CWI Tracts

CWI
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands
Telephone 31 - 20 592 9333, telex 12571 (mactr nl),
telefax 31 - 20 592 4199

CWI is the nationally funded Dutch institute for research in Mathematics and Computer Science.

CWI Tract 76

Stability, duality and decomposition in general mathematical programming

O.E. Flippo

Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

1980 Mathematics Subject Classification: 90C99.
ISBN 90 6196 398 2
NUGI-code: 811

Acknowledgements

Four years may already seem a considerable amount of time to produce a book of about two hundred pages, but it would have taken even longer if it were not for the additional support I received. The least creative contribution was financial. The main grant was awarded to me by the *Netherlands Organization for Scientific Research* (NWO; formerly ZWO) and the *Netherlands Foundation for Mathematics* (SMC); the amount of money involved was sufficient to keep me going these four years. As members of the latter foundation, *H.J.M. Wijers* and *A.R. Kloost* were assigned to monitor the research project, and this was definitely to my advantage. These two people turned out to be extremely co-operative and easy to get on with. Not once did I call on them in vain. I specifically thank them for giving me permission to leave my post at the Erasmus University Rotterdam from February till August 1988, in order to visit and work at the "Université de Paris-Dauphine", and for allowing me a trip to Tokyo in the summer of 1988, to participate in the "13^{th} International Symposium on Mathematical Programming". My stay at the Université de Paris-Dauphine, which I visited as a participant of the "European Doctoral Program in Quantitative Methods in Management" was additionally granted by the *European Community Action Scheme for the Mobility of University Students* (ERASMUS). This financial support is also gratefully acknowledged; Paris is an awfully expensive city, even if one would (unwisely) choose to live there like a monk. Finally, I am indebted to *Shell* for supporting partially my trip to Tokyo.

The second kind of support I received was professional. I acknowledge the contributions of *Antoon Kolen*, who asked me to do the job, and of *Gerard Van der Hoek*, who initiated the project, successfully applied to NWO for a grant, and acted as manager during the first two years. I am sorry that they both chose to withdraw from the project prematurely. I would also like to thank *Jaap Ponstein* for his contributions during the first year, and *Richard Wong*, who during his two months' visit to Rotterdam, collaborated with me on asymptotic convergence analysis. And even though we did not succeed in solving any open problems at that time, the current results reveal that we were in fact pretty close. During my stay at the Université de Paris-Dauphine, I was fortunate to collaborate with *Michel Minoux*. His energy and enthusiasm during our many intensive discussions just seemed to be tireless. After these sessions I always felt inspired to carry on. Finally, I am indebted to my advisor *Alexander Rinnooy Kan*. Although his time-consuming managerial activities as Rector Magnificus of the Erasmus University Rotterdam prevented him from being daily involved, his support has been substantial. Not that we sat down together to try to derive beautiful mathematical formulae; that was probably considered my job. His involvement was more at a supervisory level; he has been project leader for the last two years, he introduced me to a part of the scientific community, he raised my interest in going abroad and arranged my visit to Paris, and hopefully he will succeed in arranging another visit for me to the United States of America or Canada. I am confident that the scientific standards he taught me, will prove to be useful in any kind of environment in future, scientific or not.

I also acknowledge the moral support that was given to me, as well as any attempt at it. In this respect, I should mention my colleagues and former partners in (di)stress at the Econometric Institute; *Cees Dert*, who always found it necessary to try to cheer me up, *Stan Van Hoesel*, who then really cheered me up after Cees had left my office, by telling me the sad stories of his life, *Guus Boender*, for sharing with me his gossip and his delicious "Liga's with apple", *Auke Woerlee*, who really helped me on more than one occasion with the computer and with LaTeX, and who never ceased to amaze me with the "subtleties" of his (witty) remarks, and *Albert Wagelmans*, who got his master's degree and started his doctoral thesis on the very same days as I did; despite this long period together, I still do not share his taste of music, but I do thank him for having committed himself to make himself useful on December 14. I would also like to mention my *colleagues at LAMSADE*, Université de Paris-Dauphine, who with their cordiality and hospitality made me feel at home in this wonderful, but sometimes difficult city called Paris. There have also been many *friends*, *not* affiliated with any university, who have shown their interest in what I was doing. I apoligize for my apparent inability to clearly explain to them what my subject was all about; I know that it still strikes them as odd that one can earn a living just by reading, musing over, and writing down arcane mathematical formulae. Finally, I owe special thanks to *my parents*, who had to endure my presence for such a long time. They have always shown interest in my pursuits (though not always fully agreeing with them), and have always had a willing ear for my stories. It must have taught them a lot of patience.

