = 0$, we have a unique solution $v^0 = v(0)$ and, at $t = 1$, the solution $v(t)$ solves the original problem.

In general, given some assumptions, the path is differentiable and goes from one starting point v^0 exactly to one endpoint $v(1)$. It does not turn back to a solution $v(0)$, bifurcates to multiple solutions or diverges to infinity.

Coming back to the question how EEP can be transformed into a system of equations, we first catch the feasibility and complementarity condition (c) in the Definition 1.4 of EEP in an artificial price agent, defined by the following problem; as usual Δ denotes the unit simplex from (1.11).

Definition 2.3 *(Price agent, [33, (6.2.2)]) Given the overall excess e (supply minus demand), the price agent chooses a price p solving*

$$\min_{p \in \Delta} e^T p. \tag{2.3}$$

The price agent can be interpreted as the 'invisible hand' postulated by Smith. To see that (2.3) can indeed replace condition (c) in Definition 1.4 the following (Nash) equilibrium problem is studied.

Definition 2.4 *(Equilibrium problem, EP, [33]) (x^*, y^*, p^*) is called an equilibrium solution of EP if simultaneously x^* solves problem (1.5), y^* solves problem (1.4), and p^* solves problem (2.3).*

The following lemma uses ideas of Garcia and Zangwill [33, p. 118 f].

Lemma 2.7 *(x^*, y^*, p^*) is an economic equilibrium in the sense of Definition 1.4 if and only if it solves the related EP in Definition 2.4.*

Let (x^*, y^*, p^*) be a solution to EEP; then, by condition (c) in Definition 1.4, we know $e \geq 0$ and $p^{*T}e = 0$, $p^* \in \Delta$. Thus p^* solves problem (2.3) and consequently (x^*, y^*, p^*) is a solution to EP.

To prove the reverse implication assume (x^*, y^*, p^*) is a solution to EP. To show that it also solves EEP, the non-negativity of e and complementarity with p^* must be deduced. Suppose firstly that $e \not\geq 0$, then $\exists j : e_j < 0$. Because p^* solves problem (2.3) and the j-th unit vector is in Δ, we must have

$$p^{*T}e \leq e_j < 0. \tag{2.4}$$

On the other hand the budget constraint for each consumer requires

$$0 = M_i - p^{*T}x_i^* = p^{*T}\left[(x_i^0 + \sum_{j \in J} \lambda_{ij} y_j^*) - x_i^*\right],$$

which, summed over all consumers, yields

$$0 = p^{*T}\left[\sum_{i \in I}(x_i^0 - x_i^*) + \sum_{j \in J} \underbrace{\sum_{i \in I} \lambda_{ij}}_{=1} y_j^*\right] = p^{*T}e. \tag{2.5}$$

But this contradicts (2.4), and hence $e \geq 0$ is proven. Complementarity of p^* and e is an immediate consequence of $e \geq 0$, $p^* \geq 0$ and (2.5).

Note that there is almost no structure required for this lemma to hold; it simply suffices to have agents (consumers and producers) which—given a price signal from the price agent—reveal their optimal choice (consumption or production) and that consumers obey their budget constraint. Specifically, problem (1.5) representing consumers may contain arbitrary additional constraints. It is exactly this property of EP which makes it applicable to MM^{nr}.

Formally, EP consists of a set A (economic *a*gents) of convex maximization problems which are connected by common variables; in view of the problems (1.5), (1.4) and (2.3) we can think of $|A| = |I| + |J| + 1$. We are looking then for a simultaneous solution to these maximization problems and accomplish this by transforming all problems into their corresponding KKT-systems of equations.[3] Given some constraint qualifications are fulfilled, we have thereby reformulated EEP into a nonlinear equation problem (NEP).

The notation $v = (v_a, v_{\bar{a}})$ used in the sequel symbolizes a decomposition of the overall variable v into a part v_a which is in the realm of agent a, and the rest $v_{\bar{a}}$

[3]Usually the KKT conditions are formulated with equations and inequalities; the inequalities appear, because the sign of the Lagrange multipliers for the inequalities in the problem must be fixed. A simple transformation not only eliminates these inequalities, but makes the resulting equations even differentiable, cf. Garcia and Zangwill [33, p. 66].

not under control of a. Considering the set of problems (1.4), (1.5) and (2.3), v can be interpreted as (x, y, p), where $v_a = x_i$ if a represents consumer i, $v_a = y_j$ if a represents producer j, and $v_a = p$ if a is the price agent.

For the ease and generality of exposition we consider for all $a \in A$ the following maximization problem encompassing the aforementioned problems.

$$\left.\begin{aligned} &\max_{v_a} f_a(v_a, v_{\bar{a}}) && (2.6)\\ &\text{s.t. } g_a(v_a, v_{\bar{a}}) \leq 0, && (2.7)\\ &\qquad h_a(v_a, v_{\bar{a}}) = 0. && (2.8) \end{aligned}\right\} \quad (2.9)$$

For a specific $a \in A$ problem (2.9) can be seen as a parametric optimization problem.

To apply both the KKT-conditions and the path-following concept we assume for the rest of this section the following.

Assumption 2.2 *For all $a \in A$ the following holds:*

(a) f_a is concave and three times continuously differentiable in v_a;

(b) g_a is quasi-convex and three times continuously differentiable in v_a;

(c) h_a is affine in v_a;

(d) there exists a v^0 such that $g_a(v^0) < 0$ and $h_a(v^0) = 0$.

The affinity claimed for h in (c) might be relaxed while preserving the characterization of optimal points by the KKT-conditions, cf. Theorem A.1 and A.2 in Appendix A; in that case, however, linear independence of the vectors $\{\nabla_{v_a} h_a \mid a \in A\}$ for all feasible v is needed. An important aspect in Assumption 2.2 is that they apply only to v_a, the part of the variable under control of agent a. For example, the consumer budget constraint is linear in $v_a = x_i$, even though nonlinear in the overall variable which comprises y and p; similarly the objective of the price-agent is linear in p, because the excess e is determined exogenously by the consumers and producers.

To apply the path-following approach for solving the KKT-system corresponding to problem (2.9), we first extend the objective in problem (2.9).

$$\left.\begin{aligned} &\max_{v_a} f_a^t(v_a, v_{\bar{a}}) := t \cdot f_a(v_a, v_{\bar{a}}) - (1-t)\frac{1}{2}\|v - v^0\|^2\\ &\text{s.t. } g_a(v_a, v_{\bar{a}}) \leq 0\\ &\qquad h_a(v_a, v_{\bar{a}}) = 0 \end{aligned}\right\} \quad (2.10)$$

Note that v^0 is the unique solution for (2.10) at $t = 0$, and that due to the strict concavity of $f_a^t(v_a, v_{\bar{a}})$ in v_a for all $t \in [0, 1)$ the solution is unique for $t \in [0, 1)$. Based on (2.10) the set of all KKT-equations for all agents in A defines therefore a homotopy $H(t)$ onto which the path-following concept can be applied.

Before stating the main theorem, the notion 'regular' must be explained. In order to have a differentiable path by applying the Implicit Function Theorem, the Hessian of H with respect to the variable v and the Lagrange multipliers must be regular, that is, of full rank. This is in principle required for all $t \in [0, 1]$. By using a perturbation technique due to Sard (cited in Garcia and Zangwill [33]), however, this requirement can be overcome.

Theorem 2.8 *([33], Theorem 4.5.2) Let $D = \{v \mid g_a(v) \leq 0$ and $h_a(v) = 0\ \forall a \in A\}$ be the feasible set; if D is compact, Assumption 2.2 holds, and if $H(t)$ is regular, then the path starting at $t = 0$ reaches for $t = 1$ a simultaneous solution to (2.9) for all $a \in A$.*

In view of the later application to MM^{mr} a comment on Condition (d) in Assumption 2.2 is in order. On the one hand this condition allows to characterize all optima by KKT-points (cf. Appendix A). On the other hand Condition (d) makes $H^{-1}(0)$ a singleton, that is, v^0 is the unique primal solution, and at the same time the dual solution is also unique. This uniqueness-requirement is crucial in order to prevent the path from turning back to $t = 0$. In case of MM^{mr} both requirements are met by proving linear independence of $\{\nabla_{v_a} g_a\}$ on a subset of variables of v_a which are not used in the affine part h_a. The argument is straightforward: the affine part can always be made regular by eliminating linear dependent equations, and together with linear independence of $\{\nabla_{v_a} g_a\}$ on another subset of variables, the KKT-equations yield a unique solution for the dual multipliers.

Chapter 3

> Solche Schwierigkeiten hat der Mann vom Lande nicht erwartet; das Gesetz soll doch jedem und immer zugänglich sein, denkt er, aber als er jetzt den Türhüter in seinem Pelzmantel genauer ansieht, seine grosse Spitznase, den langen, dünnen, schwarzen tatarischen Bart, entschliesst er sich, doch lieber zu warten, bis er die Erlaubnis zum Eintritt bekommt. Der Türhüter gibt ihm einen Schemel und lässt ihn seitwärts von der Tür sich niedersetzen. Dort sitzt er Tage und Jahre. Er macht viele Versuche, eingelassen zu werden, und ermüdet den Türhüter durch seine Bitten. F. K. [58]

Solving EEP Using the VIP-Approach

This chapter discusses some mathematical background to VIPs which is relevant for solving the excess-based (VIP) posed by the primal integration, cf. Section 1.4. As part of this discussion two algorithms are presented. For a good general introduction to VIPs see e.g. Kinderlehrer and Stampacchia [65]; a summary is presented in Appendix A.4. Another excellent survey focusing equilibrium problems has been given by Harker and Pang [47]. The chapter is organized as follows.

The crucial role of the different monotonicity properties is investigated in Section 3.1 and is brought into relation with the first algorithmic concept, the cutting plane method (CPM).

Section 3.2 seeks to present the subtle structural differences between pseudo-monotonicity and monotonicity causing fundamental differences in the complexity of the algorithms deduced. While a direct application of CPM to pseudo-monotone or monotone problems does not yield a polynomial complexity, Nesterov and Vial [87] found a homogenized reformulation which is pseudo-polynomial for monotone problems. The resulting homogenized CPM for monotone VIPs is presented and discussed.

The first two sections highlight the necessity of (pseudo-)monotonicity for applying a cutting plane method. On that background Section 3.3 motivates (pseudo-) monotonicity of the excess map from an economic point of view; even though no (pseudo-)monotonicity-proof can be expected, there is quite some evidence to observe an 'almost' pseudo-monotone excess if the number and heterogeneity of agents is sufficiently large.

Contributions in Section 3.1 include the necessity-discussion of pseudo-monotonicity in the course of CPM (cf. Lemma 3.3), the equivalence of local and global (pseudo-)monotonicity (cf. Lemma 3.4) permitting the use of sensitivity analysis, or the concept of ε-pseudo-monotonicity. In Section 3.2 original work lies in the examples, and finally Section 3.3 brings in the specificities of MM^{mr}.

3.1 Monotonicity Reconsidered

The basic Variational Inequality Problem can be stated as follows.

Definition 3.1 *(Variational Inequality Problem, VIP(f, D), [65, problem 4.1]) Let $D \subset \mathbb{R}^n$ be convex and $f : D \subset \mathbb{R}^n \to \mathbb{R}^n$; find $x \in D$, such that*

$$f(x)^T(y - x) \geq 0 \qquad \forall y \in D. \tag{3.1}$$

For such a point-to-point operator f the following properties are fundamental.

Definition 3.2 *(cf. [91]) f is called*
monotone *over D, if $(f(x) - f(y))^T(x - y) \geq 0 \ \forall x, y \in D$, and*
pseudo-monotone *over D, if $[\, f(y)^T(x - y) \geq 0 \Rightarrow f(x)^T(x - y) \geq 0\,] \ \forall x, y \in D$.*

To support intuition, note that under suitable assumptions the notion of monotonicity of f is equivalent with convexity of some F where $f \equiv \nabla F$.

Proposition 3.1 *Let D be convex, $U \supset D$ be open, F be once continuously differentiable on U and set $f := \nabla F$; then*

1. *([79]) f is monotone on D if and only if F is convex on D;*
2. *([62]) f is pseudo-monotone on D if and only if F is pseudo-convex on D.*

Based on a suitable definition of subdifferentials Aussel, Corvellec and Lassonde [5] and Aussel [4] extend the above two equivalences onto semi-continuous functions F.

The set of solutions of VIP(f, D), abbreviated by $(f, D)^*$, is intimately related to a second set of 'solutions'

$$(f, D)^{**} \ := \ \{x \mid f(y)^T(y - x) \geq 0 \ \forall y \in D\}. \tag{3.2}$$

If we introduce the cut set

$$C_y := \{x \in D \mid f(y)^T(y - x) \geq 0\}$$

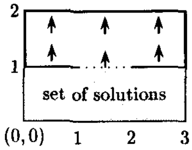
(3.2) can be written in the form $(f, D)^{**} = \cap_{y \in D} C_y$. A possible interpretation of $(f, D)^{**}$ is based on Proposition A.11, where $(f, D)^{**}$ appears as (part of the) set of minima of the minimization problem $\min_{x \in D} F(x)$, and where $f = \nabla F$ for a continuously differentiable function F. Because $(f, D)^{**}$ is built by the intersection of (possibly infinitely many) linear cuts, it is always convex and thereby forbids isolated solutions. The relation between $(f, D)^*$ and $(f, D)^{**}$ is stated in the following proposition.

Proposition 3.2 *([65, 57]) Let D be convex, then the following relations hold, where '$\boxtimes$' allows any set relation:*

	arbitrary f	*pseudo-monotone f*
arbitrary f	$(f,D)^{**} \boxtimes (f,D)^{*}$	$(f,D)^{**} \supset (f,D)^{*}$
continuous f	$(f,D)^{**} \subset (f,D)^{*}$	$(f,D)^{**} = (f,D)^{*}$

Note how $(f,D)^{**}$ approximates $(f,D)^{*}$ from inside or outside, depending on the properties given. Consequently $(f,D)^{**}$ is called set of *weak* solutions if the focus lies on pseudo-monotonicity; in this case $(f,D)^{*}$ is called set of *strong* solutions. If, however, continuity is put forward, a Lyapunov function can be given which makes all points in $(f,D)^{**}$ stable solutions of the corresponding dynamical system, motivating the name set of *stable* solutions for $(f,D)^{**}$.

To illustrate that the set relations in Proposition 3.2 are strict, a first non-continuous but pseudo-monotone example is depicted in Figure 3.1; here $(f,D)^{*}$ is non-convex and thus $(f,D)^{*} \subsetneqq (f,D)^{**}$.



$$f(x) = \begin{cases} (0, 0.3) & \text{if } (x_2 > 1) \vee (x_2 = 1 \wedge (x_1 < 1 \vee x_1 > 2)), \\ (0,0) & \text{else.} \end{cases}$$

Figure 3.1: A pseudo-monotone and non-continuous map with a non-convex solution set $(f,D)^{*}$.

An even simpler example shows $(f,D)^{**} \subsetneqq (f,D)^{*}$ for a continuous mapping; choose $D = [-1,1]$ and set $f(x) := -x$. Then $(f,D)^{**} = \emptyset \subsetneqq \{-1,0,1\} = (f,D)^{*}$.

In view of Proposition 3.2, an algorithm for VIPs with pseudo-monotone operators can be set up exploring the fact that $(f,D)^{*} \subset C_y$ for all $y \in D$. Thus, an obvious scheme for finding a point in (or close to) $(f,D)^{**}$ follows Algorithm 2.

Algorithm 2 Cutting plane method (CPM) for pseudo-monotone VIPs.

(i) Set k=0, choose an inner point $x^0 \in D$ and set $D^0 := D$.

(ii) Stop if x^k satisfies a stopping criterion, otherwise proceed.

(iii) Reduce the feasibility set: $D^{k+1} := D^k \cap C_{x^k}$.

(iv) Choose an inner point $x^{k+1} \in D^{k+1}$, set $k \leftarrow k+1$ and go back to (ii).

We will use the notion 'inner point' in the steps (i) and (iv) synonymical with 'center'. There are basically two classes of centers. On the one hand geometrically defined centers (like the prominent center of gravity, or less used ones like the center of the largest inscribed (smallest circumscribed) ellipsoid) which are independent of the analytic description of the set D^k. On the other hand so called analytic centers which do depend on the analytic representation of D^k (see Kaiser [59]), and for which 'the' analytic center became very popular in the last years. If D^k is a polytope the analytic center is defined as follows.

Definition 3.3 *(Analytic center, [97]) Let $\{x \mid Ax \leq b\}, where A \in \mathbb{R}^{m \times n}$, be a compact polytope with nonempty interior; then the unique maximizer of $\sum_{i=1}^{m} \log(b - Ax)_i$ over $\{x : Ax < b\}$ is called analytic center.*

As a convenient abbreviation we use ACCPM (analytic center cutting plane method, cf. Goffin and Vial [38]) for Algorithm 2 where the analytic center is used as inner point in step (i) and (iv).

Proposition 3.2 underlines the relevance of pseudo-monotonicity; as well as it is usually required for a minimization problem to fulfill some convexity properties in order to be 'reasonably' solvable, the same applies to monotonicity with respect to VIP. The reason is that these conditions allow to draw conclusions about the global behavior of a map given local information only.

To understand better the role of pseudo-monotonicity in the cutting plane method, we assume continuity of f and ask if pseudo-monotonicity is necessary for the equality $(f, D)^* = (f, D)^{**}$. The example with $D = [0,1]^2$ and $f(x) = (0.01 - 0.1x_1x_2, 1)$ exhibits $(f, D)^* = (f, D)^{**} = \{(0,0)\}$, but pseudo-monotonicity is not given as can be seen from the test-points $x = (0, 0.5)$ and $y = (1, 0.5)$. Hence, pseudo-monotonicity is not necessary for this cutting plane construction, it is only sufficient.

In a sense made explicit in the following lemma, however, the identity $(f, D)^* = (f, D)^{**}$ is equivalent with pseudo-monotonicity under the assumption of continuity. Or in other words: pseudo-monotonicity, under the assumption of continuity, is the weakest possible condition guaranteeing $(f, D)^* = (f, D)^{**}$.

Lemma 3.3 *Given a compact, nonempty and convex set $\tilde{D} \subset \mathbb{R}^n$ and a continuous map $f : \tilde{D} \to \mathbb{R}^n$; then the following two statements are equivalent:*

(i) f is pseudo-monotone on $\tilde{D}$;

(ii) $\forall D \subset \tilde{D}$, D convex and closed, it holds $(f, D)^* = (f, D)^{**}$.

The implication '(i) $\Rightarrow$ (ii)' is stated in Lemma 3.2; the reverse is shown as 'not (i) implies not (ii)'. Not (i) implies the existence of $x, y \in \tilde{D} :\ f(y)^T(x - y) \geq 0$ and $f(x)^T(x - y) < 0$. Set $D := [y, x]$, that is, the line segment from y to x.

Then y is obviously a solution, that is, $y \in (f, D)^*$. But by $f(x)^T(x - y) < 0$ we have $y \notin (f, D)^{**}$. Therefore (ii) does not hold. ⌋

Other similar results can be found in John [57].

Having established the necessity of pseudo-monotonicity for any cutting plane method, the question arises, under what conditions (pseudo-)monotonicity of the excess map can be expected. From an economic point of view Dafermos [16] claims that the excess map of a reasonable economic equilibrium problem should be at least 'nearly' monotone. This qualitative standpoint is mathematically analyzed by Hildenbrand [51] and will be discussed in Section 3.3. Mathematically seen, the overall excess is the sum of (part of) the solution of the individual utility and profit maximization problems; the dependency on the price signal can thus be understood as a sensitivity analysis of the underlying maximization problems. Based on such an analysis, one can either try to resolve the question on the level of the overall excess or, simpler, observe that (pseudo-)monotonicity of the individual consumers $e_i(p)$ implies overall (pseudo-)monotonicity. Here we do not address the topic of sensitivity analysis which has a rich literature (for a good presentation see Gauvin [35]). Instead, we point to the fact that sensitivity analysis gives *local* information, whereas the cutting plane algorithm requires *global* (pseudo-)monotonicity. But, as is seen below, the local and global behavior coincides.

Definition 3.4 *Let $U_\delta(x)$ be the open ball with radius δ centered at x. Then $f : D \subset \mathbb{R}^n \to \mathbb{R}^n$ is called* locally (pseudo-)monotone around $x \in D$ *if $\exists \delta > 0$, such that f is (pseudo-)monotone in $U_\delta(x) \cap D$. Furthermore, f is called* everywhere locally (pseudo-)monotone *if f is locally (pseudo-)monotone around all $x \in D$.*

Lemma 3.4 *Let $D \subset \mathbb{R}^n$ be a convex set and $f : D \to \mathbb{R}^n$ be everywhere locally (pseudo-)monotone, then f is (pseudo-)monotone on D.*

⌈First the monotone case is treated. Chose any two points $x \neq y$ in D. From the local monotonicity property we can cover the compact line segment $[x, y] \subset D$ with a finite covering $\mathcal{U}$ of (relative) open sets, where f is monotone on U for all $U \in \mathcal{U}$. From $\mathcal{U}$ we extract another covering with corresponding set of points $\{x^i\}_{i=0,\dots,N+1} \subset [x, y]$ such that all pairs of consecutive points are in the same U^i.

To do this the following algorithm is used which selects the 'good' sets from $\mathcal{U}$ and defines a first auxiliary set $\{z^k\} \subset [x, y]$:

(i) Set $k := 0$ and $z^0 := x$.

(ii) Choose any $U^k \in \mathcal{U}$ such that $z^k \in U^k$.

(iii) If $y \in U^k$ define $N := k$ and stop.
Else set $k := k + 1$, define $z^k := \underset{z \in [x,y] \setminus \cup_{j<k} U^j}{\operatorname{argmin}} \|x - z\|$, and goto (ii).

The 'else'-case in step (iii) is well defined, because the set over which the minimization is done is compact and non-empty. Finiteness of the algorithm is an immediate consequence of finiteness of $\mathcal{U}$ together with the property that no U^i can be chosen twice (because $z^k \notin U^j$ for $0 \leq j < k$). Finally, the chosen sets $\{U^i\}_{i=0,\dots,N}$ cover $[x, y]$, and we can assume $N > 0$ (otherwise x and y are in U^0 and then monotonicity would hold).

Now the interesting set $\{x^i\}$ is defined. Due to openness of the sets in $\{U^i\}_{i=0,\dots,N}$ our construction guarantees non-empty intersection $U^i \cap U^{i+1} \cap [x, y]$ for $i = 0, \dots, N-1$. Hence we can choose $x^{i+1} \in U^i \cap U^{i+1} \cap [x, y]$ such that $(x^{i+1} - x^i)^T(y - x) > 0$ for $i = 0, \dots, N-1$. Additionally we define $x^0 := x$, $x^{N+1} := y$. The resulting sequence has the property that every consecutive pair belongs to the same set where monotonicity of f holds, i.e. $x^i \in U^i$ and $x^{i+1} \in U^i$ for $i = 0, \dots, N$.

After this lengthy construction we know now

$$[f(x^{i+1}) - f(x^i)]^T[x^{i+1} - x^i] \geq 0 \quad \text{for } i = 0, \dots N.$$

The relative length

$$\lambda_i := \frac{\|x^{i+1} - x^i\|}{\|y - x\|} \quad \text{for } i = 0, \dots N$$

must be in the interval $(0, 1]$, and so from non-negativity of all summands involved we have

$$\begin{aligned}
0 &\leq \sum_{i=0}^{N} [f(x^{i+1}) - f(x^i)]^T[x^{i+1} - x^i] \\
&= \sum_{i=0}^{N} \lambda_i [f(x^{i+1}) - f(x^i)]^T[y - x] \\
&\leq [y - x]^T \left[\sum_{i=0}^{N} [f(x^{i+1}) - f(x^i)] \right] \\
&= [y - x]^T[f(y) - f(x)].
\end{aligned}$$

Now we prove the pseudo-monotone case by contradiction. Suppose f is not pseudo-monotone at $x, y \in D$, $x \neq y$, that is,

$$f(y)^T(x - y) \geq 0 \quad \text{and} \quad f(x)^T(x - y) < 0.$$

Denote by $x(t) := y + t(x - y)$ and set

$$t^* := \inf\{t \in [0, 1] \mid f(x(t))^T(x - y) < 0\}.$$

In the sequel, the following implications derived from scaling are used: If $t \in (0, 1]$, then $f(y)^T(x - y) \geq 0$ implies $f(y)^T(x(t) - y) \geq 0$; and if $1 \geq t_x > t_y > 0$, then $f(x(t_y))^T(x - y) \geq 0$ implies $f(x(t_y))^T(x(t_x) - x(t_y)) \geq 0$. The same applies to the case '<'. Now there are three possible cases concerning t^*:

(i) $t^* = 0$: This implies that for any $\delta \in (0,1)$ there is a $t_x \in (0, \delta/\|x - y\|)$ such that $f(x(t_x))^T(x - y) < 0$. Furthermore $x(t_x) \in U_\delta(y)$, hence $f(y)^T(x(t_x) - y) \geq 0$ and $f(x(t_x))^T(x(t_x) - y) < 0$, that is, f is not locally pseudo-monotone around y.

(ii) $t^* \in (0,1)$: Here we have for any $\delta \in (0,1)$ a $t_y \in (t^* - \delta/(2\|x-y\|), t^*) \cap [0,1]$ and a $t_x \in [t^*, t^* + \delta/(2\|x - y\|)) \cap [0,1]$ fulfilling

$$f(x(t_y))^T(x - y) \geq 0 \quad \text{and} \quad f(x(t_x))^T(x - y) < 0.$$

Because $t_y < t_x$ we have $\|x(t_x) - x(t_y)\| > 0$, so $f(x(t_y))^T(x(t_x) - x(t_y)) \geq 0$ and $f(x(t_x))^T(x(t_x) - x(t_y)) < 0$. From the construction follows $x(t_x) \in U_\delta(x(t_y))$ and therefore f is not locally pseudo-monotone around $x(t_y)$.

(iii) $t^* = 1$: Similar to case (i). ⌋

If convexity of D is dropped, however, the claim of Lemma 3.4 can be falsified by simple counter-examples; note also that continuity of f is not required.

So far we insisted on 'global' pseudo-monotonicity in order to guarantee that the set of solutions $(f, D)^*$ is contained in the cut set $(f, D)^{**}$. If we are satisfied by one solution $x \in (f, D)^*$, however, we can replace the condition of pseudo-monotonicity over the whole set D by the condition of pseudo-monotonicity at x alone.

Definition 3.5 *f is called pseudo-monotone at $x \in D$ over D if the implication $\big(f(x)^T(y - x) \geq 0 \implies f(y)^T(y - x) \geq 0\big)$ is true for all $y \in D$.*

Lemma 3.5 *Assume $x \in (f, D)^*$ and f is pseudo-monotone at x over D, then x is not cut away by any cut, i.e. $\forall y \in D : x \in \{z \mid f(y)^T(y - z) \geq 0\}$.*

The proof is an immediate consequence of the definition.

A second kind of relaxation is motivated by practical problems where one always has to deal with numerical imperfections. Looking at the geometry of a cut in Figure 3.2, where an additional strip of width $\varepsilon > 0$ (gray shaded) is included in the cut set $C_y^\varepsilon := \{x \in D \mid f(y)^T(y - x) \geq -\varepsilon\|f(y)\|\}$ demonstrates that the corresponding notion of ε-pseudo-monotonicity is a true relaxation of pseudo-monotonicity:

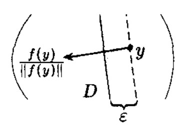


Figure 3.2: ε-relaxation.

Definition 3.6 *f is ε-pseudo-monotone over D if*

$$\big(f(x)^T(y - x) \geq 0 \implies f(y)^T(y - x) \geq -\varepsilon\|f(y)\|\big) \quad \forall x, y \in D.$$

It is interesting to note that in our practical economical examples we usually observed ε-pseudo-monotonicity for some appropriate $\varepsilon > 0$, whereas pseudo-monotonicity occured less frequently.

Let $Y^k = \{y^0, \dots, y^k\} \subset D$ be such that $\cap_{y^i \in Y^k} C^0_{y^i}$ has non-empty (relative) interior, which is exactly the situation of Algorithm 2. Then the point to set map $\varepsilon \in [0, \infty) \mapsto \cap_{y^i \in Y^k} C^{\varepsilon}_{y^i}$ is continuous. Namely continuity also holds at the boundary $\varepsilon = 0$. This is a consequence of [30, Theorem 3.2] together with the fact that non-empty *interior* of $\cap_{y^i \in Y^k} C^0_{y^i}$ implies non-emptiness of $\cap_{y^i \in Y^k} C^{\varepsilon}_{y^i}$ for some $\varepsilon < 0$.

In the course of an algorithmic process one can vary ε, say proportional to the radius of the largest inscribed sphere at the given iterate y^k, which, in case of a polytope D, is easy to compute. If such a relaxation-scheme is applied in a situation where the operator is strongly f-monotone and the volume reduction for non-relaxed cuts is bounded by a constant smaller than 1 (cf. [71]), such a relaxation is always possible while maintaining polynomial complexity.

3.2 Complexity of the VIP-Approaches

In the last 30 years the computational complexity has become an ever increasingly important aspect in the analysis of problems and corresponding algorithms. For a good survey on the historical development see for example Cook [14]. Closer to our problem is the monograph from Nemirovsky and Yudin [84]. In our context the basic question is: Given a class of problems defined by certain properties, is there a scheme which, for every $\varepsilon > 0$, finds an 'ε-close' solution after a polynomial number of arithmetic operations A (or iterations I).

To sharpen the question, the notion of an ε-close solution must be fixed; for convex minimization problems one usually considers an iterate x^k to be an ε-close solution if $|F(x^k) - F(x^*)| \leq \varepsilon$, where F denotes the objective and x^* is a minimizer.[1] Lacking an objective for VIPs, alternate measures must be used, see Definition 3.7 and 3.8 below.

Next, to make the notion 'polynomial' operational the arguments must be specified. Possible arguments comprise the dimension n of the involved Euclidean space, a characterization of the feasible set D like the diameter δ of D and the radius ρ of the largest sphere contained in D, and as characterization of f both a Lipschitz constant L and an upper bound M of $\|f(x)\|$ for $x \in D$. We call a problem class and the corresponding algorithm *polynomial*, if we can find a polynomial $p(n, \delta, \rho, L, M, \log(\frac{1}{\varepsilon}), \dots)$, such that $A \leq p(\dots)$ (or $I \leq p(\dots)$) for any problem instance in this class of problems. If the polynomial p depends on $\frac{1}{\varepsilon}$ rather than $\log(\frac{1}{\varepsilon})$, the complexity is called *pseudo-polynomial*.

[1] More popular is a *relative* ε-solution concept; if V is a good lower bound on the variation of F over D, i.e. $V \approx \max_{x \in D} F(x) - \min_{x \in D} F(x)$, a point $x \in D$ is called an ε-close solution if $|F(x^k) - F(x^*)|/V \leq \varepsilon$.

For our discussion it is important to emphasize the difference between the statement 'problem class Z is polynomial (solvable)' and 'algorithm Y which solves all problems of class Z is polynomial'. While the former is not specific about the algorithm (it requires only the existence of a polynomial algorithm), the latter claims polynomial complexity with a given algorithm.

Research in the last 10 years has revealed polynomial complexity for many important classes of convex minimization problems and consequently monotone VIPs (see e.g. [86, 98]); the underlying algorithms are so called path-following (interior point) methods which need a continuously differentiable f together with an 'easily' computable barrier, which must fulfill some additional conditions. If derivatives are not available—as it is the case for the MM^{mr}-based VIP—, and hence we apply a strategy following Algorithm 2, the example in Figure 3.5 below proves non-polynomial complexity for a monotone VIP. Because a monotone operator is *pseudo-monotone the same non-polynomial complexity of Algorithm* 2 holds for pseudo-monotone VIPs, see also Figure 3.4 below.

Strikingly enough, Nesterov [85] proves pseudo-polynomial complexity using AC-CPM for convex minimization problems, where f is the gradient map of the objective. Furthermore, for an adapted ACCPM where superfluous constraints are dropped even polynomial complexity can be shown, see Atkinson and Vaidya [3]. In view of Proposition 3.1 we have the unsatisfactory situation that ACCPM is polynomial for monotone VIPs only if f is the gradient of some convex function.

Dropping the assumption of a convex integral, a first polynomial complexity result based on Algorithm 2 using centers of circumscribed ellipsoids was proven 1985 in Lüthi [70] for strongly monotone VIPs. A generalization (in a certain restricted sense) for strongly f-monotone VIPs appeared 1996, see Magnanti and Perakis [71]. This case is of interest because it represents the weakest assumption known to date for polynomial complexity based on Algorithm 2. For monotone problems an ingenious breakthrough due to Nesterov and Vial [87] was finally achieved by homogenization of the problem and applying ACCPM in this extended setting. As an interesting detail, however, it turns out that the iterates y^k do not represent a sequence converging to a solution. Only after a clever weighting, $\bar{y}^k = \sum_{i=0}^k w_i y^i$, a sequence $\bar{y}^k$ is obtained which converges with pseudo-polynomial complexity to a solution.

To give an insight in the subtleties of the structures involved, we briefly outline in Section 3.2.1 the behavior of pseudo-monotone VIPs when solved by Algorithm 2. In Section 3.2.2 the monotone case is discussed where a homogenized ACCPM is used.

3.2.1 The Pseudo-Monotone Case

The notion of strong and weak solutions (see page 23) leads to two measures reflecting the 'closeness' of any $x \in D$ to the two solution sets $(f, D)^*$ and $(f, D)^{**}$

respectively. Examining first the strong solution property defining the true solution set $(f,D)^* = \{x \in D \mid f(x)^T(y-x) \geq 0\ \forall y \in D\}$ of VIP(f,D), we find for any $x \in D$

$$\min_{y \in D} f(x)^T(y-x) \leq 0,$$

where $\min_{y\in D} f(x)^T(y-x) = 0$ if and only if $x \in (f,D)^*$. By changing the sign we observe $\max_{y\in D} f(x)^T(x-y) \geq 0$ for any $x \in D$ and '$= 0$' if and only if $x \in (f,D)^*$ motivating the following notion.

Definition 3.7 *(Primal gap function, [48])* $g_p(x) := \max_{y\in D} f(x)^T(x-y)$.

A specific interpretation of g_p can be given in case of convex minimization problems, where we have $F(x^*) \geq F(x) + \nabla F(x)^T(x^*-x)$ for a solution x^* and any $x \in D$. We then see

$$\begin{aligned}|F(x)-F(x^*)| &= F(x)-F(x^*) \leq \nabla F(x)^T(x-x^*) \\ &\leq \max_{y\in D} \nabla F(x)^T(x-y) = g_p(x),\end{aligned}$$

that is, $g_p(x)$ gives an upper bound for $|F(x)-F(x^*)|$. More generally, $g_p(x)$ can be geometrically interpreted as measuring the width of $C_x := \{\, y \in D \mid f(x)^T(x-y) \geq 0 \,\}$ along $f(x)$ scaled by $\|f(x)\|$. Next, the weak solution property defining the weak solution set $(f,D)^{**} = \{x \in D \mid f(y)^T(y-x) \geq 0\ \forall y \in D\}$ can analogously be understood as $\min_{y\in D} f(y)^T(y-x) \leq 0$ for all $x \in D$, and with '$= 0$' if and only if x is a weak solution, that is, $x \in (f,D)^{**}$. Changing the sign yields the definition of the dual gap function:

Definition 3.8 *(Dual gap function)* $g_d(x) := \max_{y\in D} f(y)^T(x-y)$.

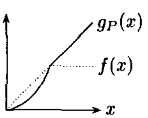


Figure 3.3: Non-convexity of g_p for a monotone VIP.

From the definitions of the gap functions follows an obvious way how a VIP can simply be stated as minimization problem. But while we can prove in Lemma 3.6 convexity of g_d without any assumptions, this is not true for the primal gap function; even under the assumption of (strong-)monotonicity g_p can be non-convex as is demonstrated by the following example. Think of a half-funnel defined over $D = [0,2]$ with $F = \frac{1}{2}x^2$ for $x \in [0,1]$ and $F(x) = x - \frac{1}{2}$ for $x \in (1,2]$. F is convex and once continuously differentiable with $f(x) := \nabla F(x) = x$ for $x \in [0,1]$ and constant 1 for larger x, see Figure 3.3. For $g_p(x)$ we find x^2 in the unit interval $[0,1]$ and x for $x \in (1,2]$. Thus, even though g_p *is convex both within* $[0,1]$ *and* $(1,2]$, *it is obviously non-convex on whole* D. By twisting slightly F on $(1,2]$ the same construction can be extended to strong monotone maps derived from F. But at least g_p is continuous. The situation with the dual gap function is nicer as is shown in the following lemma.

Lemma 3.6 *g_d is convex and continuous on D.*

g_d is the supremum of linear functions and therefor convex on D. From convexity follows continuity in the interior of D. In general, a convex function may be non-continuous on the boundary of its domain; in case of g_d, however, we can exploit its specificity. First note that the extended map $\bar{g}_d(x) := \max_{y \in D} f(y)^T(x-y)$ is well defined and convex on $D + \varepsilon$ for any $\varepsilon > 0$. Thus $\bar{g}_d$ is continuous on $\mathrm{int}(D+\varepsilon)$; because $\bar{g}_d \equiv g_d$ on D and additionally $D \subset \mathrm{int}(D+\varepsilon)$, g_d must be continuous on D.

To tackle the question, if a cutting plane concept following Algorithm 2 can have polynomial complexity for pseudo-monotone VIPs, we first have to agree on how the centers are chosen. As is demonstrated by the example presented in Figure 3.4, all 'reasonable' centers like the center of gravity, the analytic center, the center of (inscribed or circumscribed) ellipsoids or the volumetric center, behave qualitatively equal and we can choose whatever we like. Next, we have to agree on how the quality of an approximate solution x^k is measured; in this example we restrict ourselves to g_d and take as a second measure the Euclidean distance from the set of solutions, for which we have the trivial characterization $d(x^k, (f,D)^*) = 0$ if and only if $x^k \in (f,D)^*$.

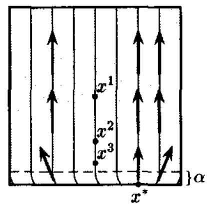


Figure 3.4: Complexity of a cutting plane algorithm for a pseudo-monotone VIP.

Looking at Figure 3.4, we first note that the vector-field is pseudo-monotone and continuous implying $(f,D)^* = (f,D)^{**} = \{x^*\}$. Next, we see that by choosing $\alpha > 0$ sufficiently small, an arbitrary number of iterates can be kept on the central vertical line. To estimate $g_d(x^k)$ and $d(x^k, x^*)$, we assume that the cube has sides of length 1, $x^* = (\frac{3}{4}, 0)$, that the length of the vectors amounts constantly to $\frac{1}{5}$ at all feasible points, and that the vector-field has an angle with the bottom line of $\frac{2\pi}{3}$ for all points on the bottom line which are by at least some small $\varepsilon > 0$ to the left of x^* (with this ε-gap we can interpolate the vector-field continuously to match the picture and have a unique solution). Choosing in the evaluation for $g_d(x^k)$ a y at the bottom line and by ε to the left of x^*, we find as a lower bound $g_d(x^k) \gtrsim \cos(\frac{\pi}{3}) \cdot \frac{1}{5} \cdot \frac{1}{4} = \frac{1}{40}$ for an arbitrary number of iterations k. Similarly we see $d(x^k, (f,D)^*) = \|x^k - x^*\| \geq \frac{1}{4}$. Hence a cutting plane concept following Algorithm 2 does not possess polynomial complexity with respect to g_d or the Euclidean distance for pseudo-monotone problems.

Such a non-polynomial behavior can also be observed if the primal gap function g_p is chosen as measure. Consider again the unit-square $\{0 \leq x_1 \leq 1\} \times \{0 \leq x_2 \leq 1\}$ and define $f(x) = (0,1)$ if $x_2 > \alpha$ and $f(x) = (1,0)$ if $x_2 \leq \alpha$. By choosing a sufficiently small $\alpha > 0$ we can provoke an arbitrary number of cuts parallel to the x_1-axis for reasonable centers (like the analytic center or the center of gravity), but finally a center will lie for the first time in the α-strip along the x_1-axis. For the analytic center or the center of gravity this will happen on the vertical line $x_1^k = 1/2$, and so $g_p(x^k) = 1/2$ for an arbitrary large k. The operator in this example could also be made differentiable by inserting an appropriate 'small' transitional strip.

To gain polynomial complexity in a cutting plane concept following Algorithm 2, strong f-monotonicity is the weakest known condition to date, cf. Magnanti and Perakis [71]. If the problem is solved in a homogenized reformulation, however, monotonicity is sufficient for pseudo-polynomial complexity. The next section outlines this approach.

3.2.2 The Monotone Case

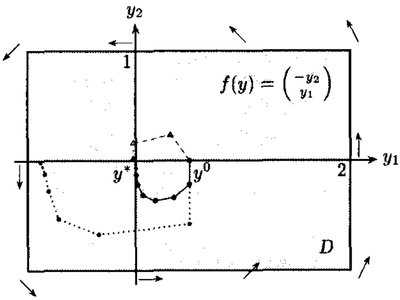


Figure 3.5: A monotone operator with diverging ACCPM.

Until recently this case resisted a theoretical satisfactory treatment. Nesterov and Vial [87] finally suggested a homogenization concept and proved pseudo-polynomial complexity for solving monotone VIPs without further conditions (except bounded $||f||$ on D, which is not demanding). In Figure 3.5 an example due to Nesterov and Vial [87] is given, where $f(y) = (y_2, -y_1)$ is monotone, $D = [-1, 2] \times [-1, 1]$, and the unique solution lies in the origin 0. For an improved presentation the vectors of the vector-field f are depicted slightly shortened along the boundary of D. Note that f is not integrable because its derivative is not symmetric. Considering Algorithm 2 all cuts pass through the origin and hence the analytic center is, starting from $y^0 = (1/2, 0)$, driven away from the solution towards $(-1, 0)$ when ACCPM is applied directly (dotted line). The hard nature of this problem is further underlined by using other centers like the center of gravity or the center of maximal inscribed ellipsoids which converge to $(-2/3, 0)$ and $(-1, 0)$ respectively, i.e. far away from the true solution.

On the background of this general difficulty with Algorithm 2, the result due to Nesterov and Vial [87] fascinates even more; its iterates are depicted along the solid line in Figure 3.5 showing convergence to the solution. To compare also with path-following methods, the triangles connected by a dashed line represent the iterates of an algorithm described in Ralph and Wright [93]. Because of the quick convergence only the first few iterates are shown, after 10 iterations the absolute value of the components of the iterates are already below 10^{-20}.

In the next section the basic analytic center cutting plane solution concept in the frame of a homogeneous feasibility problem is given, and in a subsequent section applied to monotone VIPs. For an improved presentation we denote by y the variable in the original space and by x the variable in the homogenized space.

The Homogeneous Feasibility Problem

The problem investigated in this section is the so called homogeneous feasibility problem. Given is a closed convex cone K with non-empty interior, and a second closed convex cone X^*. The feasibility problem then is to find $x \in K \cap X^*$ with $x \neq 0$, or to approximate $K \cap X^*$ in a sense specified later. The following definitions are necessary for the exposition.

Definition 3.9 *([86, Definition 2.1.1]) Let $Q \subset \mathbb{R}^n$ be open, convex and non-empty and let $\alpha > 0$. Then $F : Q \to \mathbb{R}$ is called α-self-concordant ('α-scc') on Q if $F \in C^3$ is convex, and for all $x \in Q$ and all $h \in \mathbb{R}^n$ the following inequality holds:*

$$|\nabla^3 F(x)[h,h,h]| \leq \frac{2}{\sqrt{\alpha}}(\nabla^2 F(x)[h,h])^{3/2}. \tag{3.3}$$

By the Mean Value Theorem we know $(\nabla^2 F(x^1) - \nabla^2 F(x^2))[h,h] = \nabla^3 F(\xi)[h,h,x^1 - x^2]$ for some $\xi \in [x^1, x^2]$; thus, condition (3.3) states Lipschitz continuity of $\nabla^2 F$ with respect to its own local norm $\nabla^2 F$. Note that from the definition follows 'stability under summation' ([86], Prop. 2.1.1 (ii)), i.e. if F_i is α_i-scc on Q_i, $i = 1,2$, and $Q := Q_1 \cap Q_2 \neq \emptyset$, then $F_1 + F_2$ is α-scc on Q with $\alpha = \min\{\alpha_1, \alpha_2\}$.