Last but not least, I would like to express my deepest feelings of gratitude to the *typist* for a time-consuming, sometimes boring, in the end rewarding, but in any case, excellent job. Of course, the responsibility for all remaining (typing) errors are entirely his...

Contents

Prologue

From the late forties onwards, after George B. Dantzig had developed the simplex algorithm for the solution of linear programmes, *optimization problems* have been the subject of extensive research. Basically, this research aimed at either developing and analysing *algorithms* for specially structured optimization problems, or at strengthening existing and developing new *theoretical concepts*. The contents of our contribution belong to the latter category. As the title already indicates, it involves the three related notions of *stability, duality* and *decomposition*. Before discussing why and how these notions interact, we will first clarify the title by providing an intuitive idea of what each notion stands for separately.

In the sequel, *general mathematical programming* problems in finitely dimensional Euclidean spaces will be considered. A formulation of such a problem reads

$$\begin{array}{ll} \underset{x}{\text{maximize}} & f(x) \\ \text{subject to} & G_i(x) \diamond_i 0, \quad i = 1, \ldots, m \\ & x \in X \end{array}$$

where $f(\cdot)$ and $G_i(\cdot)$ are given (extended) real-valued functions, X is a given subset of the n-dimensional Euclidean space $\mathbb{R}^n$ and $\diamond_i$ is a given equality or inequality sign ($i = 1, \ldots, m$). Note that the finite dimensionality here, refers to both the solution space X and the constraint space $\mathbb{R}^m$. This implies for example, that problems in variational calculus and optimal control are not included in our analysis.

The notion of *stability* has been used by numerous authors to indicate a variety of things. In our terminology, stability refers to the continuity of the optimal objective function value of a mathematical programme with respect to changes in its right-hand-side. This optimal objective function value considered as a function of the right-hand-side, is denoted by the term *value-function*. So in what follows, stability will be identified with the *continuity of the value-function*.

The notion of *duality* is somewhat more difficult to explain. Suppose a vector of arguments is available — either from an oracle or from an algorithm — that satisfies the constraints of a given maximization problem. How can the quality of such a feasible solution be measured? A natural measure of quality would be based on a comparison between the associated objective function value and the *optimal* objective function value. Unfortunately, such a comparison can only be made in the ideal situation in which the latter value is known. In the more realistic situation, in which a device exists that is only able to generate *upper bounds* for the optimal objective function value, only *lower bounds on the quality* of the given feasible solution can be derived. So, in order to appraise the quality of a given feasible solution, it is crucial to have (tight) upper bounds. In a duality theory, a second optimization problem, called the *dual programme*, is defined, with the property that all of its feasible solutions yield upper bounds for the optimal objective function value of the original programme, which for discriminating purposes is usually referred to as the *primal programme*. This property is denoted by *weak duality*. Weak duality is a desirable property, but unfortunately, it does not always suffice. Suppose the primal feasible solution we started with happens to be an optimal solution, in the sense that its objective function value equals the optimal primal objective function value. We would only be able to recognize this solution as being optimal, if we actually knew the optimal primal objective function value. Such knowledge can be derived from the dual programme, only if for at least one of its solutions the gap between the upper bound it generates and the optimal primal objective function value disappears. This property is called *strong duality*. So, weak duality always yields lower bounds on the quality of any primal feasible solution, whereas only in combination with strong duality it becomes possible to recognize optimality. A *duality theory* is now defined as the *collection of interactions* between such a primal and dual programme.