Definition 3.10 *([86, Definition 2.3.2]) Let $K \subset \mathbb{R}^n$ be a closed convex set and proper cone (i.e., $K \neq \mathbb{R}^n$) with non-empty interior, and let $\nu \geq 1$. Then $F : \text{int } K \to \mathbb{R}$ is called a ν-logarithmically homogeneous barrier for K (notation: $F \in B_\nu(K)$) if F is a C^2-smooth convex function on $\text{int } K$ such that $F(x_i) \to \infty$ for each sequence $\{x_i \in \text{int } K\}$ that converges to a boundary point of K, and, for each $x \in \text{int } K$ and each $t > 0$ we have*

$$F(tx) = F(x) + \nu \log t. \tag{3.4}$$

If in addition F is 1-self-concordant on $\text{int} K$ then F is called a ν-normal barrier for K (notation: $F \in NB_\nu(K)$).

Both ν-logarithmically homogeneous barriers and ν-normal barriers enjoy 'stability under summation': If $F_i \in B_{\nu_i}(K_i)$, $i = 1,2$ and $\text{int}(K_1 \cap K_2) \neq \emptyset$ then $F_1 + F_2 \in B_{\nu_1+\nu_2}(K_1 \cap K_2)$. Furthermore, if $F_i \in NB_{\nu_i}(K_i)$, $i = 1,2$, then $F_1 + F_2 \in NB_{\nu_1+\nu_2}(K_1 \cap K_2)$.

As an important example consider a convex cone defined by m hyperplanes: $K = \{x \mid a_i^T x \geq 0, i = 1, \ldots, m\}$. It possesses the m-normal barrier $F(x) = -\sum_{i=1}^m \log(a_i^T x)$ and therefore $F(x)$ is also 1-scc. More generally, let $F(x)$ be a ν-normal barrier for a cone K, and assume $K \cap \{x \mid a^T x \geq 0\}$ has nonempty interior; then the function $F(x) - \log(a^T x)$ is a $\nu + 1$-normal barrier for the cone $K \cap \{x \mid a^T x \geq 0\}$.

Definition 3.11 *([87, Definition 1]) $g(x)$ is called a homogeneous separation oracle for X^* on $\text{int} K$ if*

(i) $g(x)^T(x - x^*) \geq 0$ for all $x^* \in X^*$ and all $x \in \text{int } K$;

(ii) $g(tx) = g(x)$ for all $x \in K$;

(iii) $g(x)^T x = 0$ *for all* $x \in K$.

By (i) we do not lose any $x^* \in X^*$ by making cuts at any $x \in K$, and by (iii) the hyperplane returned by the separation oracle passes through the origin maintaining thereby the cone-property. In the sequel we assume the following.

Assumption 3.1 *([87, Assumption 1])*

(i) There is a ν*-normal barrier* $F(x)$ *for* K*;*

(ii) there is a homogeneous separation oracle $g(x)$ *for all* $x \in \operatorname{int} K$*, with* $\|g(x)\| = 1$.

Based on these assumptions the separation oracle is used in Algorithm 3 to solve the feasibility problem.

Algorithm 3 Homogeneous cutting plane method ([87, (2.5)]).

(i) Set $k = 0$, $F_0 := \frac{\nu}{2}\| x \|^2 + F(x)$.

(ii) Compute the analytic center $x_k = \operatorname{argmin}_x F_k(x)$, and set $F_{k+1}(x) = F_k(x) - \log(g(x_k)^T(x_k - x))$.

(iii) Stop if x_k satisfies a stopping criterion, otherwise set $k := k+1$ and return to (ii).

Due to the quadratic term $\frac{\nu}{2}\| x \|^2$ in F_0 the centers are called *proximal* analytic centers. In fact the factor $\frac{\nu}{2}$ can be replaced by any positive number without changing the iterates in the projective geometry. This freedom is also present in the later application to monotone VIPs described in Algorithm 4. In view of a real world implementation the accuracy in the computation of the analytic center and the oracle-response is to be clarified. At least the precision of computing the analytic center can be considerably relaxed to the standard approximation

$$\|\nabla F_k(x_k)\|_{[\nabla^2 F_k(x_k)]^{-1}} < \frac{3-\sqrt{5}}{2}$$

while preserving qualitatively the complexity analysis below.

To measure the convergence of the iterates the following weighted average of the slacks is studied:

$$\mu_k(x) = \frac{1}{S_k} \sum_{i=0}^{k-1} \lambda_{ik} g(x_i)^T (x_i - x), \tag{3.5}$$

where

$$\lambda_{ik} = \frac{1}{g(x_i)^T(x_i - x_k)} \quad \text{for } i = 0, \dots, k-1, \text{ and } \quad S_k = \sum_{i=0}^{k-1} \lambda_{ik}.$$

Note that analytic centers are always in the interior of the cone, implying $x_i \neq x_j$ for all possible i and j where $i \neq j$.

With the constants $\theta_1 := \frac{1}{2}(\sqrt{5}-1) - \log(\frac{\sqrt{5}+1}{2}) \approx 0.137$, $\theta_2 := \frac{\sqrt{5}+1}{2} \approx 1.62$, and $\theta_3 := \frac{1}{\theta_2} e^{\theta_1 - 1/2} \approx 0.43$ the following main theorem holds for the conic feasibility problem when Algorithm 3 is applied.

Theorem 3.7 *([87, Theorem 1]) For any $x \in K$ the following bound on $\mu_k(x)$ holds:*

$$\mu_k(x) \leq \frac{\sqrt{k+\nu}}{k\theta_3} e^{(F(x_k)-F(x_0))/k} \| x \|.$$

Furthermore, $F(x_k) - F(x_0) \leq k\sqrt{\nu}\theta_2$ yielding the bound

$$\mu_k(x) \leq \frac{\sqrt{k+\nu}}{k\theta_3} e^{\theta_2\sqrt{\nu}} \|x\|. \tag{3.6}$$

Solving Monotone VIPs

Consider a monotone VIP(f, D) with single valued f, where D and f are bounded, i.e. there exist constants R and L respectively such that $\| y \| \leq R$ and $\| f(y) \| \leq L$ for all $y \in D$. From monotonicity we know that $g_d(y) = 0$ if and only if $y \in (f, D)^{**}$, and moreover $g_d(y)$ is convex and continuous on D. Similar to the cutting plane method outlined in Algorithm 2 we do not intend to solve VIP(f, D), but want to find a point y which is close to $(f, D)^{**}$ in the sense of the dual gap function g_d.

Definition 3.12 *Given the operator f and the compact set D with nonempty interior, we call a point y an ε-close solution to VIP(f, D) if $g_d(y) \leq \varepsilon$.*

In a first step the VIP is transformed into a conic feasibility problem by the following embedding:

$$\begin{aligned} X^* &:= \{x := (ty, t) \mid y \in (f, D)^{**},\, t > 0\}, \\ K &:= \{x := (ty, t) \mid y \in D,\, t > 0\}. \end{aligned}$$

Next we need a ν-normal barrier for K; this can be constructed in a straightforward way if a ν-scc barrier for D is known, cf. Nesterov and Nemirovskii [86, Proposition 5.1.4]. Here we restrict ourselves to the relevant case of a feasible set D defined by linear inequalities, $\{y \mid a_i^T y \leq b_i,\, i = 1, \ldots, m\}$. This includes the unit simplex Δ which is the feasible set in our equilibrium problem. In such a case, with linear constraints only, an $m+1$-normal barrier for K is given by

$$F(x) = -\sum_{i=1}^{m} \log(t(b_i - a_i^T y)) - \log t = -\sum_{i=1}^{m} \log(b_i - a_i^T y) - (m+1)\log t,$$

where, as defined above, $x = (ty, t)$. In order to apply the machinery from the previous section we have to impose the following assumption.

Assumption 3.2 *([87, Assumption 3]) The origin* 0 *is the analytic center of* D*, i.e.* $\nabla_y F(0,t) = 0$.

Consequently, the minimizer of $F_0(x) = \frac{\nu}{2}\|x\|^2 - F(x)$ is $x_0 = (0,1)$ yielding $\|x_0\| = 1$. In order to meet this requirement a translation of the whole problem from the analytic center of D into the origin has to be done; this is also true in the case of Δ, the feasible set of our equilibrium problem.

As a final ingredient for Algorithm 3 we need a homogeneous separation oracle following Definition 3.11; with

$$\hat{g}(x) := (f(y), -f(y)^T y)$$

the conditions (ii) and (iii) of Definition 3.11 are obvious. The inequality in item (i) of Definition 3.11 is equivalent with $f(y)^T(y^* - y) \leq 0$, and this contains the set $(f,D)^{**}$. Hence we have a homogeneous separation oracle, where, at $\bar{x} = (\bar{t}\bar{y}, \bar{t})$, we have

$$\hat{g}(x)^T(x - \bar{x}) = \bar{t}(f(y), y - \bar{y}). \tag{3.7}$$

Theorem 3.8 *([87, Theorem 2]) Algorithm 3 yields an* ε*-approximate solution in the sense of Definition 3.12 in at most* k *iterations, where* k *satisfies*

$$\frac{k}{\sqrt{k+\nu}} \leq \frac{L(1+R^2)}{\varepsilon\theta_3} e^{\theta_2\sqrt{\nu}}. \tag{3.8}$$

$\ulcorner$Assume $\{x_i\} = (t_i y_i, t_i)$ is the sequence generated by Algorithm 3; define

$$\pi_{ik} = \frac{\lambda_{ik}}{\|\hat{g}(x_i)\|}, \quad P_k = \sum_{i=0}^{k-1} \pi_{ik},$$

and

$$\bar{y}_k = \frac{1}{P_k}\sum_{i=0}^{k-1} \pi_{ik} y_i. \tag{3.9}$$

Choose an arbitrary $y \in D$; from monotonicity of f we conclude

$$f(y)^T(\bar{y}_k - y) = \frac{1}{P_k}\sum_{i=0}^{k-1} \pi_{ik} f(y)^T(y_i - y) \leq \frac{1}{P_k}\sum_{i=0}^{k-1} \pi_{ik} f(y_i)^T(y_i - y). \tag{3.10}$$

Let $x = (y,1)$ be the corresponding canonic element in the cone K; then from (3.7) we deduce further

$$\begin{aligned}
\frac{1}{P_k}\sum_{i=0}^{k-1} \pi_{ik} f(y_i)^T(y_i - y) &= \frac{1}{P_k}\sum_{i=0}^{k-1} \pi_{ik}\hat{g}(x_i)^T(x_i - x) \\
&= \frac{1}{P_k}\sum_{i=0}^{k-1} \lambda_{ik} g(x_i)^T(x_i - x) \\
&= \frac{S_k}{P_k}\mu_k(x).
\end{aligned}$$

Note that

$$\|\hat{g}(x)\| = \sqrt{\|f(y)\|^2 + (f(y)^T y)^2} \leq L\sqrt{1+R^2}$$

and so

$$P_k = \sum_{i=0}^{k-1} \frac{\lambda_{ik}}{\|\hat{g}(x_i)\|} \geq \frac{S_k}{L\sqrt{1+R^2}}.$$

With $\|x\| \leq \sqrt{1+R^2}$ we derive from (3.6)

$$\begin{aligned} g_d(\bar{y}_k) &\leq \frac{S_k}{P_k} \max_{\substack{x=(y,1),\\ y\in D}} \mu_k(x) \;\leq\; \frac{S_k}{P_k} \frac{\sqrt{k+\nu}}{k\theta_3} e^{\theta_2\sqrt{\nu}} \sqrt{1+R^2} \\ &\leq \frac{\sqrt{k+\nu}}{k\theta_3} e^{\theta_2\sqrt{\nu}} L(1+R^2). \end{aligned}$$

Replacing finally $g_d(\bar{y}_k)$ by ε proves the claim. ⌋

The above theorem also proves implicitly the existence of a weak solution for monotone, not necessarily continuous maps f. In view of Theorem 3.8 this is a consequence of continuity of g_d together with compactness of D.

Note that the reduction of the dual gap function is not achieved by the direct iterates $\{x_i\}$ or $\{y_i\}$ respectively, but by the iterates $\{\bar{y}_i\}$ which are computed as weighted mean of all previous direct iterates $\{y_i\}$. The weight $1/(\hat{g}(x_i)^T(x_i - x_k))$ attached to y_i when computing $\bar{y}_k$ is large in two cases: (i) if the new iterate y_k comes close to the hyperplane through y_i measured by $1/(g(x_i)^T(x_i - x_k))$, and (ii) if the absolute value of the operator $\|\hat{g}(x_i)\|$ is small. Both conditions express the fact that y_i is a good approximate solution.

The homogeneous barrier exploits strongly the problem-structure. A remarkable outcome of this is the non-presence of the dimension of the problem in the complexity-bound; however, ν is influenced by the number and nature of the constraints, and is therefore a measure for both the dimensionality and general 'hardness' of the problem. Furthermore, ν appears in an exponential term which can become a serious drawback for higher-dimensional problems.

As indicated in Nesterov and Vial [87] for the case of constrained minimization, the quadratic dependency in R can be reduced to a linear dependency by a suitable scaling of D. In case of a VIP we construct a scaled problem with a parameter $\kappa > 0$ by $D \rightsquigarrow \frac{1}{\kappa}D$ and consequently we have to replace $f(y)$ by $f(\kappa y)$ to solve the old problem. In this scaled problem L stays unaffected, whereas R becomes $\frac{1}{\kappa}R$ and ε transforms to $\frac{1}{\kappa}\varepsilon$. If we choose $\kappa = \gamma R$ proportional to the original R, the complexity relation (3.8) of the scaled problem to regain an ε-solution of the unscaled problem is

$$\frac{k}{\sqrt{k+\nu}} \leq \frac{L\left(1+\left(\frac{R}{\kappa}\right)^2\right)}{\frac{\varepsilon}{\kappa}\theta_3} e^{\theta_2\sqrt{\nu}} = \frac{LR\left(\gamma+\frac{1}{\gamma}\right)}{\varepsilon\theta_3} e^{\theta_2\sqrt{\nu}}.$$

The factor $\gamma + \frac{1}{\gamma}$ is minimal for $\gamma = 1$, i.e. the scaling is optimal if $\kappa = R$ with the complexity relation

$$\frac{k}{\sqrt{k+\nu}} \leq \frac{LR}{\varepsilon\theta_3} e^{\theta_2\sqrt{\nu}}. \tag{3.11}$$

Let us conclude this section by expliciting the homogeneous cutting plane method for monotone VIPs.

Algorithm 4 Homogeneous CPM for monotone VIPs (cf. [87]).

(i) Choose $\varepsilon > 0$ and compute the related maximal iteration bound $\bar{k}$ following (3.8) or (3.11).

Set k=0, shift the initial feasible set at its analytic center into the origin, let $F(x)$ be the ν-normal barrier for the related cone, and define $F_0 := \frac{\nu}{2}\| x \|^2 + F(x)$.

(ii) If $k > \bar{k}$ goto step (iv).

(iii) Compute the analytic center $x_k = \operatorname{argmin}_x F_k(x)$, set $F_{k+1}(x) = F_k(x) - \log(g(x_k)^T(x_k - x))$, set $k := k + 1$ and goto step (ii).

(iv) Compute the solution $\bar{y}_k$ following (3.9).

Of course, instead of fixing the number of iterations in the beginning of Algorithm 4, it can be reasonable to replace step (ii) by step (iv) and then to judge the quality of $\bar{y}_k$ in every iteration. Note, however, that the explicit evaluation of $g_d(\bar{y}_k)$ is in general not tractable. Only the primal gap function can easily be handled as linear programming problem in case of a polytopal initial feasible set. And in view of (3.10) monotonicity helps to use the primal gap function as bound for the dual gap function.

As a final aspect note that the accuracy required in the computation of the analytic center in Algorithm 4 seems to limit the attainable final quality of solution.[2] This is a fundamental difference to Algorithm 3 for solving the feasibility problem. The question of accuracy with respect to the *oracle* seems to be still open.

3.3 Economic Evidence for (Pseudo-)Monotonicity of Market Demand

The previous two sections clarified the necessity of (pseudo-)monotonicity for applying a cutting plane method. Following the notation in Section 1.1 we want to justify here (pseudo-)monotonicity of the aggregate demand function $d(p) =$

[2] Personal communication with J.-P. Vial

$\sum_{i \in I} x_i(p)$. This is not yet the aggregate excess function which is the vector valued difference between supply and demand as a function of the price, but if the aggregate supply is monotone (e.g. independent of the price), the property of monotonicity of demand and excess coincides.[3] Dafermos [16] states for the excess function that "monotonicity assumptions, though restrictive, are in the spirit of the 'law of demand' ".

To support intuition a typical one dimensional supply and demand curve with resulting excess is depicted in Figure 3.6. The supply s is usually expected to increase when prices rise (nonnegative slope of $s(p)$), whereas the demand d decreases (non-positive slope of $d(p)$). Hence we observe not only monotonicity of the demand, $(d(p^1) - d(p^2))(p^1 - p^2) \leq 0$, but even more plausible monotonicity of the excess $e(p) = s(p) - d(p)$. The relevance and applicability of this one dimensional argument to higher dimensional cases, however, is limited. Major reasons are so called cross-price effects, where changes in the price of a commodity i influence demand or supply of a different good j, which is a common phenomenon. Given such cross-price effects together with monotonicity for each component, it depends then on the amount and sign of the cross-price effects, whether or not the excess for prices differing in more than one component is (pseudo-)monotone.

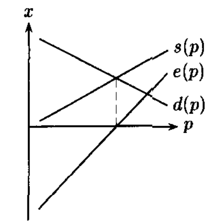


Figure 3.6: Monotonicity of demand and excess.

Note that in general the individual demand not only depends on the price, but on the monetary endowment w as well. If $w = p^T x^0$ for fixed x^0, the additional variable w is not needed, but in general w has to be included and then the individual demand function will be written in the form $f^i(p, w^i)$, $i \in I$. If there is only one individual the index i is dropped.

In the sequel a number of cases will be investigated with respect to (pseudo-)monotonicity of demand.

3.3.1 Pseudo-Monotonicity of Excess in the Case of Utility Maximizing Consumers

As indicated above we focus the discussion on demand only. This might irritate, because it is well known in the economic literature that for utility maximizing *consumers* the resulting individual demand $x_i(p)$ is already pseudo-monotone. To justify this claim we consider problem (1.6) but exclude production, i.e. $y_i = 0$, and drop the index i for convenience. In a first step we claim that the resulting individual demand function $x(p)$ fulfills WARP (weak axiom of revealed preference):

Definition 3.13 *(WARP, weak axiom of revealed preference, [94]) The demand function $f(p, w) : \mathbb{R}_+^{n \times 1} \to \mathbb{R}_+^n$, with price p and monetary endowment w (outlay, income), is said to fulfill the weak axiom of revealed preference if for every pair (p^1, w^1), $(p^2, w^2) \in \mathbb{R}_+^n \times \mathbb{R}_+$, $p^{2T} f(p^1, w^1) \leq w^2$ implies $p^{1T} f(p^2, w^2) \geq w^1$.*

[3] We use here the convention of the economic literature, where monotonicity of demand is defined with reversed sign (inequality) compared to our definition 3.2. In the case of supply or the excess, however, our definition applies.

In order to show WARP for consumers based on (1.6) we first bridge the gap to our notation by setting $x(p) = f(p, p^T x^0)$, abbreviate $x^i = x(p^i) = f(p^i, p^{iT} x^{i0}) = f(p^i, w^i)$ and observe that our consumers behavior follows $\max U(x)$ s.t. $p^T x \leq w$. To verify WARP we see from the left side of the implication that x^1 is feasible at (p^2, w^2), therefore we must have $U(x^2) \geq U(x^1)$. If we assume that the implication in the definition of WARP does not hold we get the contradiction that x^2 is strictly feasible at (p^1, w^1) and thus $U(x^1) > U(x^2)$.

From WARP we derive in a second step pseudo-monotonicity of the individual excess. The relation

$$p^{2T} f(p^1, w^1) \leq w^2 \;\Rightarrow\; p^{1T} f(p^2, w^2) \geq w^1$$

can be written in the notation of (1.6) as

$$p^{2T}(x(p^1) - x^0) \leq 0 \;\Rightarrow\; p^{1T}(x(p^2) - x^0) \geq 0, \tag{3.12}$$

or, using the excess $e(p) = x^0 - x(p)$, we find $p^{2T} e(p^1) \geq 0 \Rightarrow p^{1T} e(p^2) \leq 0$. With the budget identity $p^T e(p) = 0$ we finally conclude

$$e(p^1)^T (p^2 - p^1) \geq 0 \;\Rightarrow\; e(p^2)^T (p^2 - p^1) \geq 0.$$

Essential in this derivation of pseudo-monotonicity of $e(p)$ is the price-independence of production. This holds e.g. for a pure exchange economy, but as soon as production is allowed, the identity $w = p^T x^0$ is extended to $w = p^T(x^0 + y(p))$ and consequently we find instead of (3.12)

$$p^{2T}(x(p^1) - x^0 - y(p^2)) \leq 0 \;\Rightarrow\; p^{1T}(x(p^2) - x^0 - y(p^1)) \geq 0,$$

where $x(p^1) - x^0 - y(p^2)$ can no more be interpreted as $-e(p^1)$ except for the case when $y(p)$ is in fact independent of the price.

Numerically we find—due to the production in problem (1.6)—price pairs violating pseudo-monotonicity of the excess in case of MM^{mr}. Nevertheless, WARP still holds if $w = w(p) = p^T(x^0 + y(p))$ *is* dependent on the price, i.e. for problem (1.6) in its full generality. But because an agent of the form (1.6) includes production, pseudo-monotonicity of the individual excess can not be deduced.

3.3.2 Monotonicity of Demand in the Case of a Continuum of Equal Consumers

Coming back to the question of monotonicity of the demand $f(p, w)$ we find in Hildenbrand [50] a first positive answer. Let f be an individual demand function, i.e. f fulfills WARP and the budget identity $p^T f(p, w) = w$ holds for all $p > 0$ and $w \geq 0$. Consider the following situation; there is a continuum of consumers with respect to w described by the same individual demand function $f(p, w)$, and furthermore the distribution of individual expenditure described by the density

$\rho : \mathbb{R}_+ \to \mathbb{R}_+$ is a decreasing function with $\int_0^\infty \rho(w)dw = 1$ and $\int_0^\infty w\rho(w)dw < \infty$. Then the mean (market) demand function is defined by

$$F(p) = \int_0^\infty f(p,w)\rho(w)dw. \tag{3.13}$$

Note that in this formulation ρ and w are independent of the price. The following result can be shown.

Theorem 3.9 *([50], Theorem 1) For every individual demand function f and for every decreasing density ρ, the mean demand function F from (3.13) is monotone, i.e, $(p^1 - p^2)^T(F(p^1) - F(p^2)) \leq 0$ for every $p^1 > 0$, $p^2 > 0$.*

Hence, under the given assumptions we can achieve a stronger property by aggregation than the individual demand functions enjoy. Note also that part of the assumptions can be relaxed, e.g. ρ may be increasing in the beginning to some extend, or there may be a finite set of different demand functions.

But comparing this situation with problem (1.6) we find nevertheless first of all only finitely many consumers (quite few indeed). Secondly, the demand functions are usually mutually different, and thirdly w depends strongly on p. Thus, even though $x_i(p)$ derived from (1.6) is an individual demand function, we cannot derive monotonicity of demand.

3.3.3 Non-Monotonicity of the Slutsky Compensated Demand Function

Contrary to the other parts this section documents *non-monotonicity* of an approximate demand function called Slutsky compensated demand function (or 'Slutsky demand' for short). Two reasons motivate this; first the so called Slutsky decomposition (of demand), which underlies the Slutsky demand, is needed in the next section, and secondly the Jacobian of the Slutsky demand is negative semidefinite (n.s.d.). Because this led various authors erroneously to the conclusion that the Slutsky demand function is monotone (cf. Eatwell, Milgate and Newman [22, pp. 544]), the subtleties involved will briefly be outlined.

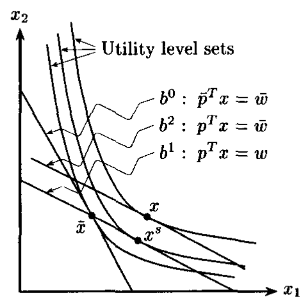


Figure 3.7: Slutsky decomposition of demand.

The object under investigation is an individual, continuously differentiable demand function $f(p,w)$ with a constant (price independent) outlay w. Assume $f(p,w)$ is analyzed around a given point $\bar{x} = f(\bar{p},\bar{w})$ where $\bar{p}^T\bar{x} = \bar{w}$. Then the Slutsky demand function, defined by $x^s_{\bar{x}} : p \mapsto f(p, p^T\bar{x})$, reflects the change in demand when the price changes under the assumption of constant purchasing power, that is, the outlay is adapted to keep $\bar{x}$ just affordable. Graphically this corresponds to a rotation of the budget constraint around $\bar{x}$ in Figure 3.7. While $\bar{x} \to x^s$ can be interpreted as *(Slutsky) substitution effect*, $x^s \to x$ is constructed by a parallel shift of the budget constraint $b^1 \to b^2$ to regain the old outlay in the new price setting p and called *income effect*. Looking at our model (1.6) we have $w = p^T(x^0 + y(p))$; to simplify assume that production y is price independent yielding a demand function $f(p, p^T\bar{x}^0)$ with $\bar{x}^0 := x^0 + y$. Thus at $\bar{x}^0$ the Slutsky demand function $x^s_{\bar{x}^0}(p)$ and the simplified demand function $f(p, p^T\bar{x}^0)$ stemming from (1.6) coincide, and one is motivated to use $x^s_{\bar{x}}(p)$ as an approximation for $f(p, p^T\bar{x}^0)$ in a neighborhood of $\bar{p}$, where $\bar{x} = f(\bar{p}, \bar{p}^T\bar{x}^0)$. Differentiating the Slutsky compensated demand function $x^s_{\bar{x}}(p)$ gives

$$\nabla_p x^s_{\bar{x}}(p)\big|_{\bar{p}} = \nabla_p f(p, p^T\bar{x})\big|_{\bar{p}} = \nabla_p f(p, \bar{w})\big|_{\bar{p}} + \nabla_w f(\bar{p}, w)\big|_{\bar{w}}\bar{x}^T, \tag{3.14}$$

where $Sf := \nabla_p x^s_{\bar{x}}(p)\big|_{\bar{p}}$ and $Af := \nabla_w f(\bar{p}, w)\big|_{\bar{w}}\bar{x}^T$ are called 'Slutsky substitution matrix' and 'matrix of income effects' respectively. Now Hildenbrand [51, p. 176] proves equivalence of WARP and n.s.d. of Sf under the assumption of budget identity, i.e. at points $(\bar{p}, \bar{x})$ where $x^s_{\bar{x}}(\bar{p}) = \bar{x}$. Based on the first equality in (3.14) one might hope to exploit n.s.d. of Sf to prove monotonicity of $f(p, p^T\bar{x})$ for all $\bar{x}$ and all p in a neighborhood of $\bar{p}$ by using the Mean Value Theorem,

$$(f_{\bar{x}}(p) - f_{\bar{x}}(\bar{p}))^T(p - \bar{p}) = (p - \bar{p})^T \nabla_p f_{\bar{x}}(p)\big|_{\xi}^T (p - \bar{p}),$$

where the notation $f_{\bar{x}}(p)$ abbreviates $f(p, p^T\bar{x})$, and ξ is chosen appropriately from the interval $[p, \bar{p}]$. It turns out, however, that n.s.d. of $\nabla_p f_{\bar{x}}(p)\big|_{\xi}$ can not be deduced from n.s.d. of Sf because the latter holds only at points $(\bar{p}, \bar{x})$ where $x^s_{\bar{x}}(\bar{p}) = \bar{x}$, and hence can be false at $p = \xi$. It is even possible to construct individual demand functions of the form $f(p, p^T\bar{x}^0)$ which are not monotone, see Eatwell et al [22, p. 545].

3.3.4 Monotonicity of Demand in the Case of a Large Population of Sufficiently Heterogeneous Consumers

In the previous section we approximated the real demand function $f(p, w(p))$ stemming from problem (1.6) by requiring price-independence of production y and thereby relating it to the Slutsky demand function. But we can as well approximate $f(p, w(p))$ by the usual demand function $f(p, w)$ where w is price-independent, that is constant. Given a finite population of consumers with individual continuously differentiable demand functions $f^i(p, w^i)$, where w^i is price-independent, the mean market demand function $F(p)$ is defined by

$$F(p) = \frac{1}{|I|} \sum_{i \in I} f^i(p, w^i), \tag{3.15}$$

and again we ask if $F(p)$ is monotone. A positive answer, based on the main ideas in Hildenbrand [51], is outlined in the following. The reasoning starts with the Slutsky decomposition stated in (3.14). The goal is to raise evidence for n.s.d. of $\nabla_p F(p)$ and we rewrite for that purpose (3.14) in the form

$$\nabla_p F(p) = S(p) - A(p), \tag{3.16}$$

where $S(p)$ and $A(p)$ represent the corresponding mean over the set of consumers of $Sf(p)$ and $Af(p)$ respectively. From the last section we know $Sf(p)$ and thus $S(p)$ is n.s.d. and concentrate therefore here on conditions implying positive semi-definiteness (p.s.d.) of $A(p)$. To begin with note that $A(p)$ is p.s.d. if and only if $M(p) := A(p) + A(p)^T$ is. From (3.14) and (3.15) we deduce

$$\begin{aligned} M(p) &= \frac{1}{|I|} \sum_{i \in I} \nabla_{w^i} \left[f^i(p, w^i) f^i(p, w^i)^T \right] \\ &= \lim_{h \to 0} \frac{1}{h} \left[m^2\{f^i(p, w^i + h)\} - m^2\{f^i(p, w^i)\} \right], \end{aligned}$$

where m^2 is the second moment of a cloud of vectors with components defined by

$$m^2_{jk}\{f^i(p, w^i)\} := \frac{1}{|I|} \sum_{i \in I} f^i_j(p, w^i) f^i_k(p, w^i).$$

Observe that the second moment of any cloud of vectors is p.s.d. (if in the definition of m^2 every summand is p.s.d. then also the sum, and the former is equivalent to requiring that uu^T is p.s.d. for any $u \in \mathbb{R}^n$ which is trivial). Therefore we mean by 'increasing spread of consumers demand' that for every sufficiently small $h > 0$ the matrix $m^2\{f^i(p, w^i + h)\} - m^2\{f^i(p, w^i)\}$ is p.s.d., and this is a sufficient condition to have a p.s.d. $A(p)$, and thus monotonicity of the mean market demand.

One plausible argument why $M(p)$ should be p.s.d. stems from the observation that an increase in income also increases the variance of demand (heteroscedasticity). However, it must be fundamentally acknowledged that empirical evidence is needed here. Interestingly enough this shows also that the mean market demand can enjoy properties that are non-existent in any of the underlying individual demand functions. In Hildenbrand [51] this concept is much more elaborated and made applicable to real world data, and empirically verified using data sets from the United Kingdom and France.

3.3.5 Monotonicity of Demand Implied by a Sufficiently Small Curvature of Utility

Another possibility, due to Mitjuschin and Polterovich (1978, cited in [51]), is based on the insight that by imposing certain conditions on the utility U, the resulting individual substitution effect $Sf(p)$, and thus $S(p)$, can be made sufficiently negative definite in order to guarantee n.s.d. of $\nabla_p F(p)$.

Proposition 3.10 *(cf. [51]) Let $f(p,w)$ denote a C^1 demand function that is derived from a C^2, monotone, and concave utility function U. If*

$$-\frac{x^T\nabla^2 U(x)x}{x^T\nabla U(x)} < 4 \qquad \forall\, x \in \mathbb{R}^n_{++},$$

then the function $f(p,w)$ is strictly monotone in p for $w > 0$, that is,

$$(f(p^1,w) - f(p^2,w))^T(p^1 - p^2) < 0 \qquad \forall\, p^1 \neq p^2 \in \mathbb{R}^n_{++}.$$

3.3.6 Monotonicity of Demand Implied by Homothetic Individual Demand Functions

Instead of making Sf sufficiently negative definite we ask now for conditions implying Af to be p.s.d. Assume for that purpose that an individual demand function $f(p,w)$ is homogeneous of degree one in w, that is, $f(\,.\,,\lambda w) = \lambda f(\,.\,,w)$ for all $\lambda \geq 0$. Using Euler's equation $w\nabla_w f(\,.\,,w) = f(\,.\,,w)$ we then have collinearity of $\nabla_w f(\,.\,,w)$ and $f(\,.\,,w)$ which is equivalent with p.s.d. of $\nabla_w f(\,.\,,w) f(\,.\,,w)^T = Af$. This can be extended to so called homothetic functions which are produced by applying a strictly increasing transformation to a homogeneous function. It is exactly this class of homothetic demand functions, characterized by the collinearity property of $\nabla_w f$ and f, which guarantee Af to be p.s.d.

3.3.7 Conclusions

The cases above can be seen as different approximations to the 'real' demand (or excess) function resulting from a set of agents of the form (1.6). Though none of them is equivalent to (1.6), they support the claim of Dafermos [16] cited in the beginning. Specifically, it is interesting to find pseudo-monotonicity of the individual excess (and thereby the aggregated excess) given price-independence of production. Further reaching, however, are the concepts relying on sufficiently large sets of heterogeneous consumers where the mean variance of demand increases with rising w. This justifies hope to gain (pseudo-)monotonicity of the excess for models comprising an increasingly number of agents of the form (1.6).

Der Türhüter stellt öfters kleine Verhöre mit ihm an, fragt ihn über seine Heimat aus und nach vielem andern, es sind aber teilnahmslose Fragen, wie sie grosse Herren stellen, und zum Schluss sagt er ihm immer wieder, dass er ihn noch nicht einlassen könne. Der Mann, der sich für seine Reise mit vielem ausgerüstet hat, verwendet alles, und sei es noch so wertvoll, um den Türhüter zu bestechen. Dieser nimmt zwar alles an, aber sagt dabei: "Ich nehme es nur an, damit du nicht glaubst, etwas versäumt zu haben." F. K. [58]

Chapter 4

Solving EEP Using the Negishi-Approach

In this chapter the conceptual Negishi algorithm (Algorithm 1 page 16) is discussed more in depth. First, Section 4.1 starts by comparing the VIP- with the Negishi-view. Next, in Section 4.2 two strategies are presented on how the weight vector α^k can be updated, thereby concreting the conceptual Negishi Algorithm. Because the algorithms solve in each iteration a Negishi-welfare problem, Section 4.3 treats two undesirable properties of the Negishi-welfare problem: (i) to actually built it requires in general a global reformulation of all underlying individual utility maximization problems, and (ii) to solve the resulting large welfare problem may be intractable. Both problems are simultaneously resolved by a technique called decomposition and for which an algorithm is given.

Contributions comprise the suggestion of the δ-Negishi-Algorithm, the analysis of the t-Negishi-Algorithm in case of $\mathrm{MM}^{\mathrm{mr}}$ (see Appendix E.2.2), and the discussion of unboundedness of the Lagrangian in Section 4.3.2.

4.1 Comparing the Negishi- and VIP-View

As stated in Theorem 2.5 there are relations between the VIP- and the Negishi problem[1] motivating the notion of 'primal' problem for the VIP-approach and 'dual' for the Negishi problem.

Figure 4.1 depicts symbolically the Negishi- and the VIP-approach; the Negishi-view is located to the left, whereas the VIP-view is presented to the right. U_r represents the objective function (utility) and K_r the set of constraints (feasibility set) for $r \in R$ from problem (1.5), or, more generally, from (1.6), *without*

[1]While the Negishi-welfare problem is given in Definition 1.2, the 'Negishi problem' is then to find a weight vector α, such that a solution of the related welfare problem represents an equilibrium, cf. Theorem 2.5.

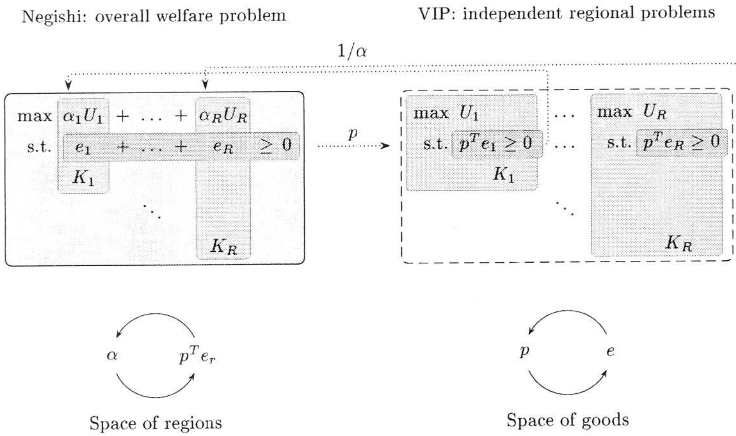


Figure 4.1: Dual relationship between the VIP- and Negishi-approach.

the budget constraint. Here the index set R instead of I is chosen in view of our later application to a multiregional problem. There are three levels of gray: light gray is the Negishi-welfare problem integrating all individual (regional) problems in one large optimization problem, medium gray are the individual structures which—in case of the VIP-view—are essentially the individual utility maximization problems (1.5) or (1.6), and finally, emphasized dark-gray are the excess-related constraints. The latter constraints account on the one hand for the main difference between the individual problem (1.6) and the EEP and, on the other hand, are central for the dual relationship between the Negishi- and the VIP-view. In the VIP-part the surrounding box is only dashed and not shaded to underline its consistence of independent subproblems.

Now Theorem 2.5 states that in an equilibrium the dual multiplier vector p of the excess constraint in the welfare problem is exactly an equilibrium price, motivating the dotted arrow from left to right. Reversing the view, the inverse of the dual multipliers δ of the budget constraint in the VIP-sub-problems form a set of equilibrium Negishi weights. This is indicated by the dotted arrows from right to left labeled '$1/\alpha$'. Note that scaling α or p by any positive scalar does not affect its equilibrium properties; this permits keeping both α and p in the unit simplex Δ of appropriate dimensionality.

From Definition 1.4 it follows that p is an equilibrium price if and only if the resulting excess e from solving all regional problems in the VIP-box is non-negative and complementarity with the price holds. This is suggested by the bottom right circle in Figure 4.1, where in one iteration $e(p)$ is computed, and depending on its outcome p adjusted. As for the Negishi-view Theorem 2.5 together with Definition 1.4 claims that α is an equilibrium weight vector if and only if $p^T e_r = 0$ for all $r \in R$, where p is the dual multiplier vector of the excess constraint. This leads to the bottom left circle where α is judged and updated following the outcome of $p(\alpha)^T e_r(\alpha)$ for all $r \in R$.

Denoting by the index g any good, an intuitive approach for updating α_r or p_g simply looks at the corresponding test-quantities $p(\alpha)^T e_r(\alpha)$ and $e_g(p)$ respectively: increase α_r if $p(\alpha)^T e_r(\alpha) > 0$ and decrease it otherwise, and do the analogue with reversed inequality in case of p_g and $e_g(p)$. Economically this means in the welfare view that a region (consumer) should have more weight if it did not use all the wealth it has, and decrease the weight of those exceeding there budgets ($p(\alpha)^T e_r(\alpha) < 0$). In the VIP-world a price-component p_g is increased if demand exceeds supply ($e_g(p) < 0$) and vice versa.

This is the basic idea behind the so called 'tâtonnement'-process which was one of the first algorithmic concepts used in computational economics, cf. Ginsburgh and Waelbroeck [36]. A direct application of such a tâtonnement-process, however, is not only theoretically unsatisfactory, but can yield poor results in practice too (mainly because it ignores cross-effects). As for the VIP-approach, Algorithm 2 based on the analytic center or center of gravity is more robust. Similarly, the δ-Negishi algorithm extends also the tâtonnement-process and convinces on our practical problem MM^{mr}. Nevertheless, both the VIP- and the Negishi-algorithm used in this work are under the given structures only heuristics; for a possible exact algorithm based on a fixed point approach see e.g. Taheri [99].

Note that on the first level the Negishi-approach leads to an algorithm in the space of the regions or agents, whereas the VIP-view works in the space of goods. Because the dimensionality of the problem influences strongly the computational efficiency, this can be of determining importance. However, if the Negishi-welfare problem is solved using decomposition, the dimensionality of the goods reappears, cf. Algorithm 6. If this can be handled efficiently, e.g. in that the number of goods increases the computational burden comparably slowly, then the Negishi-approach seems preferable if the number of goods exceeds significantly the number of regions or agents.

4.2 Two Algorithms

4.2.1 The δ-Negishi-Algorithm

As a direct outcome of the previous discussion the δ-Negishi-algorithm updates the weight vector by explicitly computing the dual multiplier of the budget constraint in the underlying regional problem.

The heuristic fixed point method given in Algorithm 5 proved to be very contractive and additionally very robust with respect to starting points. In case of our

Algorithm 5 δ-Negishi-algorithm

(i) Choose a set of initial weights α^0 and set $k = 0$.

(ii) Solve the Negishi-welfare problem and compute thereby the dual price p^k.

(iii) Stop if the solution satisfies a stopping criterion, otherwise proceed.

(iv) With the price p^k solve the utility maximization problem of all economic agents ((1.5) or (1.6)) and retrieve the dual multipliers δ_r^k of the budget constraint.

(v) Set $k := k + 1$ and update the weights by

$$\alpha_r^k = \frac{1}{\delta_r^{k-1} \sum_{r \in R} 1/\delta_r^{k-1}};$$

return to (ii).

examples it was advantageous to start with a suitable p^0 in step (iv). A more in depth discussion is presented in Appendix E.2.1. Numerical results are presented in Appendix F.4 and F.5.

4.2.2 The t-Negishi-Algorithm

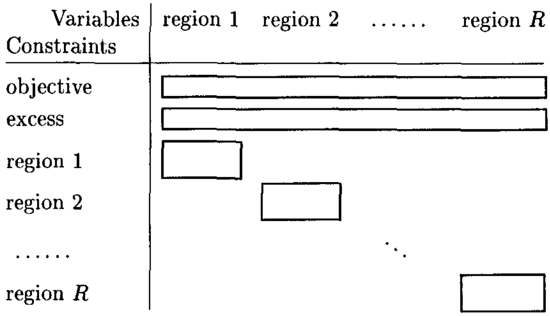
Another approach to update the Negishi-weights was successfully used e.g. in $5R$, a five region model based on simplified Markal-Macro models and described in Manne and Rutherford [75]. The central idea is to estimate δ, the dual multipliers of the budget constraint, from a solution of the welfare problem. Such a scheme avoids solving the regional models (1.6) in every iteration of Algorithm 5.

Because deriving such estimators depends on the concrete structure of the model, it will be done in Appendix E.2.2 for MM^{mr}.

A final note on the naming; t is motivated from the fact that the resulting scheme is a 'tâtonnement' strategy, where the old weight is essentially updated by the (weighted) addition of the budget excess. That is, $\alpha_r^{k+1} = \alpha_r^k + wp^{kT}e_r^k$ with an appropriate weight w.

4.3 Decomposing the Negishi-Welfare Problem

In every iteration of the $[\delta, t]$-Negishi-algorithm we face a Negishi-welfare problem (1.3). It exhibits a typical block-diagonal structure with a few connecting constraints:



The overall problem can—depending on $|R|$—grow exceedingly large. If we consider 10 regions with 5000 variables each, the Hessian of the objective contains $50'000 \times 50'000 = 2.5 \cdot 10^9$ elements; supposing 10 % non-zeros using 8 byte precision, we need for storing alone about 2 gigabyte computer memory, a number which beats state of the art workstations by a factor of 10.

Besides the need to keep the Negishi-welfare problem tractable for real world computers we face in the next step the complexity observed in practice (i.e. the time needed to find an approximate solution). Theoretically optimal local methods for solving nonlinear convex problems require at least $O(n \log \frac{1}{\varepsilon})$ iterations to find an ε-approximate solution, where n is the number of variables (dimension of the problem), see Elster [63]. This implies that at best real world solvers based on Newton-kind of methods exhibit an increase of computation time of an order of n^3. Even though this is a low order polynomial satisfying theoreticians it can already forbid to solve large models; to see this assume that it takes 15 minutes to solve one regional problem (a typical value in our case), then the same solver needs $10^3 \cdot 15 = 15000$ minutes or about 250 hours—more than 10 days!—to solve the Negishi welfare problem with 10 regions.

A final obstacle specific to our situation is that the overall Negishi-welfare problem can hardly be set up; each region is a large and complex piece of GAMS-code consisting of over 100 files, and to put together several regions would require to extend the whole code by a regional index—an enormous task taking months of work.[2] Even worse, one would run into update problems: every time the original regional code is changed this has to be followed up in the Negishi-code.

For all these reasons we are thus seeking for a procedure which allows to solve the Negishi-welfare problem by solving the underlying independent and (almost) unchanged regional problems. It is exactly this integrative aspect which obliged us to use decomposition techniques. The other side of the medal—the usual one

[2] At least in case of GAMS there are attempts to extend the system to handle such situations in a transparent way, i.e. without requiring to change the underlying model code. But in general the problem still exists: how to design an algorithm for solving the Negishi-welfare problem where the underlying regional problems are left as much as possible unchanged.

where large problems are split into pieces and thereby the computational burden reduced—attracts increasingly attention, cf. [37, 30, 31].

The basic technique of dualizing common constraints is presented sufficiently general in Appendix A.2. In the rest of this chapter we apply this to the separable structure of the Negishi-welfare problem and subsequently solve it by using ACCPM, a cutting plane method, cf. Goffin, Haurie and Vial [38, 37]. The section concludes with a complexity result due to Nesterov [85].

4.3.1 The Lagrangian Dual Problem

The Negishi-welfare problem (1.3) (or (7.9) in case of MM^{mr}) can be written as

$$\left.\begin{array}{llll}
\max \alpha_1 U_1(v_1) & + \ldots\ldots + & \alpha_R U_R(v_R) & \\
\text{s.t.}\ \ e_1(v_1) & + \ldots\ldots + & e_R(v_R) & \geq 0, \\
g_1(v_1) & & & \leq 0, \\
h_1(v_1) & & & = 0, \\
& \ddots & & \\
& & g_R(v_R) & \leq 0, \\
& & h_R(v_R) & = 0.
\end{array}\right\} \tag{4.1}$$

The constraints of the underlying regional subproblems define $K_r := \{v_r \mid g_r(v_r) \leq 0,\, h_r(v_r) = 0\}$; abbreviating further the objective $o_r(v_r) := \alpha_r U_r(v_r)$ and the overall feasible set by $K := \Pi_{r \in R} K_r$, problem (4.1) can be written in the form

$$\begin{array}{lllll}
\max & o_1(v_1) & + \ldots\ldots + & o_R(v_R) & \\
\text{s.t.} & e_1(v_1) & + \ldots\ldots + & e_R(v_R) & \geq 0, \\
& v_1 \in K_1, & & & \\
& & \ddots & & \\
& & & v_R \in K_R. &
\end{array}$$

The Lagrangian dual function (cf. Appendix A.2) is then defined as

$$\begin{aligned}
\theta(p) &:= \max_{v \in K} \left[\sum_{r \in R} o_r(v_r) + \sum_{r \in R} p^T e_r(v_r) \right] \\
&= \sum_{r \in R} \max_{v_r \in K_r} \left[o_r(v_r) + p^T e_r(v_r) \right],
\end{aligned} \tag{4.2}$$

where p has the dimension d of the image set of $e(v)$. The Lagrangian dual problem following (A.5) is

$$\min_{p \geq 0} \theta(p). \tag{4.3}$$

Given Condition (A.6) is satisfied, Theorem A.3 and A.4 state that it is equivalent either to solve the primal problem (4.1) or the dual problem (4.3).[3] In view of (4.2) and the introduction we observe that the evaluation of $\theta(p)$ for a given p results in $|R|$ independent maximization problems of the form

$$\max_{v_r \in K_r} \left[o_r(v_r) + p^T e_r(v_r)\right]$$

which coincide with the original regional models up to the Negishi multiplier hidden in o_r and the additional term $p^T e_r(v_r)$ in the objective. The remaining question is how to solve the dual problem (4.3). Of course, if we want to be successful we need in general convexity of θ together with information about (sub-)gradients. For that purpose let $P := \{p \geq 0 \,|\, \theta(p) < \infty\}$ denote the set where θ is finite and write succinctly $o(v) = \sum_{r \in R} o_r(v_r)$, $e(v) = \sum_{r \in R} e_r(v_r)$ and finally $\phi(p, v) := o(v) + p^T e(v)$. The next lemma proves convexity of θ, and so P is also convex.

Lemma 4.1 *(for similar results see Bazaraa and Shetty [7]) θ is convex; furthermore, if $p \in P$ and $v^* \in \operatorname{argmax}_{v \in K} \phi(p, v)$, then $e(v^*)$ is a subgradient of θ at p.*

To show convexity of θ choose $\lambda \in (0, 1)$ and take any prices $p, q \geq 0$; we then have

$$\begin{aligned}
\theta(\lambda p + (1-\lambda)q) &= \max_{v \in K}\{o(v) + [\lambda p + (1-\lambda)q]^T e(v)\} \\
&= \max_{v \in K}\{\lambda[o(v) + p^T e(v)] + (1-\lambda)[o(v) + q^T e(v)]\} \\
&\geq \lambda \max_{v \in K} \phi(p, v) + (1-\lambda) \max_{v \in K} \phi(q, v) \\
&= \lambda\theta(p) + (1-\lambda)\theta(q).
\end{aligned}$$

To see the second claim concerning $e(v^*)$ being a subgradient of θ at p, choose any $q \geq 0$; then

$$\begin{aligned}
\theta(q) &= \max_{v \in K} \phi(q, v) \\
&\geq \phi(q, v^*) \\
&= o(v^*) + \bar{p}^T e(v^*) \\
&= o(v^*) + p^T e(v^*) - p^T e(v^*) + \bar{p}^T e(v^*) \\
&= \theta(p) + e(v^*)^T (q - p).
\end{aligned}$$

[3] To make this statement precise note that a solution p^* of the dual problem (4.3) has implicitly attached a primal 'solution' $\bar{v} \in K$ by means of (4.2). In general $\bar{v}$ is *not* a solution to (4.1), but once p^* is known it can be used within a derived linear programming problem to approximate a primal solution $v^* \in K$. For a detailed discussion see Bazaraa and Shetty [7, Section 6.5].

From this lemma we know how to linearly approximate $\theta(p)$ by (sub-)gradient cuts within P. But this is not sufficient to solve (4.3) because for a $p \geq 0$ outside P we have so far no information in what direction the minimum of $\theta(p)$ might be, or, to be more modest, where P is located. This topic is briefly discussed in the next section.

4.3.2 Unboundedness of the Lagrangian Dual

To begin with note that unboundedness of $\theta(p)$ is of practical relevance because the additional term $p^T e(v)$ in the objective can turn bounded problems into unbounded ones. This is specifically true in case of the subproblems in the Negishi-decomposition approach of MM^{mr}. Given p with $\theta(p) = \infty$, in this section we try to derive supporting hyperplanes of P passing through p. Such a process can yield an outer approximation of P.

In the sequel we make for all $r \in R$ the following assumptions:

Assumption 4.1 *$o_r(v_r)$ and $e_r(v_r)$ are continuously differentiable, concave on K_r and finite if v_r is finite. Furthermore, K_r is non-empty, convex and closed.*

From these assumptions the following lemma can be deduced allowing to focus the further discussion on a single region $r \in R$.

Lemma 4.2 *Choose any $p \geq 0$; then $\theta(p)$ is finite if and only if $\theta_r(p) := \max_{v_r \in K_r} \{o_r(v_r) + p^T e_r(v_r)\}$ is finite for all $r \in R$.*

This lemma implies that if we choose any $r \in R$ and find a (linear) constraint which cuts away a part of $\mathbb{R}^d$ (with d the dimension of P) where $\theta_r(p) = \infty$, then the remaining part contains P. Or to state it differently: every subproblem can independently generate so called feasibility-cuts which form an outer approximation of P. In the following discussion we concentrate therefore on an arbitrary $r \in R$.

From Assumption 4.1 follows boundedness of $\theta_r(p)$ for any finite $p \geq 0$ if K_r is also bounded. A possible strategy for practitioners could thus consist of imposing an overall box constraint, which is reasonable for 'real' world problems.[4]

In case of unbounded $K_r \subset \mathbb{R}^{n_r}$ we call a vector $d_r \in \mathbb{R}^n$ a *direction* of K_r if $v_r + \lambda d_r \in K_r$ for all $\lambda \geq 0$. Additionally, we call a feasible point *extremal* if it can not be represented as a proper (i.e. $\lambda \in (0,1)$) convex combination of two

[4]This strategy is also present in typical solvers for nonlinear convex optimization problems; usually the variable values are bounded to a 'reasonable' large box like $\pm 10^{20}$ as is the case for GAMS-related solvers. In such a situation one can in principle ignore any 'unboundedness-message' of the solver and simply use the subgradient at the point where the solution process was stopped which, without harm, may happen at the boundary of the large box.

different feasible points. A direction is called *extremal* if it can not be represented as a positive combination of two different directions.

Assuming K_r has directions we ask how P can be characterized. As for notation, $\nabla e_r(v_r)$ or simply ∇e_r is the differential of the vector-valued map e_r with respect to v_r; usually this matrix is called Jacobian.

Lemma 4.3 *Let both o_r and e_r be affine and Assumptions 4.1 hold; if $p \in P$ then for any direction d_r of K_r we have*

$$(\nabla o_r + \nabla e_r{}^T p)^T d_r \leq 0. \tag{4.4}$$

⌜Assume on the contrary $p \in P$ and $(\nabla o_r + \nabla e_r{}^T p)^T d_r > 0$; From affinity of o_r and e_r we then have

$$o_r(v_r + \lambda d_r) + p^T e_r(v_r + \lambda d_r) = o_r(v_r) + p^T e_r(v_r) + \lambda(\nabla o_r + \nabla e_r{}^T p)^T d_r,$$

where $\lambda(\nabla o_r + \nabla e_r{}^T p)^T d_r$ can be made arbitrarily large by increasing λ. But this contradicts the assumption $p \in P$.⌟

Because (4.4) holds for all $p \in P$, Lemma 4.3 gives an outer approximation of P if all (extremal) directions d_r of K_r are checked. To reverse the implication and construct thereby an inner approximation of P, it is necessary to strengthen (4.4) as is shown in the example illustrated in Figure 4.2 below.

The notion concave for a vector-valued function used below is defined by concavity of all its components.

Lemma 4.4 *Let both o_r and e_r be concave and Assumptions 4.1 hold; then $p \in P$ if there is a $v_r \in K_r$ such that for all directions d_r of K_r we have*

$$(\nabla o_r(v_r) + \nabla e_r(v_r)^T p)^T d_r < 0. \tag{4.5}$$

⌜The proof will given by showing the equivalent statement '$p \notin P \implies \forall v_r \in K_r$ there is a direction d_r of K_r with $(\nabla o_r(v_r) + \nabla e_r(v_r)^T p)^T d_r \geq 0$'. Choose any $v_r^1 \in K_r$; from $p \notin P$ we know there exists a continuing sequence $v_r^n \in K_r$, $n \geq 2$, such that $o_r(v_r^n) + p^T e_r(v_r^n) \to \infty$ for $n \to \infty$, and such that $v_r^1 \neq v_r^n \forall n \geq 2$. This implies (from the previous assumption that o_r and e_r are finite for finite v_r) that v_r^n must tend to infinity; defining for $n \geq 2$

$$d_r^n := \frac{v_r^n - v_r^1}{||v_r^n - v_r^1||},$$

we observe that d_r^n is a sequence on the unit-ball, and thus has a convergent subsequence which—without loss of generality—is assumed to be d_r^n with limit d_r^∞. Note that due to closedness of K_r d_r^∞ is a direction.

Setting $\lambda^n := \|v_r^n - v_r^1\|$ we derive from concavity of o_r and e_r

$$\begin{aligned} o_r(v_r^n) + p^T e_r(v_r^n) &= o_r(v_r^1 + \lambda^n d_r^n) + p^T e_r(v_r^1 + \lambda^n d_r^n) \\ &\leq o_r(v_r^1) + p^T e_r(v_r^1) + \lambda^n(\nabla o_r(v_r^1) + \nabla e_r(v_r^1)^T p)^T d_r^n, \end{aligned}$$

and thus $(\nabla o_r(v_r^1) + \nabla e_r(v_r^1)^T p)^T d_r^n > 0$ for all sufficiently large n which implies $(\nabla o_r(v_r^1) + \nabla e_r(v_r^1)^T p)^T d_r^\infty \geq 0.$ ∎

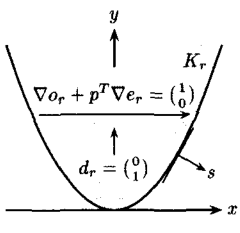


Figure 4.2: Direction and unboundedness.

To see why (4.5) can not be relaxed to '$\leq$' consider the following example, where both o_r and e_r are assumed affine. Choose as primal feasible set $K_r := \{(x, y) \mid y \geq x^2\}$ in $\mathbb{R}^2$ which is simply the epigraph of a paraboloid. The only extremal direction is $d_r = (0, 1)$; if p is such that $(\nabla o_r + \nabla {e_r}^T p) = (1, 0)$ then (4.5) holds with equality, that is, $(\nabla o_r + \nabla {e_r}^T p)^T d_r = 0$. But the sequence $d^n := (n, n^2)$ makes $(\nabla o_r + \nabla {e_r}^T p)^T d^n = n$ diverge to infinity, implying by virtue of affinity $o_r(d^n) + p^T e_r(d^n) \to \infty$.

To verify Lemma 4.4 in this example, let us now assume $(\nabla o_r + \nabla {e_r}^T p)^T d_r < 0$, which implies a negative y-component of $s := (\nabla o_r + \nabla {e_r}^T p)$. The maximum of $o_r(v_r) + p^T e_r(v_r)$ over K_r is then finite and achieved at the point where s equals the normal of a supporting plane as is indicated in Figure 4.2.

Anticipating Algorithm 6 which is used to solve (4.3), the following observation is useful. Assume P is characterized by the cuts (4.4) formed by all extremal directions, and denote this outer approximation of P by $\bar{P}$. Then any *inner* point of $\bar{P}$ satisfies in fact (4.5). Hence, once $\bar{P}$ is available Algorithm 6 which uses analytic centers suffers no more from unboundedness of the Lagrangian.

In practice, $\bar{P}$ will be iteratively built up: Starting with $\bar{P}^0 := \{p \in \mathbb{R}_+^d \mid p \leq Me\}$ for a sufficiently large M, $\bar{P}^k$ is reduced to $\bar{P}^{k+1} := \bar{P}^k \cap (4.4)$ whenever a direction d_r^k is detected, and otherwise left unchanged. Now, if $\bar{P}^k$ is a 'sufficiently close' approximation to P and our inner test points p^k are 'sufficiently far away' from the boundaries of $\bar{P}^k$, we are in the happy situation that $\theta(p^k) < \infty$ despite the possibility $\bar{P}^k \backslash P \neq \emptyset$.

Nevertheless, Lemma 4.4 can be sharpened; one possible way is by requiring boundedness of the extremal points K_r^e of K_r for all $r \in R$. Such a K_r can be seen as algebraic sum of a bounded convex set and a convex cone; to give an example think of the paraboloid in Figure 4.2 where the convex bounded set is changed to $Q = \{(x, y) \mid y \geq x^2 \wedge y \leq 1\}$ and extended by the cone $C = \{d \mid d = \lambda(-1, 2) + \mu(1, 2),\ \lambda \geq 0,\ \mu \geq 0\}$.

Denote by the operator 'conv' the convex hull of a set, let K_r^e be the (bounded) set of extremal points of K_r, and let C be a cone such that $K_r = \text{conv}(K_r^e) + C$. For any $v_r \in K_r$ we then have the existence of a $v_r' \in \text{conv}(K_r^e)$ and a $d_r' \in C$ such that $v_r = v_r' + d_r'$. With n_r the dimension of K_r we can further derive from

Carathéodorys Theorem the existence of $n_r + 1$ extremal points $v_r^i \in K_r^e$ such that v_r' can be written as convex combination of the v_r^i, i.e. there exists a vector of weights $\lambda \geq 0$, $\sum_{i=1}^{n_r+1} \lambda_i = 1$, such that $v_r' = \sum_{i=1}^{n_r+1} \lambda_i v_r^i$.

Based on these prerequisites the following characterization of P can be established.

Lemma 4.5 *Let Assumption 4.1 hold and let $K_r = conv(K_r^e)+C$ be the algebraic sum of a bounded convex set $conv(K_r^e)$ plus a cone C. We then have $p \in P$ if there is a $v_r \in conv(K_r^e)$ such that for all directions d_r of K_r*

$$(\nabla o_r(v_r) + \nabla e_r(v_r)^T p)^T d_r \leq 0. \tag{4.6}$$

We demonstrate the equivalent implication '$p \notin P \Rightarrow \forall v_r \in \text{conv}(K_r^e)$ there exists a direction $d_r \in C$ with $(\nabla o_r(v_r) + \nabla e_r(v_r)^T p)^T d_r > 0$' (cf. the proof of Lemma 4.4).

Choose any $v_r \in \text{conv}(K_r^e)$ and define

$$M := \sup_{w_r \in \text{conv}(K_r^e)} o_r(w_r) + p^T e_r(w_r).$$

which is finite from Assumption 4.1.

From boundedness of K_r^e we have a finite diameter $\delta(K_r^e)$ of the set K_r^e, and furthermore from Assumption 4.1 follows the existence of a finite upper bound L for both $||\nabla o_r(v_r)||$ and $||\nabla e(v_r)||$ if v_r is in $\text{conv}(K_r^e)$. As usual, the matrix-norm $||\nabla e(v_r)||$ is defined by the maximum of the product $|p^T \nabla e(v_r) v_r|$ over all unit-vectors p and v_r.

Now $p \notin P$ implies the existence of a sequence $v_r^n \in K_r$, $v_r^n \to \infty$, such that $o_r(v_r^n) + p^T e_r(v_r^n) \to \infty$ for $n \to \infty$. Therefore there exists an n such that

$$o_r(v_r^n) + p^T e_r(v_r^n) \ > \ L(1 + ||p||)\delta(K_r^e) + M.$$

Let $v_r^n = v_r' + d_r$ for a suitable $v_r' \in \text{conv}(K_r^e)$ and $d_r \in C$. Seen from the chosen $v_r \in \text{conv}(K_r^e)$ we then have

$$\begin{aligned}
&L(1 + ||p||)\delta(K_r^e) + M \\
&\quad < o_r(v_r^n) + p^T e_r(v_r^n) \\
&\quad = o_r(v_r + (v_r' - v_r) + d_r) + p^T e_r(v_r + (v_r' - v_r) + d_r) \\
&\quad \leq o_r(v_r) + p^T e_r(v_r) + (\nabla o_r(v_r) + \nabla e_r(v_r)^T p)^T ((v_r' - v_r) + d_r) \\
&\quad = o_r(v_r) + p^T e_r(v_r) + (\nabla o_r(v_r) + \nabla e_r(v_r)^T p)^T (v_r' - v_r) \\
&\qquad\qquad + (\nabla o_r(v_r) + \nabla e_r(v_r)^T p)^T d_r \\
&\quad \leq M + L(1 + ||p||)\delta(K_r^e) + (\nabla o_r(v_r) + \nabla e_r(v_r)^T p)^T d_r,
\end{aligned}$$

and hence $0 < (\nabla o_r(v_r) + \nabla e_r(v_r)^T p)^T d_r$ which proves the lemma.

This statement can be made more useful by observing that from Carathéodorys Theorem for cones it is sufficient to restrict condition (4.6) to *extremal* directions only.

From the Lemmata 4.3 and 4.5 the following corollary follows.

Corollary 4.6 *Let Assumption 4.1 hold, let $K_r = conv(K_r^e) + C$ be the algebraic sum of a bounded convex set $conv(K_r^e)$ plus a cone C, and let both o_r and e_r be affine. We then have $p \in P$ if and only if for all directions d_r of K_r*

$$(\nabla o_r + \nabla {e_r}^T p)^T d_r \leq 0.$$

4.3.3 Solving the Lagrangian Dual Problem

To solve (4.3) we abbreviate (sub-)gradients of $\theta(p)$ at p_k by g_k, i.e. $g_k = g(p_k) = e(v^k)$, where $v^k \in \text{argmax}_{v \in K}(o(v) + p_k^T e(v))$.

The Proximal Analytic Barrier Method

This section is based on Nesterov [85]. Choose a constant $\sigma \geq 7$, and a starting point p_0 which satisfies $\|p_0 - p^*\| \leq \rho$ for a solution p^* and a constant ρ. Assuming furthermore $B_{1.1 \cdot \rho}(p_0) \subset P$ and $\|g(p)\| \leq L$ for all $p : \|p - p_0\| \leq 1.1 \cdot \rho$ for a suitable constant L, the so called analytic barrier is defined to be

$$\begin{aligned} F_0(p) &:= \frac{\sigma}{2R^2}\|p - p_0\|^2, \\ F_k(p) &:= F_{k-1}(p) + \frac{1}{2R^2}\|p - p_0\|^2 - \log(g_{k-1}^T(p_{k-1} - p)), \qquad k \in \mathbb{N}. \end{aligned}$$

Based on this strictly convex barrier $F_k(p)$, the next iterate is then

$$p_k := \operatorname*{argmin}_p F_k(p), \qquad k \in \mathbb{N}_0,$$

where p is varied over the interior of the polytope $\{p \,|\, g_i^T(p_i - p)) \geq 0, i = 0, \ldots, k-1\}$. It can be shown that $\|p_k - p_0\| \leq 1.1 \cdot \rho$ and so g_k is well defined for all k. Using the notation $\theta_k^* = \text{argmin}_{i \leq k} \theta(p_i)$ and $\theta^* = \theta(p^*)$ it is shown in Nesterov [85] that for all $k \in \mathbb{N}_0$

$$\theta_k^* - \theta^* \leq c(\sigma) L \rho \frac{e^{\sigma/(2(k+1))}}{\sqrt{\sigma + k + 1}},$$

where $c(\sigma)$ is some constant depending on σ. Hence the gap $\theta_k^* - \theta^*$ is asymptotically decreased at a rate of at least $1/\sqrt{k}$. Also remarkable in this result is its independence of the problem dimension d. In practical problems, however, ρ and L might have to grow with d; e.g. to include the unit cube ρ grows with $\sqrt{d}$. Finally note that this convergence result must be multiplied with the cost of the oracle, that is, in our case the time required to compute $e(p)$.

The Analytic Center Cutting Plane Method (ACCPM)

This scheme described in [37, 38, 39] and used in our implementation of the Negishi-algorithm has a number of good properties: it gives an upper bound on $\theta_k^* - \theta^*$, it handles unbounded directions in case of linear objectives $o_r(v_r) + p^T e_r(v_r)$, and it speeds up convergence by using multiple (sub-)gradient cuts in each iteration. A convergence analysis for one or two simultaneous cuts can be found in Goffin and Vial [40, 41]. Let $v_r^k := \operatorname{argmax}_{v_r \in K_r} \phi_r(p_k, v_r)$, and $v^k := (v_1^k, \ldots, v_R^k)$, then from Lemma 4.1 we know that $e(v^k)$ is a (sub-)gradient of $\theta(p)$ at p_k:

$$\begin{aligned} \theta(p) &\geq \theta(p_k) + (p - p_k)^T e(v^k) \\ &= \sum_{r \in R} \left[\phi_r(p_k, v_r^k) + (p - p_k)^T e_r(v_r^k)\right]. \end{aligned}$$

An obvious lower approximation $\bar{\theta}_k$ of θ based on a set of test points $\{p_0, \ldots, p_k\} \subset P$ is therefore

$$\left.\begin{aligned} \bar{\theta}_k(p) := {} & \min z_1 + \ldots + z_R \\ & \text{s.t. } z_r \geq \phi_r(p_i, v_r^i) + (p - p_i)^T e_r(v_r^i), \quad \forall r \in R,\ i = 0, \ldots, k, \end{aligned}\right\} \quad (4.7)$$

which is a linear programming problem. Note that the approximation $\bar{\theta}_k(p)$ takes every regional excess e_r as separate subgradient into account, whereas for $\theta(p)$ only the overall excess gives one subgradient.

Setting $\bar{\theta}_k^* := \min_{p \geq 0} \bar{\theta}_k(p)$ (the optimal solution for the approximation with k test points), the difference $\theta_k^* - \bar{\theta}_k^* \geq 0$ is called duality gap. Because $\theta^* \geq \bar{\theta}_k^*$, this difference gives an upper bound on $\theta_k^* - \theta^*$ which we want to make sufficiently small. Of course $\theta_k^* - \bar{\theta}_k^*$ depends on the set of test points; if 'sufficiently many' of them are close to p^*, the approximation quality of $\bar{\theta}_k$ increases and we observe hopefully $\theta_k^* - \bar{\theta}_k^* \searrow 0$.

To make ACCPM applicable the following is assumed:

(i) There are appropriate bounds $0 \leq \underline{B}_p \leq p^* \leq \bar{B}_p$ and $\underline{B}_z \leq z^* \leq \bar{B}_z$.

(ii) $M := \max_{\underline{B}_p \leq p \leq \bar{B}_p} \theta(p) < \infty$.

In (ii), instead of the unknown exact maximum M, any upper bound can be chosen. With $z = (z_1, \ldots, z_R)$ the so called 'set of localization' F_k, $k \geq 0$, is defined to be

$$\begin{aligned} F_k := \{(p, z) \mid {} & \underline{B}_p \leq p \leq \bar{B}_p,\ \underline{B}_z \leq z \leq \bar{B}_z, \\ & \sum_{r \in R} z_r \leq \theta_k^*, & (4.8) \\ & z_r \geq \phi_r(p_i, v_r^i) + (p - p_i)^T e_r(v_r^i) \quad \forall r \in R,\ i = 0, \ldots, k.\}. & (4.9) \end{aligned}$$

Algorithm 6 ACCPM for decomposition

(i) Choose $\varepsilon > 0$ and set $k := 0$, $p_0 := \frac{1}{2}(\bar{B}_p - \underline{B}_p)$, $\theta^*_{-1} := M$.

(ii) For all $r \in R$ compute a primal solution

$$v_r^k \in \operatorname*{argmax}_{v_r \in K_r} \phi_r(p_k, v_r), \tag{4.10}$$

i.e. solve all regional problems ((7.10) in case of MM^{mr}). Add the resulting $|R|$ subgradient cuts to F_k; if $\theta_k < \theta^*_{k-1}$ update also (4.8).

(iii) If $\theta^*_k - \bar{\theta}^*_k \leq \varepsilon$ STOP.

(iv) Set $k := k+1$ and compute the new analytic center (p_k, z_k) of F_{k-1}; return to (ii).

F_k can be written in the form $\{u \mid A_k^T u - s_k = c_k,\, s_k \geq 0\}$ where, from k to $k+1$, the matrix A enlarges maximally by $|R|$ columns and the vectors s_k and c_k are extended appropriately. Based on these preparations the implemented decomposition-ACCPM is described in Algorithm 6.

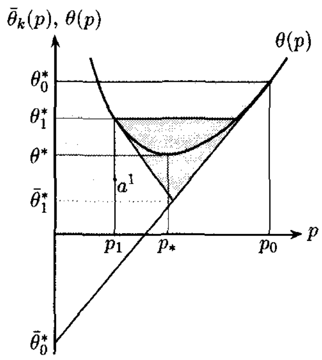


Figure 4.3: Minimizing the Lagrangian function.

Before giving some remarks on Algorithm 6, a simplifying picture is shown in Figure 4.3. Starting at p_0 the subproblems are solved, i.e. $\theta_r(p_0)$ is computed for all $r \in R$; then the linear approximation at $(p_0, \theta(p_0))$ defines $\bar{\theta}_0(p)$. The set of localization F_0 is depicted light-gray shaded; its analytic center a_1 is projected onto the p-axis yielding p_1. In the second iteration the subproblems are solved again with the new price signal p_1 and the resulting subgradient cut at $(p_1, \theta(p_1))$ is inserted in F_0 yielding F_1. Because $\theta(p_1) < \theta^*(p_0) \equiv \theta(p_0)$, the so called 'value cut' (4.8) is updated and F_1 becomes the darker shaded region. $\bar{\theta}_1(p)$ is defined as the maximum of $\bar{\theta}_0(p)$ and the subgradient cut at $(p_1, \theta(p_1))$. In these two iterations the duality gap decreases from $\theta_0^* - \bar{\theta}_0^*$ down to $\theta_1^* - \bar{\theta}_1^*$.

If in every iteration the minimizer of $\bar{\theta}_k(p)$ is taken instead of the analytic center, the method is called 'Kelley's cutting plane method', see Kelley [64], or 'Benders-decomposition', see Benders [8]. An interesting discussion comparing ACCPM with the Kelley decomposition is contained in du Merle, Goffin and Vial [21], where also some hints are given on how to improve the efficiency of ACCPM.

The following remarks are in order:

- The stopping criterion is an 'absolute' duality gap; in theoretical analysis a relative duality gap defined by $(\theta_k^* - \bar{\theta}_k^*)/(M - \theta^*)$ is preferred. In Algorithm 6 this can be attained by dividing ε with $M - \theta^*$ (or an estimate of it). Note also that determining $\bar{\theta}_k^*$ requires solving the LP (4.7), i.e. before the stopping criterion in step (iii) can be examined an LP has to be solved.
- In general the best iterate (the 'solution' returned when stopping in iteration k) is the price p_i where i is the last iteration where a value cut was performed, i.e. the last iteration i where $\theta_i < \theta_{i-1}^*$ holds.
- Practical experience shows that convergence is improved by putting more weight on the value cut compared to the subgradient cuts when computing the thereby weighted analytic center. Going further it can be useful not to check for any redundancy when inserting constraints of the form (4.8) and (4.9), but add in every iteration the full $|R| + 1$ constraints. If identical constraints are inserted several times this increases the weight of these constraints and thereby 'pushes away' the subsequent analytic centers, which is a positive effect in the practical behavior of the algorithm.
- The computation of the analytic center for a given F_k can impose some difficulties despite the fact that it is a smooth, strictly convex minimization problem. The reason lies in its high nonlinearity; to apply ordinary solvers thus requires both a good starting point and a suitable scaling. As starting point we choose the center of the maximal inscribed sphere (a linear programming problem), and scale the problem with the solution at this point. Another scaling aspect is involved with the shrinkage of F_k; after some iterations the radius of the largest inscribed sphere of F_k can drop below the feasibility tolerance of ordinary (linear or nonlinear) solvers which then reject the problem as infeasible. By multiplying each constraint with an appropriate factor $\gg 1$ this problem can be overcome. The factor may differ among the constraints because this does not influence the location of the analytic center (the factors drop out of the logarithm into an additive constant), cf. the discussion concerning implementation in Appendix E.

  Instead of computing the analytic center with an existing solver, dedicated solvers can be advantageous. One code was developed in Geneva (see http://ecolu-info.unige.ch/~logilab/software/accpm.html) and successfully applied to a wide range of problems. Specifically, it can efficiently compute the new analytic center based on the location of the old analytic center if one or several cuts are introduced.

- The dimension of the localization set F_k using this multiple cut scheme is $d + |R|$ where d denotes the number of connecting constraints which are dualized into the objective. If instead of $|R|$ regional subgradient-cuts of the form (4.9) only the overall excess cut from Lemma 4.1 is used, the dimension of F_k is reduced to $d + 1$. To state it differently: the price paid for generating $|R|$ subgradient cuts in each iteration is the increase in dimensionality of the localization set. But usually the gain outperforms clearly this price.

- Experiments on certain test problems encourage the use of Kelley decomposition for small dimensional localization sets, whereas in higher dimensional cases—say more than 20—the analytic center behaves in general better, cf. [21]. MM^{mr} is solved more efficiently using analytic centers than Kelley test points, even though the dimension was typically around 10.

- If o_r and e_r are linear for all $r \in R$, and the extremal points of K_r are bounded (as is the case for polyhedral K_r), then we do not need to require $\max_{\underline{B}_p \leq p \leq \bar{B}_p} \theta(p) < \infty$. In this case constraints of the form (4.6) are inserted into F_k whenever an unbounded subproblem is detected.

- If the original problem is a linear programming problem, there is only a finite number of extremal points and directions. Hence $\bar{\theta}_k(p)$ will eventually coincide with $\theta(p)$ and so the Kelley decomposition terminates with the exact solution. The ACCPM on the other hand does in general not produce an exact solution in a finite number of iterations due to its 'interior nature'.

Während der vielen Jahre beobachtet der Mann den Türhüter fast ununterbrochen, er vergisst die andern Türhüter, und dieser erste scheint ihm das einzige Hindernis für den Eintritt in das Gesetz. Er verflucht den unglücklichen Zufall, in den ersten Jahren rücksichtslos und laut, später, als er alt wird, brummt er nur noch vor sich hin. Er wird kindisch, und, da er in dem jahrelangen Studium des Türhüters auch die Flöhe in seinem Pelzkragen erkannt hat, bittet er auch die Flöhe, ihm zu helfen und den Türhüter umzustimmen. F. K. [58]

Chapter 5

Qualitative Comparison of the Algorithms

In Chapter 3 and 4 we have described three basic algorithms: (i) Algorithm 2 which is a general cutting plane method (CPM) and called ACCPM if the analytic center is used as inner point; (ii) Algorithm 4, a homogenized cutting plane method; and (iii) Algorithm 5 a fixed point heuristic. Here we try to briefly compare qualitatively these three algorithms and to position them with respect to other algorithms from the literature. A more in depth numerical comparison of the three algorithms, together with some variants, can be found in Appendix F.

A survey on the main algorithmic possibilities solving computable general equilibrium problems (CGE) faces the curse of specificity. Due to the hard nature of CGE-problems, algorithms usually explore as much as possible the specific structure of a given equilibrium problem. This leads to an enormous amount of different mathematical formulations and even more of different algorithms. For an incomplete impression the references [47, 46, 100, 33, 9, 20] can serve as a starting point.

Nevertheless, one ordering criterion for this plethora of concepts is simple and at the same time of practical relevance: the level of aggregation. Either an algorithm explores directly the complete structure of the economic agents (disaggregated view), or the agents are treated as 'black-box' oracle where no information about the internal structure is used by the algorithm (aggregated view).

In principle, the additional information available in the disaggregated view should make the first group of algorithms more efficient. Among them a very competitive solver is PATH[1], where the problem is formulated as a Mixed Complementarity Problem. At its heart PATH performs Newton-Iterations, and because the Newton-concept is the key-technique in most other algorithms working in a disaggregated setting, we have chosen PATH as representative for the whole group.

[1]Developed by Dirkse and Ferris [20], the code is commercially distributed as a solver with GAMS.

A central limitation of Newton-based methods is, however, that some convexity properties must be given. Otherwise, fixed point homotopal approaches can still be applied, where the only requirement is continuity of the functions involved. A prominent example of a homotopal algorithm is OCTASOLV [9]. As a consequence, the linear, super-linear or even locally quadratic convergence of Newton-methods is in case of fixed point approaches replaced by simple global convergence without polynomial complexity bounds.

Concerning algorithms working in the aggregated setting, a first question could be, why one should ignore information which is—in principle—always available. A first theoretical motivation comes from the level of abstraction obtained, allowing an application of oracle-based methods to a broad and divers set of equilibrium problems. More decisive, however, can be the practical side of the problem. The requirement to have all information, including derivatives, available implies a formulation of the equilibrium problem in a specific modeling and solving environment. If sophisticated models of agents are already existing, the effort to transform them into such a specific environment can be intractable.

On top of this, an aggregated view facilitates also a kind of 'sensitivity' analysis on the level of agents. The effort to replace an agent-model by another model, to include more agents or drop some agents can be kept small.

Besides Algorithm 2, 4 and 5 described in this work, we include in Table 5.1 another aggregated-view algorithm due to Taheri, Maxfield and Luenberger [100], which applies the homotopal solver OCTASOLV to a cleverly designed surplus function.

Given the necessary structural assumptions are fulfilled, Algorithm 2 and 4 are pseudo-polynomial only, whereas Newton-based methods working in the disaggregated ('d.-a.') setting are polynomial. Furthermore, the speed of convergence

view		**method**	**assumption**	**measure**	**convergence**
d.-a.		Newton-Methods (PATH [20])	convex	merit function	global, locally superlinear
aggregated view	cutting plane	ACCPM (Algorithm 2)	pseudo-monotone	$\|\|e(p)\|\|$	provably not given
			pseudo co-coercive	g_p	pseudo-polynomial
		conic ACCPM (Algorithm 4)	monotone	g_d	pseudo-polynomial
	fixed point	Negishi (Algorithm 5)	—	$\|\|be(\alpha)\|\|$	heuristic, empirically linear
		Homotopy ([100])	continuous	utility or price deviation	globally convergent

Table 5.1: Characteristics of different algorithmic concepts.

is significantly higher in the latter case.

After this comparison with approaches from the literature, the rest of the chapter concentrates on a qualitative comparison of the Algorithms 2, 4 and 5 discussed in this work. More details on the implementation and resulting numbers are reported in Appendix E and F.

The empirical behavior of all three algorithms is strongly influenced by the stopping criterion and the starting feasible set. The starting point is less important but can have a minor influence in some cases. Let us comment the stopping criterion first.

In Algorithm 2 we have chosen $||e(p)||$ and not, say, $|p^T e(p)|$, $g_p(p)$ or $g_d(p)$, because the norm of the excess has a direct economic interpretation as opposed to the other measures. Furthermore, $e(\lambda p) = e(p)$ for all $\lambda > 0$ and so neither g_p, g_d nor the complementarity product $p^T e(p)$ have an absolute meaning, making the interpretation of an ε-solution difficult. The choice of $||e(p)||$ requires, however, that $e(p^*) = 0$ at equilibrium prices p^*, a condition which is satisfied by all our test-problems.

In Algorithm 4 g_d is usually not computable. Nevertheless, either g_d can be bounded by g_p using monotonicity, or the norm of the excess $||e(p)||$ is used again. In the implementation of both Algorithm 2 and 4 we used simultaneously $||e(p)||$ and the number of iterations as stopping criterion.

In Algorithm 5 the notion $e(p)$ is replaced by the budget excess vector $be(\alpha)$ which appears in the space of agents. Hence, we stop if either $||be(\alpha)||$ is sufficiently small, or the number of iterations reaches a given bound.

Next, concerning the starting feasible set, the user can exploit his *a priori* knowledge about the location of the equilibrium price by choosing a tight price-set. In case of Algorithms 2 and 4 this is straightforward. In the Negishi-Algorithm this knowledge can be used in the decomposition-machinery of Algorithm 6. Because the time spent in the decomposition routine strongly dominates the overall computing time, the restriction of the price-set influences directly the total running time. One should be aware, however, that only in an equilibrium α^* the solution p^* of the decomposition coincides (up to scaling) with the equilibrium price. Or to put it the other way round: For $\alpha \neq \alpha^*$ a tight restriction of the feasible price-set around the true equilibrium price may exclude the actual solution of the decomposition. In case of our examples, though, we observed a remarkable robustness in that the solution price of the decomposition was always very close to the final equilibrium price, i.e. almost independent of the weight vector α.

The examples we tested numerically are described in some detail in Appendix F. Basically we used different variants of the energy-economy model MM^{mr}, both with respect to data and model-structure. Additionally, a set of simple agents defined by non-linear optimization problems was investigated.

Based on these examples, let us state some general conclusions, cf. Appendix F.6. Most importantly, Algorithm 2 and 5 including all variants always found the same solution. Specifically, the starting point had no noticeable influence on the final iterate. In a few cases Algorithm 2 got stuck near the equilibrium solution, in most cases, however, Algorithm 2 was surprisingly robust and reliable. Compared to Algorithm 2 Algorithm 4 reduced $||e(p)||$ quicker—in the first 30–40 iterations, but then converged to a non-equilibrium price. This phenomenon is probably due to non-monotonicity of $e(p)$, but also the limited accuracy of computing the analytic center or the excess can be responsible. In view of these experiences it can be reasonable to combine Algorithm 4 in a first stage with Algorithm 2 in a second stage. Furthermore, Algorithm 4 is dropped from the remaining discussion.

Concerning Algorithm 5 we observed empirically a linear convergence of $||be||$ as a consequence of a contractive fixed point map. This interesting result deserves future attention. Usually 5–6 Negishi-iterations are sufficient to obtain a good solution from an arbitrary starting point. Only the limited accuracy in the decomposition routine limits the attainable quality of solution.

The choice of the feasible set influences strongly the CPU-time of the algorithms, aggravating thereby a comparison of Algorithm 2 and 5. Moreover, a sound recommendation on which algorithm to use should rely on a complexity-analysis, which—in view of the structural deficiencies of $e(p)$—seems difficult. Nevertheless, based on the discussion in Chapter 3 and 4 and empirical findings in our test problems the following considerations can be of interest.

To begin with let us assume that the number of Negishi-iterations in Algorithm 5 depends only 'little' on the number of agents. Then the computational burden of Algorithm 5 is essentially determined by the decomposition subroutine which has a pseudo-polynomial complexity bound. Looking at Algorithm 2 using analytic centers the situation seems ambivalent: on the one hand we can not give a (pseudo-)polynomial complexity bound, on the other it outperformed Algorithm 5 by a factor of 2–4 on our test-problems. The test-problems included 3 agents and 6–16 goods. In the sequel we try to raise evidence why Algorithm 2 might stay superior when $|G|$, the number of goods, increases.

Let us assume that D^k, the feasible set in iteration k of Algorithm 2, stays sufficiently fat, i.e. the ratio of the largest inscribed sphere over the smallest circumscribed sphere stays sufficiently far away from 0. Moreover, based on our test-problems we assume an average volume reduction of 50 % per iteration and introduce some appropriate constants c, c' and c''. Therefore $\mathrm{vol}(D^k) \approx \mathrm{vol}(D^0)2^{-k} \approx c[\delta(D^k)]^d$, where $\delta(D^k)$ denotes the diameter of D^k, and hence $\delta(D^k) \approx c'2^{-k/d}$. If we assume $e(p)$ is Lipschitzian continuous with constant L we can further deduce $||e(p^*) - e(p^k)|| \lessapprox L||p^* - p^k|| \lessapprox c''L2^{-k/d}$. Finally, with the stopping criterion $||e(p^*) - e(p^k)|| \leq \varepsilon$ we find a polynomial complexity bound $k \leq p(\log \frac{1}{\varepsilon}, \dots)$. In such an ideal situation Algorithm 2 clearly outperforms Algorithm 5 which has a pseudo-polynomial complexity bound only. In case of our test-problems the superiority of Algorithm 2 over Algorithm 5 might therefore be even accentuated when $|G|$ increases.

Schliesslich wird sein Augenlicht schwach, und er weiss nicht, ob es ihm dunkler wird, oder ob ihn nur seine Augen täuschen. Wohl aber erkennt er jetzt im Dunkel einen Glanz, der unverlöschlich aus der Türe der Gesetzes bricht. Nun lebt er nicht mehr lange. F. K. [58]

Chapter 6

Economic Aspects of CO_2 Permits

In the last 20 years the scientific valuation of greenhouse gas (GHG)-induced temperature rise has become quite uniform: a doubling of the CO_2 concentration in the atmosphere by 2050 will rise the global mean temperature by about 2^0 C, and thereby is predicted to cause damage in the range of one to a few percent of GDP for developed countries, and several times more for developing countries (see IPCC [54]). Estimates of the marginal damage of CO_2 emission range between US$\$_{1990}$ 5 and US$\$_{1990}$ 125 per ton of carbon emitted now, where US$\$_{1990}$ denotes US dollars in 1990. As an example, recent estimations of damage costs for Switzerland (see Meier [78]) predict about US$\$_{1997}$ $2 \cdot 10^9$ or almost 1 % of GDP; major effects are expected in tourism and direct weather induced damage like floods.

The wide acceptance of this standpoint in the international scientific community is rather new. The discussion can be traced back more than 200 years; but only since the publication of Arrhenius [1] in 1896 the role of CO_2 became prominent. Until about 50 years ago most authors acknowledged positively the potential warming. For example, the title of a booklet from 1919 promises "Es winken Palmenhaine von Berlin bis Stuttgart" (Palm groves wave from Berlin to Stuttgart), and Callendar [11] postulates with respect to climate effects "the consumption of fossil fuels will probably prove useful for humanity".