In many situations, the mathematical programme under consideration contains one or several substructures which are in some sense easy to handle. A natural solution strategy is then to try to disengage these structures from each other and from the full problem, with the idea in mind that the original problem may be easier solved by exploiting their presence. Of course, a complete disengagement of such structures cannot be obtained other than in specific cases only, but even so, it may still be worthwhile to at least partially separate them from the full problem in some way. This is in fact what all *decomposition methods* aim for. In these methods, the original problem is tackled by *partially separating* useful and promising substructures from the full programme. Information gathered during the solution of these structures is then used to improve on the way they were separated from the original problem. In this way, an iterative procedure results between substructures on the one hand, and updates c.q. improvements on the way of separating them from the original problem on the other.

Now that we have explained the ingredients, we will return to the question how the aforementioned notions interact. Let us start with the relation between duality and decomposition. The decomposition methods that will be discussed here, are generalizations of two well-known decomposition methods for linear programmes, viz. Benders Decomposition and Dantzig-Wolfe Decomposition. In the latter two methods, Linear Programming duality plays a crucial role. Consequently, *generalizations* of these methods will require a similar theory. The general duality theory that will be used for this purpose, is an extension of Linear Programming duality, in the sense that the latter can be obtained from the former when applied to linear programmes. So, the relation between the two notions mentioned above is that *general duality theory is an essential prerequisite for the general decomposition methods that will be described.*

As already mentioned, the general duality theory we will consider, is an extension of Linear Programming duality. This observation, however, does not necessarily imply that all properties of the latter carry over to the former when applied to other than linear programmes. In fact, they do not. One of the properties that is lost in the general case, is the one-to-one correspondence between constraints in the primal programme and variables in the dual. The relation between stability and duality now stems from the fact that *in case of stability, this one-to-one correspondence between primal constraints and dual variables can be restored.*

The final relation concerns the one between stability and decomposition. *It can, and will, be proven that, except for pathological cases maybe, stability is an essential condition for the general decomposition methods to converge asymptotically.* The latter statement means that each accumulation point of the sequence of intermediate solutions which is generated by the iterative procedure, is in fact an optimal solution for the problem under consideration.

This monograph is subdivided into three separate parts, each one dealing with one of the three notions of stability, duality and decomposition separately. The contents of these parts are self-contained. Each part is preceded by an introduction and concluded by a summary, which also contains a list of theoretical contributions. The monograph itself is preceded by a preliminary section on notational conventions, and concluded by an epilogue, an author index, a subject index, a list of references and a summary in Dutch. The list of references is also subdivided into three parts. This has been done for reasons of clarity, although it has led to a small overlap between the lists of the second and third part. Finally, it should be mentioned that the reader is assumed to be familiar with elementary concepts in analysis (such as converging sequences, accumulation points, compactness, topological closure of a set, convexity), and with basic concepts in Linear Programming.

One final reflection is in order. As already mentioned, the approach followed is one of *abstraction*; the forthcoming analysis is definitely more conceptual than algorithmic in nature. One may of course question the use of theoretical research that does not seem to have any immediate application to the solution of real-life problems. However, one should always bear in mind that it seems impossible to encounter the limits of the scope of concepts and methods, if one remains safely *within* these limits. The only way to encounter those limits is to try to go *beyond* them. This is exactly what we have been doing here. By approaching the notion of stability and duality from a *general* point of view, and by trying to *generalize* the decomposition methods and their properties to the most general mathematical programmes, one gets a feel of where the limits of these concepts and methods lie. Of course, it will never be possible to decide whether or not these limits are artificial, in the sense that they might be overcome by looking at the problem from a different angle. However, as long as we realize their limitations, the results of a theoretical analysis may well turn out to be informative, useful and inspiring, even to those whose primary interest is in real-life applications.

Notational preliminaries

In this section some notational conventions will be agreed upon. Nearly all symbols will be generic, although there has been an attempt to minimize the number of changes in interpretation of each symbol in the course of the text. The symbols which have been defined by means of a numbered definition, however, will keep their meaning throughout the entire (sub)section in which they have been defined.