Although there are indeed positive effects for an increased global mean temperature, scientists today are more aware of potential damage caused by floods, storms, global destabilization of ecosystems, rising sea level, and loss of human life, as well as shifting climate belts which bring about droughts in fertile areas and increased precipitation on waste land. Despite the detailed knowledge of *potential* damage based on climate scenarios, there is still a large uncertainty on what the real local climate will actually be all over the world.

Thus, the economic estimation of damage costs is difficult, explaining partly why the focus of many economists lies more on the other side of the coin, the estimation of CO_2 abatement costs, which is more tractable. This focus, however, may introduce bias into the damage/abatement equation. A discussion of the present estimation of abatement costs based on energy-economy models can be found in Ekins [26] and is summarized in Section 6.2.

Besides the critical discussion of existing literature, a main contribution of this chapter include the section 'Initial distribution of permits', where the burden imposed by CO_2 abatement is compared between developing and developed countries, which, as we will see, can be understood as sharing the burden between sellers and buyers of permits.

6.1 Emission Permits

In order to understand the meaning of emission permits consider a group of countries which agree on the maximal amount of combined emissions for a given future period t. This combined amount is distributed among the countries as the initial endowment of permits. Then, at the end of period t, the actual amount of CO_2 emission, cumulated over t, must be equal to or less than the amount of permits it owns. If the initial endowment exceeds actual CO_2 emission the country can sell the excess permits to other countries, while in the opposite case the country must buy the necessary amount of permits from other countries.

Emission permits have attained a high degree of respect not only in environmental economy, but also in politics, where the US-administration—not necessarily known for an active strategy for CO_2 abatement—recently acknowledged CO_2 emission permits as a tool against climate change (see Cushman [15]). Reasons for this situation are a number of advantages of permits over alternative instruments like taxes or regulative laws. Specifically, permits are claimed to have a high ecological aptness, high economic efficiency[1], constitute a dynamic incentive for development and implementation of improved technologies (see Schubert [96]), and can be implemented with little administrative cost. In contrast to these advantages of permits financial instruments like taxes have only an indirect effect on the amount of emission and, furthermore, are in practice usually differing between the countries and therefore economically inefficient. Country-wise regulative laws can directly control the amount, but are not only often tedious to implement, but suffer in general from different marginal costs of abatement and are consequently also not efficient economically.

[1] A state is economic efficient if it can not be obtained with less cost. In our situation this is characterized by equal marginal cost of CO_2 abatement throughout all countries. To see this assume two countries with marginal costs $\delta_1 < \delta_2$ and define $\Delta := \delta_2 - \delta_1 > 0$. Then a profit of Δ can be realized by allowing region 2 to emit one unit more (costs: $-\delta_2$), and at the same time by reducing the emissions of region 1 by one unit (costs: $+\delta_1$). Assuming a liberal market economy where world market prices coincide with domestic prices, emission permits imply equal marginal abatement cost in all countries considered.

Nevertheless, CO_2 emission permits also have their shortcomings and give rise to a number of questions which are commented below: (i) what is the level of 'optimal' CO_2 concentration in the atmosphere, (ii) how shall they be distributed initially, (iii) how are the permits implemented, (iv) under what conditions can the potential advantage be realized, etcetera. In the recent polit-economic discussion the distributional effect and the level of 'optimal' emissions are most prominent.

'Optimal' Level of CO_2 Emission

Often the notion 'optimal' is interpreted only economically, designating a state of maximal net profit or welfare, whereas other branches of science may interpret it differently. Hence opinions about the optimal level of CO_2 concentration in the atmosphere are very different, ranging from todays level up to an arbitrarily high level.

Physicists and other natural scientists, being aware of the high non-linearity and uncertainty in the reaction of global climate on a rise of CO_2 concentration, *tend to hold the former, conservative position.* Indeed, IPCC [54] highlights that "There are many uncertainties and many factors currently limit our ability to project and detect future climate change. Future unexpected, large and rapid climate system changes (as have occurred in the past) are, by their nature, difficult to predict. This implies that future climate changes may also involve 'surprises'. In particular, these arise from the nonlinear nature of the climate system. When rapidly forced, nonlinear systems are especially subject to unexpected behavior." As an example of such recent 'surprises' GRIP [43] indicates that abrupt temperature changes with a magnitude of 6–14^0 C within only 30–60 years happened in Greenland. Such a change is definitely out of imagination and beyond the results of larger international models which are usually 'tuned' to show a 'reasonable' behavior, i.e. what modelers expect.

The latter position—no reduction of CO_2 emission—is closer to some economists standpoint being aware of the uncertainty in the estimation of the damage costs on the one hand, and tending to expect very high abatement costs on the other hand. If additionally a high discount rate (say $> 5\,\%$ p.a.) is assumed it becomes prohibitively expensive to invest today in CO_2 abatement, if its 'return', the avoidance of damage, happens considerably later in time. A more thorough discussion on how abatement costs are estimated by todays models can be found in Ekins [26] and will be taken up in Section 6.2.

Initial Distribution of Permits

The initial endowment with permits is intimately related to the global distributional effect of CO_2 abatement, which is of decisive importance for the political survival of this concept. Because decision makers want to know the consequences of positions beforehand in order to negotiate optimally, the crucial role of good studies which possibly rely on the kind of models presented in this work is obvious. If information about the consequences for all the participating countries is openly available, rational negotiation is strengthened.

Major open questions concern the participation of lower developed countries (LDC) and the endowment with permits over time among LDC and DC (developed countries). Facing future demographic and economic trends and potential leakage effects, it is compulsory on the long run to integrate all major emitters in such a permit community to make it work effectively. Presently, typical simplified positions are the non-willingness of LDC to accept a reduction of their economic development as a consequence of CO_2 abatement, whereas DC tend to show to some extend a willingness to pay. One possibility to overcome this problem is to make the LDC profit through a sufficiently large initial endowment which then can be sold on the permit market.

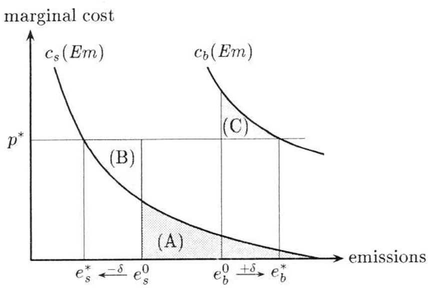


Figure 6.1: Profits from trade and distributional effects due to CO_2 permits.

Assume a representative developed country has a marginal CO_2 abatement cost curve $c_b(Em)$ as a function of the emission Em and where the subscript 'b' stands for buyer. Similarly a developing country is assumed to exhibit the characteristic $c_s(Em)$ ('seller'), cf. Figure 6.1. Here the notion buyer and seller generalize the notion DC and LDC respectively. Assume further the initial endowment is e_b^0 and e_s^0 respectively, and the equilibrium price is p^* at equilibrium emission e_b^* and e_s^* respectively. Let us analyze the situation of the seller first; the cost of participating in the permit community versus doing nothing is $\int_{e_s^0}^{\infty} c_s(Em)\, dEm = $ (A) (dark gray shaded). By moving from the initial endowment e_s^0 to the equilibrium e_s^*, the seller profits by $\int_{e_s^*}^{e_s^0} [p^* - c_s(Em)]\, dEm = $ (B) (light gray shaded). The difference

$$\int_{e_s^*}^{e_s^0} [p^* - c_s(Em)]\, dEm - \int_{e_s^0}^{\infty} c_s(Em)\, dEm \;=\; (B) - (A)$$

is thus the direct net profit of CO_2 abatement for the seller. Obviously, by shifting initial endowment from the buyer to the seller, i.e. $e_s^0 \leftarrow e_s^0 + \Delta$ and $e_b^0 \leftarrow e_b^0 - \Delta$, the net profit for the seller increases, and can even become positive in some situations, such as not too restrictive global emission bounds and a sufficient decrease of marginal cost $c_s(Em)$ in the range $[e_s^*, e_s^0]$.

Besides this 'buying the seller' aspect, Figure 6.1 shows the profit for both the buyer and the seller achieved by trading visavis obeying the single country emission bounds e_s^0 and e_b^0 respectively. In such a comparison 'trade versus non-trade', both the buyer and the seller profit by the light gray shaded area (B) and (C) respectively. The evaluation of such differing regional profits could guide negotiations for distributing the initial endowment. Besides such direct profits, a major indirect profit is to come from a reduction of the damage costs mentioned in the introduction. Furthermore, it can be expected that a reduction of CO_2 emission causes also other indirect profits called 'secondary benefits': reduced emissions of other pollutants, increased efficiency of the overall economy, improved quality due to increased process control, etcetera.

Aspects in the Implementation of Permits

Among the numerous aspects we like to comment on very few only. First some of the different possible time-structures of permits are discussed in Section 7.1.2. Next, the context of international permits rises another level of complexity by the question on how the local implementation happens, i.e. how the local agents (single consumers or firms) are taken into account. Possibilities comprise national taxes on fossil fuels, a system of domestic permits taking up the international permits, etcetera. As a non-regional possibility one can think of imposing a permit scheme on the producers of fossil fuels, theoretically bypassing all political obstacles. Or more regionally oriented, in case of crude oil the refineries could constitute the agents in the permit system.

Conditions for the Functioning of the Permit Market

Neglecting almost everything which can be found in the standard economic literature, we want to point to some specificities of $\mathrm{MM}^{\mathrm{mr}}$. First note that the regions behave myopically; they take the price signal as given and maximize their utility. By contrast, regarded as a player in an $|R|$-person game, regions might well choose a strategy where parts of the goods are held back and thereby the prices influenced (cf. Gabszewicz and Vial [32]).[2] Such oligopolistic strategies are tempting above all for larger agents owning a significant part of the total permits and producing much of the (numéraire) goods. Focusing on the market of permits, it is likely that such a behavior can be detected by the other agents, because reliable information about regional emissions are available from several independent sources. It is then up to the interregional community to establish clear rules in the beginning and consequently apply them accordingly.

[2]It is interesting that in one of the few existing real world implementations of emission permits—sulphur rights in the US—such a behavior was indeed postulated based on empirical findings (see Murphy, Sanders and Shaw [80]); but the explanation there did not assume a direct profit-increasing strategic behavior of the agents via increased prices, instead a certain conservative attitude was found where emitters were afraid of not having sufficient permits when needed. This risk-averseness obstructs the functioning of the permit market especially in the case of a small number of agents.

Similarly, the case of hiding emissions by systematically falsifying regional statistics or other cheating possibilities must be properly taken into account.

And finally, as mentioned above, a market of emission permits should comprise all major emitters, otherwise leakage effects can affect the whole economy, ruining part of the economy of the permit-trading regions on the one hand, and missing the global CO_2 emission targets on the other (cf. Manne and Rutherford [75]). It is unclear, however, to what extend such leakages could negatively affect regional economies in reality. Not only a large part of emission activities is locally fixed (heating/cooling, transportation, service sector, etcetera), but also in the industrial production sector energy costs are normally making up only a small part of the overall production costs. In the remaining cases like cement or heavy steel and aluminum, a permit induced tax-regime can be imposed in case of imports, and reversely permit induced costs can be paid back if such goods are exported. Furthermore, it is even questionable if uncorrected leakages are harmful on the long run for regions in the permit-community, because permits give anyway an incentive for the economy to shift from industrial energy-intensive products to higher level service-sector products.

The situation is comparable to countries being rich in resources vis-à-vis countries endowed with little natural resources. The former will tend to stay in a resource depletion oriented structure bringing only modest wealth, whereas economies of countries lacking such resources tend to increasingly develop value-added products and services, resulting finally in wealthier economies. This may be accepted as an empirical finding, but moreover it can be an argument against model-results, where typically the economic structure modeled is limited in anticipating such long running sectoral shifts.

6.2 Energy Economy Models

An economically optimal behavior attempts to equate the marginal damage costs of GHG emissions to the marginal cost of reducing them, and thereby prevents damage. The estimation of both cost curves—regionally differentiated or globally aggregated—are subject to a wide margin of uncertainty. But whereas an estimation of mitigation costs is calculated within the human economic system and is therefore more open to an economic analysis, the factors determining damage costs are beyond human control, and therefore aggravate significantly their analysis. This might explain why so much research concentrates on mitigation costs while only little effort is put into the estimation of damage costs. As a consequence of the uncertainty involved in estimating mitigation costs, a number of papers appeared in the last 10 years comparing and analyzing different models (see [26, 17, 34]). Not surprisingly, the assumptions on which the models are constructed have a decisive influence on the outcomes.

To start with, the base line scenario assumptions like rate of economic growth, levels of discount rate or demographic trends greatly affect the resulting abatement costs. Other influential assumptions include the value of land, the value of human life, agricultural losses, macroeconomic assumptions like substitution parameters, technological assumptions like backstop technologies or the aeei[3] factor, use of unemployed resources, revenue recycling of potential carbon taxes, existing distortions in the tax regime, how carbon taxes are set up ('double dividend'[4]), and considerations stemming from secondary benefits.

Surprisingly enough, most models show multiple severe shortcomings in these respects. For example, Pearce [92, Note 6 p. 940] expresses surprise "that most of the simulations of hypothetical carbon taxes do not consider revenue neutrality." Nordhaus [90, p. 317] states, "The importance of revenue recycling is surprising and striking. These findings emphasize the critical nature of designing the instruments and use of revenues in a careful manner. The tail of revenue recycling would seem to wag the dog of climate change policy." Finally, Gaskins and Weyant [34, p. 320] confirm the importance of revenue recycling: "Simulations with four models of the US economy indicate that from 35 % to more than 100 % of the GDP losses could ultimately be offset by recycling revenues through cuts in existing taxes". Concerning MM^{rw}, it has no explicit carbon tax, but the resulting equilibrium can equally be interpreted as a result of a tax regime with *neutral* revenue recycling.

Another influential parameter is the aeei factor; while the values in the models range from 0–1 %, Dean and Hoeller [17] in their comparative study of the six main global models note: "A difference of 0.5 % in this parameter, given compounding, can lead to an outcome in 2100 which is as much as 20 billion tons of carbon emissions different." Ekins [26] comments that "differences in baseline emissions of this magnitude would greatly affect the cost of reducing these emissions to any particular level."

Finally, concerning secondary benefits, Ekins [26, p. 261] reports that its importance in a benefit cost analysis has been recognized by many analysts of global warming, and continues: "This makes it the more surprising that neither of the two main cost-benefit analysis of global warming to appear to date make any attempt to incorporate into their assessment, even tentatively, the various estimates of secondary benefits that have so far been made."

Ekins [26, p. 271] concludes that "implementing a carbon tax sensitively with regard to issues such as these could partially or totally offset the negative economic effects deriving from increasing the price of energy." As important conditions he identifies on the one hand the gradual imposition, and on the other hand the likewise reduction of other taxes to keep the fiscal package invariant.

[3] Autonomous energy efficiency improvement; it describes the price-independent technological increase of energy efficiency over time, i.e. the increase of output over energy input.

[4] Double dividend designates the double gain from internalizing the externalities due to free CO_2 emissions on the one hand, and the gain from a reduction of distortions in the existing tax system due to the introduction of a carbon tax with a concurrent reduction of distortional taxes on the other hand.

> Vor seinem Tode sammeln sich in seinem Kopfe alle Erfahrungen der ganzen Zeit zu einer Frage, die er bisher an den Türhüter noch nicht gestellt hat. Er winkt ihm zu, da er seinen erstarrenden Körper nicht mehr aufrichten kann. Der Türhüter muss sich tief zu ihm hinunterneigen, denn der Grössenunterschied hat sich sehr zuungunsten des Mannes verändert. F. K. [58]

Chapter 7

Modeling Trade of CO_2 Permits: The Model Markal-Macro Multi-Region MM^{mr}

When 1973 the oil crisis shook the economy of almost the whole world, the energy market attracted the attention from many researchers. In the following years two institutions started to build large energy models: IIASA[1] and IEA[2]. In 1974 Häfele and Manne built a first model for IIASA, from which 1974 ETA-Macro [72] and 1981 'the' IIASA-model [45] were derived. IEA on the other hand initiated *Markal with its variations (see below) and the IEA-ORAU model [24, 25]*. A third important contribution came from Nordhaus (Yale-University), who built in the late 70's a linear energy-optimizing model [88, 89]. From ETA-Macro Manne derived Markal-Macro, Global2100 and 12RT[3].

Markal and Markal-Macro are insofar outstanding, in that they are based from the beginning on an international collaboration. As a consequence, Markal and Markal-Macro are implemented and used in more than two dozen countries all over the world. This results in a reliable and well tested model code, data availability for many different countries, and ongoing discussions and improvements of the models. A survey on both Markal and Markal-Macro can be found in Appendix C.

The chapter is organized as follows. Section 7.1 presents the extensions needed for interregional trade of CO_2 permits. Here different modeling approaches for permits are discussed and formally analyzed, and the role of the numéraire price is investigated. Section 7.1 is written from an economical standpoint focusing on a single region. In Section 7.2 this single region view is broadened to encompass the equilibrium problem posed by the set of all regions; two formulations are

[1] International Institute for Applied System Analysis

[2] International Energy Agency

[3] 12 region trade: this model deals especially with world trade and its effects on CO_2 strategies.

given: first the VIP-formulation (VIP) from page 7 and second the fixed point problem due to Negishi discussed in Section 2.2.

Contributions comprise the whole chapter.

7.1 MM^{mr}: A Multi-Regional Version of Markal-Macro Including Trade of CO_2-Permits

Each regional Markal-Macro model can be understood as partial equilibrium model in the energy sector, and the result represents a regional equilibrium, cf. Appendix C. Here we want to extend the regional models by coupling them in an interregional (international if the regions coincide with nations) model where in principle an arbitrary set of goods can be traded on a common market, see Figure 7.1. In the concrete present formulation, however, only two goods are traded in each time period (i.e. due to the dynamic nature of MM there are in fact $2|T|$ goods, with T the set of time periods). On the one hand we have CO_2 emission permits, often simply called permits, and on the other hand the aggregated good Y_t which is the numéraire and closes trade. A region in the resulting equilibrium model MM^{mr} equals the 'utility maximizing agent' (1.6).

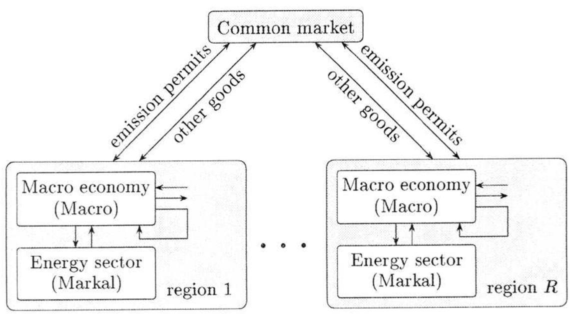


Figure 7.1: The multiregional Markal-Macro MM^{mr}.

7.1.1 Changes in the Model

In the multiregional framework every quantity needs an additional regional index r, which is, however, dropped in case of coefficients for the sake of notational convenience. Based on the notation from Appendix C, the following will be used in the sequel:

T set of periods (time horizon);

R	set of regions;
U_r	utility of region r;
$C_{r,t}$	consumption of region r in period t;
$Y_{r,t}$	production of region r in period t;
$I_{r,t}$	investment of region r in period t;
$EC_{r,t}$	energy cost of region r in period t;
$K_{r,t}$	capital of region r in period t;
$L_{r,t}$	labour of region r in period t;
$D_{r,t}$	demand of (aggregate of) energy services ... ;
$IECO2_r \in \mathbb{R}^{\|T\|}$	initial endowments of CO_2 permits of region r;
$NTXCO2_r \in \mathbb{R}^{\|T\|}$	net export (exports minus imports) of permits of region r;
$NTX_r \in \mathbb{R}^{\|T\|}$	net export of Y_r of region r;
$p_{CO2} \in \mathbb{R}_+^{\|T\|}$	price of CO_2 permits;
$p_{NTX} \in \mathbb{R}_+^{\|T\|}$	price of Y_r;
$e \in \mathbb{R}^{2\|T\|}$	excess $= (\sum_r NTX_r, \sum_r NTXCO2_r)$.

The changes of the original Macro-model comprise the following constraints which are distinguishable from the original MM-constraints by the superscript ^{mr} (multi regional). First, the use of production is corrected by the net trade,[4]

$$Y_{r,t} = C_{r,t} + I_{r,t} + EC_{r,t} + NTX_{r,t}. \qquad (\mathrm{USE}_{r,t}^{mr})$$

Second, the CO_2 emissions allowed are restricted by the initial endowment minus the net selling of permits,

$$Em_{r,t} \leq IECO2_{r,t} - NTXCO2_{r,t}. \qquad (\mathrm{EC}_{r,t}^{mr})$$

Finally, the overall budget (trade balance) must have no deficit:

$$0 \leq p_{NTX}^T NTX_r + p_{CO2}^T NTXCO2_r. \qquad (\mathrm{BC}_r^{mr})$$

All other Markal-Macro constraints in (C.5) are left unchanged except for the addition of a regional index r. In view of (1.6) and (C.5) the resulting regional problems of MM^{mr} can be written as follows:

$$\left.\begin{array}{ll} \max & U_r(C_r) \\ \text{s.t.} & (C.1)\text{–}(C.4),\ (\mathrm{PRD}_{r,t}),\ (\mathrm{L}_{r,t+1}),\ (\mathrm{CAP}_{r,t+1}),\ (\mathrm{TC}_r), \\ & (\mathrm{USE}_{r,t}^{mr}),\ (\mathrm{EC}_{r,t}^{mr}),\ (\mathrm{BC}_r^{mr}), \\ & \text{(all other Markal-Macro constraints).} \end{array}\right\} \qquad (7.1)$$

[4]As for notation we use the gams-related convention that an equation has to be repeated for each meaningful occurrence of indices which are given in the tag. As an example the tag $(\mathrm{USE}_{r,t}^{mr})$ indicates that the corresponding equation is repeated for all $t \in T$ and $r \in R$.

Important economic results of the models are the GDP (gross domestic product) and the GNP (gross national product). Based on the original MM-definition of production ($\mathrm{PRD}_{r,t}$) we have for the stand-alone MM-model

$$\mathrm{GDP}_t \;:=\; C_t + I_t \;=\; Y_t - EC_t.$$

Because the view of Macro on the energy sector does not distinguish between the value added and the cost of inputs, the interpretation of the GDP requires extra caution. One of the possibilities is to treat the total cost EC_t as input cost, and interpret the part $\sum_d b_d \, D^{\rho}_{r,d,t}$ in the production function related to energy services as total output.

In the multiregional setting of MM^{mr} the GDP takes also foreign trade into account and is defined by

$$\mathrm{GDP}^{mr}_{r,t} \;:=\; C_{r,t} + I_{r,t} + NTX_{r,t} \;=\; Y_{r,t} - EC_{r,t}.$$

If the capital-exchange is further taken into account we arrive at the GNP:

$$\mathrm{GNP}^{mr}_{r,t} \;:=\; \mathrm{GDP}^{mr}_{r,t} + \frac{p_{CO2,t}}{p_{NTX,t}} NTXCO2_{r,t} \;=\; Y_{r,t} - EC_{r,t} + \frac{p_{CO2,t}}{p_{NTX,t}} NTXCO2_{r,t}.$$

In all these relations $Y_{r,t}$ is defined by the CES-production function in (PRD_t) page 117, extended by a regional index r. Some authors prefer to call this specific GNP 'green *national product*' because *the capital-flow is based on* 'artificial' rights at the nature and not on 'proper' goods, labor or services.

Notational Generalization

The extensions presented above are specific for trade of CO_2 permits. It is obvious, however, that MM^{mr} can be generalized to model trade of a larger set of goods. Assume $G = \{0, 1, \ldots, |G| - 1\}$ is the index set of traded goods without subdivision by periods, and with the convention that the index 0 represents the numéraire good. In our case good 0 is NTX and good 1 represents the CO_2 permits. Denote by $x^0_{g,r,t}$ the initial endowment of good g in region r and period t, and by $x_{g,r,t}$ the corresponding excess ('export minus import' in the context of geographical units). The formulation of our previous model MM^{mr} can then be expressed by $x^0_{0,r,t} = 0$,

$$Y_{r,t} \;=\; C_{r,t} + I_{r,t} + EC_{r,t} + x_{0,r,t}. \qquad (\mathrm{USE}^{mr}_{r,t})$$

$$Em_{r,t} \;\leq\; x^0_{1,r,t} - x_{1,r,t}, \qquad (\mathrm{EC}^{mr}_{r,t})$$

and

$$0 \;\leq\; \sum_{g \in G} p^T_g x_{g,r}. \qquad (\mathrm{BC}^{mr}_{r})$$

To incorporate more traded goods requires simply to explicit the excess based on the corresponding constraints in the underlying model and to extend the set of goods in the budget constraint.

7.1.2 Discussion of Some Aspects of MM^{mr}

The Numéraire-Price and the Budget Constraint

Note that the price vector can be scaled by any positive scalar without affecting the primal solution of any MM-based agent. (The dual multiplier of the budget constraint, however, is linearly influenced by such a scaling.) A single value of a price component $p_{g,t}$ is therefore without information; here the function of the numéraire comes in which is defined as a good representing the monetary unit of an economy. That is, the price of the numéraire good $x_{0,r,t}$ in the local currency of region r is known (fixed exogenously), and usually taken to be one, which can always be achieved by adjusting the unit-amount. Based on that the 'real' prices of the goods are $\frac{p_{g,t}}{p_{0,t}}$ for all $g \neq 0$ and all t. In case of MM all prices are undiscounted and therefore the fraction $\frac{p_{g,t}}{p_{0,t}}$ is also undiscounted. How discounting of equilibrium prices should be done will be discussed below. But first the economic interpretation of the budget constraint deserves some clarification. Because the discussion focuses on one region, the index r will be suppressed. Consider in a first step a 'per period' budget constraint of the form

$$p_{0,t}x_{0,t} + p_{1,t}x_{1,t} \geq 0 \qquad \forall t \in T;$$

in the model-related undiscounted 'real' prices it can equivalently be written as

$$x_{0,t} + \frac{p_{1,t}}{p_{0,t}}x_{1,t} \geq 0 \qquad \forall t \in T.$$

In our case, though, we have chosen a scalar budget constraint which balances the monetary exchange over all time periods. An economically consistent approach to derive this takes the budget excess of each time period, multiplies it by an appropriate discount factor a_t, and balances it over the whole time horizon:

$$\sum_{t\in T} a_t \left(x_{0,t} + \frac{p_{1,t}}{p_{0,t}}x_{1,t} \right) \geq 0. \tag{7.2}$$

It is proven in the next lemma that we regain our model formulation

$$\sum_{t\in T}(p_{0,t}x_{0,t} + p_{1,t}x_{1,t}) \geq 0 \tag{7.3}$$

if and only if $a_t = \lambda p_{0,t}$ $\forall t \in T$ and some $\lambda > 0$.

Lemma 7.1 *There exists $\lambda > 0$ such that $a_t = \lambda p_{0,t}$ $\forall t \in T$ if and only if*

$$\left\{ x \,:\, \sum_{t\in T} a_t \left(x_{0,t} + \tfrac{p_{1,t}}{p_{0,t}}x_{1,t} \right) \geq 0 \right\} = \left\{ x \,:\, \sum_{t\in T} p_{0,t}x_{0,t} + p_{1,t}x_{1,t} \geq 0 \right\}.$$

The sets—a halfspace whose defining hyperplane contains the origin—are equal if and only if the corresponding normal vectors

$$\begin{aligned} n^1 &:= \left(a_1, \ldots, a_T, \tfrac{a_1}{p_{0,1}} p_{1,1}, \ldots, \tfrac{a_T}{p_{0,T}} p_{1,T}\right) \quad \text{and} \\ n^2 &:= (p_{0,1}, \ldots, p_{0,T}, p_{1,1}, \ldots, p_{1,T}) \end{aligned}$$

differ by a positive scalar only. We then have from the first half of the vector equality $n^1 = \lambda n^2$ the relation

$$a_t = \lambda p_{0,t} \quad \forall t \in T,$$

and in the second half $\frac{a_t}{p_{0,t}} p_{1,t} = \lambda p_{1,t}$, which, by the previous relation $a_t = \lambda p_{0,t}$, gives the trivial identity $\lambda p_{1,t} = \lambda p_{1,t}$, and so we have verified $n^1 = \lambda n^2$.

Economically interpreted the (endogenously determined) price for the numéraire p_0 is *exactly proportional* to the discount factor a_t in the budget constraint (7.2), that is, we have

$$\frac{a_t}{a_{t+1}} = \frac{p_{0,t}}{p_{0,t+1}} \qquad t \in \{1, \ldots, T-1\}.$$

Hence the equilibrium discount factor is given by $a_t = p_{0,t}/p_{0,1}$ and equals the sum of inflation plus net (real) return on capital. As for MM there is no inflation. Note that the discount factor of the traded goods cannot be obtained by information contained in the agents; this knowledge can only be extracted from the equilibrium solution. It seems therefore reasonable to give always the *discounted* equilibrium prices, otherwise decision-makers might be tempted to erroneously take the discount rate of their regional agent.

The Numéraire-Price and the Regional Utility Discount Factor

Here the equilibrium discount rate represented by the numéraire price p_0 is put into correspondence with the utility discount factor b_t in the MM objective (C.1). A simplified MM-model reflecting the direct influence of the budget constraint on the objective caused by trade of the numéraire x_0 is

$$\begin{aligned} \max & \sum_{\tau \in T} b_\tau \log(C_\tau - x_{0,\tau}) \\ \text{s.t.} & \sum_{\tau \in T} p_{0,\tau} x_{0,\tau} + \sum_{\tau \in T} p_{1,\tau} x_{1,\tau} \geq 0. \end{aligned}$$

Obviously we can assume equality in the budget constraint. From economic theory it follows that arbitrage over time is zero in all equilibrium solutions, or formally interpreted that the dual multiplier of the budget constraint in

$$\left.\begin{aligned} & \max b_t \log(C_t - x_{0,t}) \\ & \text{s.t.} \sum_{\tau \in T} (p_{0,\tau} x_{0,\tau} + p_{1,\tau} x_{1,\tau}) = 0 \end{aligned}\right\} \tag{7.4}$$

is the same for all $t \in T$. Note that the maximization is done only with respect to one time period t. For this model the Lagrangian is

$$L(x_{0,t}, \lambda_t) = b_t \log(C_t - x_{0,t}) + \lambda_t \sum_{\tau \in T} (p_{0,\tau} x_{0,\tau} + p_{1,\tau} x_{1,\tau}),$$

and we derive from the first order optimality conditions $\partial L / \partial x_{0,t} = 0$ the relation

$$\lambda_t = \frac{b_t}{p_{0,t}(C_t - x_{0,t})}. \tag{7.5}$$

From the arbitrage argument we expect to find in an equilibrium solution $\lambda_t = \lambda_{t+1}\ \forall t \in \{1, \dots, T-1\}$. The relation (7.5) is based on an extremely simplified model but numerical tests in equilibrium points of the full-fledged MM^{mr} confirmed constant Lagrange multipliers following (7.5) with a high accuracy of 8–10 digits (based on the solver gams-minos). The reason for this coincidence lies in the fact that our simplified model comprises all occurrences of the numéraire. Therefore we have found the proportionality of the components of the numéraire price to be

$$\frac{p_{0,t+1}}{p_{0,t}} = \frac{b_{t+1}}{b_t} \cdot \frac{(C_t - x_{0,t})}{(C_{t+1} - x_{0,t+1})}, \tag{7.6}$$

that is, the fraction

$$\frac{C_t - x_{0,t}}{C_{t+1} - x_{0,t+1}}$$

appears additionally in the transmission of the utility discount factor b onto the equilibrium discount factor p_0.

Free Permits

In the presentation of Section 7.1.1 it was assumed that permits are valid only in one period. This comes probably close to todays usual political intention, but it is straightforward to model permits which are free to be used in any period. The idea is to allow an arbitrary exchange of permits in all periods, and to require a balanced 'CO_2-budget' only at the end of the time horizon. As a consequence, such a free world allows to emit CO_2 prior to buying the corresponding permits.[5] To capture this free behavior the emission constraint ($EC^{mr}_{r,t}$) in the model MM^{mr} is replaced by

$$\sum_{t \in T} Em_{r,t} \leq \sum_{t \in T} [IECO2_{r,t} - NTXCO2_{r,t}]. \tag{EC^{free}_r}$$

[5]This might irritate, but one can interpret such free permits in that there is no periodic structure in the permits at all; the whole amount of permits for the whole time horizon is simply given at the beginning, and the regions are free to use them in whatever way they like. This supports actors who like to do the whole job of emission abatement towards the end of the time horizon.

In view of the previous section, it is straightforward to expect a formal dependency among the permit prices if the emission constraint is scalar; this helps both to check economic consistency of the results, and to reduce the problem dimension. The starting presumption is the same 'no-arbitrage'-argument as in the previous section; the dual multiplier (shadow price) of ($\mathrm{EC}_r^{\mathrm{free}}$) must be constant over all time periods. Technically this means that if a solution is given, choose $t \in T$, freeze all variables for $t' \neq t$, and compute the dual multiplier μ_t of ($\mathrm{EC}_r^{\mathrm{free}}$). Then all such-wise computed multipliers must be equal, i.e. $\mu_t = \mu_{t'}$ must hold for all $t' \neq t$. Economically seen constant μ_t means that the marginal utility of permits is constant over time, that is, arbitrage possibilities are excluded. To analyze the situation problem (7.4) is extended to

$$\begin{aligned} &\max b_t \log(C_t - x_{0,t}) \\ &\text{s.t.} \sum_{\tau \in T} (p_{0,\tau} x_{0,\tau} + p_{1,\tau} x_{1,\tau}) = 0, \\ &\qquad \sum_{\tau \in T} (x^0_{1,\tau} - x_{1,\tau}) = \sum_{\tau \in T} Em_\tau. \end{aligned}$$

Fixing the emissions $\sum_\tau Em_\tau =: c$, the following Lagrangian is obtained:

$$\begin{aligned} L(x_{0,t}, x_{1,t}, \lambda_t, \mu_t) &= b_t \log(C_t - x_{0,t}) \\ &\quad + \lambda_t \sum_{\tau \in T} (p_{0,\tau} x_{0,\tau} + p_{1,\tau} x_{1,\tau}) + \mu_t \Big[\sum_{\tau \in T} (x^0_{1,\tau} - x_{1,\tau}) - c \Big]; \end{aligned}$$

resolving for μ_t we derive from the first order optimality conditions $\partial L / \partial x_{1,t} = 0$ the relation $\mu_t = \lambda_t p_{1,t}$. The no-arbitrage argument requires that both μ_t and λ_t are constant over time, thus $p_{1,t}$ must also be constant over time, and so the dimensionality of independent price components is reduced by $|T| - 1$. Note that the 'real' (undiscounted) permit prices $p_{1,t}/p_{0,t}$ are then proportional to the equilibrium discount rate $1/p_{0,t}$, which equals the situation of natural resources analyzed by Hotelling [52].

The constant price components for p_1 give raise to non-unique solutions due to redundant variables. To see this, first note that x_1 appears only in the budget constraint and the (scalar) emission constraint ($\mathrm{EC}_r^{\mathrm{free}}$); in both constraints x_1 shows up only as a whole sum $\sum_{t \in T} x_{1,t}$, and therefore the feasible set and the objective does not change if for two distinct time periods $t, t' \in T$ we perturb $x_{1,t} + \delta$ and $x_{1,t'} - \delta$ for any $\delta \in \mathbb{R}$. This ambiguity aggravates the solution behavior of both the ACCPM and the regional models. In case of ACCPM this is caused by almost arbitrarily large and 'jumpy' excesses occuring throughout the iterations which disturb to focus the solution; furthermore, these large excesses can happen arbitrarily close to a solution and even in the solution itself, because in an equilibrium only $p_{1,t} \cdot \sum_{\tau \in T} x_{1,\tau} = 0$ holds for an arbitrary $t \in T$, and so the components $x_{1,\tau}$ can still be large. The regional models are negatively affected in that solving the models using restart techniques is slowed down due to possible bigger distances between successive solutions. All these problems can be overcome and the solution made unique by restricting trade to a single period, say the last one, which is achieved by fixing all permit trade except for the last period to zero, $x_{1,t} \equiv 0\ \forall t < |T|$. This unique solution is obviously also a (maximal) solution of the regional problem where x_1 is not restricted such-like. Furthermore, this strategy allows to reduce the size of the equilibrium problem by $|T|-1$. Of course, instead of fixing all but the last component to zero, and to make the numbers 'nicer', one could as well distribute the trade equally among all periods by imposing $x_{1,t} = x_{1,t'}$ for all t and t' in T, or by simply redistributing it after computing the equilibrium by setting $x_{1,t} := \frac{1}{|T|}x_{1,T}\ \forall t \in T$.

Floating Permits

The previous two concepts for trading permits—totally period dependent in the original MM^{mr}-model versus total freedom in the section above—are two extreme cases. As an intermediate case we consider here so called floating permits; the idea is to allow the use of unused permits in later periods, but—opposed to free permits—it is not allowed to use permits from future periods. As in the case of free permits two ways of exchanging permits are possible: exports/imports and internal savings. While there was no need to explicitly model the internal saving of permits in case of free permits, this must be done in case of floating permits, because now savings from future periods are no more allowed to flow backwards, i.e. credits are not allowed. To that end a new variable called $SCO2_{r,t}$ (Savings of CO_2-permits) is introduced; it represents the change of the regional (internal) stock of permits in period t. Thus, at the end of every period t the total amount of permit savings is $\sum_{\tau=1}^{t} SCO2_{r,\tau}$ which must be non-negative for all $t \in T$:

$$\sum_{\tau=1}^{t} SCO2_{r,\tau} \geq 0. \qquad (\mathrm{SCO2}_{r,t})$$

Based on these savings the emissions are bounded by the following set of $|T|$ constraints replacing $(\mathrm{EC}^{mr}_{r,t})$ from above:

$$Em_{r,t} \leq IECO2_{r,t} - NTXCO2_{r,t} - SCO2_{r,t}. \qquad (\mathrm{EC}^{\mathrm{float}}_{r,t})$$

Because permits are only allowed to move into future periods, the dual multipliers of $(\mathrm{EC}^{\mathrm{float}}_{r,t})$ are non-increasing over time. As a consequence we observe

$$p_{1,1} \geq p_{1,2} \geq \cdots \geq p_{1,T}. \qquad (7.7)$$

Because MM is a growth model and the initial endowment with permits is linearly decreasing over time we observe usually an overall scarcity of permits. Hence $p_{1,T} > 0$ and so $p_1 > 0$.

The role of $\mathrm{SCO2}_{r,t}$ in $(\mathrm{EC}^{\mathrm{float}}_{r,t})$ is close to that of a slack; a major difference is that some of the components may be negative due to requiring only $\sum_{\tau=1}^{t} SCO2_{r,\tau} \geq 0$

instead of $SCO2_{r,t} \geq 0$ for each period t. By putting all positive slack appearing in ($EC^{float}_{r,t}$) into $SCO2_{r,t}$ we can require equality in ($EC^{float}_{r,t}$). Substituting ($EC^{float}_{r,t}$) into ($SCO2_{r,t}$) leads then to equivalence of the set of constraints described by ($SCO2_{r,t}$) and ($EC^{float}_{r,t}$) with

$$\sum_{\tau=1}^{t} Em_{r,\tau} \leq \sum_{\tau=1}^{t} [IECO2_{r,\tau} - NTXCO2_{r,\tau}]. \qquad (EC^{float2}_{r,t})$$

Hence, in the real implementation the set of constraints ($EC^{float2}_{r,t}$) suffices and it is not necessary to introduce explicitly the savings $SCO2_{r,t}$ with the corresponding constraints ($SCO2_{r,t}$).

In principle we have the same problem of non-uniqueness due to the introduction of redundant variables as we had with free permits. To see this, note that if in a period permits are at least as scarce as in the preceding period, the equilibrium price of the corresponding components of p_1 are equal. As discussed in case of free permits in the previous section, trade is non-unique within a certain (sub-)set of time periods $T' \subset T$ if the corresponding price components of p_1 are equal. By imposing $p_{1,t} \geq p_{1,t+1} + \varepsilon$, $t = 1, \ldots, |T| - 1$, with a sufficiently small $\varepsilon > 0$, trade $x_{1,r}$ becomes unique while still being arbitrarily close to an unperturbed solution of the floating permit equilibrium problem.

In our scenarios we usually observe the same equilibrium solutions p using either free or floating permits. To justify this note that a growing economy together with a linear decrease of the endowment with CO_2-permits leads to a situation where permits are typically more scarce in period T than in all preceding periods. This implies already $p_{1,t} = p_{1,t'}$ $\forall t, t' \in T$, and for all $t < |T|$ the constraint ($EC^{float2}_{r,t}$) is non-binding and can be dropped. Therefore a practical way to solve MM^{mr} with floating permits is to solve it actually with free permits and then check the resulting equilibrium price p if it is also an equilibrium price for floating permits. This seemingly obscure way to solve the floating permit equilibrium problem is actually of practical relevance, because ACCPM has severe problems solving it directly. If, in a first attempt, (7.7) is not required, the same large and jumpy excesses appear as discussed in case of free permits. If, in a second attempt, (7.7) is obeyed, two undesirable effects are present: (i) non-unique solutions and therefore again large and jumpy excesses (remember that in an equilibrium solution we usually have $p_{1,t} = p_{1,t'}$ $\forall t, t' \in T$); (ii) the restriction (7.7) for feasible prices leads to poorly distributed excess vectors, eventually pushing the iterates away from an interior solution point. Imposing in a third attempt $p_{1,t} \geq p_{1,t+1} + \varepsilon$ $\forall t = 1, \ldots, |T| - 1$ resolves (i) but does not improve on (ii).

Consistency

Integrating consistent (regional) models which were developed independently raises the question of consistency of the overall equilibrium model. Trivial aspects cover equal units, more importantly are, however, the underlying model assumptions. In principle all regional models should make the same 'global' assumptions, whereas regional differentiation should still be possible. For example, it is reasonable to require the same crude oil price on the world market for all regions, but it is also meaningful to find different growth rates of the population in the different regions. And here the discussion starts: how far should the model-assumptions be equalized when integrating regional models? A relevant example in the case of MM^{mr} is the question of available technologies and the corresponding prices. Should different regional prices for photo-voltaic power generation be allowed, even though there is an almost homogeneous world market for panels? Maybe yes due to differing tax regimes or different local construction costs, but one can as well reject it as being inconsistent. Similar questions can be asked for almost all technologies. It is also questionable to what extend certain technologies should be exclusive to certain regions. On the macroeconomic level more questions are added: are different depreciation rates acceptable? What about growth rates, elasticities, or the autonomous energy efficiency improvement factor? As well as it is obvious that consistent regional models can produce an inconsistent overall model, it is also clear that there are in case of complex regional models too many details involved to adjust all of them by a central authority. In case of such complex regional models the only tractable way is to bring together the local modelers and make them discuss directly there local assumptions. In fact, ETSAP[6], running now for more than 15 years, is exactly a forum where this kind of adjustment has been done excessively. For this reason, and in view of the above discussion, we accepted the regional models as is without any changes except an adaption of units, cf. Appendix E.3. We are aware, however, that a more thorough discussion is still lacking, and that this issue becomes increasingly important once developing countries are integrated.

7.2 MM^{mr} as VIP and Negishi-Problem

7.2.1 The VIP-Formulation

Based on (VIP) from page 7, MM^{mr} can be written as a variational inequality problem as follows. Let $p = (p_{NTX},\ p_{CO2}) \in \mathbb{R}_+^{2|T|}$, define the regional excess $e_r(p) = (NTX_r(p),\ NTXCO2_r(p)) \in \mathbb{R}^{2|T|}$ to be the outcome of the regional utility maximization problem (7.1) at price signal p, and denote the aggregation by $e(p) := \sum_{r\in R} e_r(p)$. Because $e(p) = e(\lambda p)$ for any $\lambda > 0$, we can restrict the set of feasible prices to $\Delta \subset \mathbb{R}_+^{2|T|}$, cf. (1.11), and formulate the concrete VIP for MM^{mr} as

$$\text{find } p^* \in \Delta \text{ such that } e(p^*)^T(p - p^*) \geq 0, \ \forall p \in \Delta. \tag{7.8}$$

[6]ETSAP abbreviates Energy Technology Systems Analysis Project which was founded by the IEA, the International Energy Agency.