Sets will usually be denoted by a *capital roman* letter, except for some special cases, in which they will be denoted by a *capital greek* or *capital calligraphic* letter.

Variables (vectors as well as scalars) will be denoted by *small roman* or *small greek* letters. Indices are usually denoted by one of the letters i, j, k, ℓ, m, n.

(Extended) functions are single-valued mappings from a given domain to a given codomain. They will be denoted by a *roman* or *greek* letter, followed by a matching pair of parentheses enclosing a period. So, a function $f(\cdot)$ will be described as: $f(\cdot) : X \to Y$, where X and Y are the domain and codomain of $f(\cdot)$ respectively.

If the codomain of a function equals $(\mathbb{R} \cup \{\pm\infty\})^m$, where m is some positive integer, then the function is called *extended vector-valued*; in that case, it will be denoted by a *capital roman* or *capital greek* letter. In the case that $m = 1$ the function is called *extended real-valued* and is denoted by a *small* letter. If the codomain of a function equals $\mathbb{R}^m$, then the function is called *vector-valued*, or *real-valued* if $m = 1$.

Point-to-set maps are multi-valued mappings from a given domain to a given codomain. They relate to each point of the domain, a *subset* of the codomain. Point-to-set maps will be denoted by *small greek* letters, followed by a matching pair of parentheses enclosing a period. So, a point-to-set map $\alpha(\cdot)$ will be described as: $\alpha(\cdot) : X \rightarrow\rightarrow Y$, where X and Y are the domain and codomain of $\alpha(\cdot)$ respectively. A double arrow is used to distinguish point-to-set maps from functions.

If $\alpha(\cdot) : Y \rightarrow\rightarrow Z$ and $\beta(\cdot) : X \rightarrow\rightarrow Y$ are two point-to-set maps, then the *composed map* $\alpha(\beta(\cdot)) : X \rightarrow\rightarrow Z$ is defined as $\alpha(\beta(x)) = \cup_{y \in \beta(x)} \alpha(y) \; \forall x \in X$.

Sequences will be denoted by a variable, supplied with a superscript (usually k), and enclosed by a matching pair of parentheses; a subscript is added to indicate the possible values for the superscript. For instance, the infinite sequence $(x^1, x^2, \ldots)$ will be denoted by $(x^k)_{\mathbb{N}}$.

Subsequences can be described by defining a suitable *monotonically increasing function* from the set of possible values for the superscript of the given sequence to itself. In most cases, the notation $p(\cdot)$ will be used for this purpose. A subsequence of $(x^k)_{\mathbb{N}}$ is thus denoted by $(x^{p(k)})_{\mathbb{N}}$.

To describe the *limiting behaviour* of an infinite sequence, there will be no ambiguity with respect to which quantity approaches infinity. Therefore, any indication of this quantity will be suppressed. For instance, $\limsup x^k = x^\infty$ means

$$\limsup_{k \to \infty} x^k = x^\infty$$

A similar notation is used with respect to the *limes inferior* (lim inf) and the *limit* (lim).

Special symbols

- $\mathbb{N}$ is the set of positive integers.
- $\mathbb{Z}$ is the set of integers.
- $\mathbb{R}$ is the set of real numbers.
- $\mathbb{R}^m_+ = \{r \in \mathbb{R}^m \mid r_i \geq 0,\ i = 1, \ldots, m\}$ $(m \in \mathbb{N})$.
- $\mathcal{P}$ denotes a primal programme.
- $\mathcal{D}$ denotes a dual programme.
- $\mathcal{F}$ is the dual solution space.
- Γ denotes a subset of $\mathcal{F}$.
- $cl(\cdot)$ is the topological closure of a given set.
- ∇ denotes the Jacobian matrix of a given function.
- $\mathcal{C}^1(\cdot)$ is the set of all continuously differentiable functions on a given set.
- $\varphi(\cdot)$ denotes the optimal solution value of a given optimization problem (possibly $\pm\infty$).