Assuming non-redundant variables, the strict concave regional objective functions imply a single valued $e(p)$. Based on the map $e(p)$, Algorithm 2 or Algorithm 4 can be applied, where in the latter case an initial transformation of the problem together with a final transformation of the iterates is required.

7.2.2 The Negishi Problem

Here, Definition 1.2 is the starting point; to relate MM^{mr} to a welfare problem denote by K_r the feasible set of region $r \in R$ defined by all constraints in (7.1) and where the budget constraint (BC_r^{mr}) is omitted. By $(C_r, NTX_r, NTXCO2_r, \dots) \in K_r$ we then mean that the corresponding components fulfill simultaneously all constraints in (7.1) except the budget constraint. In view of the VIP-formulation above, we define again the regional excess $e_r = (NTX_r,\ NTXCO2_r)$ which is, in the context of the Negishi welfare-problem, simply the compound of the two vectors NTX_r and $NTXCO2_r$ and hence not price dependent. For any weight vector $\alpha \in \mathbb{R}_+^{|R|}$, $\sum_{r \in R} \alpha_r = 1$, the welfare problem ('Negishi-welfare problem') is then defined by

$$\left.\begin{aligned} &\max \sum_{r \in R} \alpha_r U_r(C_r) \\ &\text{s.t.} \sum_{r \in R} e_r \ \geq \ 0, \\ &\quad (C_r, NTX_r, NTXCO2_r, \dots) \ \in \ K_r \ \ \forall r \in R. \end{aligned}\right\} \qquad (7.9)$$

Based thereupon Algorithm 5 can be applied, using a decomposition scheme like the one described in Algorithm 6. To apply the latter we conclude this section by expliciting (4.8) and (4.9), i.e. the interpretation of θ_k^*, $\phi_r(p^k, v_r^k)$ and $e_r(v_r^k)$ is given in the context of MM^{mr}.[7] To start with, in iteration k the regional problem defining $\max \phi_r(p^k, v_r^k)$ in (4.10) is

$$\left.\begin{aligned} &\max \ \alpha_r U_r(C_r) + {p_{NTX}^k}^T NTX_r + {p_{CO2}^k}^T NTXCO2_r \\ &\ \text{s.t.} \ (C.1)\text{-}(C.4),\ (\mathrm{PRD}_{r,t}),\ (\mathrm{L}_{r,t+1}),\ (\mathrm{CAP}_{r,t+1}),\ (\mathrm{TC}_r), \\ &\qquad (\mathrm{USE}_{r,t}^{mr}),\ (\mathrm{EC}_{r,t}^{mr}), \\ &\qquad (\text{all other Markal-Macro constraints}). \end{aligned}\right\} \qquad (7.10)$$

Comparing (7.10) with (7.1) reveals two differences in the regional problem: (BC_r^{mr}) is dropped and the objective is extended by the penalty-term $p^{kT} e_r = {p_{NTX}^k}^T NTX_r + {p_{CO2}^k}^T NTXCO2_r$. Next, θ_k^* is the sum of the objective values over all such regional problems in iteration k. Finally, $e_r(v_r^k) = (NTX_r,\ NTXCO2_r)$ where NTX_r and $NTXCO2_r$ are taken from the solution of (7.10) in iteration k.

[7] Opposite to (4.8) and (4.9) the iteration-index is here a superscript again.

"Was willst du denn jetzt noch wissen?" fragt der Türhüter, "du bist unersättlich." "Alle streben doch nach dem Gesetz", sagt der Mann, "wieso kommt es, dass in den vielen Jahren niemand ausser mir Einlass verlangt hat?" Der Türhüter erkennt, dass der Mann schon an seinem Ende ist, und, um sein vergehendes Gehör noch zu erreichen, brüllt er ihn an: "Hier konnte niemand sonst Einlass erhalten, denn dieser Eingang war nur für dich bestimmt. Ich gehe jetzt und schliesse ihn." F. K. [58]

Chapter 8

Economic Results of MM^{mr} for Three Countries

The regional MM models draw a sophisticated picture of the related energy sectors and produce additionally a large amount of macro-economic data. A presentation and discussion of these results requires an in-depth knowledge of all underlying regional models which is beyond the scope of this study. We therefore restrict the presentation to results which are a direct outcome of the equilibrium model MM^{mr}, most notably equilibrium prices, dual multipliers of the CO_2 emission constraints, the GNP and the amount of trade. All these economic results are to be considered cautiously, as an in-depth discussion with economists of the involved countries is not included. Nevertheless, the underlying regional data sets are up to date (summer 1996 to spring 1997), and comprise the three countries Sweden (SW), the Netherlands (NL) and Switzerland (CH).

Although this is definitely a very small equilibrium model, the results can represent the equilibrium of a larger group of countries; to see this, note that the resulting trade can be interpreted in two ways. On the one hand the equilibrium solution forces the overall excess to zero (given positive price components, which is always the case in our scenarios), that is, trade among the three countries is implicitly modeled as if it is closed, which is far away from reality. On the other hand the only external information of a single region is the price signal given exogenously. It makes no difference, therefore, from the perspective of a single region if there are 3 or 30 regions included in the trade model, its (excess) reaction is solely determined by the price signal.

If we assume that the three countries are representative of a larger part of Europe 'E', in the sense that the resulting equilibrium prices of the three countries correspond to the E trade model, then we can equally well interpret trade of the three countries as trade within the larger E community. In view of the structural variation of the three countries discussed below we claim that they might indeed come near being representative in this equilibrium price sense, and hence their results could—to some extend—be interpreted in such a sense. Specifically, trade of numéraire could be seen in the context of such a larger virtual E community.

The Regional Models

The energy-related specificities of the three countries can be summarized as follows, cf. Bahn, Büeler, Kypreos and Luethi [6]. The three countries have high living standards. For instance, the 1993 gross domestic product (GDP) per capita was (in thousand US$) 20.9 for the Netherlands, 24.7 for Sweden and 35.7 for Switzerland, cf. [53]. However, both the structure and efficiency of their energy systems are rather different, particularly in the case of CO_2 emission, see Kram [66].

The Netherlands is a major exporter of natural gas, and its own energy system relies heavily on gas. In 1993, 98 % of all houses were connected to the natural gas grid, and around 50 % of electricity production came from gas power plants and 40 % from coal. Furthermore, fossil fuels accounted in 1990 for 97 % of the total primary energy use (TPE) resulting in CO_2 emissions of 161.3 million tons, that is 10.8 tons per capita.

In contrast in Switzerland, 60 % of electricity production is hydro-generated and 38 % nuclear. Under the current nuclear moratorium (valid until 2000), nuclear capacity is not allowed to increase. Fossil fuels accounted in 1990 for 54 % of the TPE, but the use of coal is very low. In 1990, CO_2 emissions from combustion were 43 million tons or 6.4 tons emitted per capita. The main contributors of CO_2 were transportation and heating activities.

Sweden has large hydroelectric resources. Its electricity production is primarily based on hydro-power (52 %) and nuclear power (42 %), the rest is produced from fossil fuels. This situation is due to change, as the Swedish Parliament has decided in 1980 to phase-out nuclear energy by the year 2010, starting in 1995. Fossil fuels accounted in 1990 for only 34 % of the TPE, resulting in CO_2 emissions of 54 million tons, that is 6.3 tons emitted per capita.

These differences lead to significant variations of CO_2 abatement costs among the three countries and for this reason constitute an incentive for cooperating on CO_2 emission abatement through an international market of emissions permits.

Parameter	CH	NL	SW
pot. GDP growth rate [%]	1990: 0.57, 2000: 1.97, 2010: 1.75, 2020–40: 1.2	2000: 2.19, 2010: 2, 2020: 1.74, 2030: 1.5, 2040: 1.25	2
discount rate [%]	5	5	5
ESUB	0.2	0.25	0.4
US$/local currency	0.8333	0.6	0.15

Table 8.1: Economic parameters of the regions in MM^{mr}.

Both the Swiss and Swedish data cover initially the periods 1990–2030, whereas the data representing the Netherlands are from spring 1997 and cover the periods 2000–2040. To extend trade, in the case of Switzerland and Sweden the data from period 2030 were duplicated and adjusted, yielding data sets covering the periods 1990–2040. In period 1990, Switzerland and Sweden behave as isolated regions without CO_2 emission bounds; this is achieved by forcing the numéraire exchange variable $NTX_{1990} = 0$ together with setting the initial endowment of CO_2 permits to a sufficiently large constant, i.e. formally $IECO2_{1990} = \infty$. A few relevant economic parameters are presented in Table 8.1.

Emission Cases and Scenarios

There are basically five cases to consider, see Table 8.2: (A) the regional models are isolated, i.e. are identical to the original formulation, without any trade; (B) the countries trade only the aggregated product *NTX* but no emission permits; and (C), the countries trade in addition CO_2 emission permits *NTXCO2*. Case (A) and (B) can be further combined with CO_2 emission limits which then have to be fulfilled in each country. Case (B) is important as reference for the full trade case (Cl), because trade by itself influences considerably the results and hence it would be misleading to compare only case (A) and (C). The above abbreviations

	No trade	Trade of *NTX* only	Trade of *NTX* and *NTXCO2*
Unlimited emissions	(Au)	(Bu)	—
Limited emissions	(Al)	(Bl)	(Cl)

Table 8.2: Abbreviations used for the different cases.

are extended in Table 8.2 and will be used in the rest of the chapter. In principle, trade of permits can be further subdivided into 'per period', 'floating' and 'free' permits following Section 7.1.2. For this chapter we have chosen 'per period' permits noting any exceptions.

Differing from the notion 'case', we designate by 'scenario' the CO_2 reduction target. The reference values are the approximate CO_2 emission values in period 1990 and set to 42 million tons for Switzerland, 62 for Sweden and 160 for the Netherlands. Based on these reference emissions the reduction scenarios are built; starting from the reference emissions in period 2000, the initial endowment decreases linearly between 2000 and 2040 by the prescribed percentage. As an example, the endowment in the $-20\,\%$-scenario is given in Table 8.3. In the special period 1990 the emissions of Switzerland and Sweden are unbounded. There are three scenarios used in this study: Stabilization (i.e. $0\,\%$-scenario), the $-20\,\%$- and $-40\,\%$-scenario. In the (Al) and (Bl) case the initial endowment with permits is exactly the emission bounds of each country; only in case (Cl) permits are traded and consequently the realized regional emissions may differ from the endowment.

Country	2000	2010	2020	2030	2040
CH	42	39.9	37.8	35.7	33.6
SW	62	58.9	55.8	52.7	49.6
NL	160	152.0	144.0	136.0	128.0

Table 8.3: Initial endowment of CO_2 permits for the −20 %-scenario, in [Mtns/y] CO_2.

Effects of Numéraire Trade

As a first result Table 8.4 shows the unlimited emissions in the case (A) of isolated countries and case (B) where only the numéraire goods are traded. It is interesting to see that regional emissions can increase non-monotonically (NL, 2010 → 2020), and that trade of numéraire does not significantly influence emissions but can slightly increase *or* decrease them. This is remarkable in view of the significant increases of utility and GDP induced by trade presented in Table 8.5 below. Finally, the aggregated emissions increase smoothly by about 60 % from period 2000 to 2040.

Country	Case	1990	2000	2010	2020	2030	2040
CH	(Au)	42.8	42.5	48.9	50.4	52.5	55.4
	(Bu)	42.8	42.9	49.6	51.1	54.1	56.2
NL	(Au)	—	162.9	176.0	175.9	176.0	194.2
	(Bu)	—	162.9	177.4	176.8	178.0	197.2
SW	(Au)	62.2	67.1	108.5	130.0	164.0	186.8
	(Bu)	62.2	64.9	102.1	124.6	156.9	178.9
$\sum$	(Au)	—	272.5	333.4	356.3	392.5	436.4
	(Bu)	—	270.7	329.1	352.5	389.0	432.3

Table 8.4: CO_2 emissions (Mt/year) for the cases (Au) and (Bu).

While the utility-index—being the objective of the regional optimization problem—is a scalar, results like GDP, consumption[1], etcetera are period-dependent. The presentation of such period-dependent results stemming from dynamic models can be done in two ways; either the results are given for each period separately, or they are aggregated. We have chosen to aggregate results in order to give

[1]For a definition of all economic terms see Chapter 7. 'Consumption' denotes total macroeconomic consumption.

a clearer overview. However, the choice of an appropriate discount rate is not 'canonical'. In order to emphasize the effects in the later periods, a small discount rate of 2.5 % p.a.—and not say 5 %—is chosen, and based on this discount rate the quantities are aggregated over the periods where trade takes place, i.e. 2000–2040.

The last period 2040 poses a final problem; to damp terminal effects, the utility discount factor of the last period is increased following a geometrical series, see (C.1) page 117.[2] This tends to make consumption in the last period higher than in the other periods and so the weight attached to this last period influences significantly the aggregated outcome. But in our interpretation the results of the last period are simply representatives of period 2040 and therefore discounted back by $0.975^{40} \approx 0.363$ to the base period 2000. This works fine for Switzerland and the Netherlands, but Sweden increases the utility discount factor of the last period considerably more than the other regions, resulting in a significant shift of consumption into period 2040. For example, in case (Bu) Sweden increases its consumption from US\$ $3.3 \cdot 10^{12}$ in 2030 to US\$ $6.7 \cdot 10^{12}$ in 2040. This effect is not visible in the other two regions, explaining why only Sweden exhibits a decrease in the aggregated consumption if trade is allowed, cf. Table 8.5.

	CH		NL		SW	
Measure	(Au)	(Bu)	(Au)	(Bu)	(Au)	(Bu)
Utility-index	2555.86	2559.66	3230.78	3236.89	5091.78	5157.34
GDP	12.16	12.41	18.48	18.63	10.69	10.43
Consumption	9.25	9.55	15.93	16.48	9.02	8.14

Table 8.5: Benefit of trade without emission limits; GDP and consumption are in [10^{12} US\$] with aggregation discount factor 2.5 %.

If in the case of Sweden the internal utility discount factors are used in the aggregation, then the consumption, which defines the objective function, increases significantly from US\$ $20.70 \cdot 10^{12}$ in the (Au)-case to US\$ $26.67 \cdot 10^{12}$ in the (Bu)-case. Using this internal utility discount factor also for aggregating the GDP, however, we observe a slight drop from US\$ $24.68 \cdot 10^{12}$ for (Au) down to US\$ $24.05 \cdot 10^{12}$ for (Bu). Using Swedish data these numerical examples show how the discount factor can shape the result of aggregation.

In the rest of this chapter we restrict ourselves to an aggregation discount factor of 2.5 %. As a first main result Table 8.5 demonstrates nicely the expected 'mutual benefit from trade' through the increase of the utility-index for all regions when

[2] A look at (C.1) reveals that for all but the last period T the utility discount factor 'udf_t' is the only discount factor attached to $\log C_t$. In the last period T, however, an additional $1/(1-(1-udr_T)^{10})$ factor accompanies udf_T. For linguistic simplicity we denote in this chapter the whole expression $udf_T \frac{1}{1-(1-udr_T)^{10}}$ as utility discount factor of the last period.

trade is allowed.[3] Surprisingly, having the greatest gain in the utility-index, Sweden has also the only drop in aggregated consumption and GDP. But as indicated by the word 'index', the quantitative gain/loss in utility can not be easily interpreted economically. Instead, measures like compensated or equivalent variations are more appropriate for its interpretation.

Country	2000	2010	2020	2030	2040
CH	−0.32821	−0.21302	−0.20958	−0.11786	0.9949
NL	−0.48481	−0.49876	−0.30145	−0.29456	1.7142
SW	0.81302	0.71178	0.51103	0.41241	−2.7091

Table 8.6: (Bu); Undiscounted *NTX* trade without emission limits, in [10^{12} US$].

The discussion above indicates another fundamental behavior of such dynamic equilibrium models; the amount of trade—above all of the monetary numéraire good—tends to show a 'bang-bang' behavior as seen in Table 8.6. For example, in the case of Sweden, the very high utility discount factor attached to the last period formes an incentive to export numéraire in all prior periods but the last, and then to import in the last period, when it counts most from the perspective of the objective function, as much numéraire as needed in order to satisfy the budget constraint. Ginsburgh and Waelbroeck [36, Chapter 4], however, strongly argue against the temptation to correct such 'anomalous' outcomes by inserting ad hoc constraints.

Another possibility would simply be to ignore the problematic last period in the aggregation process. This, however, would compromise for example the Swedish results where all imports would be left out. Also interesting is that the amount of traded numéraire—the aggregated monetary foreign trade—can exceed 20 % of GDP. This is, as indicated in the introduction, only plausible if trade is interpreted as happening within a larger group of countries.

Full Trade of Numéraire and Permits

'Per Period' Permits

To start the presentation of the equilibrium results with one of its most interesting facets, the undiscounted equilibrium prices resulting from case (Cl) together with the corresponding marginal costs of isolated CO_2 abatement from case (Bl) are given in Figure 8.1. The large difference in the marginal abatement costs constitute an incentive to join such a permit community and to trade certificats.

[3] To prevent misunderstandings the mutual benefit from trade assures an improvement for each region only if the non-trade-case (perfectly isolated regions) is compared with a case where trade (of some goods) is possible. Specifically, economic theory does not predict a mutual increase of utility if, starting from a situation where some goods are traded, additional goods are traded. E.g., the utility-index of a region may drop when trade is extended from case (Bl) to (Cl).

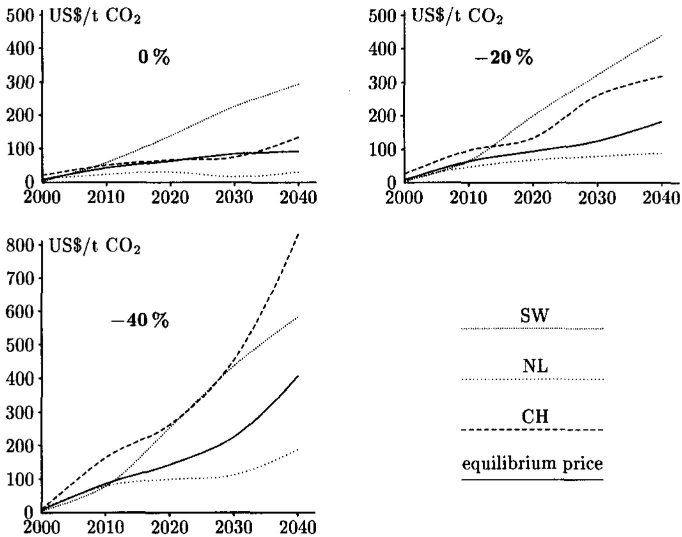


Figure 8.1: Undiscounted marginal reduction costs and equilibrium permit prices.

Figure 8.1 shows in particular that the permit prices lie between the lowest and highest national marginal reduction costs. As can be expected from the differing structures in the energy sector, Switzerland and Sweden have siginificantly higher marginal abatement costs than the Netherlands. If the marginal costs and equilibrium prices are linearly extrapolated from the interval [+60 %, 0 %], where +60 % denotes a scenario without emission limits, we observe considerably *nonlinearly increased costs and prices in the* −20 % and −40 % *scenarios.*

The above undiscounted equilibrium prices, which exceed US$ 400 per ton CO_2 in period 2040 for the −40 %-scenario, seem to be high, but discounting them back to the base period 2000 by means of the numéraire price component (cf. Section 7.1.2 page 77), show in fact the opposite. Based on p_{NTX} the anual equilibrium discount factor is around 4.7 %.

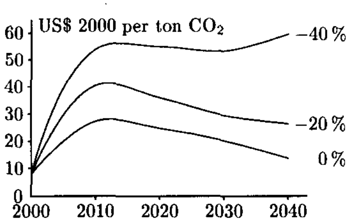


Figure 8.2: Discounted equilibrium prices.

The discounted prices, depicted in Figure 8.2, are very reasonable even for demanding scenarios like a reduction of 40 % until 2040. To give an example, a tax of US$ 60 per ton of CO_2 measured in dollars from period 2000 equals approximately 13–14 US cents per liter of gasoline. Also interesting is the bump in period 2010 which is possibly due to the inertia of the energy sector. It takes time to implement cost effective alternatives, suggesting an early but comparably slow emission reduction instead of doing nothing now and leave harder reduction targets for later time periods.

'Free' versus 'Per Period' Permits

One possible strategy to investigate how permits are optimally distributed over time is to leave the choice to the model, i.e. to look at the solution generated from 'free permits' (FP), cf. Section 7.1.2 page 79. Whereas in the previous section 'per period' (PP) permits were assumed, Figure 8.3 shows the CO_2 emissions for both the 'per period' and the 'free' permits for each country. The 'free permit' curves exhibit a characteristic raise at the end and are thereby easily distinguished from the 'per period' curves. The aggregated emissions are only shown for the 'free permit' option, because the aggregated 'per period' emissions form a straight line with a slope given by the reduction scenario. The aggrageted curves as well as all regional curves show most clearly the optimal saving behavior: save permits in all periods but the last, i.e. emit less, and use more permits in the last period. These results underline the potential overall improvement achieved by early reduction of emissions, instead of no reduction in early periods and severe reductions later. It is still unclear, however, to what extend the shift of emissions into the last period is influenced by the increased utility discount factor attached to period 2040.

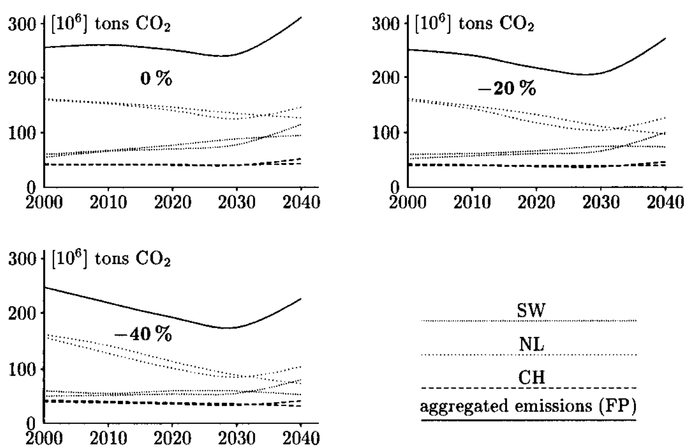


Figure 8.3: CO_2 emissions in case of free or 'per period' permits.

The utility-index for the individual regions, however, may decrease in the equilibrium solution if permits are free to move in time. Mathematically this is not surprising, because even though the $|T|$ periodical emission constraints are relaxed by one overall constraint and thereby the regional feasibility sets enlarged, the equilibrium price may change, affecting the feasibility sets. Thus, by changing the equilibrium, some parts of the feasibility sets can expand, while others can shrink. Table 8.7 shows the utility for the different reduction scenarios in case (Cl), where either the permits are restricted to periods (PP) or are free to move among the periods (FP). In the first two reduction scenarios there are only minor differences, showing a slight improvement of Switzerland and Sweden on the one hand and a deterioration for the Netherlands on the other hand. The −40 %-scenario is more drastic in that only Sweden profits from free permits, whereas Switzerland and the Netherlands both lose utility-index.

	CH		NL		SW	
Scenario	PP	FP	PP	FP	PP	FP
0 %	2559.365	2559.375	3237.232	3237.095	5151.147	5152.304
−20 %	2558.984	2559.014	3236.769	3236.541	5147.358	5149.767
−40 %	2559.058	2558.381	3239.869	3235.560	5143.758	5146.171

Table 8.7: Comparison of per period versus free permits expressed by utility-index, case (Cl).

Finally the undiscounted equilibrium prices resulting from free permits are given in Figure 8.4. In view of the 'free permit' discussion in Section 7.1.2 one expects an exponential increase over time. This exponential growth of prices over time is confirmed by the results in all but the last period. The drop in period 2040 is again due to the higher weight given to period 2040 in the objective, which impels the regions to shift emissions into that period, decreasing thereby the marginal abatement costs, resulting finally in a lower permit price.

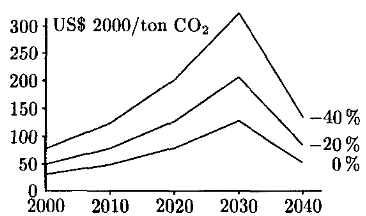


Figure 8.4: Undiscounted free permit prices.

As discussed in Section 7.1.2 the discounted prices are constant over time, avoiding irregularities like the bump in Figure 8.2. But the drop in period 2040 aggravates the estimation of the discounted equilibrium prices for the different scenarios. As a first simple strategy the undiscounted prices of period 2000 can be taken as the discounted prices. But this is an overestimation due to the shift of permits into the last period and the resulting lower undiscounted prices in period 2040. An improved estimation can be obtained by discounting the prices back to period 2000 and then building the weighted average, where the utility discount factors are used as relative weight. An educated estimate of the discount factor is based on the numéraire price components of the periods 2000–2030 and yields approximately 4.7 % per year. The weighted average then produces discounted equilibrium prices in the range of 22.9, 36.9 and 58.4 US$ per ton CO_2

respectively for the three reduction scenarios.

These discounted equilibrium prices, based on free permits, seem to be comparable with the prices of 'per period' permits depicted in Figure 8.2.

'Per Period' Permits: Amount of Foreign Trade and GNP for Different Emission Scenarios

The influence of the different reduction scenarios (0 %, −20 % and −40 % CO_2 emissions) on numéraire trade are marginal only, i.e. Table 8.6 remains in priciple valid for all scenarios. The same applys less strictly to trade of permits, as can be seen from Table 8.8, where, as usual, a positive value designates exports and a negative value imports. In all scenarios, Sweden trys to buy sufficient permits to keep emissions at least at the 2000 period level, whereas the Netherlands is a major net seller of permits. In conclusion, the further the overall emissions are reduced, the more Switzerland increases its imports, the Netherlands reduces exports and Sweden reduces imports of permits.

Scenario	Country	2000	2010	2020	2030	2040
	CH	−0.8	−0.5	0.3	0.6	−1.6
0 %	NL	−0.6	6.4	14.7	26.3	34.2
	SW	1.5	−5.8	−15.0	−26.5	−32.6
	CH	−0.8	−1.5	−1.7	−3.7	−6.8
−20 %	NL	−0.6	4.6	12.2	25.5	31.1
	SW	1.4	−3.2	−10.8	−22.2	−24.2
	CH	−0.8	−2.7	−4.2	−6.6	−7.0
−40 %	NL	−1.4	2.2	14.8	23.2	22.8
	SW	1.7	0.6	−10.7	−16.5	−15.8

Table 8.8: Net export of CO_2 emission permits (Mt/year).

As a final result, the effect on the aggregated GNP of bounding emissions and trading permits is presented in Figure 8.5.[4] As before, the aggregation is done over the periods 2000–2040 with a discount rate of 2.5 %. The origin of the x-axis is marked by '+60 %' to indicate that no emission limits are imposed; starting

[4]In the usual definition of GNP the repair costs due to climate-change-induced damages (floods, storms, etcetera) must be *added* to the GNP. Ecologically oriented economists criticize this definition of GNP because then damages are considered positively as increasing the GNP. In our Macro model, the GNP does not explicitly include the repair costs caused by climate change, which implies a larger GNP in the baseline (Bu)-case. Consequently, this larger GNP increases the GNP losses when the baseline case (Bu) is compared with a situation where CO_2 emissions are limited and hence repair costs are reduced. Because IPCC estimate the damage costs in the range of one to a few percents of GNP (cf. the introduction of Chapter 6) the above baseline effect could significantly influence the results in figure 8.5.

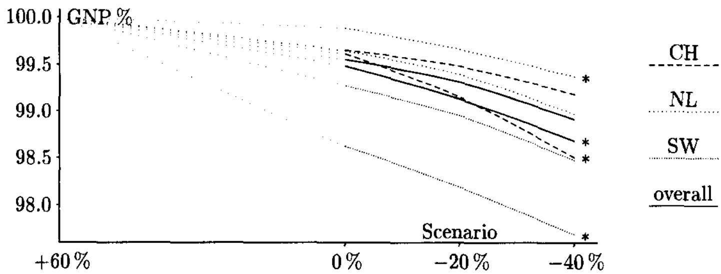


Figure 8.5: Decrease of aggregated GNP in %; reference (= 100 %) is the case (Bu), * designate (Bl) cases and non-* designate (Cl) cases.

from there and going to the right, the emissions are increasingly reduced. On the y-axis 100 % represent the aggregated GNP in case (Bu), i.e. when numéraire is traded, but CO_2 emissions are not bounded. The decrease in GNP caused by bounding emissions follows either case (Bl), i.e. no tradeable emission permits, or case (Cl) where the burden can be shared by trading permits. The (Bl) curves, marked by a star, are generally lower than the corresponding (Cl) curves with the exception of the Netherlands where trade of permits brings about slightly lower GNP. As seen in Figure 8.5 the GNP losses are not equally distributed. Sweden's GNP suffers most in both cases and all scenarios, whereas the Dutch economy is less affected. A possible explanation is the high costs involved in the phase-out of nuclear power in Sweden compared with the lower costs involved in phasing out coal plants in the Netherlands. It is also Sweden which gains most (reduces its losses most) from trading permits if the non-trade case (Bl) is considered as reference. In view of this the negotiation of initial endowments may improve benefits/losses equity among the countries.

The GNP summed over all countries ('overall') decreases by a little more than one percent, where trade can reduce the decrease by about 0.2 %-points, that is, trade reduces the GNP losses by roughly 20 %. All GNP losses are small; indeed, one percent of aggregated GNP by the year 2040 corresponds to an average yearly growth rate of approximately 0.045 %-points,[5] something which is beyond statistical measurability. In view of the criticism leveled at the definition of the GDP in MM, however, the above values should be interpreted cautiously.

[5]Assuming a yearly growth rate $g = 0.02$ and a time horizon $T = 40$ years then a first approximate answer is given by the equation $(1+g-l)^T = 0.99 \cdot (1+g)^T$ yielding $l = 0.026\,\%$. Because the losses are aggregated over the whole period T an improved estimation starts from the equation $\int_0^T (1+g-l)^t dt = 0.99 \cdot \int_0^T (1+g)^t dt$. For its approximation $\sum_0^T (1+g-l)^t = 0.99 \cdot \sum_0^T (1+g)^t$ a numerical solution based on bisectioning yields $l = 0.045\,\%$.

Summary

The results can be summarized as follows. First the discounted equilibrium prices are very low; even for drastic reduction targets like −40 % they are only around 13–14 US cents per liter of gasoline. The corresponding loss of aggregated GNP is around 1–1.5 %, which is also very low, corresponding to an anual growth rate reduction of approximately 0.06 %-points. Trading permits reduce the overall GNP losses by about 20 % when compared with unilateral reductions. Clearly, this GNP 'gain' from trading permits is strongly related to the distribution of the initial endowment. Finally, the results indicate that it is profitable to start reducing emissions early and to leave emission reserves for the future.

We believe the results presented in this chapter contribute to the understanding of important aspects of CO_2 emission permits. However, they emanate from the particular models and data chosen, and are in that sense specific. Nevertheless, we are convinced that the economic and mathematical reasoning developed in this work can be profitably applied to a variety of situations.

Appendix A

Notation and Basic Theorems

The chapter summarizes some basic mathematical material which is relevant to the main part of this work. After some convexity definitions the characterization of optima due to Karush, Kuhn and Tucker is given in Section A.1. The Lagrangian dual problem, used in the decomposition of the Negishi-welfare problem, is discussed in Section A.2. Because both the Karush-Kuhn-Tucker characterization of optima and the connection to the Lagrangian dual problem requires differentiability, we give in Section A.3 a well known relaxation to continuous functions. Finally Section A.4 gives a brief introduction to VIPs.

Definition A.1 *(cf. [7]) A set $C \subset \mathbb{R}^n$ is called* **convex** *if $\forall \lambda \in [0,1]$ and $\forall x, y \in C$ we have $\lambda x + (1-\lambda)y \in C$. A function $f : D \subset \mathbb{R}^n \to \mathbb{R}$ is called*

- **convex on** *D if $\forall \lambda \in [0,1]$ and $\forall x, y \in D$ we have $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$;*
- **pseudo-convex on** *D if f is differentiable on (an open set containing) D and $\forall x, y \in D$ with $\nabla f(x)^T(y-x) \geq 0$ we have $f(y) \geq f(x)$;*
- **quasi-convex on** *D if $\forall \lambda \in [0,1]$ and $\forall x, y \in D$ we have $f(\lambda x+(1-\lambda)y) \leq \max\{f(x), f(y)\}$.*

The same definitions with concavity replacing convexity are obtained by exchanging f with $-f$. Given differentiability of f, convexity implies pseudo-convexity which in turn implies quasi-convexity. The latter is equivalent with convexity of all level sets $L_\alpha := \{x \mid f(x) \leq \alpha\}$, thus level sets of convex and pseudo-convex functions are always convex. The following two sections are based on [7].

A.1 The Karush-Kuhn-Tucker Characterization of Optima

The problem under consideration can be described as follows. Let X be a nonempty, open subset of $\mathbb{R}^n$ and let f, g_i for $i \in I$ and h_j for $j \in J$ be functions from X to $\mathbb{R}$, where both I and J are finite. The problem is to find a solution of

$$\left.\begin{array}{ll}\min & f(x) \\ \text{s.t.} & g_i(x) \leq 0 \;\; \forall i \in I, \\ & h_j(x) = 0 \;\; \forall j \in J, \\ & x \in X.\end{array}\right\} \tag{A.1}$$

Under suitable conditions such optimization problems can be replaced by a set of equalities and inequalities. This is the content of the following theorems developed in the fifties and sixties, and which are today one of the most often used mathematical tools in economic theory.

Theorem A.1 *(Karush-Kuhn-Tucker, necessary conditions; [7, Theorem 4.3.6]) Consider the problem (A.1), suppose x^* is a local solution and define the set of binding indices at x^*, $I^b = \{i \in I \mid g_i(x^*) = 0\}$. If f and g_i $\forall i \in I^b$ are differentiable at x^*, g_i $\forall i \notin I^b$ are continuous at x^*, h_j $\forall j \in J$ are continuously differentiable at x^*, and finally the binding gradients $\{\nabla g_i(x^*), i \in I^b\} \cup \{\nabla h_j(x^*), j \in J\}$ are linearly independent, then there exist scalars u_i for $i \in I^b$ and v_j for $j \in J$ such that*

$$\begin{gathered}\nabla f(x^*) + \sum_{i \in I^b} u_i \nabla g_i(x^*) + \sum_{j \in J} v_j \nabla h_j(x^*) = 0 \\ u_i \geq 0 \quad \forall i \in I^b.\end{gathered} \tag{A.2}$$

If g_i $\forall i \notin I^b$ are also differentiable at x^* condition (A.2) is equivalent to

$$\begin{gathered}\nabla f(x^*) + \sum_{i \in I} u_i \nabla g_i(x^*) + \sum_{j \in J} v_j \nabla h_j(x^*) = 0 \\ \left.\begin{array}{rcl} u_i g_i(x^*) & = & 0 \\ u_i & \geq & 0 \end{array}\right\} \quad \forall i \in I.\end{gathered} \tag{A.3}$$

The linear independence of the binding gradients $\{\nabla g_i(x^*), i \in I^b\} \cup \{\nabla h_j(x^*), j \in J\}$ is called Kuhn-Tucker constraint qualification. Depending on the problem it can be advantageous to replace this linear independence condition by other conditions. A popular one is the so called Slater condition which implies the Kuhn-Tucker constraint qualification.

Definition A.2 *(Slater constraint qualification) If there exists an $\bar{x}$ such that*

$$\begin{array}{rcl} g_i(\bar{x}) & < & 0 \quad \forall i \in I, \\ h_j(\bar{x}) & = & 0 \quad \forall j \in J, \\ & \bar{x} \in X, & \end{array}$$

we say the constraints fulfill the Slater conditions.

While Theorem A.1 states conditions under which a so called KKT-point *exists*, that is, a point x^* where (A.2) holds, it is interesting to know when such a KKT-point is an optimum.

Theorem A.2 *(Karush-Kuhn-Tucker, sufficient conditions; [7, Theorem 4.3.7]) Suppose x^* is feasible to problem (A.1), i.e. $x^* \in X$, $g_i(x^*) \leq 0$ $\forall i \in I$ and $h_j(x^*) = 0$ $\forall j \in J$. Assume further f and g_i $\forall i \in I^b$ are differentiable at x^*, g_i $\forall i \notin I^b$ are continuous at x^*, h_j $\forall j \in J$ are continuously differentiable at x^*, and that x^* fulfills (A.2). If f is pseudo-convex at x^*, g_i is quasi-convex at x^* for $i \in I^b$, h_j is quasi-convex at x^* for $j \in J : v_j > 0$ and h_j is quasi-concave at x^* for $j \in J : v_j < 0$, then x^* is a global optimal solution to (A.1).*

Simplifying those sophisticated structural properties, we have the following equivalence: Assume f and g_i for $i \in I$ are convex, h_j for $j \in J$ is affine, and either the KKT constraint qualification at a feasible point x^* or the Slater conditions are fulfilled; then x^* is a global solution to (A.1) if and only if it fulfills the KKT-condition (A.2).

A.2 The Lagrangian Dual Problem

Consider problem (A.1) where X may be closed. Then the Lagrangian dual function θ is defined as follows:

$$\theta(u, v) := \inf_{x \in X} \left[f(x) + \sum_{i \in I} u_i g_i(x) + \sum_{j \in J} v_j h_j(x) \right]. \tag{A.4}$$

Note that θ can attain $-\infty$; based on θ the following dual problem can be posed:

$$\max_{u \geq 0, v} \theta(u, v). \tag{A.5}$$

The following Theorem A.3 states that under the condition

$$\exists \bar{x} \in X \text{ such that } g(\bar{x}) < 0,\ h(\bar{x}) = 0,\ 0 \in \operatorname{int}\{h(x) \mid x \in X\} \tag{A.6}$$

the so called primal problem (A.1) has the same objective value as the derived Lagrangian dual problem (A.5).

Theorem A.3 *([7, Theorem 6.2.4]) Let $X \subset \mathbb{R}^n$ be a nonempty convex set, $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^{|I|}$ be convex, and let $h : \mathbb{R}^n \to \mathbb{R}^{|J|}$ be affine. Suppose the constraint qualification (A.6) holds. Then*

$$\inf\{f(x) \mid x \in X,\, g(x) \leq 0,\, h(x) = 0\} \quad = \quad \sup\{\theta(u, v) \mid u \geq 0,\, v\}.$$

Note that the constraint qualification (A.6) is closely related to the Slater condition. As a consequence of the previous theorem the following so called 'saddle point' criteria can be derived which is based on

$$\Phi(x,u,v) \;:=\; f(x) + \sum_{i\in I} u_i g_i(x) + \sum_{j\in J} v_j h_j(x). \tag{A.7}$$

The notion 'saddle point' is motivated from concavity of Φ in (u,v) for fixed x and convexity in x for fixed (u,v) forming the image of a saddle; the 'saddle points' are under appropriate conditions simultaneous solutions to both the primal and the dual problem.

Theorem A.4 *([7, Theorem 6.2.5]) Let $X \subset \mathbb{R}^n$ be a nonempty convex set, and let $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^{|I|}$ and $h : \mathbb{R}^n \to \mathbb{R}^{|J|}$. Suppose that there exist $\bar{x} \in X$ and $(\bar{u},\bar{v})$ with $\bar{u} \geq 0$, such that*

$$\Phi(\bar{x},u,v) \;\leq\; \Phi(\bar{x},\bar{u},\bar{v}) \;\leq\; \Phi(x,\bar{u},\bar{v}) \quad \forall x \in X,\ \forall (u,v) \text{ with } u \geq 0. \tag{A.8}$$

Then $\bar{x}$ solves the primal problem (A.1) and $(\bar{u},\bar{v})$ solves the dual problem (A.5).

Conversely, suppose that f and g are convex and that h is affine. Further, suppose that the constraint qualification (A.6) is satisfied. If $\bar{x}$ solves the primal problem (A.1) then there exists $(\bar{u},\bar{v})$ with $\bar{u} \geq 0$, such that (A.8) holds true.

Note that the set X may be chosen freely, that is, depending on the situation it can be $\mathbb{R}^n$ or can contain a (sub-)set of the constraints g_i, $i \in I$ or h_j, $j \in J$, a circumstance which is very useful when using the Lagrangian dual problem for decomposition. This situation differs from the Karush-Kuhn-Tucker-characterization of optima where openness of X is required. It is exactly here where equivalence of KKT-points and saddle-points holds: under convexity of f and g and affinity of h we have to require $\bar{x} \in \text{int}X$ in order to have equivalence of KKT-points and saddle-points (SP). Differently stated, if the constraint qualification (A.6) holds we have

$$\{(\bar{x},\bar{u},\bar{v})_{\text{KKT}} \mid \bar{x} \in \text{int}X\} \;=\; \{(\bar{x},\bar{u},\bar{v})_{\text{SP}} \mid \bar{x} \in \text{int}X\},$$

where the index KKT refers to Karush-Kuhn-Tucker points (i.e. points satisfying (A.2)) and SP to saddle-points (i.e. points satisfying (A.8)).

A.3 Differentiability and Continuity

Even though continuous functions are almost always almost nowhere differentiable (that is, in the set of continuous maps the ones which are differentiable on more than a set of measure zero have measure zero), they are in a sense arbitrarily 'close' to differentiable functions. Considering functions defined on a closed set D let us define the distance of two functions $f : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ by

$$||f-g||_\infty := \max_{x\in D} ||f(x)-g(x)||.$$

There is actually a plethora of differentiable functions which approach in this sense any (non-differentiable) continuous function. In the following theorem due to Weierstrass polynomials are chosen:

Theorem A.5 *(Weierstrass' approximation theorem) Let $D \subset \mathbb{R}^n$ be compact and $f : D \to \mathbb{R}^m$ be continuous. Then given any $\varepsilon > 0$ there is a polynomial $g : D \to \mathbb{R}^m$ such that*

$$||f-g||_\infty < \varepsilon.$$

A.4 An Introduction to Variational Inequality Problems (VIP)

As a start, two examples give a first intuition on variational inequalities (cf. [65]):

1. Let $a, b \in \mathbb{R}$, and $f : [a,b] \to \mathbb{R}$ be continuously differentiable. The points x_0: $f(x_0) = \min_{x\in[a,b]} f(x)$ have to be determined. Three cases may occur:
   (a) if $x_0 \in]a,b[$, then $f'(x_0) = 0$;
   (b) if $x_0 = a$, then $f'(x_0) \geq 0$;
   (c) if $x_0 = b$, then $f'(x_0) \leq 0$.

   These three cases can be understood as x_0 solves the variational inequality problem $f'(x_0)(x - x_0) \geq 0 \;\forall\, x \in [a,b]$.

2. Let $D \subset \mathbb{R}^n$ be a closed, convex set and $f : D \to \mathbb{R}$ be continuously differentiable with minimum $x_0 \in D$. Let $x \in D$ be an arbitrary point; then the function

   $$\Phi(t) := f(x_0 + t(x - x_0)), \qquad 0 \leq t \leq 1,$$

   attains its minimum at $t = 0$. From the first example follows

   $$\Phi'(0) := \nabla f(x_0)^T(x - x_0) \geq 0 \qquad \forall x \in D.$$

   Hence, any minimum x_0 fulfills the variational inequality $\nabla f(x_0)^T(x-x_0) \geq 0 \;\forall x \in D$.

A.4.1 Existence and Uniqueness of Solutions for VIPs

Definition A.3 *(Variational Inequality Problem, VIP(f, D), [65, problem 4.1]) Given $f : D \subset \mathbb{R}^n \to \mathbb{R}^n$; find $x \in D$, such that*

$$f(x)^T(y - x) \geq 0 \qquad \forall y \in D. \tag{A.9}$$

The existence of solutions to VIP(f, D) can be shown by means of fixed point arguments. We call h a *contraction mapping*, if $\|h(x) - h(y)\| \leq \alpha\|x - y\|$ for all $x, y \in D$, and some $\alpha \in [0, 1)$. In the case of $\alpha = 1$, h is called *non-expansive*. With this notion the following obvious contraction lemma can be formulated:

Lemma A.6 *([65, Theorem 1.2]) If $h : D \subset \mathbb{R}^n \to D$ is a contraction mapping, then there exists a unique fixed point of h.*

Note that D may be non-compact in the previous lemma. h being contracting can be replaced by continuity if D is additionally compact, cf. Brouwer's Fixed Point Theorem 2.2.

Finally, to bridge the gap between fixed points and variational inequalities, the concept of projection is needed. If $D \subset \mathbb{R}^n$ is a closed, convex subset, then for each $x \in \mathbb{R}^n$ there is a unique $y \in D$ with

$$\|x - y\| = \inf_{z \in D} \|x - z\|,$$

called the projection of x onto D, and written $y = P_D x$. Obviously, $P_D x = x \; \forall x \in D$. The following lemma characterizes projections.

Lemma A.7 *([65, Theorem 2.3]) If $D \subset \mathbb{R}^n$ is closed and convex, then $y = P_D x$ is the projection if and only if*

$$y^T(z - y) \geq x^T(z - y) \qquad \forall z \in D.$$

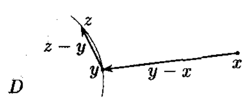


Figure A.1: Characterization of projection.

This can be written in the equivalent form $(y - x)^T(z - y) \geq 0 \; \forall z \in D$ and interpreted as 'y is the outermost point in D towards x', see Figure A.1. From Lemma A.7 it follows that the projection map is nonexpansive, i.e. $\|P_D x - P_D y\| \leq \|x - y\|$ for all $x, y \in \mathbb{R}^n$, and thus Lipschitz-continuous. After these preparations, a first existence result for VIP can be stated.

Theorem A.8 *([65, Theorem 3.1]) If $D \subset \mathbb{R}^n$ is convex and compact and $f : D \to \mathbb{R}^n$ is continuous, then there exists a solution to VIP(f, D).*

The proof of Theorem A.8 gives an idea of the relevance of projection and fixed points, therefore it is outlined here:

Multiplying (A.9) by -1 and adding $x^T(y - x)$ on both sides yields the equivalent relation

$$x^T(y - x) \geq (x - f(x))^T(y - x) \qquad \forall y \in D.$$

With $\mathbb{1}$ the identity mapping, $P_D \circ (\mathbb{1} - f) : D \to D$ is continuous. Hence, by Proposition 2.2, there exists a fixed point $x \in D$, i.e. $x = P_D(x - f(x))$. Together with the characterization of projection, this can be written as

$$x^T(y - x) \geq (x - f(x))^T(y - x) \qquad \forall y \in D$$

which is exactly the relation to be proved.

The proof shows that every VIP(f, D) is equivalent with the fixed point problem $x = P_D \circ (\mathbb{1} - f)(x)$, see Figure A.2.

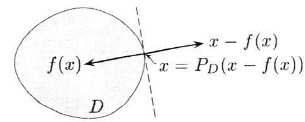


Figure A.2: VIP as fixed point problem.

In Theorem A.8, the existence of a solution was proven for bounded D. The example $f(x) = \exp(x)$ with $D = \mathbb{R}$ shows that this condition can not be dropped without further considerations.

Define $B_r(0)$, the ball with center 0 and radius r, and set $D_r := D \cap B_r(0)$. Then the following lemma gives a sharp condition for the existence of a solution in the case of unbounded D, see Figure A.3.

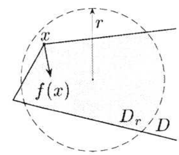


Figure A.3: Characterization of a solution to VIP(f, D) in case of unbounded D.

Lemma A.9 *([65, Theorem 4.2]) If $D \subset \mathbb{R}^n$ is convex and closed, and if $f : D \to \mathbb{R}^n$ is continuous, then there exists a solution to VIP(f, D) if and only if $\exists r > 0$, such that a solution x_r of VIP(f, D_r) satisfies $\|x_r\| < r$.*

A more useful condition is based on a notion called *coerciveness*; $f : D \subset \mathbb{R}^n \to \mathbb{R}^n$ is called coercive, if $\exists x_0 \in D$ with

$$\frac{(f(x) - f(x_0))^T(x - x_0)}{\|x - x_0\|} \to +\infty \qquad \forall x \in D, \|x\| \to \infty. \tag{A.10}$$

Rewriting (A.10) in the form $f(x)^T(x-x_0)/\|x-x_0\| - f(x_0)^T(x-x_0)/\|x-x_0\| \to +\infty$ shows that the projection of $f(x)$ onto the unit vector $(x - x_0)/\|x - x_0\|$ must grow to infinity, as $x \in D$ tends to infinity. As a simplified picture one can think of a vectorfield with all vectors $f(x)$ going away from x_0 and becoming longer the further away x is from x_0 (a star with growing rays).

Theorem A.10 *([65, Corollary 4.3]) If $f : D \subset \mathbb{R}^n \to \mathbb{R}^n$ is coercive, D is convex and closed, then VIP(f, D) has a solution.*

A solution to VIP(f, D) may not be unique. There is, however, a natural condition which ensures uniqueness. If x, x' are two solutions of VIP(f, D), we have

$$\left.\begin{array}{r} f(x)^T(y - x) \geq 0 \\ f(x')^T(y - x') \geq 0 \end{array}\right\} \quad \forall y \in D.$$

Setting $y := x'$ in the first and $y := x$ in the second relation and adding them yields

$$(f(x) - f(x'))^T(x - x') \leq 0.$$

Hence the existence of several solutions implies this last relation, or reversely, if this last relation is denied, there is at most one solution. This denial is called strict monotonicity and is defined below.

Definition A.4 *(cf. [91, 71]) f is called*
monotone *over D, if* $(f(x) - f(y))^T(x - y) \geq 0 \;\; \forall x, y \in D$,
pseudo-monotone *over D, if* $[\, f(y)^T(x - y) \geq 0 \Rightarrow f(x)^T(x - y) \geq 0\,] \; \forall x, y \in D$,
strictly monotone *over D, if* $(f(x) - f(y))^T(x - y) > 0 \;\; \forall x, y \in D$,
strongly monotone *over D, if* $(f(x) - f(y))^T(x - y) \geq \alpha\|x - y\|^2 \;\; \forall x, y \in D$, $\alpha > 0$, *and*
strongly f-monotone *over D, if* $(f(x) - f(y))^T(x - y) \geq \alpha\|f(x) - f(y)\|^2 \;\; \forall x, y \in D$, $\alpha > 0$.

To get some intuition, note that under suitable assumptions the notion of monotonicity of f is equivalent with convexity of some F where $f \equiv \nabla F$, see Proposition 3.1.

A.4.2 Optimization and VIP

This section relates the problem of minimizing a function $F : D \subset \mathbb{R}^n \to \mathbb{R}$ to $\mathrm{VIP}(\nabla F, D)$.

Proposition A.11 *Let D be convex, let $F : D \to \mathbb{R}$ be once continuously differentiable, and set $f(x) := \nabla F(x)$; then we have*

(i) ([65, Proposition 5.1]) x solves VIP(f, D) if $F(x) = \min_{y \in D} F(y)$;

(ii) ([65, Proposition 5.2]) suppose F is convex and x solves VIP(f, D), then x satisfies $F(x) = \min_{y \in D} F(y)$;

(iii) (cf. [56, Satz 4.2]) $F(x) = \min_{y \in D} F(y)$ *if* $f(y)^T(y - x) \geq 0 \; \forall y \in D$.

While part (i) and (ii) are rather obvious as can be seen from the first example in this section, the sufficiency condition presented in (iii)—which can be difficult to check—is only rarely found in the literature. The points x satisfying $f(y)^T(y - x) \geq 0 \; \forall y \in D$ are called 'weak solutions' of VIP(f, D).

In the previous proposition optimization problems are related to certain equivalent VIPs. Using so called gap functions, discussed in Section 3.2, *every* VIP can be transformed into an optimization problem. Due to the specificities of the gap functions, however, these optimization problems are in general not tractable.

Appendix B

Proving the Existence of Equilibria using VIPs

For point-to-point and continuous excess maps Theorem 2.1 and Proposition 2.2 guarantee the existence of an equilibrium. Here we want to discuss the non-continuous or multi-valued case, which allows to regain the proof of existence of a solution to EEP (see Definition 1.4) under relaxed assumptions. It requires, however, to leave the aggregated view of an excess map and to analyze the underlying structures of the economic agents.

This chapter is based on Yao [103], where proofs of existence of solutions to—in a certain sense extended—VIPs are discussed and subsequently applied to economic equilibrium problems.

The contribution lies in the bridging between the general formulation due to Yao, and the specific situation of our economy described in Chapter 1.

As usual in the context of VIPs, the notion 'generalized' applies to the situation where the operator is set-valued; a further extension is achieved with the so called generalized *quasi*-variational inequality problems (GQVIP), where the feasibility set is variable.[1] To clarify the notation, all v- and a-related quantities have the meaning from Section 2.3, whereas f, x and y are general quantities and do not refer to any previous usage. Otherwise the situation described in Section 1.1 is assumed.

Definition B.1 *(Generalized quasi-variational inequality problem (GQVIP(f, K, D)), [12]) Let $D \subset \mathbb{R}^n$, $f : D \to 2^{\mathbb{R}^n}$, and $K : D \to 2^D$; find $x \in D$ and $\xi \in f(x)$, such that*

$$\xi^T(y - x) \geq 0 \qquad \forall y \in K(x). \tag{B.1}$$

[1] Note that there is also a notion called quasi-monotone which relates to quasi-convexity in the same way as pseudo-monotonicity to pseudo-convexity, cf. Proposition 3.1.

If the operator f is single valued the problem is called QVIP(f, K, D). The relevance of this formulation can be seen from the consumer problem (1.5), where the feasibility set depends on the price and production, which are determined by other agents—the producers and the price agent in the more general setting of problem (2.9). Recalling the splitting of variables $v = (v_a, v_{\bar{a}})$ for an agent $a \in A$, we assume that the feasible set depends on the action of all other agents, i.e. we require $v_a \in K_a(v_{\bar{a}})$, which is a notational simplification for (2.7) and (2.8). Let us further assume that for each agent $a \in A$ there exists V_a such that $K_a(v_{\bar{a}}) \subset V_a$ for all $v_{\bar{a}} \in V_{\bar{a}}$.

In view of Lemma 2.7, rewriting (2.9), we call a point v^* an *equilibrium* of this abstract economy if for all $a \in A$

$$v_a^* \in K_a(v_{\bar{a}}^*) \quad \text{and} \quad f_a(v_a^*) = \max_{v_a \in K_a(v_{\bar{a}}^*)} f_a(v_a, v_{\bar{a}}^*). \tag{B.2}$$

Consequently we abbreviate such an abstract economy by $[f_a, K_a, V_a]_{a \in A}$, and write furthermore $F = (-\nabla_{v_a} f_a)_{a \in A}$, $K := \Pi_{a \in A} K_a : V \to 2^V$ and $V := \Pi_{a \in A} V_a$.

In the sequel f is a general notion for a mapping and can be real or vector valued, its meaning should always be clear from the context. F on the other hand is reserved for the compound negative gradient of the pseudo-concave objective functions of the agents. To use the VIP-based machinery, the following lemma, which slightly generalizes Lemma A.11, relates optimization problems to VIPs.

Lemma B.1 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be pseudo-concave and differentiable, $D \subset \mathbb{R}^n$ nonempty, closed and convex;*

(a) if x^* solves VIP$(-\nabla f, D)$ it is also a solution to $\max_{x \in D} f(x)$*;*

(b) if ∇f is continuous and x^* is a solution to $\max_{x \in D} f(x)$*, then x^* solves also VIP$(-\nabla f, D)$.*

To see (a) note that a solution to VIP$(-\nabla f, D)$ is characterized by $-\nabla f(x^*)^T(z - x^*) \geq 0 \; \forall z \in D$ and that by pseudo-concavity $f(x^*) \geq f(z) \; \forall z \in D$.

To prove (b) assume $x^* \in \operatorname{argmax}_{x \in D} f(x)$ and x^* does not solve VIP$(-\nabla f, D)$, that is, there exists $z^* \in D : \; \nabla f(x^*)^T(z^* - x^*) > 0$. By continuity of ∇f and convexity of D there exists $\delta \in (0, 1]$ such that $\nabla f(x(t))^T(z^* - x^*) > 0 \; \forall t \in [0, \delta]$ where $x(t) := x^* + t(z^* - x^*)$. But then

$$f(x(\delta)) - f(x^*) \; = \; \int_0^\delta \nabla f(x(t))^T(z^* - x^*) \, dt \; > 0$$

and so $x^* \notin \operatorname{argmax}_{x \in D} f(x)$. This contradiction proves that x^* has to solve also VIP$(-\nabla f, D)$. ∎

Using this lemma it is sufficient to find a solution to VIP in order to have a solution to the maximization problem. The reverse requires continuity of the gradient map.

The next lemma gives a trivial characterization of the solution set of VIPs which will be used in the theorems below.

Lemma B.2 *x^* is a solution to VIP($-\nabla f, D$) if and only if*

$$x^* \in \{x \in D \mid \sup_{z \in D} -\nabla f(x)^T(x - z) \leq 0\}.$$

x^* is a solution to VIP($-\nabla f, D$) if and only if

$$\begin{aligned} -\nabla f(x^*)^T(z - x^*) \geq 0 \quad \forall z \in D &\iff \inf_{z \in D} -\nabla f(x^*)^T(z - x^*) \geq 0 \\ &\iff \sup_{z \in D} -\nabla f(x^*)^T(x^* - z) \leq 0. \end{aligned}$$

The situation is now as follows; in order to find for a specific $a \in A$ a solution to the maximization problem (B.2), the corresponding VIP($-\nabla_{v_a} f_a(v_a; v_{\bar{a}}), K_a(v_{\bar{a}})$) can be solved. Based on that we expect that a solution of QVIP(F, K, V) is a simultaneous solution of the maximization problems (B.2) for all $a \in A$, that is, an equilibrium of the abstract economy.

Lemma B.3 *If v^* solves QVIP(F, K, V), where $F(v) := (-\nabla_{v_a} f_a(v))_{a \in A}$, then v^* is a simultaneous solution to (B.2) for all $a \in A$.*

Let v^* be a solution to QVIP(F, K, V); assume that there exists $a \in A$ for which v_a^* is not a solution to VIP($-\nabla_{v_a} f_a, K_a(v_{\bar{a}}^*)$), i.e. $\exists z_a' \in K_a(v_{\bar{a}}^*)$ such that $-\nabla_{v_a} f_a(v^*)^T(z_a' - v_a^*) < 0$. Let $z := (z_a', v_{\bar{a}}^*)$; then we have

$$\sum_{a \in A} -\nabla_{v_a} f_a(v^*)^T(z_a - v_a^*) \;=\; -\nabla_{v_a} f_a(v^*)^T(z_a' - v_a^*) \;<\; 0$$

which contradicts v^* being a solution. This contradiction proves that a solution to QVIP(F, K, V) is also a solution to all its subproblems VIP($-\nabla_{v_a} f_a, K_a(v_{\bar{a}}^*)$) and with Lemma B.1 the claim follows.

A function $g : \mathbb{R}^n \to \mathbb{R}$ is said to be lower semi-continuous (l.s.c.) if the set $\{x \mid g(x) \leq \alpha\}$ is closed for any $\alpha \in \mathbb{R}$, or equivalently, for all convergent sequences $x^n \to x$ we have $g(x) \leq \liminf_{n \to \infty} g(x^n)$. Next, we say a (set-valued) function $f(v)$ has convex values (or is convex valued), if the set $f(v)$ is convex.

Now the prerequisites are ready for proving existence of an equilibrium based on the following general existence result for GQVIP.

Theorem B.4 *([103, Theorem 3.3]) Let $D \subset \mathbb{R}^n$ be nonempty compact and convex, $f : D \to 2^{\mathbb{R}^n}$ and $K : D \to 2^D$. Suppose that the following conditions are satisfied:*

1. *The mapping f has nonempty, compact and convex values, and for every fixed $y \in D$ the mapping*

$$x \mapsto \inf_{\xi \in f(x)} \xi^T(x-y)$$

is lower semi-continuous on D;

2. *K has nonempty, closed and convex values, and for every fixed $p \in \mathbb{R}^n$ the set $\{x \in D \mid p^T x \leq \sup_{y \in K(x)} p^T y\}$ is closed;*

3. *the so called interacting set $\{x \in D \mid \sup_{y \in K(x)} \inf_{\xi \in f(x)} \xi^T(x-y) \leq 0\}$ is closed.*

Then there exists a solution to GQVIP(f, K, D).

Assuming $|A| < \infty$, the following existence theorem for the abstract economy can be deduced from the previous theorem.

Theorem B.5 *([103, Theorem 6.1]) Consider an abstract economy $[f_a, K_a, V_a]_{a \in A}$ where $V = \Pi_{a \in A} V_a \subset \mathbb{R}^n$; suppose that the following conditions are satisfied:*

1. *V_a is nonempty, compact and convex for all $a \in A$;*

2. *f_a is pseudo-concave with respect to v_a for all $a \in A$;*

3. *for every fixed $y \in V$ the mapping $x \mapsto F(x)^T(x-y)$ is lower semi-continuous on V, where $F(v) = (-\nabla_{v_a} f_a(v))_{a \in A}^T$;*

4. *$K_a(v_{\bar{a}}) \subset V_a$ is nonempty, closed and convex for all $a \in A$, and for every fixed $p \in \mathbb{R}^n$ the set $\{v \in V \mid p^T v \leq \sum_{a \in A} \sup_{z \in K_a(v_{\bar{a}})} p_a^T z\}$ is closed, where p_a denotes the part of p related to agent a;*

5. *the interacting set $\mathcal{I} := \{v \in V \mid \sup_{z \in K(v)} F(v)^T(v-z) \leq 0\}$ is closed.*

Then there exists an equilibrium.

The proof is a direct application of Theorem B.4 using the relation between optimization problems and VIP given in Lemma B.1. Note that Condition 3. is trivially satisfied if f_a is not only differentiable but *continuously* differentiable, and that the so called interacting set is exactly the set of solutions to QVIP(F, K, V).

To apply this general framework to the situation in Section 1.1 we have to relate closedness in Condition 4. and 5. to properties of K. First note, however, that for all fixed $v \in V$ we have from $|A| < \infty$ and additivity of the objective

$$\sup_{z \in K(v)} p^T z = \sup_{z \in K(v)} \sum_{a \in A} p_a^T z_a = \sum_{a \in A} \sup_{z_a \in K_a(v_{\bar{a}})} p_a^T z_a.$$

Lemma B.6 *Let $V \in \mathbb{R}^n$ be compact and $K : V \to 2^V$ have nonempty, closed and convex set values for all $v \in V$. If K is closed then for every $p \in \mathbb{R}^n$ the set $\mathcal{P} := \{v \in V \mid p^T v \leq \sup_{z \in K(v)} p^T z\}$ is closed.*

Assume K is closed (cf. Definition 2.1); take any $p \in \mathbb{R}^n$ and convergent sequence $\{v^n\} \subset \mathcal{P}$, $v^n \to v^\infty$. We have to show that $v^\infty \in \mathcal{P}$. Choose a sequence $z^n \in \text{argmax}_{z \in K(v^n)} p^T z$. Because V is compact $\{z^n\}$ has a convergent subsequence which, without loss of generality, is assumed to be $\{z^n\}$ and converges to z^∞. We then have the following chain of inequalities,

$$p^T v^\infty = \lim_{n \to \infty} p^T v^n \leq \lim_{n \to \infty} \sup_{z \in K(v^n)} p^T z = \lim_{n \to \infty} p^T z^n = p^T z^\infty \leq \sup_{z \in K(v^\infty)} p^T z,$$

where the second inequality is due to $v^n \in \mathcal{P}$, and the last inequality follows from closedness of K which implies $z^\infty \in K(v^\infty)$. Thus $v^\infty \in \mathcal{P}$ and so $\mathcal{P}$ is closed.

Lemma B.7 *Let $V \in \mathbb{R}^n$ be compact and $K : V \to 2^V$ have nonempty, closed and convex values for all $v \in V$. If K is open and F continuous, then the interacting set*

$$\mathcal{I} := \{v \in V \mid \sup_{z \in K(v)} F(v)^T (v - z) \leq 0\}$$

is closed.

Take any convergent sequence $\{v^n\} \subset \mathcal{I}$, $v^n \to v^\infty$; we have to show that $v^\infty \in \mathcal{I}$. Let $z^\infty \in \text{argmax}_{z \in K(v^\infty)} F(v^\infty)^T (v^\infty - z)$ which exists due to nonempty and compact values of $K(v^\infty)$ plus continuity of F. From openness of K follows the existence of a sequence $\{z^n\}$, such that $z^n \in K(v^n)$ $\forall n \in \mathbb{N}$ and z^n converges to z^∞. We then have the following chain of inequalities:

$$\begin{aligned}
\sup_{z \in K(v^\infty)} F(v^\infty)^T (v^\infty - z) &= F(v^\infty)^T (v^\infty - z^\infty) \\
&= \lim_{n \to \infty} F(v^n)^T (v^n - z^n) \\
&\leq \lim_{n \to \infty} \sup_{z \in K(v^n)} F(v^n)^T (v^n - z) \\
&\leq 0.
\end{aligned}$$

The second relation follows from convergence of both $\{v^n\}$ and $\{z^n\}$ together with continuity of F, the last is a consequence of $v^n \in \mathcal{I}$ $\forall n \in \mathbb{N}$.

Recalling the comments at the end of Section 2.1, we see that by the previous two lemmata we ensure a continuous feasible set map K which, together with continuity of f_a for all $a \in A$, implies a closed optimal set map, and this is exactly what is required for applying Kakutani's Theorem. In that sense Theorem B.5 is a generalization of Kakutani's existence Theorem for this class of problems. If we assume that K is defined by sets of (in-)equalities, closedness of K follows already from continuity of the (in-)equalities. However, openness of K requires usually constraint qualifications to hold, see Flippo [30, Theorem 3.1–3.4]. In order to verify openness of K the following lemma reduces openness of the overall feasibility set map K to openness of the feasibility set map K_a for each agent $a \in A$.

Lemma B.8 *$K: V \to 2^V$ is open if and only if $K_a: V \to 2^{V_a}$ is open for all $a \in A$.*

'$\Longrightarrow$': Choose $a \in A$ and a convergent sequence $v^n \to v^\infty$ in V, and choose $z_a^\infty \in K_a(v^\infty)$. Because $K = \Pi_{a \in A} K_a$ there exists $z^\infty \in K(v^\infty)$ satisfying $(z^\infty)_a = z_a^\infty$, where $(\cdot)_a$ symbolizes the vector of a-related components; from openness of K we conclude that there is a convergent sequence $z^n \to z^\infty$ with $z^n \in K(v^n)\ \forall n \in \mathbb{N}$. But by the definition of K we have then $(z^n)_a \in K_a(v^n)\ \forall n \in \mathbb{N}$ and from convergence of $z^n \to z^\infty$ we conclude convergence of its part $(z^n)_a \to (z^\infty)_a = z_a^\infty$.

'$\Longleftarrow$': Choose a convergent sequence $v^n \to v^\infty$ in V, and choose a $z^\infty \in K(v^\infty)$. Because K_a is open for all $a \in A$ there exist convergent sequences $z_a^n \to (z^\infty)_a$ where $z_a^n \in K_a(v^n)$. From this construction we conclude $z^n := \Pi_{a \in A} z_a^n \in K(v^n)$ and $z^n \to z^\infty$.

The analogue, where closedness of K is equivalent with closedness of K_a for all $a \in A$, is straightforward. One might hope that, due to the specific structure of our K deduced from (2.9), closedness of the interacting set $\mathcal{I}$ is given. This is in general not true as is demonstrated by the example depicted in Figure B.2. But if we are more modest and restrict our consumer and producer agents to (1.4) and (1.5) we succeed.

Lemma B.9 *The feasibility map K_a is open for the problems (1.4), (1.5) and (2.3).*

The problems (1.4) and (2.3) have a constant feasibility map and so they are open. Problem (1.5)—though most simple in the structure of its feasibility map—is not so obvious. Without loss of generality we assume there is exactly one consumer and producer and drop therefore the corresponding indices; the feasibility map is

$$K(y,p) = \{x \mid p^T(x - x^0 - y) \leq 0\}.$$

Let $(y^n, p^n) \to (y^\infty, p^\infty)$ be a convergent sequence and choose $x^\infty \in K(y^\infty, p^\infty)$. Then the distance between x^∞ and $K(y^n, p^n)$ is

$$\delta(x^\infty, K(y^n,p^n)) = \max\{0, \frac{p^{nT}}{\|p\|}(x^\infty - x^0 - y^n)\}.$$

Because $p \in \Delta$ (the unit simplex), δ is continuous in (y^n, p^n), and by choosing the minimizer of the distance function in $K(y^n, p^n)$ as x^n, we have found a convergent sequence with $x^n \in K(y^n, p^n)$ and $x^n \to x^\infty$.

From the previous lemma we have an open feasibility map, closedness follows from continuity of the underlying restrictions in the setting of Section 1.1, and so we have the following corollary of Theorem B.5.

Corollary B.10 *The equilibrium problem based on Definition 1.4 and its related definitions and assumptions in Section 1.1 has a solution.*

The adverse of this bright view shows up as soon as the feasibility maps are more complicated. Assume for example that a consumer has some additional constraints which are even independent of prices or production quantities. To demonstrate the difficulties and highlight the intimate relation of l.s.c. of the value map and closedness of the feasible set map, Figure B.2 shows a 2-dimensional simple example. The feasibility set (gray shaded) is $[0,1]^2$, further cut by a rotating constraint $((x_1, x_2) - (1, \frac{1}{2}))(-1, p)^T \leq 0$. The problem is to maximize $\frac{1}{4} - [(x_1-1)^2 + (x_2-1)^2]$ whose level curves form circles and are partly dotted drawn. Let $p^n = \frac{1}{n}$ form a sequence converging to 0; the resulting feasible set $K(p^n)$ shrinks continuously maintaining always a non-empty interior. The solution of the maximization problem for all p^n stays at $x^n = (1, \frac{1}{2})$ with constant objective value 0. In the limit $p^\infty = 0$, however, $K(p)$ expands non-continuously to $1 \times [0,1]$, and the solution value jumps to $\frac{1}{4}$. Openness of K forbids exactly this non-continuous growth of K, and as can be seen from this example, this condition can in general not be weakened.

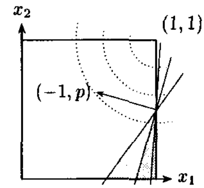


Figure B.2: A non-open $K(p)$.

Appendix C

The Original Markal and Markal-Macro Models

For some general comments on energy economy models see Section 6.2. Introductory notes are also given at the beginning of Chapter 7. Here we restrict ourselves to giving an overview on the linear energy-model Markal in the first section and a more detailed description of the nonlinear macroeconomic model Macro in the second section.

C.1 Markal

The linear model Markal (**Mar**ket **Al**location) was mainly developed in the late seventies by Fishbone *et al* [27, 28] at BNL[1] in collaboration with and ordered by the ETSAP-group[2], which is itself an outcome of IEA. Originally written in the language OMNI, Goldstein [42] translated the model in the early nineties into GAMS[3]. The purpose of this section is to give an idea of the structure of Markal without going into details which can be found e.g. in Kypreos [68].

The fundamental concept of this model is the so called Reference Energy System (RES), cf. Figure C.1. RES is a process-oriented flow-chart covering all possible connections between primary energy sources and final energy services. It contains a rich set of intermediate nodes for transformations, and on its edges all kind of economical, technical and ecological information are attached. The solution process in Markal can thus be interpreted as a search for a minimal cost flow in the feasible set of paths admitted by RES, satisfying the given demand in the sinks. The resulting Markal is a dynamic linear programming (LP) problem, which, given exogenously the demand for a number of different energy services for the different time periods, searches for the cheapest possibility to satisfy the

[1]Bruckhaven National Laboratory in the USA.
[2]Energy Technology Systems Analysis Project.
[3]General Algebraic Modeling System.

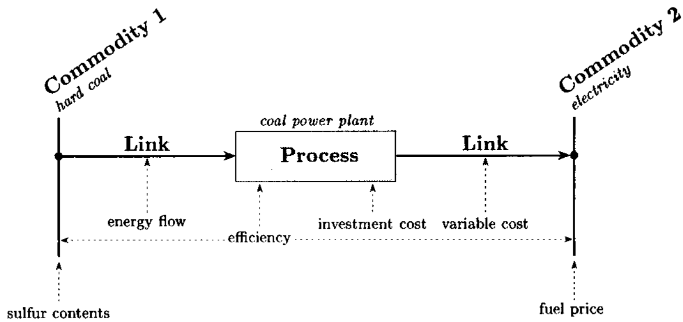


Figure C.1: Objects of the Reference Energy System.

demand. It comprises a description of the available and—in the near future—expected technologies to become available. Depending on the number of time periods and the details in the technological description, it can contain up to 5,000–10,000 activities and constraints respectively.

The required demand for energy services can be generated exogenously by appropriate models, or endogenously by linking Markal to a macroeconomic model like Macro. In case of Switzerland the demand generator used is called SMEDE[4] and is an adaption and implementation of MEDEE-S, originally developed by B. Lapillonne at the *Institut Economique et Juridique de l'Energie* at *Université des Sciences Sociales de Grenoble.* SMEDE computes, based on technological, economical and social data, the demand for the different energy services for a specified time period. Hence its results depend crucially on the scenarios given, like the economic growth rate in the future, demographic development, and many more. The model Macro, which is an alternative way to generate the demand endogenously, is described in the next section.

The activities of Markal can be subdivided into three groups:

Capacities: Reserves (e.g. oil) and capacities of various technologies or plants.

Activities: Annual production of all processes (e.g. electricity or heat).

Energy-resources: Annual consumption of energy carriers.

The restrictions can be grouped as follows:

Capacity-transfer: These inter-period constraints connect the available capacities with foregoing investments and depreciation.

[4]Swiss MEDE, see Kypreos [67].

Demand: Supply must be greater or equal demand for each energy service in each period.

Fuels budget: The sum of imports, depletion of reserves or resources, and the production must meet the consumption of the plants and of other demand sources. This holds for all fuels and time periods.

Electricity budget: Analogous to the above budget of fuels.

Heat budget: Extending the previous two budgets the regional structure is taken into account for the heat.

Load-restrictions: Guarantee coverage of peak demand with existing capacity.

Plant structure: Models time for maintenance of plants, or period dependent changes in capacities like river power plants.

Investment- and resource-usage: Period-specific investment- and plant-capacity-restrictions are in this category as well as time related constraints for introducing new technologies.

Rest: All other restrictions; this can be region specific, like the usage of electric heaters in Switzerland which is regulated by law and restricted by the capacity of the net.

The overall (discounted) cost include all variable costs (e.g. for buying fuels), fixed costs (e.g. for building plants), and finally the so called 'salvage costs' which account for the problem when a plant is not yet at the end of its life time when the model reaches its time horizon.

C.2 Markal-Macro (MM)

Markal-Macro is a synthesis between the bottom-up engineering model Markal described in the previous section and a top-down macroeconomic model called Macro. Analog to the previous section the presentation of Macro is limited; more details can be found in [76, 77, 73, 75, 74]. MM allows to investigate the relationship between economic growth, demand for energy services and the structure of the energy sector. Macro is supply-oriented, i.e. it is assumed that the (aggregated) production is fully consumed. The author of Macro motivates it as follows:

> [...] macroeconomic models, with their descriptions of effects within the total economy but fewer technical details on the energy system, tend to overestimate future energy demands[.] Conversely, [...] engineering models, ignoring feedbacks to the general economy and non technical market factors but containing rich descriptions of technology options, tend to take to optimistic a view of conservation and the use of renewable energy resources [...] *Manne and Wene [76]*

Of special interest with regard to such a hybrid model is the influence of changing energy/fuel prices or limiting ecological effects on macroeconomic indicators like the GDP (gross domestic product). Figure C.2 sketches the overall structure of MM.

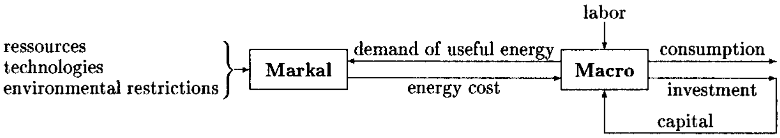


Figure C.2: The model Markal-Macro following Manne and Wene [76].

Markal can thus be understood as an oracle which, given the demand for useful energy by Macro, returns the cost in each period. Because the costs are minimal while all energy services are fully 'consumed', MM can be seen as a partial equilibrium model in the energy sector. In this setting Macro is the master-program representing

> [...] a macroeconomic model with an aggregated view of long-term economic growth. The basic input factors of production are capital, labor and individual forms of energy. The economy's outputs are used for investment, consumption and interindustry payments for the cost of energy. Investment is used to build up the stock of capital. The model clearly distinguishes between autonomous and price–driven conservation. *Manne and Wene [76]*

In the following detailed description we assume 10 years per period t and denote by T the set of time periods. The *objective function* of (Markal-)Macro is called utility and defined as logarithm of consumption. One of the most relevant *exogenously* determined quantity is labor L_t, further exogenous coefficients not discussed here include ρ, α, a, b_d, grow, k, $\mathrm{aeeifac}_{d,t}$, $\mathrm{supply}_{j,d,t}$, $\mathrm{cost}_{j,t}$, c, c_χ, expf, ... On the top level of Macro we find the *variables* C_t (consumption), I_t (investment) and EC_t (energy-cost), followed by K_t (capital) and $D_{d,t}$ (demand for energy-service of kind d). Finally, the variable $\mathrm{XCAP}_{\chi,t}$ permits to use the technology χ beyond its availability but penalizes it by additional cost.

Even though all quantities have a period index $t \in T$, they relate always to one year. E.g. C_t is the mean consumption per year of period t, and I_t is the investment per year in the average of period t.

Macro consists of only 4 *restrictions:*[5] USE_t (usage of production), PRD_t (production), CAP_t (capital accumulation) and TC (terminal condition). In addition to these economic constraints, relations describing the link between Markal and Macro are needed. While the first $T-1$ periods comprise 10 years, the last period

[5] But—as in Markal—each restriction is repeated for every period, thus there are theoretically about $4|T|$ restrictions. In the implementation, however, this number is reduced by using some obvious algebraic simplification possibilities.

T represents the rest of the time horizon to infinity. The objective is to maximize 'utility'[6]

$$U(C) := \sum_{t=1}^{T-1} udf_t \log C_t + udf_T \frac{1}{1-(1-udr_T)^{10}} \log C_T, \tag{C.1}$$

where $udf_t = \prod_{\tau=0}^{t-1}(1-udr_\tau)^{10}$ is the utility discount factor for period t computed from the average annual utility discount rate in period τ, udr_τ. The exponent represents the number of years per period. The fraction in the last summand stems from the summation of a geometric sequence $\sum_{n=0}^{\infty} q^n = 1/(1-q)$. Hence this implicit terminal condition assumes a constant growth in the future leading to a higher weight of consumption in the last period of the optimization problem.

Next, the constraint USE_t distributes production Y_t on consumption, investment and energy cost:[7]

$$Y_t = C_t + I_t + EC_t. \tag{USE$_t$}$$

The production is determined by a nested CES-function[8] of the form

$$Y_t = \left[aK_t^{\rho\alpha}L_t^{\rho(1-\alpha)} + \sum\nolimits_d b_d D_{d,t}^{\rho}\right]^{1/\rho}. \tag{PRD$_t$}$$

a, b_d, ρ and α are coefficients, K_t is the capital stock accumulated up to period t, L_t is the Labor(-potential) in period t, and $D_{d,t}$ is the demand for energy services of form d in period t. Thus, production is determined on the first level by a capital-labor-aggregate and different energy services. On the next lower level the capital-labor-aggregate connects capital and labor in a Cobb–Douglas function fixing the elasticity of substitution between capital and labor to 1. Here, α can be interpreted as optimal share of capital in the aggregate. Price-induced energy savings are essentially determined by σ, the elasticity of substitution between energy and the capital-labor-aggregate. It holds $\sigma = 1/(1-\rho)$, see Chiang [13].

The previous two relations, (USE_t) and (PRD_t), assume implicitly that the gross value of energy services is captured in Y_t, whereas the outlay EC_t must be subtracted explicitly in a second step to gain the net production from energy services.

The long-term economic growth is mainly determined by the exogenously given labor supply L_t and its productivity, cf. Figure C.2. Initially, L_0 is set to 1 and subsequently increased following

$$L_{t+1} = (1+\mathrm{grow})^{10} L_t, \tag{L$_{t+1}$}$$

where grow is the potential growth rate of the economy. As mentioned, we have chosen for the ease of exposition 10 years per period, hence the exponent of 10.

[6]More precisely, $U(C)$ is a utility *index*; the absolute quantity $U(C)$ has no direct economic interpretation, hence $U(C)$ is no cardinal utility but only an ordinal index.

[7]As a notational convenience we use the gams-related notion that an equation has to be repeated for each meaningful occurrence of indices which are given in the tag. As an example the tag (USE_t) indicates that the corresponding equation is repeated for all $t \in T$.

[8]*Constant elasticity of substitution.*

Having chosen α and ρ, the quantities a and b_d are determined by calibrating the model using real data.

On the one hand, capital K_t is accumulated by investments I_t, on the other hand it is depreciated by a given annual capital depreciation factor k:

$$K_{t+1} = (1-k)^{10}K_t + 5((1-k)^{10}I_t + I_{t+1}). \qquad (\mathrm{CAP}_{t+1})$$

The quantity $5((1-k)^{10}I_t + I_{t+1})$ permits a better estimation of the *mean* investment in period t. Initially, we set $I_0 = (\mathrm{grow} + k)K_0$.

Finally, the following terminal condition guarantees reasonable investments in the last period:

$$K_T(grow + k) \leq I_T. \qquad (\mathrm{TC})$$

This 'primal' terminal condition reduces some of the effects which are caused by the finiteness of the time horizon.

The model description is completed by outlining the connection between Markal and Macro. As mentioned above, Markal requires the energy services to be given exogenously, or to state it reversely: the link between Macro and Markal has to generate the demand for energy services from the state of Macro. Let X_j be an activity of Markal supplying useful energy of the form d proportional to $\mathrm{supply}_{j,d}$. With the 'autonomous energy efficiency improvements factor' $\mathrm{aeeifac}_d$,[9] the demand constraints for Markal are for all meaningful combinations (d,t) given by

$$\sum\nolimits_j \mathrm{supply}_{j,d,t} X_{j,t} = \mathrm{aeeifac}_{d,t} D_{d,t}. \qquad (\mathrm{C.2})$$

To transfer the costs from Markal to Macro the link computes for each activity and period the cost $\mathrm{cost}_{j,t}$ per unit of activity $X_{j,t}$. A first approach is

$$\sum\nolimits_j cost_{j,t} X_{j,t} = EC_t;$$

because the accelerated introduction of technological capacities is possible but penalized, a quadratic term is added,

$$\sum\nolimits_j cost_{j,t} X_{j,t} + c \sum\nolimits_\chi c_\chi XCAP_{\chi,t}^2 = EC_t. \qquad (\mathrm{C.3})$$

Here $\mathrm{XCAP}_{\chi,t}$ is the amount of capacity installed beyond the capacity expansion factor expf. Therefore the last constraint needed is

$$CAP_{\chi,t+1} \leq expf\ CAP_{\chi,t} + XCAP_{\chi,t+1}. \qquad (\mathrm{C.4})$$

[9] $\mathrm{aeeifac}_d$ allows to account for energy saving effects induced by the general technological or social development. As mentioned above there is also a price-induced energy saving possibility available by the substitution of energy with the capital-labor-aggregate in the production function.

Putting the pieces together, MM can be written as follows:

$$\left.\begin{array}{l} \max U(C) \\ \quad \text{s.t. } (C.1)\text{–}(C.4),\ (\mathrm{USE}_t),\ (\mathrm{PRD}_t),\ (\mathrm{L}_{t+1}),\ (\mathrm{CAP}_{t+1}),\ (\mathrm{TC}), \\ \qquad \text{(all other Markal-Macro constraints).} \end{array}\right\} \quad (C.5)$$

By substituting (PRD_t) and (USE_t) into the objective function (C.1), and by further linearizing (C.3), the whole set of constraints of MM can be kept linear. This supports efficient solution techniques and is exploited e.g. in Minos5, the solver used. Convexity of Markal-Macro is shown in Appendix D

C.2.1 Discussion of Some Aspects

In the objective function the logarithm of consumption $\log(C_t)$ is chosen and not, say, simply C_t, GDP or GNP, because choosing a linear function as the objective has a strong tendency to produce 'bang-bang' solutions. For example, consume nothing and invest everything in all periods prior to the end of the horizon; then consume everything at the end of the horizon. Furthermore, most applied general equilibrium modelers focus on consumption rather than GDP because GDP includes investment as well as consumption. Investment is like other costs of doing business usually viewed as a means to an end—not an end in itself. Next, a nested Cobb-Douglas function within a CES production function is chosen instead of just a one-level CES-function where all production factors are equally treated, because the former permits the handling of two basic 'facts': (i) the elasticity of substitution between capital and labor is usually estimated as something close to unity; and (ii) the price elasticity of demand for energy is usually estimated as something a good deal lower than unity. Such a production function is therefore chosen not only in Markal-Macro, but also in ETA-Macro and Global2100.

Among the shortcomings of Macro are (i) the aggregation of the economy which inhibits the analysis of distributional effects, e.g. between economic sectors or social groups; (ii) a disproportion between the detailed energy-technology modeling part and the aggregated macroeconomic part; (iii) investments are not modeled in a so called vintage-like manner, but are free to be de-invested in later periods; (iv) the determination of the GDP suffers from not differencing input cost and added value in the energy sector.

Appendix D

Proof of Existence of an Equilibrium for MM^{mr}

The existence of an equilibrium solution to MM^{mr} will be given using the path-following approach described in Section 2.3. The other two possibilities discussed in Section 2.2 and Appendix B can not be directly applied for the following reasons. On the one hand the Negishi-concept requires Assumptions 1.1 and 2.1 to hold which can not be easily verified in case of MM^{mr}, even though Section 1.2 relates the structure in Definition 1.4 to the formulation of MM^{mr}. On the other hand, the VIP-based approach from Appendix B requires closedness of the so called interacting set $\mathcal{I}$ (Theorem B.5) which can be achieved by openness of the feasible set map K (Lemma B.7 and B.8). On the background of the related discussion in Flippo [30] one can not expect to easily verify this openness for general agents like the ones in MM^{mr}, cf. the example page 111 depicted in Figure B.2.

Contributions of this chapter include the adapted application of the general path-following concept to MM^{mr} for proving the existence of an economic equilibrium.

The proof of existence based on path-following relies on Theorem 2.8 and requires Assumption 2.2 to hold which will be checked in the sequel. To start with, the notation in Chapter 7 is related to the formalism in Section 2.3. The set of agents A is replaced by $R+1$ where R denotes the set of regions and is augmented by the price agent (2.3). The MM^{mr}-variables of one region r are caught in

- p price vector,
- x_r traded goods (numéraire and permits),
- y_r residual Macro-variables not appearing in Markal (C_t, I_t, K_t, EC_t, L_t, $D_{...}$, XCAP_t), and
- z_r all Markal variables;

they are further put together following (2.9):

$$
\begin{aligned}
v_r &= (x_r, y_r, z_r),\\
v_{\bar r} &= (v_1, \dots, v_{r-1}, v_{r+1}, \dots, v_R, p), \text{ and}\\
v &= (v_1, \dots, v_R, p) = (v_r, v_{\bar r}).
\end{aligned}
$$

So each region has its own variable v_r, which it controls, and the other variable $v_{\bar r}$ it has to accept exogenously. The special case p is subsumed into v_{R+1}.

Concerning the constraints, the equality part of the linear model Markal plus the linear equations of Macro is represented by h_r, whereas the inequality part can be represented by a part of the concave function g_r. All nonlinear equality-constraints in Macro can be written as inequalities based on the specific economic meaning. *Thus every regional problem of MM^{mr} can be expressed in the form of* problem (2.9). The remaining price agent $R+1$ is obviously a simple linear programming problem and poses no difficulties. Now the formal foundation is ready for checking conditions (a)–(d) of Assumption 2.2.

Concerning condition (a) concavity of the objective follows from concavity of the log function plus the fact that the positively weighted sum of concave functions remains concave.

As for condition (b) the linear part is acceptable; there are two nonlinear relations left, which could destroy convexity: the production function $Y_t = [aK_t^{\rho\alpha} L_t^{\rho(1-\alpha)} + \Sigma_d b_d\, D_{d,t}^{\rho}]^{1/\rho}$, and the quadratic constraint concerning investment $\sum_j \text{cost}_{j,t} X_j + \sum_\chi c_\chi \text{XCAP}^2_{\chi,t} = \text{EC}_t$. In both cases '=' must logically be replaced by '$\le$'; the quadratic investment constraint is convex, thus only concavity of the production function is required. To that end the production function is broken into a Cobb-Douglas function

$$f_1(K, L) := cK^\alpha L^{(1-\alpha)}$$

and a CES–function

$$f_2(A, D) := (a_1 A^\rho + a_2 D^\rho)^{1/\rho}.$$

The term A is the aggregate of capital and labour formed by f_1. If both functions are concave and if f_2 is monotone in A, that is in f_1, then the nested function $f(K, L, D) := f_2(f_1(K, L), D)$ is concave:

$$
\begin{aligned}
\lambda f(K^0, L^0, D^0) &+ (1-\lambda) f(K^1, L^1, D^1)\\
&= \lambda f_2(f_1(K^0, L^0), D^0) + (1-\lambda) f_2(f_1(K^1, L^1), D^1)\\
&\le f_2(\lambda f_1(K^0, L^0) + (1-\lambda) f_1(K^1, L^1), \lambda D^0 + (1-\lambda) D^1)\\
&\le f_2(f_1(\lambda K^0 + (1-\lambda)K^1, \lambda L^0 + (1-\lambda)L^1), \lambda D^0 + (1-\lambda)D^1)\\
&= f(\lambda K^0 + (1-\lambda)K^1, \lambda L^0 + (1-\lambda)L^1, \lambda D^0 + (1-\lambda)D^1).
\end{aligned}
$$

The monotonicity mentioned is given if $\rho > 0$. Furthermore non-negativity of all variables has to be assumed. So the problem of concavity of the production function is reduced to show concavity for each of those two functions. This is performed by computing the eigenvalues of the Hessian of both, resulting for f_1 in $\lambda_1 = 0$ and

$$\lambda_2 = c\frac{1}{LK^2}\alpha\left(\frac{K}{L}\right)^{\alpha}(\alpha-1)(L^2+K^2).$$

In MM^{mr} K and L are always positive as well as c, and economic theory demands $\alpha \geq 0$. Therefore the sign of λ_2 equals the sign of $\alpha - 1$. That is, under the given assumptions f_1 is concave in K and L if $\alpha \in [0,1]$. The eigenvalues of the Hessian of f_2 are $\lambda_2 = 0$ and

$$\lambda_1 = \frac{a_1a_2(a_1A^{\rho}+a_2D^{\rho})^{1/\rho}(AD)^{\rho}(\rho-1)(A^2+D^2)}{A^2D^2(a_1^2A^{2\rho}+2a_1a_2(AD)^{\rho}+a_2^2D^{2\rho})}.$$

Also in this case A and D can be assumed positive as well as the constants a_1 and a_2. Hence the sign of λ_1 coincides with the sign of $\rho - 1$, that is f_2 is concave in A and D if $\rho \in (0,1]$. In the sequel it is assumed that $\alpha \in [0,1]$ and $\rho \in (0,1]$, and so condition (b) in Assumption 2.2 is satisfied.

Condition (c) in Assumption 2.2 is trivial.

Before checking condition (d) some comments are required. A careful look at the proof of Theorem 2.8 reveals that (d) is needed exactly for two reasons: (i) as a constraint qualification (CQ) in order to guarantee the existence of KKT-points, and (ii) to make the primal and dual solution of (2.10) unique. The second property is required to inhibit the path from turning back. As mentioned in the comments following Theorem 2.8 both can be achieved by requiring v^0 to be any *feasible* point (not necessarily interior) together with the well known linear independence of the (binding) gradients as CQ (see Appendix A). We assume therefore all regional MM-models to be feasible under the additional constraint

$$Em_{r,t} \leq IECO2_{r,t} \quad \forall r \in R. \tag{D.1}$$

In the sequel we denote by D the global feasible set

$$D := \{v | g_r(v) \leq 0, h_r(v) = 0, r = 1, \ldots, R+1\},$$

and restrict ourselves to verifying the linear independence of the gradient vectors, cf. Theorem A.1.

(CQ) *System (2.9) fulfills that the gradients with respect to v_r of all components of g_r and h_r are linear independent for each r and any $v \in D$.*

The proof will be given step by step.

The whole linear part, consisting of the complete MARKAL-part and parts of Macro, can be assumed linear independent by means of elimination. The overall linear independence follows then if the gradients of all Macro relations localized on the set of pure Macro variables are linear independent. For the ease of exposition the Macro-relations are reproduced here while dropping the regional index r; note that (PRD_t) and (USE_t) are not present in the Macro-constraints because they are directly substituted into the objective.

$$
\begin{aligned}
L_0 &= 1, && \text{(D.2)}\\
L_t &= (1+\mathrm{grow})^{10}\, L_{t-1}, && \text{(D.3)}\\
0 &\le \sum_{t\in T}(p_{0,t}x_{0,t} + p_{1,t}x_{1,t}), && \text{(D.4)}\\
K_0 &= c, && \text{(D.5)}\\
I_0 &= (\mathrm{grow}+k)K_0, && \text{(D.6)}\\
K_t &= (1-k)^{10}K_{t-1} + 5((1-k)^{10}I_{t-1} + I_t), && \text{(D.7)}\\
I_T &\ge K_T(\mathrm{grow}+k), && \text{(D.8)}\\
\sum\nolimits_j \mathrm{supply}_{j,d,t}X_{j,t} &= \mathrm{aeeifac}_{d,t}D_{d,t}, && \text{(D.9)}\\
\mathrm{EC}_t &\ge \sum\nolimits_j \mathrm{cost}_{j,t}X_{j,t} + c\sum\nolimits_\chi c_\chi \mathrm{XCAP}^2_{\chi,t}, && \text{(D.10)}\\
\mathrm{CAP}_{\chi,t} &\le \mathrm{expf}\,\mathrm{CAP}_{\chi,t-1} + \mathrm{XCAP}_{\chi,t}, && \text{(D.11)}\\
\mathrm{Em}_t &\le x^0_{1,t} - x_{1,t}. && \text{(D.12)}
\end{aligned}
$$

The gradients of those relations, restricted to pure Macro variables only, are analyzed in the following tables. The basic idea is to group the constraints according to the variables they contain, and such that those variables do not appear in the other groups; if the constraints have linear independent gradients with respect to these group variables, the overall set of gradients is linear independent. Note that some constraints appear only once ((D.2), (D.4), (D.5), (D.6) and (D.8)), while the others are repeated for $t \in T$. To have nice checkable tables we assume, without loss of generality, $|T| = 3$.

Group 1 contains the constraints (D.2) and (D.3) where L_t is the group variable. We then have the following table of gradients obviously implying linear independence of the set of corresponding gradients:

	L_0	L_1	L_2	L_T
$\nabla(D.2)$	1			
$\nabla(D.3)_{t=1}$	$1+\mathrm{grow}$	-1		
$\nabla(D.3)_{t=2}$		$1+\mathrm{grow}$	-1	
$\nabla(D.3)_{t=T}$			$1+\mathrm{grow}$	-1

Group 2 contains the constraints (D.4) and (D.12) with the group variables x. The corresponding table of gradients is:

	$x_{0,1}$	$x_{0,2}$	$x_{0,T}$	$x_{1,1}$	$x_{1,2}$	$x_{1,T}$
$\nabla(D.4)$	$p_{0,1}$	$p_{0,2}$	$p_{0,T}$	$p_{1,1}$	$p_{1,2}$	$p_{1,T}$
$\nabla(D.12)_{t=1}$				-1		
$\nabla(D.12)_{t=2}$					-1	
$\nabla(D.12)_{t=T}$						-1

Combining any column with positive $p_{0,t}$ from the first half of the array with the second half yields a non-singular matrix and thus the set of corresponding gradients is linear independent. The third group comprises (D.5), (D.6), (D.7) and (D.8), where K and I are the exclusive variables. The table of gradients is:

	K_0	K_1	K_2	K_T	I_0	I_1	I_2	I_T
$\nabla(D.5)$	1							
$\nabla(D.6)$	grow$+k$				-1			
$\nabla(D.7)_{t=1}$	$(1-k)^{10}$	-1			$5(1-k)^{10}$	5		
$\nabla(D.7)_{t=2}$		$(1-k)^{10}$	-1			$5(1-k)^{10}$	5	
$\nabla(D.7)_{t=3}$			$(1-k)^{10}$	-1			$5(1-k)^{10}$	5
$\nabla(D.8)$				grow $+ k$				-1

To see that the following subset of columns forms a non-singular matrix one has to recall the rule for calculating determinants of block-diagonal matrices, $\det\left(\begin{smallmatrix} A & 0 \\ * & B \end{smallmatrix}\right) = \det(A)\det(B)$.

	K_0	I_0	I_1	I_2	I_T	K_T
$\nabla(D.5)$	1					
$\nabla(D.6)$	grow $+ k$	-1				
$\nabla(D.7)_{t=1}$	$(1-k)^{10}$	$5(1-k)^{10}$	5			
$\nabla(D.7)_{t=2}$			$5(1-k)^{10}$	5		
$\nabla(D.7)_{t=3}$				$5(1-k)^{10}$	5	-1
$\nabla(D.8)$					-1	grow $+ k$

We conclude that the gradients in group 3 are also linear independent with respect to K and I given $5(grow + k) \neq 1$. To see linear independence of the gradients with respect to the other constraints note that $D_{...}$ appears only in (D.9), EC_t appears only in (D.10), and $CAP_{...}$ appears only in (D.11).

We have proven now (CQ) under two conditions: (i) there is a positive numéraire price component $p_{0,t} > 0$, and (ii) $5(grow + k) \neq 1$. Condition (ii) is merely technical and easy to fulfill. Condition (i) is very reasonable, because already a single zero-valued numéraire price component implies an unbounded objective in problem (7.1). Without loss of generality we assume therefore that the feasible set of the price agent is altered by requiring $p_0 \geq \varepsilon$ for some $\varepsilon > 0$ instead of requiring $p_0 \geq 0$.

Besides Assumption 2.2 Theorem 2.8 requires compactness of the global feasible set D. From the economical-technological background it is always possible to derive lower and upper bounds on all quantities and we can assume therefore compactness of D without loss of generality.

Having verified all requirements of Theorem 2.8 the following corollary is the ultimate answer to the question of existence of an equilibrium to MM^{mr}:

Corollary D.1 *There exists an equilibrium solution of the model MM^{mr} if the following conditions are satisfied: all regional MM-models (C.5) are feasible under the additional emission constraint (D.1), D is compact,* $5(grow + k) \neq 1$, $\rho \in (0, 1]$, $\alpha \in [0, 1]$, *and* $K_0 > 0$.

Appendix E

Implementation of Both Algorithms

In this chapter we describe the implementation of the VIP-algorithms (Algorithm 2 and 4) and the $[\delta, t]$-Negishi-algorithms (cf. Algorithm 5). The program-codes are freely available from the author ([bueeler, root]@ifor.math.ethz.ch).

Two aspects were considered most important in the implementation of both the VIP- and the Negishi-approach: (i) leave the regional gams-models as far as possible unchanged, and (ii), solve the regional models in a transparent parallel way. The first point carries over to an overall 'lazy' implementation where the major work is done by existing programs or solvers; the second point allowed both the VIP- and the Negishi-algorithm to run in a number of different settings: on a single processor machine, on two or three single processor machines which were geographically distributed solving one or two regional models each, and finally on a multiprocessor machine. Operating systems include hp-ux and aix which are unix derivatives of Hewlett Packard and IBM respectively. The main work to integrate additional machines is the installation of gams and the Markal-Macro model, whereas the changes in the equilibrium code are comparatively simple.

Contributions of this chapter include the overall coding and the implementation of parallel solving techniques for both the VIP- and Negishi-algorithm.

E.1 VIP-Based Cutting Plane Methods (CPM)

The cutting plane methods for solving VIPs can be applied either directly in the original space $\Delta \subset \mathbb{R}^{2|T|}$ (Algorithm 2), or indirectly in an extended homogenized space (Algorithm 4). While Algorithm 2 from page 23 was implemented and tested with a variety of centers, Algorithm 4 from page 38 was used with the analytic center only. In both cases the overall algorithmic structure is identical, see Figure E.1. The heart of the algorithm is a small program written in C ('main') which alternates between calling the appropriate center computing unit ('centers:') and the excess computing unit ('MM (7.1)' standing for Markal-Macro regions of the form (7.1)).

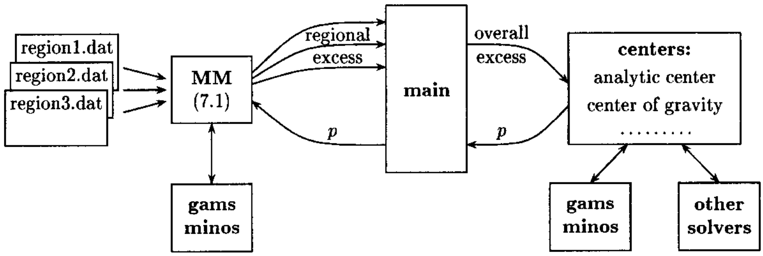


Figure E.1: Scheme of cutting plane methods.

This main routine coordinates furthermore the communication using files and accounts for starting and stopping specificities. The actual work to compute the excess and the (analytic) center is done by the solver minos through appropriate gams problems (boxes at the bottom of Figure E.1). Parallel execution of the regional problems (left part) is implemented in 'main' using `fork()`[1] in connection with `execl()` to generate a process for each regional problem and `waitpid()` to coordinate termination. In case of different (geographically distributed) computers `rcp` (remote copy) maintains communication and `rsh` (remote shell) cares for the execution of the regional models on remote machines.

To speed up the solution process the feasibility set Δ is further restricted to some $\Delta^0 \subset \Delta$ before starting. In case of Algorithm 2 an inner point $x^0 \in \Delta^0$ can be chosen freely as starting point; in contrast to this the starting point of Algorithm 4 is determined by Δ^0 and represents essentially its analytic center, cf. Section 3.2.2.

Stopping criterions for both algorithms are either the number of iterations or $||e(p)|| \leq \varepsilon$ for some chosen $\varepsilon > 0$. If Algorithm 2 stops due to exhausting the number of iterations, the iterate with minimal $||e(p)||$ is returned. The criterion $||e(p)||$ is chosen, and not say $|p^T e(p)|$ (typical for complementarity problems) or the dual gap function g_D (typical for variational inequality problems), because the former depends on the arbitrary scaling of prices, and the latter is not computable. Furthermore, $e(p)$ has a direct economic interpretation which—due to $e(p^*) = 0$ in all our equilibria—is well reflected by $||e(p)||$, making it an attractive measure for the quality of approximate solutions. In the sequel questions on how to compute centers are briefly touched.

[1]This and all subsequent verbatim terms are unix system-functions or shell commands; all of them are standard avoiding portability problems.

Computing the Analytic Center

There is a rich literature on how to compute the analytic center, cf. for example [86, 2, 39]. The problem is especially 'good-natured' and can be solved very efficiently if an inner starting point is given, see Nesterov [86, Theorem 2.2.3]. These theoretical findings are also strongly supported by computational experiences. Our goal was, however, to program as little as possible and to use existing solvers to do the work.[2] Because the CPU-time required for computing the analytic center is much below 1 % of the overall CPU-time, the following brute force method was used. Given in iteration k the feasibility set $\Delta^k = \{p \mid A^k p \leq a^k,\ B^k p \leq b^k\}$, the center of a maximal inscribed sphere, which is a solution p^c of the following LP,

$$\begin{aligned} \max\ & \lambda \\ \text{s.t.}\ & \lambda e \leq a^k - A^k p, \\ & B^k p \leq b^k, \end{aligned}$$

is computed. Here e denotes the vector of all 1's in the appropriate dimension. The constraints $B^k p \leq b^k$ are used for several purposes, e.g. to keep p on the affine hyperplane $\sum_i p_i = 1$, or to impose various proportionality relations, cf. the discussion about free and floating permits in Section 7.1.2. They share the common property that they should not be included in the barrier, either because they represent a hyperplane or to allow the center fulfilling the constraints tight. Starting at p^c, the solver minos is then used to determine the maximum of $\sum_i \log(s_i^k)$, where $s^k = a^k - A^k x$ and the inequality $B^k x \leq b^k$ is obeyed, cf. Definition 3.3. To make this scheme work down to tiny Δ^k (the diameter should decrease to zero), an appropriate scaling is of crucial importance. Note that the analytic center is invariant with respect to x if $a^k - A^k x$ is replaced by $(a^k - A^k x) \cdot \mathrm{diag}(f)$, where f is a vector of scaling factors and $\mathrm{diag}(f)$ the related matrix of diagonal elements. It is therefore easy to keep the feasibility set sufficiently large by using an appropriate scaling vector f.

In case of the homogeneous ACCPM there is no bounded polytope, instead the analytic center is defined as minimum of the proximal barrier F_k defined in Algorithm 4 page 38. Hence the center of a maximal inscribed sphere can no more be computed; a reasonably good starting point is nevertheless available by implicitly fixing $t = 1$, i.e. by simply ignoring the conic structure and treating the start-problem in the original space as described above. The result turned out to be a very satisfying starting point for minos to derive the analytic center of the conic barrier.

[2]In fact, in the mean time very robust and efficient codes for computing the analytic center were developed. One is due to a group in Geneva, see http://ecolu-info.unige.ch/~logilab/software/accpm.html.

Computing the Center of Gravity

Computing the center of gravity (cog) is closely related to volume computation as can be seen from its definition, cog $= \int_P x dx / \int_P dx$. Here P denotes the appropriate bounded convex polytope under consideration in iteration k of Algorithm 2. It is know that exact, deterministic volume computation based on a hyperplane-representation, $Ax \leq b$, can not be done in polynomial time, and furthermore, given any representation, volume computation is $\#\mathcal{P}$-hard, cf. Gritzmann and Klee [44].

The simple idea we implemented to compute the cog is to triangulate P into a set of simplices, compute the cog of each simplex and build the volume-weighted average over all simplicial cog's. There are a number of different triangulation-possibilities described in the literature. An analysis of some of these concepts together with numerical tests can be found in [10].

Despite the fact that the cog is hard to compute, and as a matter of fact could not compete with the analytic center in our examples, a cutting plane algorithm using the cog has two attractive aspects. On the one hand it is 'optimal' if algorithms are ranked according to the criterion 'worst rate of convergence (independently of f) in the case of many variables', where only first order information from an oracle can be used, see [83, p. 551]. On the other hand it is optimal if the effort for evaluating the oracle grows sufficiently large.

In the last ten years, however, hope arose that the situation can be significantly improved. The basis is an interesting randomized approximation-scheme for volume computation, see Gritzmann and Klee [44] with the references therein, and specifically Kannan, Lovász, Miklós and Simonovits [60]. Whereas an exact and deterministic computation of the volume of a polytope is hard, and also the *approximate deterministic* computation is hard, the situation changes drastically if *randomized* approximation algorithms are considered. Choose $\beta \in (0,1)$ and $\varepsilon > 0$, then the randomized volume approximation $\mu_r(P)$ is defined as follows.

$$\text{Prob}\left\{\left|\frac{\mu_r(P)}{\text{vol}(P)} - 1\right| \leq \varepsilon\right\} \geq 1 - \beta.$$

Based on random walks which are analyzed by rapidly mixing Markov chains a polynomial randomized volume approximation scheme is described in Kannan, Lovász and Simonovits [60], where the complexity is bounded by

$$O\left(\frac{d^5}{\varepsilon^2}\left(\log\frac{1}{\varepsilon}\right)^3 \log\frac{1}{\beta} \log^5 d\right).$$

In case of Algorithm 2 a randomized scheme where some of the approximate cogs are in fact outside the feasibility set is no problem, because the hyperplane-representation is always available, and so points which are not 'sufficiently' interior can simply be rejected. Thus, a valid CPM can be set up where all iterates are in the interior of P and represent most of the time the cog sufficiently well.

A Quadratic Cut Method

Convergence of Algorithm 2 can be improved if higher order information is additionally used. As an example consider a strongly monotone operator f, i.e. we observe for an $\alpha > 0$ and for all x, y in the polyhedral feasibility set D the relation $(f(y) - f(x))^T(y - x) \geq \alpha\|y - x\|^2$. In such a case the quadratic cut set $C^q_{x^k} := \{x \in D \mid -f(x^k)^T(x - x^k) - \alpha\|x^k - x\|^2 \geq 0\}$ contains the solution of VIP(f, D) for any $x^k \in D$. That is, for strongly monotone operators we can reduce the feasibility set D^k in step (iii) of Algorithm 2 further by using $C^q_{x^k}$ instead of C_{x^k}. If more precise second order information is available, $C^q_{x^k}$ can be defined more generally $C^q_{x^k} := \{x \in D \mid -f(x^k)^T(x - x^k) - (x^k - x)^T H^k(x^k - x) \geq 0\}$, where H^k is a suitable positive semi-definite matrix.

If higher order information is not available with full certainty one can nevertheless improve Algorithm 2 in the following way. Compute in iteration k the analytic center x^{k+1} of $D^k \cap C^q_{x^k}$, where H^k is chosen according to the underlying problem. Then update the feasibility set *without the quadratic term*, i.e. set $D^{k+1} := D^k \cap \{x \in D \mid f(x^k)^T(x^k - x) \geq 0\}$. That is, the uncertain quadratic information at x^k is only used to position x^{k+1}, but it does not reduce the feasibility set D^{k+1}. This procedure can therefore be used whenever f is pseudo-monotone and a reasonable hypothesis about ∇f is at hand. Two ways to approximate ∇f are the well known rank one update scheme and the Davidson-Fletcher-Powell method, cf. Fletcher [29]. This 'use-and-forget'-quadratic cut method is discussed in depth in Denault [19], where also its application to a variety of problems, including MM^{mr}, is presented. Empirically it improved convergence significantly in many instances. It's theoretical properties, however, are not yet fully understood.

E.2 Negishi-Based Methods

Two algorithms are discussed in this section: the δ-Negishi-algorithm which exploits the relation $\alpha_r = 1/\delta_r$ (see Theorem 2.5) and which is also the default Negishi-algorithm throughout this work. And secondly an alternative updating scheme called t-Negishi-algorithm ($t$âtonnement) is studied.

E.2.1 The δ-Negishi-Algorithm

The implementation of the δ-Negishi-algorithm (Algorithm 5 page 48) is depicted in Figure E.2. The large gray shaded box contains its main component, the ACCPM-decomposition machinery from Algorithm 6 page 58 to solve the Negishi-welfare problem (4.1). The input to this welfare problem is the Negishi-weight α, the ultimate output are the dual multipliers of the common excess constraint $e \geq 0$ which are called 'dual price' throughout the figure (see also Figure 4.1

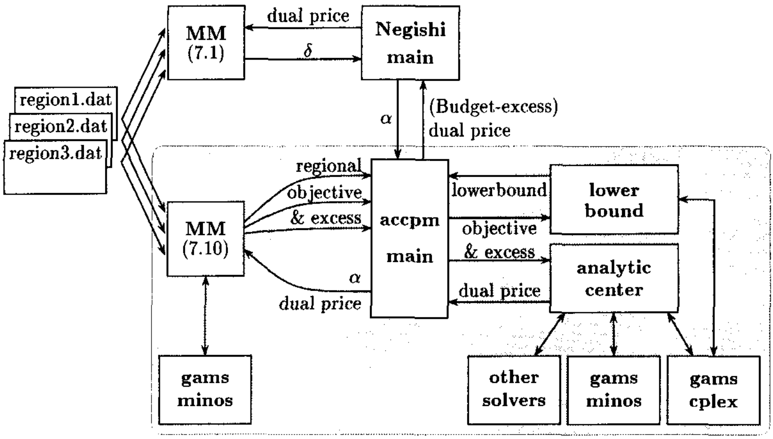


Figure E.2: The δ-Negishi-algorithm using decomposition.

page 46). This (dual) price p is then used in 'Negishi main' to compute the dual multipliers δ of the budget constraint (top left). Note that there are two slightly different regional MM-models present in the Negishi algorithm, but which use the same data-set 'region?.dat'. The MM-model (7.10) on the one hand contains a penalty term in the objective based on the excess, but has no budget constraint. MM-model (7.1) on the other hand has the original objective together with a budget constraint. The inverse of the dual multiplier δ of this budget constraint is, after inverting and normalizing, taken as new approximation of the Negishi weight.

The following remarks are in order:

- To start the δ-Negishi-algorithm one can either choose a first Negishi weight α^0 which is then used in the decomposition machinery; or a first (dual) price is chosen and δ^0 retrieved by solving (7.1), which in turn yields a first α^0. The second strategy is significantly superior because solving (7.1) is much cheaper than solving the Negishi welfare problem (gray shaded box), and at the same time the resulting δ^0 comes already very close to the true solution of the equilibrium problem even if the price signal is rather far away from the equilibrium price.

- This 'robustness' mentioned in the previous item was observed in fact in both directions, and on the whole feasible set of p and α respectively. Taking any $p \in \text{int}\Delta$ and solving (7.1) yields δ which—taken inverse and normalized—is close to an equilibrium weight α^*. And conversely: Choosing an arbitrary feasible α and solving (7.9) yields dual multipliers of the excess constraints which are close to an equilibrium price p^*. Mathematically spoken one such Negishi-iteration represents a contraction mapping (cf. Section A.4.1) with a small contraction constant around 0.01 in our examples. This remarkable behavior deserves future attention and is the basement of successfully solving equilibrium problems using the Negishi-approach.

- In the decomposition-machinery the overall performance can greatly be influenced by choosing an appropriate small box around the solution. Additionally it turned out that choosing the starting point p_0 too close to the true solution slows down the decomposition because the related first cut pushes away subsequent iterates.

- The Negishi welfare-problem is very sensitive to minor changes of the weights α. This sensitivity makes numerical 'rounding' effects in the decomposition machinery a major reason for limited convergence of the overall δ-Negishi-algorithm. The regional problems with the dual multipliers are only solved approximately by the solver minos, yielding approximate sets of localization. In the implementation the decomposition code returns the dual price once a minimal duality gap is reached. Due to the numerical difficulties the minimal absolute duality gap can not be lowered below 10^{-6} (corresponding to a relative duality gap of 10^{-9}), which in turn limits the attainable accuracy in the overall δ-Negishi-algorithm.

- To further speed up the algorithm, the duality gap limit in the stopping criterion of the decomposition machinery is dynamically reduced. In the first Negishi-iteration the decomposition is carried out rather approximately, whereas in the following Negishi-iterations the decomposition is performed increasingly accurate.

- Looking closer at the excess and the dual price in the last few iterations of the decomposition algorithm reveals a surprisingly large 'jumping around'. The main strategy used in the decomposition machinery is to return the quantities of the last iteration where a value-cut was performed. Another successful strategy takes the average over the last few iterations of those quantities.

E.2.2 The t-Negishi-Algorithm

This algorithm can also be visualized by Figure E.2, if the top left box (MM (7.1)) together with the related arrows is dropped. The basic idea to estimate δ, the dual multipliers of the budget constraint in (7.1), is taken from the model $5R$, cf. Manne and Rutherford [75]. In case of MM^{mr}, the derivation is based on

(7.5):

$$\delta_r \approx \delta_r^a := \frac{b_{r,t}}{p_{NTX,t}(C_{r,t} - NTX_{r,t})} \tag{E.1}$$

at an arbitrarily chosen time period (cf. the discussion on page 79); here '≈' means 'approximately modulo scaling'. Because the same equilibrium solution is obtained in the VIP- and the Negishi-approach, the quantities from the solution of the Negishi welfare problem can be used to compute such an approximate δ_r^a. The updating scheme of the Negishi weights uses then the inverse of δ_r^a and adds a scaled amount of the budget excess following the tâtonnement concept described page 47. Empirically we found that the numerical behavior depends crucially on the scaling of the budget excess added, motivating a closer look at the underlying quantities.

Amazingly enough, we observed in the fullfledged MM^{mr}-model that in every iteration k the relation

$$\alpha_r^k \equiv \frac{1}{\delta_r^{ak} \sum_r (1/\delta_r^{ak})}$$

holds with an accuracy of more than 10 digits. Here the quantities in (E.1), which define δ_r^{ak}, are taken from the solution of (7.10) at α_r^k. That is, δ_r^{ak} derived from the solution of (7.10) at α_r^k does not estimate δ_r of (7.1), but is—taken inverse and normalized—simply α_r^k again.

To analyze this phenomenon the machinery of Section 7.1.2 is used, applying it to the following simplified Negishi welfare problem:

$$\begin{aligned} \max \ & \sum_{r \in R} \alpha_r \sum_{t \in T} b_{r,t} \log(C_{r,t} - NTX_{r,t}) \\ \text{s.t.} \ & \sum_{r \in R} NTX_{r,t} = 0 \quad \forall\, t \in T. \end{aligned}$$

Using $p_{NTX,t}$ as Lagrange multiplier the corresponding Lagrange function reads

$$\begin{aligned} L(C_{r,t}, NTX_{r,t}, p_{NTX,t}) = & \sum_{r \in R} \alpha_r \sum_{t \in T} b_{r,t} \log(C_{r,t} - NTX_{r,t}) \\ & + \sum_{t \in T} p_{NTX,t} \sum_{r \in R} NTX_{r,t}. \end{aligned}$$

The first order optimality condition $\partial L / \partial NTX_{r,t} = 0$ yields

$$\alpha_r = \frac{p_{NTX,t}(C_{r,t} - NTX_{r,t})}{b_{r,t}}. \tag{E.2}$$

Due to the non-arbitrage argument this is independent of t. Comparing the inverse of (E.1) with (E.2) we find coincidence explaining why the ostensible guess for the new Negishi weight, $1/\delta_r^a$, returns in fact simply the old Negishi weight.

The update of the Negishi weight α from iteration k to $k+1$ can thus be understood as *tâtonnement* process (therefore *t*-Negishi-algorithm), and be described as follows. Let $be_r^k = p^T e_r$ be the budget excess of region r in iteration k, where p is the dual multiplier of the common excess constraint, set $be^k = \sum_r |be_r^k| + 1$, and let $redfac > 0$ be an appropriate reduction factor. Then the update scheme (step (iii) in Algorithm 1) proceeds as follows.

$$\alpha'^{k+1}_r = \max\{0, \alpha_r^k + redfac^k \cdot be_r^k / be^k\}, \tag{E.3}$$

$$\alpha''^{k+1}_r = \alpha'^{k+1}_r / \sum_r \alpha'^{k+1}_r, \tag{E.4}$$

$$\alpha_r^{k+1} = \sigma \cdot \alpha_r^k + (1-\sigma) \cdot \alpha''^{k+1}_r. \tag{E.5}$$

In (E.3) the basic updating step is performed which consists of adding the budget excess to the old weight. Then the weights are normalized again (E.4) and finally in (E.5) they are smoothed with the former weight, where $\sigma \in [0,1]$ is chosen appropriately. In the original scheme used by Manne and Rutherford (E.5) was not present, i.e. σ was set to 0. The reduction factor *redfac* is reduced in each iteration. A good starting value together with a suitable reduction are of critical importance to come sufficiently close to a solution and to overcome the non-contractive nature of the map. This heuristic has to be adjusted for each problem.

In the solution process of $5R$ the scaling factor *redfac* was implicitly fixed by replacing (E.3) with (F.1).

E.3 Adaptions Needed in the Regional MM-Models

First the common changes from (C.5) to (7.1) and (7.10) are described, secondly the specific changes are reported. Note that no explicit regional index r is introduced in the gams-code of the regional models. Common changes comprise the introduction of the parameter *IECO2* (initial endowment with CO_2 emission permits), the variables *NTXCO2* (net exchange of permits) and *NTX* (net exchange of numéraire), and the equation *EMC* (CO_2 emission constraint). To adapt the regional models to the same monetary and emission units, the scalars *CO2EMFAC* (regional emission units per megatons CO_2) and *LCU_FACT* (US$ per regional monetary unit) are further defined. Based thereupon the following common changes are made to (C.5): (i) replace $\log(C_t)$ by

$$\log(LCU\_FACT \cdot (C_t - NTX_t)),$$

and (ii) bound the emissions by

$$EM(TP,'CO2')/CO2EMFAC \leq CCO2(TP).$$

In (ii) it is in the case of the Negishi-algorithm important to put the *CO2EMFAC* on the left side of the inequality, otherwise incompatible dual multipliers are produced and the decomposition machinery does not work.

Specific for (7.1), additional price parameters p_{NTX} and p_{NTXCO2} and the budget constraint are declared.

As for (7.10) specific declarations comprise the dual price parameters p_{NTX} and p_{NTXCO2}, and the Negishi weight α_r. Finally, the objective is extended to

$$\alpha_r \left[\sum_t \log(LCU\_FACT \cdot (C_t - NTX_t)) \right] + \sum_t p_{NTX,t} NTX_t + \sum_t p_{NTXCO2,t} NTXCO2_t.$$

Appendix F

Numerical Comparison of the Algorithms

This chapter presents numerical comparisons among pairs of algorithms; specifically neither economic results are reported, nor are the data the same throughout all comparisons. Only within a comparison the underlying data are kept identical. Reasons are the different numerical requirements of the algorithms, the evolution of the data-situation in time, and technical convenience. To begin with, in Section F.1 (pseudo-)monotonicity is tested using a very simplified (nonlinear) MM^{mr}-model. Next, concerning Algorithm 2 the analytic center is compared with the center of gravity cutting plane method in Section F.2. Then in Section F.3 the classic non-conic ACCPM versus the new conic ACCPM (Algorithm 4) is investigated. The δ-Negishi-algorithm vis-à-vis the ACCPM is presented in Section F.4. Finally, the δ-Negishi-algorithm (Algorithm 5) is compared with the t-Negishi-algorithm.

As a general remark, all algorithms found the same equilibrium solution, whereas—due to the algorithmic specificities—the accuracy can differ. To obtain the true permit prices in US\$/ton CO_2, the price components related to CO_2 permits must be divided by the components of the numéraire NTX and multiplied by 1000 to account for internal scaling. Because all algorithms work in the full price space, some algorithmic comparisons will be made presenting untransformed prices.

F.1 Almost Pseudo-Monotonicity of the Excess in a Simplified Model

The following, very simplified MM^{mr}-model was set up, and the resulting overall excess examined. There are three regions $r_1, r_2, r_3 \in R$, three time periods $t_1, t_2, t_3 \in T$, and three energy demand forms $w, n, f \in D$ standing for water, nuclear and fossil respectively. The arguments in the following presentation of the variables or data structures are the index set over which the corresponding quantities are defined. The variables in the model are

$TDC(R)$	total discounted consumption,
$C(R,T)$	consumption,
$X(R,T)$	consumption of CO_2 emission certificates,
$P0(T)$	price for X(.,T),
$P1(T)$	price for NTX(.,T),
$Em(R,T)$	emission of CO_2,
$EC(R,T)$	energy cost,
$E(R,T,D)$	energy demand (consumption),
$NTX(R,T)$	net exchange of (abstract) products,
$NTXCO2(R,T)$	net exchange of CO_2-emission-certificates.

The variables C, X, Em, EC and E are nonnegative. The data are as follows:

ρ	discount rate in objective $= .2$,
σ	elasticity in CES-production function $= .6$,
$EC0(R)$	base energy cost $= (2, 1.2, 1)$,
$EEC(D)$	energy emission coefficient in emission function $= (.01, .02, 1)$,
$X0(R,T)$	emission certificate endowment:

	t_1	t_2	t_3
r_1	150	60	60
r_2	25	15	10
r_3	510	250	120

$EPC(R,D)$ energy production coefficient in production function:

	w	n	f
r_1	50	38	90
r_2	50	38	95
r_3	50	38	110

$ECC(R,D)$ energy cost coefficient in energy cost function:

	w	n	f
r_1	5	25	9
r_2	3	15	7
r_3	9	15	7

The model consists of the following equations for each region $r \in R$:

OBJ	objective,
BCon	budget constraint,
USE(T)	usage constraint,
NTXCO2Def(T)	defining constraint for $NTXCO2$,
ECCon(T)	energy cost constraint,

EMCon(T)	emission constraint,
EMB(T)	emission budget,
MFEW(T)	minimal fossil energy for producing 'water' energy,
MFEN(T)	minimal fossil energy for producing 'nuclear' energy,
EUB(T, D)	energy consumption growth upper bound,
ELB(T, D)	energy consumption growth lower bound.

They are defined as follows:

$$TDC = \sum_{t \in T} (1+\rho)^{-t} \log C_t, \tag{OBJ}$$

$$0 \leq \sum_t P1_t NTX_t + P0_t NTXCO2_t, \tag{BCon}$$

$$NTXCO2_t \leq X0_t - X_t, \tag{NTXCO2Def}$$

$$C_t \leq 0.0005 \left[\sum_{d \in D} EPC_d E_{t,d}^{\sigma} \right]^{1/\sigma} - EC_t - NTX_t, \tag{USE}$$

$$EC_t = EC0 + \sum_{d \in D} ECC_d E_{t,d}, \tag{ECCon}$$

$$Em_t \leq X_t, \tag{EMCon}$$

$$Em_t = \sum_{d \in D} EEC_d E_{t,d}, \tag{EMB}$$

$$E_{t+1,d} \leq 1.3 \cdot E_{t,d}, \tag{EUB}$$

$$E_{t,d} \geq 0.5 \cdot E_{t-1,d}, \tag{ELB}$$

$$E_{t,'w'} \leq 100 \cdot E_{t,'f'}, \tag{MFEW}$$

$$E_{t,'n'} \leq 10 \cdot E_{t,'f'}. \tag{MFEN}$$

CO_2 emissions are free in the first period and hence $p_1^0 = 0$ is fixed, leaving 5 tradable goods. Furthermore, instead of requiring $p \geq 0$ together with $\sum_i p_i = 1$ (i.e. $p \in \Delta$), each of the 5 price-components is logarithmically distributed from 0.0001 to 10, giving raise to 6745 valid and different price-vectors, for which the solver minos could compute the excess. Though the excess $e(p)$ is invariant under price-scaling, the monotonicity-product $(p-p')^T(e(p)-e(p'))$ depends on the scaling of the prices, requiring a rescaling of the prices onto Δ. We found about 5 % of the price pairs where monotonicity is violated (560926 out of 22744140). The maximal monotonous product was 59998, the minimal -787, hence the extend of violation of monotonicity is around 1.3 %.

Then pseudo-monotonicity was tested, that is, we checked whether $e(p)^T(p'-p) \geq 0$ implies $e(p')^T(p'-p) \geq 0$. The first relation, $e(p)^T(p'-p) \geq 0$, was given for $4.6 \cdot 10^6$ price pairs, and only in case of 18 pairs we then had $e(p')^T(p'-p) < 0$, i.e. pseudo-monotonicity was violated only extremely rarely. Looking to what extend it was violated, the product $e(p')^T(p'-p)$ attained at least -0.14 and at most 25,000 among the pairs where $e(p)^T(p' - p) \geq 0$ holds. Hence pseudo-monotonicity is practically fulfilled. On top of that the equilibrium price p^* was never cut away by any of the above prices, i.e. the minimum of $e(p)^T(p - p^*)$ over all prices p on the grid is positive.

The fullfledged MM^{mr}-model can not be tested in such a way, because the numerical effort exceeds todays computer-capacities. Nevertheless, the numerous runs of MM^{mr} using ACCPM cut away the solution only rarely, and if so it happened already close to an equilibrium price. That is, only if $\|e(p)\|$ is small, ACCPM sometimes cut away p^*. This indicates also non-pseudo-monotonicity of $e(p)$ in case of MM^{mr}.

F.2 Analytic Center versus Center of Gravity CPM

In the previous section both the data and the model were very simplified compared to the fullfledged MM^{mr}. In this section, the full gams code for the regions is used, but the data-set is simplified to shrink the size of the resulting regional problems to about 10 % of its original value. At the same time the number of time periods is increased to 5, where, as before, trade of permits in the first period is omitted in order to have the starting period under control.

The first four components in the price-vector (`P0` in the Tables F.2 and F.3) relate to the CO_2 permits of the periods $t_2, \ldots, t_5$, the final 5 price-components (`P1`) correspond to the numéraire good of the periods $t_1, \ldots, t_5$. Hence the number of traded goods is $4 + 5 = 9$.

There are three different regions in this example; region 1 has simplified energy data from Switzerland, region 2 is about five times larger than region 1 and electricity generation from nuclear power is cheaper, and finally region 3 is about 10 times larger than region 1 with cheaper fossil fuels. The differences in the data are shown in Table F.1. In the cases where values are given for 1990 and 2030 only, the values for the intermediate periods are linearly interpolated.

In the Tables F.2 and F.3 '`absolute vol.`' is the volume of the feasibility set in 8 dimensions, '`v. red.`' is the volume reduction from the current cut, and '`|excess|`' is $\|e(p)\|$. Comparing those two tables the following remarks are in order.

- Both cutting plane methods produce the same approximate solution while the iterates can differ considerably.

	Region 1	Region 2	Region 3
ECO (initial energy cost)	13.5	67.44	144.2
GDPO (initial GDP)	300	1500	3000
IECO2.1990 (certificates endowment)	28	90	270
IECO2.2000	30	105	300
IECO2.2010	35	120	350
IECO2.2020	40	130	400
IECO2.2030	40	140	450
DDAT.RH.PREF *(reference price for residential heating)*	30.68	37.43	41.86
DM_DEMAND.RH.1990 (Maximal demand for residential heat)	100	500	1000
DM_DEMAND.RH.2030 (Maximal demand for residential heat)	185	925	1850
DM_DEMAND.TX.1990 (Maximal demand for transportation)	50	250	500
DM_DEMAND.TX.2030 (Maximal demand for transportation)	105	500	1050
SEP_COST1.IMP.HCO.1.1990*2030 (import price for hard coal)	5	5	3
SEP_COST1.IMP.OIL.1.1990*2030 (... oil)	8	8	5
SEP_COST1.IMP.URN.1.1990*2030 (... uranium)	1	0.8	1
TCH_FIXOM.E21.1990*2030 (fix cost of LWR)	500	200	500
TCH_INVCOS.E21.1990*2030 (investment cost of LWR)	5000	3000	5000

Table F.1: Differences in the regions of Utopia.

- The overall volume reduction coefficient is in case of the analytic center 0.52, in case of the center of gravity, slightly better, 0.49. But whereas the volume reduction factor in case of the center of gravity has a small variation—it must stay in the interval $(1/e, 1 - 1/e)$—the reduction rate in case of the analytic center varies broadly between 0.1 and 0.9.

- After 100 iterations the norm of the excess $||e(p)||$ has decreased non-monotonically by a factor of about 10^{-4}, where a similar accuracy is obtained using either center.

- The unit simplex in 9 dimensions represents in fact an 8-dimensional volume computation problem. The burden to compute the center of gravity was around 2/3 of the overall computation time of Algorithm 2. Compared to the analytic center the center of gravity is already in this low dimensional problem much harder to compute, requiring roughly 10–1000 times more computation time.

F.3 Non-Conic versus Conic ACCPM

It must be emphasized that the crucial assumption for the conic ACCPM—monotonicity of $e(p)$—is not given in case of MM^{mr}. Furthermore, the accuracy is limited in the concrete implementation. In that sense this comparison must be interpreted cautiously. Here both the fullfledged MM^{mr}-model and the full

it.	P0(t2)	P0(t3)	P0(t4)	P0(t5)	P1(t1)	P1(t2)	P1(t3)	P1(t4)	P1(t5)	absolute vol.	v. red.	|excess|
0	0.050000	0.050000	0.050000	0.050000	0.160000	0.160000	0.160000	0.160000	0.160000	2.480159e-05		8788.88
1	0.099376	0.099062	0.098354	0.097980	0.190441	0.169567	0.080104	0.059884	0.105231	1.233960e-05	0.49753	4714.76
2	0.082595	0.081338	0.079867	0.078650	0.249025	0.115271	0.150037	0.039446	0.123771	5.597990e-06	0.45366	7466.42
3	0.069377	0.067613	0.065815	0.064206	0.240089	0.201283	0.102509	0.042828	0.146280	2.989210e-06	0.53398	4730.11
4	0.062463	0.057324	0.055359	0.056142	0.353376	0.133243	0.094083	0.083071	0.104938	1.252500e-06	0.41901	5728.31
5	0.055469	0.048848	0.046810	0.048434	0.255536	0.187554	0.124118	0.077067	0.156163	6.903200e-07	0.55115	1855.57
6	0.050942	0.043077	0.039032	0.043816	0.357642	0.209395	0.081411	0.055278	0.119406	2.467530e-07	0.35745	3830.81
7	0.047599	0.036980	0.032144	0.039328	0.336219	0.178559	0.112231	0.060990	0.155950	1.446350e-07	0.58616	2848.46
8	0.042874	0.032612	0.026548	0.035219	0.278608	0.249963	0.116445	0.084022	0.133708	5.156330e-08	0.35651	4513.17
9	0.040376	0.031047	0.024084	0.034034	0.315452	0.192790	0.090735	0.076473	0.195010	3.236420e-08	0.62766	5790.56
10	0.038270	0.029611	0.022722	0.032957	0.313777	0.244811	0.152528	0.065814	0.099509	1.481740e-08	0.45783	5175.66
11	0.036631	0.028378	0.021897	0.031529	0.311118	0.218020	0.121962	0.070641	0.159826	8.266050e-09	0.55786	669.55
12	0.032818	0.026080	0.020064	0.030710	0.365829	0.169370	0.124815	0.086678	0.143635	4.237340e-09	0.51262	3257.15
13	0.030687	0.024529	0.019132	0.030848	0.363302	0.200050	0.112703	0.070004	0.148746	2.534920e-09	0.59823	1121.45
14	0.027318	0.022045	0.016849	0.030403	0.327582	0.202210	0.136728	0.077288	0.159577	1.007660e-09	0.39751	2606.09
15	0.025572	0.020640	0.016602	0.031960	0.336123	0.202938	0.117952	0.075821	0.172392	6.847670e-10	0.67956	1335.54
16	0.024741	0.019651	0.016021	0.031847	0.344677	0.225886	0.130434	0.067870	0.138872	3.962730e-10	0.57870	2021.06
17	0.024829	0.018869	0.015598	0.032014	0.340384	0.211711	0.125108	0.073293	0.158193	2.486910e-10	0.62757	344.80
18	0.025882	0.014906	0.013478	0.043966	0.340461	0.205840	0.120180	0.078173	0.157114	1.568430e-10	0.63068	1689.51
20	0.026058	0.012829	0.011697	0.040425	0.342227	0.206526	0.122905	0.074535	0.162798	2.846260e-11	0.82272	251.89
22	0.028317	0.011143	0.010253	0.038263	0.348376	0.209870	0.124247	0.074009	0.155523	9.080220e-12	0.64653	250.64
24	0.028426	0.015490	0.010427	0.035157	0.343083	0.209779	0.123834	0.073537	0.160266	3.010250e-12	0.85014	112.86
26	0.011589	0.012090	0.011959	0.025386	0.356158	0.216755	0.125800	0.075814	0.164447	1.751320e-13	0.15789	252.56
28	0.019980	0.015650	0.010799	0.030822	0.340508	0.216206	0.127168	0.078040	0.160828	7.362570e-14	0.69197	621.53
30	0.021531	0.009894	0.011210	0.033962	0.344662	0.213218	0.127186	0.076057	0.162282	7.897050e-15	0.13983	209.54
32	0.029439	0.009301	0.014014	0.035575	0.341233	0.212012	0.123964	0.074385	0.160076	3.550250e-15	0.61664	200.72
34	0.028580	0.010586	0.010589	0.041032	0.340112	0.211366	0.123944	0.073125	0.160667	7.453410e-16	0.59214	155.62
39	0.030986	0.012574	0.011498	0.043341	0.340252	0.207797	0.122547	0.073404	0.157601	3.564690e-17	0.53053	130.56
44	0.027998	0.013519	0.011105	0.038138	0.342510	0.210410	0.123689	0.073513	0.159117	1.395220e-18	0.52178	42.11
49	0.026688	0.012634	0.011637	0.038857	0.343164	0.209771	0.123844	0.073832	0.159574	6.202520e-20	0.56667	37.17
54	0.028392	0.012594	0.011198	0.037094	0.342033	0.210693	0.124277	0.074129	0.159590	2.599250e-21	0.60879	54.06
59	0.028215	0.012428	0.011757	0.037685	0.342950	0.209878	0.123672	0.073641	0.159774	1.042240e-22	0.60942	21.47
64	0.028177	0.012452	0.011279	0.038199	0.342698	0.210031	0.123711	0.073694	0.159760	4.617240e-24	0.54547	7.85
69	0.028043	0.012427	0.011233	0.038038	0.342897	0.210258	0.123675	0.073636	0.159793	2.823030e-25	0.73971	15.21
74	0.027728	0.012413	0.011526	0.038183	0.342701	0.210114	0.123704	0.073744	0.159888	1.210680e-26	0.76901	9.00
79	0.028169	0.012425	0.011469	0.038131	0.342739	0.210000	0.123646	0.073665	0.159757	3.530320e-28	0.57611	6.33
84	0.028257	0.012411	0.011492	0.038115	0.342568	0.210033	0.123678	0.073695	0.159752	1.454850e-29	0.42856	4.90
89	0.028002	0.012402	0.011410	0.038136	0.342768	0.210056	0.123702	0.073716	0.159808	5.605520e-31	0.49910	4.73
94	0.028233	0.012391	0.011374	0.038099	0.342690	0.210041	0.123681	0.073689	0.159802	2.468840e-32	0.46151	12.10
99	0.028138	0.012394	0.011396	0.038120	0.342674	0.210063	0.123702	0.073710	0.159804	1.198360e-33	0.56199	2.67

Average volume reduction is 0.51880

Table F.2: ACCPM applied to Utopia.

data set (state spring 1997) is used. Trade is possible in 5 time periods (2000–2040), and also CO_2 permits are traded in all periods forming a 10-dimensional price-space. The two Figures F.1 and F.2 show the behavior of the direct, non-conic ACCPM versus the conic ACCPM for the CO_2 scenarios 0 %-reduction and -20 %-reduction. As measure for the quality of a solution the natural logarithm of the norm of the excess, $\log(\|e(p)\|)$, is chosen, see the related discussion in Chapter 5.

There are three curves in each figure: (i) iterates from the direct ACCPM (solid lines, Algorithm 2) as a reference, (ii) untransformed iterates of the conic AC-

it.	P0(t2)	P0(t3)	P0(t4)	P0(t5)	P1(t1)	P1(t2)	P1(t3)	P1(t4)	P1(t5)	absolute vol.	v. red.	|excess|
0	0.050000	0.050000	0.050000	0.050000	0.160000	0.160000	0.160000	0.160000	0.160000	2.480159e-05		8788.88
1	0.110340	0.109948	0.109057	0.108582	0.153381	0.149428	0.084273	0.057855	0.117135	1.233962e-05	0.49753	3566.36
2	0.096061	0.093992	0.091827	0.089893	0.232699	0.116865	0.121833	0.049816	0.107015	5.511762e-06	0.44667	5763.28
3	0.085538	0.082634	0.079908	0.077356	0.252383	0.161163	0.077795	0.056912	0.126311	2.661961e-06	0.48296	4089.86
4	0.075400	0.071388	0.068162	0.065216	0.201125	0.202905	0.129600	0.041327	0.144877	1.166308e-06	0.43814	6113.12
5	0.070612	0.063745	0.060337	0.057416	0.286912	0.167304	0.107155	0.059437	0.127082	5.403280e-07	0.46328	2375.72
6	0.061549	0.049908	0.046491	0.043871	0.261571	0.189157	0.104032	0.072942	0.170480	2.532707e-07	0.46874	3260.18
7	0.060542	0.044990	0.041493	0.043472	0.295252	0.233187	0.112962	0.073639	0.094463	1.328834e-07	0.52467	5239.49
8	0.058911	0.041239	0.037061	0.040942	0.337445	0.165429	0.101174	0.058944	0.158856	5.730536e-08	0.43125	2290.88
9	0.058126	0.041272	0.034811	0.043300	0.271777	0.210337	0.127340	0.077197	0.135839	2.948009e-08	0.51444	1604.69
10	0.054650	0.041024	0.030251	0.043883	0.327473	0.201118	0.101787	0.062849	0.136965	1.573465e-08	0.53374	1714.00
11	0.051357	0.036170	0.025571	0.041039	0.322399	0.154135	0.126507	0.088510	0.154312	7.486591e-09	0.47580	5040.12
12	0.052492	0.035781	0.024225	0.042493	0.293324	0.205396	0.123842	0.062878	0.159568	3.600419e-09	0.48092	2885.19
13	0.052672	0.034664	0.022954	0.044107	0.330023	0.181925	0.112177	0.079543	0.141934	1.670651e-09	0.46402	3337.73
14	0.049421	0.033116	0.022530	0.041640	0.320556	0.197370	0.118866	0.068918	0.147582	8.942357e-10	0.53526	697.70
15	0.038531	0.027412	0.020235	0.045187	0.307018	0.206703	0.111778	0.072195	0.170941	4.306215e-10	0.48155	1886.28
16	0.031298	0.025880	0.019993	0.042341	0.331898	0.201516	0.117829	0.072047	0.157197	1.934980e-10	0.44935	417.15
17	0.025098	0.020032	0.015853	0.034289	0.337279	0.219531	0.125647	0.072069	0.150202	9.020694e-11	0.46619	615.61
18	0.023576	0.018711	0.016033	0.034997	0.345646	0.195565	0.127167	0.074914	0.163392	4.522843e-11	0.50139	1614.70
20	0.026844	0.018727	0.016627	0.037566	0.314845	0.213126	0.131299	0.082788	0.158179	1.145773e-11	0.49941	1331.10
22	0.027487	0.016947	0.015481	0.038358	0.317462	0.216633	0.126638	0.074009	0.166985	2.805355e-12	0.48539	558.31
24	0.026963	0.013013	0.012282	0.048435	0.344300	0.203421	0.122827	0.072139	0.156618	6.403990e-13	0.49281	526.92
26	0.022652	0.012952	0.012515	0.032447	0.343977	0.213344	0.126879	0.074155	0.161079	1.842663e-13	0.50960	417.86
28	0.024139	0.012698	0.010917	0.039503	0.344695	0.208669	0.123181	0.073627	0.162573	4.936661e-14	0.57182	133.70
30	0.020096	0.013774	0.012346	0.036818	0.346205	0.212300	0.124623	0.074541	0.159296	1.339383e-14	0.50288	121.25
32	0.030565	0.014052	0.011138	0.035293	0.344020	0.208178	0.123512	0.073769	0.159474	3.344152e-15	0.52789	88.33
34	0.022098	0.012432	0.012102	0.041702	0.342943	0.208882	0.124656	0.075120	0.160066	7.715153e-16	0.48754	218.37
39	0.027620	0.012654	0.011244	0.038506	0.343302	0.210085	0.123473	0.073509	0.159608	2.369837e-17	0.54462	28.81
44	0.024977	0.012441	0.010901	0.039112	0.344024	0.210144	0.124137	0.074155	0.160108	6.462169e-19	0.48313	50.17
49	0.027428	0.012938	0.011805	0.037404	0.343136	0.210082	0.123902	0.073756	0.159548	1.999213e-20	0.47081	37.40
54	0.028057	0.012868	0.011875	0.038052	0.342382	0.209841	0.123586	0.073781	0.159559	5.594735e-22	0.53605	32.53
59	0.027669	0.012422	0.010626	0.037609	0.343488	0.210483	0.123920	0.073905	0.159879	1.285555e-23	0.48780	23.81
64	0.027947	0.012408	0.011664	0.037819	0.342741	0.210144	0.123760	0.073739	0.159776	4.526520e-25	0.48791	6.00
69	0.028100	0.012432	0.011396	0.038208	0.342438	0.210181	0.123779	0.073755	0.159711	1.203271e-26	0.48188	11.05
74	0.028142	0.012427	0.011318	0.038084	0.342800	0.210009	0.123673	0.073735	0.159812	3.105132e-28	0.46748	9.90
79	0.028175	0.012408	0.011406	0.038122	0.342722	0.209994	0.123689	0.073688	0.159796	8.834833e-30	0.48839	7.13
84	0.028205	0.012392	0.011428	0.038098	0.342647	0.210058	0.123698	0.073702	0.159773	2.639925e-31	0.47656	33.49
89	0.028210	0.012406	0.011404	0.038102	0.342685	0.210047	0.123692	0.073696	0.159758	6.666574e-33	0.47913	3.10
94	0.028119	0.012404	0.011393	0.038116	0.342709	0.210062	0.123698	0.073707	0.159791	2.199996e-34	0.55645	2.63
99	0.028177	0.012402	0.011389	0.038115	0.342696	0.210048	0.123689	0.073700	0.159786	6.883292e-36	0.52018	2.67

Average volume reduction: 0.49129

Table F.3: CoGCPM applied to Utopia.

CPM (dotted), and (iii) the 'real' (transformed) solution-iterates of the conic ACCPM (Algorithm 4) formed by a weighted sum of the direct iterates of the conic ACCPM (dashed).

In both figures the same qualitative behavior can be observed:

- ACCPM (solid line, Algorithm 2) decreases slower, but gets finally further down; furthermore it is less smooth.

- The iterates of both the direct ACCPM and the conic ACCPM (without transformation of the solutions (dotted)) are comparable for the first 40 iterations. Then the conic iterates converge on a certain level of $e(p)$, while the direct ACCPM iterates decrease further.

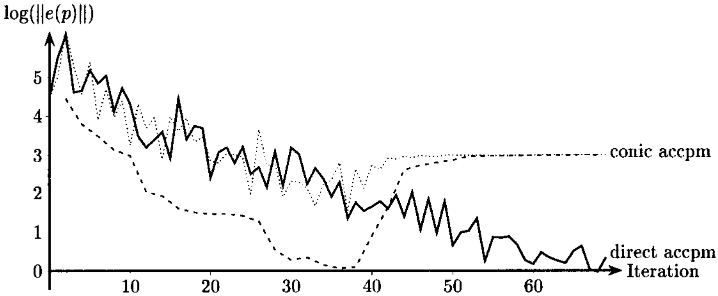


Figure F.1: Conic versus non-conic ACCPM, 0 %-reduction scenario.

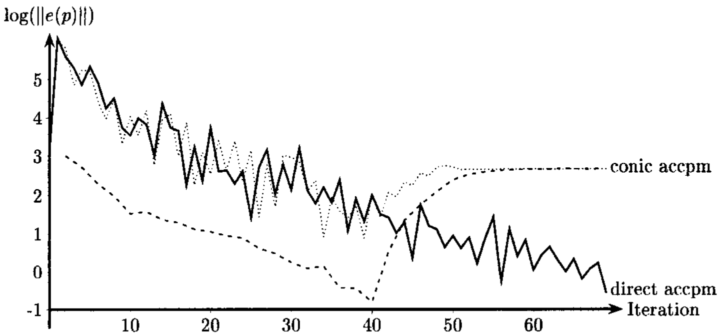


Figure F.2: Conic versus non-conic ACCPM, -20 %-reduction scenario.

- The conic ACCPM with the transformed iterates behaves much smoother, even yet in the first iterations, and it reduces the excess much quicker in the first 30–40 iterations. Afterwards, the transformed iterates converge to a non-equilibrium price, leaving the norm of the excess on a higher level. Among the possible reasons, why it does not converge to a true solution, we find non-monotonicity of $e(p)$ and a limited numerical precision for both computing the analytic center and the oracle-response. In that sense the conic version seems to be more sensitive to precision problems and monotonicity than the direct ACCPM.

- For the conic ACCPM, $||e(p)||$ of the direct (i.e. non-transformed) iterates converge towards the norm of the excess of the solution-iterates. The reason for this coincidence is that both price iterates in the conic case, i.e. the transformed solution iterates and the direct iterates, converge toward the same (non-equilibrium) price.

- It seems that the conic ACCPM 'smells' quickly the approximate location of the solution with its weighted iterates, but looses it finally. A possible improvement of the direct ACCPM could therefore consist of using the conic version with its transformed iterates, and once $||e(p)||$ starts to raise the direct ACCPM is used in a neighborhood of the best conic solution.

F.4 δ-Negishi-Algorithm versus ACCPM

Here the full MM^{mr}-model code together with the full regional data-sets of Sweden, The Netherlands and Switzerland is present (state summer 1996). The prices given are already in the unit US$/ton CO_2; trade is allowed in the 4 periods 2000–2030.

it.	P(2000)	P(2010)	P(2020)	P(2030)	$\|\|e(p)\|\|$
0	8.18	16.00	30.00	61.29	617.94
5	85.62	117.79	144.84	140.14	291.50
10	57.33	69.68	88.98	97.53	54.41
15	26.23	45.56	68.60	109.22	86.91
20	13.16	44.03	63.33	79.87	56.45
30	5.08	42.10	51.46	124.52	28.59
40	3.25	42.96	64.33	114.24	4.02
50	2.98	43.91	65.46	113.73	1.00
60	2.86	44.26	63.63	112.89	0.67
70	3.15	44.16	63.67	113.45	0.13
80	3.04	44.26	63.50	113.44	0.30

Table F.4: Permit prices [US$/t CO_2] in the ACCPM-iterations and the corresponding norm of the excess vector $||e(p)||$.

Table F.4 shows convergence of the direct ACCPM to a satisfying solution of the equilibrium problem. As usual, the norm of the excess $||e(p)||$ does not go down monotonically.

In the first 4 iterations of Algorithm 5 a fast convergence can be observed, see Table F.5. Each iteration reduces the norm of the budget excess vector $||b.e.||$ by about two order of magnitude, achieving an overall reduction of 5 order of magnitude within 4 iterations. Then, in the last 2 iterations, a plafond is reached by the norm of the budget excess, which is due to the limited accuracy in the decomposition, measured as duality gap in the approximation of the Lagrangian. While in the first iteration the stopping criterion in the decomposition is an absolute duality gap of 10^{-4}, it is set in the rest of the iterations to 10^{-6}; this corresponds to a relative duality gap of 10^{-7} and 10^{-9} respectively. Due to numerical problems it was not possible to decrease it further. The results show

	Negishi weight			CO_2 permit price					
it.	Switzerland	Sweden	Netherlands	2000	2010	2020	2030	$\|\|b.e.\|\|$	#it.
1	0.45000000	0.20000000	0.35000000	3.10	44.41	66.82	112.33	60.207	43
2	0.39019676	0.15302284	0.45678040	3.05	44.23	63.49	113.36	2.0496	56
3	0.39345156	0.15387045	0.45267800	3.11	44.20	63.59	113.39	0.0557	59
4	0.39336264	0.15384863	0.45278873	3.13	44.23	63.54	113.36	0.0026	60
5	0.39336458	0.15384931	0.45278611	3.11	44.20	63.58	113.38	0.0034	58
6	0.39336439	0.15384915	0.45278646	3.03	44.23	63.52	113.37	0.0018	61

Table F.5: The δ-Negishi-algorithm for the first 6 iterations. The number of iterations in the decomposition method is given in the last column #it.

the crucial influence of the accuracy in the decomposition, both on the overall number of iterations and on the convergence of the δ-Negishi-algorithm.

Note also that the price vector obtained in the first Negishi iteration based on a starting weight vector is already near the true equilibrium price, or to say it more generally, the prices are quite independent of the Negishi weights whereas the budget excess is very sensitive to small changes in the weights. In the VIP-approach a dual observation can be made (cf. Figure 7.1); the dual multipliers of the budget constraints are quite independent of the prices, whereas the excess $e(p)$ is sensitive to changes in the price.

This dual robustness—on the one hand is the dual price of the excess constraint in the Negishi welfare problem close to the equilibrium price and only weakly influenced by the Negishi weight, and on the other hand the inverse of the budget constraint in the regional VIP-problems is close to an equilibrium Negishi weight and only weakly influenced by the price—is the reason for this remarkable convergence of the δ-Negishi-algorithm.

F.5 δ- versus t-Negishi-Algorithm

The Negishi update scheme in iteration $k+1$ used in the t-Negishi-algorithm is

$$
\begin{aligned}
\alpha'^{k+1}_r &= \sum_{t\in T} \frac{p_{NTX,t}(C_{r,t} - NTX_{r,t})}{b_{r,t}} + p^T e_r, \\
\alpha^{k+1}_r &= \alpha'^{k+1}_r / \sum_r \alpha'^{k+1}_r .
\end{aligned}
\tag{F.1}
$$

Here p is the dual multiplier associated with the excess constraint $\sum_{r\in R} e_r \geq 0$, and so $p^T e_r$ is the budget excess of region r. The expression $\sum_{t\in T} p_{NTX,t}(C_{r,t} - NTX_{r,t})/b_{r,t}$ is motivated from (E.2) where this is shown to approximate (modulo

it.	Negishi weight Switzerland	Sweden	Netherlands	CO_2 permit price 2000	2010	2020	2030	$\|\|b.e.\|\|$	#it.
1	0.45000000	0.20000000	0.35000000	2.85	44.66	66.33	112.33	60.1984	36
2	0.42403936	0.14299279	0.43296785	2.91	44.03	63.12	113.36	14.5142	51
3	0.40364823	0.15470586	0.44164591	3.07	44.14	63.69	113.39	5.6157	53
4	0.39760614	0.15329384	0.44910002	3.05	44.19	63.61	113.36	2.0015	53
5	0.39495138	0.15379185	0.45125678	3.12	44.22	63.57	113.38	0.7880	51
6	0.39398597	0.15379777	0.45221626	3.07	44.23	63.54	113.37	0.2998	52
7	0.39360284	0.15383450	0.45256266	3.12	44.24	63.52	113.37	0.1163	56
8	0.39345687	0.15384259	0.45270055	3.06	44.22	63.56	113.37	0.0446	52
9	0.39340062	0.15384617	0.45275321	3.05	44.20	63.57	113.37	0.0173	55
10	0.39337861	0.15384816	0.45277323	3.09	44.21	63.57	113.37	0.0070	53

Table F.6: *The t-Negishi-algorithm for the first 10 iterations. The number of iterations in the decomposition method is given in the last column #it.*

scaling) the existing weight; to make it numerically more robust[1] the sum over all time periods is taken. The reason why not simply the old Negishi-weight α_r^k is used has to do with the different scaling of the quantities. This heuristic seems to be very sensitive on how much weight is put to the old weight α_r^k and how much the budget excess $p^T e_r$ contributes. The implicit weighting in (F.1) did perform most satisfying in our experiments and is therefore reported in Table F.6.

First note that the accuracy in the decomposition machinery was identical with the case shown in Table F.5. To speed up the iterations, however, the feasibility set in the decomposition was reduced, explaining why the number of decomposition iterations is slightly below the ones of Table F.5. The major difference between Table F.6 and F.5 is the speed of convergence; while the true computation of the dual multipliers yields a very quick decrease of the budget excess $||b.e.||$, the reduction in Table F.6 is considerably slower. More aggravating even, this heuristic depends sensitively on the implicit weighting when the budget excess is added to the old Negishi weight. As a detail note that the Negishi weight is considerably more 'jumpy' here compared to Table F.5, but the quality of the final solution is similar.

[1]Based on the argument of no-arbitrage in a solution of the Negishi welfare problem it follows that $p_{NTX,t}(C_{r,t} - NTX_{r,t})/b_{r,t}$ is constant over time periods. However, in reality where concrete solvers are at work, slight differences are possible. In fact, a coinciding accuracy of 10–12 digits was found; this summation has therefore two effects: increase further the accuracy, and increase the weight in the sum (F.1).

F.6 Conclusions

Based on the examples presented above together with further experiences the following conclusions can be drawn.

- The cutting plane methods (Algorithm 2 and 4) used for solving VIPs are much easier to implement and manage than the Negishi-approaches (Algorithm 5). This is mainly due to the decomposition required in the latter case.

- In our tests ACCPM (Algorithm 2) required only about 1/3 to 1/2 of the computation time as compared with the δ-Negishi-algorithm (Algorithm 5). But this depends on the relation #regions/#goods, and on the behavior of the decomposition method (Algorithm 6 in our case) when the number of goods or regions changes.

- The center of gravity is much too costly in our examples. Moreover, the analytic center shows in all examples tested a very good average volume reduction.

- The new conic ACCPM (Algorithm 4) improves on ACCPM (Algorithm 2) considerably—in the first 40 iterations. Here a two stage scheme may prove useful in practice, where Algorithm 4 is used in the beginning, and once $\|e(p^k)\|$ starts to raise, the scheme switches to Algorithm 2. A crucial open question would then be, how the feasible set in Algorithm 2 is chosen after such a switch. A reasonable strategy could take the set Δ^k defined by cutting $\{t = 1\}$ with the cone K^k stemming from the kth iteration of Algorithm 4, because this intersection guarantees to contain $(f, D)^{**}$ for pseudo-monotone operators.

Bibliography

[1] S. Arrhenius. On the influence of carbonic acid in the air upon the temperature of the ground. *Philosophical Magazine*, 41:237–276, 1896.

[2] D. S. Atkinson and P. M. Vaidya. A scaling technique for finding the weighted analytic center of a polytope. *Mathematical Programming*, 57:163–192, 1992.

[3] D. S. Atkinson and P. M. Vaidya. A cutting plane algorithm for convex programming that uses analytic centers. *Mathematical Programming*, 69:1–43, 1995.

[4] Didier Aussel. Subdifferential properties of quasiconvex and pseudoconvex functions: A unified approach. *JOTA*, 97(1), 1998.

[5] Didier Aussel, Jean-Noël Corvellec, and Marc Lassonde. Mean value property and subdifferential criteria for lower semicontinuous functions. *Transactions of the American mathematical society*, 347(10):4147–4161, 1995.

[6] O. Bahn, B. Büeler, S. Kypreos, and H.-J. Luethi. Modeling an international market of CO_2 emission permits. *Global Energy Issue*, to appear, 1997.

[7] M. S. Bazaraa and C. M. Shetty. *Nonlinear Programming.* Wiley, 1979.

[8] J. F. Benders. Partitioning procedures for solving mixed-variables programming problems. *Numerische Mathematik*, 4:238–252, 1962.

[9] M. N. Broadie. An introduction to the octahedral algorithm for the computation of economic equilibria. *Math. Programming Study*, 23:121–143, 1985.

[10] Benno Büeler, Andreas Enge, and Komei Fukuda. Exact volume computation for polytopes: A practical study. Paper submitted to *Polytopes: Combinatorics and Computation*, Günter Ziegler, editor, DMV-seminar volume, Birkhäuser Verlag, January 1998.

[11] G. S. Callendar. The artificial production of carbon dioxide and its influence on temperature. *Quarternary Journal of the Royal Meteorological Society*, 64:223–242, 1938.

[12] D. Chan and J. S. Pang. The generalized quasi-variational inequality problem. *Math. Oper. Res.*, 7:211–222, 1982.

[13] Alpha C. Chiang. *Fundamental Methods of Mathematical Economics.* McGraw Hill, 1984.

[14] Stephen A. Cook. An overview of computational complexity. *Communications of the ACM*, 26(6):400–408, 1983.

[15] John H. Cushman. U.S. will seek pact on global warming. *The New York Times International*, CXLV:A6, July 17 1996.

[16] Stella C. Dafermos. Exchange price equilibria and variational inequalities. *Mathematical Programming*, 46:391–402, 1990.

[17] A. Dean and P. Hoeller. Costs of reducing CO_2 emissions; evidence from six global models. Technical Report 19, OECD Economic Studies, 1992.

[18] Gerard Debreu. *Theory of Value: An Axiomatic Analysis of Economic Equilibrium.* Cowles Foundation for Research in Economics at Yale University, Monograph 17. John Wiley & Sons, 1959.

[19] Michel Denault. *ACCPM-VI: A Cutting-Plane Algorithm for Variational Inequalities.* PhD thesis, McGill University, Montréal, Québec, Canada, 1998.

[20] Steven P. Dirkse and Michael C. Ferris. The PATH solver: A non-monotone stabilization scheme for mixed complementarity problems. *Optimization Methods and Software*, 5:123–156, 1995.

[21] Olivier du Merle, Jean-Louis Goffin, and Jean-Philipe Vial. On the comparative behavior of Kelley's cutting plane method and the analytic center cutting plane method. Technical report, HEC, Université de Genève, March 1996.

[22] John Eatwell, Murray Milgate, and Peter Newman, editors. *The new Palgrave: A Dictionary of Economics.* Macmillan London, 1987.

[23] B. C. Eaves. On the basic theorem of complementarity. *Mathematical Programming*, 1:68–75, 1971.

[24] J. A. Edmonds and J. M. Reilly. A long-term global energy-economic model of carbon dioxide release from fossil fuel use. *Energy Economics*, 5(2):74–88, 1983.

[25] J. A. Edmonds and J. M. Reilly. *Global Energy—Assesing the Future.* Oxford University Press, 1985.

[26] Paul Ekins. Rethinking the costs related to global warming: A survey of the issues. *Environmental and Resource Economics*, 6:231–277, 1995.

[27] L. G. Fishbone and H. Abilock. MARKAL, a linear programming model for energy systems analysis: Technical description of the BNL version. *International Journal of Energy Research*, 5:353–375, 1981.

[28] L. G. Fishbone, G. Giesen, G. Goldstein, H. A. Hymmen, K. J. Stocks, H. Voss, D. Wilde, R. Zolcher, C. Balzer, and H. Abilock. User's guide for MARKAL (BNL/KFA version 2.0). Technical Report BNL–51701, Brookhaven National Laboratory, Upton, N. Y. 11973 and KFA Jülich, BRD, 1983.

[29] Roger Fletcher. *Practical Methods of Optimization.* John Wiley & Sons, second edition, 1993.

[30] O. E. Flippo. *Stability, Duality and Decomposition in General Mathematical Programming.* CWI, 1991.

[31] E. Fragniere, J. Gondzio, R. Sarkissian, and J.-P. Vial. Structure exploiting tool in algebraic modeling languages. Technical Report 97.2, Logilab, Section of Management Studies, University of Geneva, June 1997.

[32] Jean-Jaskold Gabszewicz and Jean-Philippe Vial. Oligopoly 'A la Cournot' in a general equilibrium analysis. *Journal of Economic Theory*, 4:381–400, 1972.

[33] C. B. Garcia and W. I. Zangwill. *Pathways to Solutions, Fixed Points, and Equilibria.* Prentecie Hall, 1981.

[34] D. W. Gaskins and J. P. Weyant. Model comparision of the costs of reducing CO_2 emissions. *American Economic Review*, 83(2):318–323, 1993.

[35] Jaques Gauvin and Robert Janin. Directional behavior of optimal solutions in nonlinear mathematical programming. *Mathematics of Operations Research*, 13(4):629–649, 1988.

[36] V. A. Ginsburgh and J. L. Waelbroeck. *Activity Analysis and General Equilibrium Modelling.* North–Holland, 1981.

[37] J.-L. Goffin, A. Haurie, and J.-P. Vial. Decomposition and nondifferentiable optimization with the projective algorithm. *Management Science*, 38:284–302, 1992.

[38] J.-L. Goffin and J.-P. Vial. Cutting planes and column generation techniques with the projective algorithm. *Journal of Optimization Theory and Applications*, 65:409–429, 1989.

[39] J.-L. Goffin and J.-P. Vial. On the computation of weighted analytic centers and dual ellipsoids with the projective algorithm. *Mathematical Programming*, 60:81–92, 1993.

[40] Jean-Louis Goffin and Jean-Philippe Vial. Shallow, deep and very deep cuts in the analytic center cutting plane method. Technical Report Logilab Technical Report 96-3, Department of Management Studies, University of Geneva, 1996. Revised June 1997.

[41] Jean-Louis Goffin and Jean-Philippe Vial. A two cut approach in the analytic center cutting plane method. Technical Report Logilab Technical Report 97-20, Department of Management Studies, University of Geneva, 1997.

[42] G. A. Goldstein. PC-MARKAL and the MARKAL Users Support System — User's Guide. Technical Report BNL-46319, Brookhaven National Laboratory, Upton, N. Y. 11973, 1991.

[43] Greenland Ice-core Project (GRIP) Members. Climate instability during the last interglacial period recorded in the GRIP ice core. *Nature*, 364:203–207, July 1993.

[44] Peter Gritzmann and Victor Klee. On the complexity of some basic problems in computational convexity: II. volume and mixed volumes. In T. Bisztriczky, McMullen, R. Schneider, and A.I. Weiss, editors, *Polytopes: Abstract, Convex and Computational (Scarborough, ON, 1993)*, NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci., 440, pages 373–466. Kluwer Acad. Publ., Dordrecht, 1994.

[45] W. Häfele. *Energy in a Finite World: A Global System.* Ballinger Publishing Company, Cambridge, 1981.

[46] Patrick T. Harker. *Lectures on Computation of Equilibria with Equation-Based Methods.* Core Lecture Series. Core Foundation, Louvain La Neuve, Belgium, 1993.

[47] Patrick T. Harker and Jong-Shi Pang. Finite-dimensional variational inequality and nonlinear complementarity problems: A survey on theory, algorithms and applications. *Mathematical Programming*, 48:161–220, 1990.

[48] D. W. Hearn. The gap function of a convex program. *Operations Research Letters*, 1:67–71, 1982.

[49] H. Heuser. *Lehrbuch der Analysis, Teil 2.* B. G. Teubner, Stuttgart, 1988.

[50] Werner Hildenbrand. On the "law of demand". *Econometrica*, 51(4):997–1019, 1983.

[51] Werner Hildenbrand. *Market Demand: Theory and Empirical Evidence.* Princeton University Press, 1994.

[52] H. Hotelling. The economics of exhaustible resources. *Journal of Political Economy*, 39:137–175, 1931.

[53] World Resources Institute. *World Resources 1996-1997.* Oxford University Press, 1996.

[54] Intergovernmental Panel on Climate Change WMO/UNEP, IPCC. *The IPCC Second Assessment Synthesis of ScientificTechnical Information Relevant to Interpreting Article 2 of the UN Framework Convention on Climate Change.* IPCC Secretariat, World Meteorological Organization, Geneva, Switzerland, 1996.

[55] Reinhard John. Abraham Wald's equilibrium existence proof reconsidered. Technical Report A-540, Rheinische Friedrich-Wilhelms-Universität Bonn, Sonderforschungsbereich 303, February 1997.

[56] Reinhard John. Skript zur Vorlesung 'Mathematische Methoden der Wirtschaftstheorie'. Rheinische Friedrich-Wilhelms-Universität Bonn, Sonderforschungsbereich 303, SS 1997.

[57] Reinhard John. Variational inequalities and pseudomonotone functions: Some characterizations. Technical Report A-530, Rheinische Friedrich-Wilhelms-Universität Bonn, Sonderforschungsbereich 303, April 1997.

[58] Franz Kafka. *Vor dem Gesetz,* in *'Das Urteil'.* Fischer Taschenbuch Verlag, 1980.

[59] M. J. Kaiser, T. L. Morin, and T. B. Trafalis. Centers and invariant points of convex bodies. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science. THE VICTOR KLEE FESTSCHRIFT.*, 4:367–385, 1991.

[60] Ravi Kannan, Lászlo Lovász, and Miklós Simonovits. Random walks and an $O^*(n^5)$ volume algorithm for convex bodies. Technical report, Carnegie-Mellon and Yale University, January 1996.

[61] S. Karamardian. Generalized complementarity problem. *J. Opt. Theory and Appl.*, 8:161–167, 1971.

[62] S. Karamardian. Complementarity problems over cones with monotone and pseudomononote maps. *J. Opt. Theory and Appl.*, 18:445–454, 1976.

[63] Karl-Heinz Elster, editor. *Modern Mathematical Methods of Optimization.* Mathematical Topics. Akademie Verlag, Berlin, 1993.

[64] J. E. Kelley. The cutting-plane method for solving convex programs. *Journal of the SIAM*, 8:703–712, 1960.

[65] David Kinderlehrer and Guido Stampacchia. *An Introduction to Variational Inequalities and their Applications.* Academic Press, 1980.

[66] T. Kram. National energy options for reducing CO_2 emissions: A report of the IEA-ETSAP/Annex IV. Technical Report ECN-C-93-046, Netherlands Energy Research Foundation ECN, Petten, the Netherlands, 1993.

[67] Socrates Kypreos. Energienachfrage–Szenarien bearbeitet mit dem Energiesimulations–Modell SMEDE. Technical Report TM–13–87–12, EIR, CH–5303 Würenlingen, 1987.

[68] Socrates Kypreos. Energy scenarios for Switzerland and emission control, estimated with a normativ model. Technical report, Paul Scherrer Institut, CH–5232 Villingen, 1990.

[69] C. F. Lemke and J. T. Howson. Equilibrium points of bimatrix games. *Journal of the Society of Industrial Applied Mathematics*, 12(2):413–423, 1964.

[70] H.-J. Lüthi. On the solution of variational inequalities by the ellipsoid method. *Mathematics of Operations Research*, 10(3):515–522, 1985.

[71] T. Magnanti and G. Perakis. A unifying geometric solution framework and complexity analysis for variational inequalities. *Mathematical Programming*, 71(3), 1996.

[72] Alan S. Manne. ETA: A model for energy technology assessment. *Bell Journal of Economics*, 7:379–406, 1976.

[73] Alan S. Manne. Global 2100; an almost consistent model of CO_2 emission limits. *Swiss Journal of Economics and Statistics*, 127(2):181–197, 1991.

[74] Alan S. Manne and Richard G. Richels. *Buying Greenhouse Insurance: The Economic Costs of Carbon Dioxide Emission Limits.* The MIT Press, Cambridge, Massachusetts, 1992.

[75] Alan S. Manne and Thomas F. Rutherford. International trade in oil gas and carbon emission rights: An intertemporal general equilibrium model. *Energy Journal*, 15(1):57–76, 1994.

[76] Alan S. Manne and Clas-Otto Wene. MARKAL–MACRO: A linked model for energy-economy analysis. Technical Report BNL–47161, Brookhaven National Laboratory, Upton, N. Y. 11973, February 1992.

[77] Alan S. Manne, Clas-Otto Wene, et al. MARKAL–MACRO: An Overview. Technical Report BNL–48377, Brookhaven National Laboratory, Upton, N. Y. 11973, November 1992.

[78] Ruedi Meier. Auswirkungen der Klimaänderungen für die Schweiz, und, Die potentiellen Schadenskosten der Klimaänderungen für die Schweiz. In Ulrich Roth, editor, *Sozioökonomische Aspekte von Klimaänderungen und Naturkatastrophen*, Nationales Forschungsprogramm 31. Presse- und Informationsdienst des Schweizerischen Nationalfonds, 1997.

[79] G. J. Minty. On the monotonicity of the gradient of a convex function. *Pacific J. Math.*, 14:243–247, 1964.

[80] F. Murphy, R. Sanders, and S. Shaw. Modelling tradeable sulphur rights in the national energy modelling system. In *Energy models for policy and planning*, First Joint International Symposium. London Business School/IFORS, July 1995.

[81] Anna Nagurney. *Network Economics: A Variational Inequality Approach*, volume 1 of *Advances in Computational Economies.* Kluwer Academic Publishers, 1995.

[82] T. Negishi. *General Equilibrium Theory and International Trade.* North–Holland Publishing Company, 1972.

[83] G. L. Nemhauser, A. H. G. Rinnooy Kaan, and M. J. Todd. *Optimization.* Handbooks in Operations Research and Management Science, Volume 1. North-Holland, 1989.

[84] A. S. Nemirovsky and D. B. Yudin. *Problem Complexity and Method Efficiency in Optimization (original title: Sloznost' zadac i effektivnost' metodov optimizacii, 1979).* Wiley-Interscience Series in Discrete Mathematics. Wiley, 1983.

[85] Yu. Nesterov. Complexity estimates of some cutting plane methods based on the analytic barrier. *Mathematical Programming*, 69:149–176, 1995.

[86] Yu. Nesterov and A. S. Nemirovsky. *Interior-Point Polynomial Algorithms in Convex Programming.* SIAM, Studies in Applied Mathematics, 1994.

[87] Yu. Nesterov and Jean-Philippe Vial. Homogeneous analytic center cutting plane methods for convex problems and variational inequalities. Technical Report 9767, CORE discussion paper, Université Catholique de Louvain, September 1997.

[88] W. D. Nordhaus. Economic growth and climate: the carbon dioxide problem. *American Economic Review*, 67(1):341–346, 1977.

[89] W. D. Nordhaus. *The Efficient Use of Energy Resources*, volume 26. Cowles Foundation Monograph, 1979.

[90] W. D. Nordhaus. Optimal greenhouse gas reductions and tax policy in the 'DICE' model. *American Economic Review*, 83(2):313–317, 1993.

[91] J. M. Ortega and W. C. Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables.* Academic Press, 1970.

[92] D. Pearce. The role of carbon taxes in adjusting to global warming. *Economic Journal*, 101:938–948, 1991.

[93] Daniel Ralph and Stephen Wright. Superlinear convergence of an interior-point method for monotone variational inequalities. Technical Report 3, Department of Mathematics, University of Melbourne, January 1996.

[94] Paul Anthony Samuelson. *Foundations of Economic Analysis*, volume 80 of *Harvard Economic Studies.* Harvard University Press, Cambridge, 1947.

[95] Herbert Scarf. *The Computation of Economic Equilibria.* Cowles Foundation Monograph. Yale University Press, 1973.

[96] Renate Schubert. International handelbare Emissionszertifikate: Ihre Bedeutung für Entwicklungsländer. In Hans-Bernd Schäfer, editor, *Die Entwicklungsländer im Zeitalter der Globalisierung*, number 245 in Schriften des Vereins für Socialpolitik. Duncker und Humblot, Berlin, 1996.

[97] G. Sonnevend. An "analytic center" for polyhedrons and new classes of global algorithms for linear (smooth, convex) programming. In A. Prekopa, J. Szelezsan, and B. Strazicky, editors, *Proceedings of 12th IFIP conference; Budapest, September 2-6, 1985*, number 84 in Lecture Notes in Control and Information Sciences, pages 866–876. Springer, Berlin, 1985.

[98] Jie Sun and Gongyun Zhao. A long-step quadratically convergent interior point method for nonlinear monotone variational inequality problems. Technical report, National university of Singapore, 1996.

[99] Hadi Taheri. *Fast Computation of General Equilibrium.* PhD thesis, Stanford University, 1995.

[100] Hadi Taheri, Robert Maxfield, and David Luenberger. Fast computation of general equilibrium based on the zero-minimum principle. Technical report, Stanford University, 1996.

[101] Abraham Wald. Über die eindeutige positive Lösbarkeit der neuen Produktionsgleichungen; Ergebnisse eines mathematischen Kolloquiums 6 1933-34. 12-18, 1935.

[102] Leon Walras. *Eléments d'économie politique pure ou Théorie de la richesse sociale.* Corbaz, Lausanne, Switzerland, 1874.

[103] Jen-Chih Yao. Generalized quasi-variational inequality problems with discontinuous mappings. *Mathematics of Operations Research*, 20(2):465–478, 1995.

[104] W. I. Zangwill and C. B. Garcia. Equilibrium programming: The path-following approach and dynamics. *Mathematical Programming*, 21:262–289, 1981.

Curriculum Vitae

Personal data

Name:	Benno Paul Büeler
Place of birth:	Uzwil (SG), Switzerland
Date of birth:	13. 12. 1961
Marital status:	Single
Children:	Anna (4.2.87) and Joel (1.1.90)

Education

88–93:	Study of mathematics at ETH Zürich, Switzerland *Degree: Dipl. Math. ETH*
82–87:	Study of agricultural science at ETH Zürich *Degree: Dipl. Ing.-Agr. ETH*
73–81:	High school in Münchenstein, Switzerland
69–73:	Primary school in Aesch, Switzerland

Working experiences

93–97:	Assistance at the Institute for Operations Research, ETH Zürich (Prof. Dr. H.-J. Lüthi). Writing a doctoral thesis about economic equilibrium problems and related solution concepts.
92–93:	Teaching mathematics at a high school in Wetzikon, Switzerland.
88–92:	Assistance at the Institute for Agricultural Economics, ETH Zürich (Prof. Dr. D. Onigkeit).
87–90:	Teaching computer science for unemployed people at VFBW in Zürich.
85:	3 months practice at the Institute for Organic Farming (FIBL), Oberwil, Switzerland.
84:	6 months practice on a farm.