Part I

Stability

Section 1

Introduction

In mathematical programming the notion of *stability* has been used by numerous authors to indicate a variety of things. The connection between all these different usages seems to be the *smoothness* of certain *characteristics* of mathematical programmes which are subject to *perturbation*. One such characteristic is described by the *value-function*, also denoted by *optimal value-function*, *extremal value-function*, *marginal function* and *perturbation function*. A general definition of this function reads

$$v(r) = \begin{cases} \sup_x \{\tilde{f}(x,r) \mid x \in \alpha(r)\} & \text{if } \alpha(r) \neq \emptyset \\ -\infty & \text{otherwise} \end{cases} \tag{1.1}$$

where $\tilde{f}(\cdot\,,\cdot)$ is an (extended) real-valued function and $\alpha(\cdot)$ is a point-to-set map which identifies with each choice of the parameter r, a set of feasible solutions for the resulting optimization problem; usually $\alpha(\cdot)$ is described by means of (in)equalities. Another characteristic, for instance, is the curve of (unique) optimal solutions of a family of perturbed optimization problems (if such a curve exists). With respect to the notion of smoothness, a similar variety of usages seems to exist.

Usually, a function is called *smooth* if it is (once) continuously differentiable, but occasionally, smoothness is associated with other properties, like continuity, or the existence of directional derivatives. In any case, the notion of *stability*, though not well-defined, is intimately related to the notion of *sensitivity*.

In what follows we will confine ourselves to the *continuity* of the *value-function*. In addition, we will only study the special case of (1.1) in which

$$\forall r \in \mathbb{R}^m : \tilde{f}(\cdot, r) = f(\cdot) \wedge \alpha(r) = \{x \in X \mid G(x) \diamond r\} \tag{1.2}$$

Here, X is a subset of $\mathbb{R}^n$, $G(\cdot)$ is an (extended) vector-valued function, and $\diamond \in \{\leq, =\}^m$ $(m, n \in \mathbb{N})$. Note that $G(x) \diamond r$ is a shorter notation for $G_i(x) \diamond_i r_i$, where $G_i(\cdot)$, r_i and $\diamond_i$ are the i-th components of $G(\cdot)$, r and $\diamond$ respectively $(i = 1, \ldots, m)$. So, we will restrict attention to those instances of the value-function for which the underlying optimization problems differ through their right-hand-sides only, and for which the domain as well as the codomain of the feasible set map $\alpha(\cdot)$ are subsets of *finitely* dimensional Euclidean spaces.

Some of the results that will be stated in the sequel are easily shown to hold for more general cases than the one we consider. However, two good reasons exist why we may avert from generality with a clear conscience. On the one hand, the case of *right-hand-side perturbations* is general enough in view of the results to come in Part II and III. On the other hand, (1.2) is not so restrictive as one might suspect at first sight. The loss of generality in (1.2) as compared to (1.1) is only twofold.

- The perturbations in (1.1) are described by means of a finite number of (extended) real-valued functions, their common domain being the Cartesian product of a given solution space and a given parameter space. More formally, $\alpha(r) = \{x \in X \mid H_i(x, r) \diamond_i 0,\ i \in I\}$, where X is some solution set, $H_i(\cdot, \cdot)$ is an (extended) real-valued function, $\diamond_i \in \{\leq, =\}$ and I is some finite index set.

- Both the domain and codomain of $\alpha(\cdot)$ are subsets of finitely dimensional Euclidean spaces.

It can easily be shown that under these two assumptions, (1.1) is just as general as the seemingly more restrictive case of right-hand-side perturbations considered in (1.2); the value-function is not affected if the parameter r is replaced by a new variable y and the constraint $y = r$ is added (cf. [Fiacco, 1983]). The result of these manipulations, however, is a mathematical programme which is perturbed through the right-hand-side only. Whether it is beneficial from an *analytical* point of view to actually perform these manipulations, heavily depends on the question to be resolved. The above argument only implies that, under the aforementioned assumptions, it is not overly restrictive from a *conceptual* point of view to consider only right-hand-side perturbations.

The outline of part I is as follows. In Section 2 it will be argued that the continuity of the value-function is essentially equivalent to the continuity of the corresponding feasible set map $\alpha(\cdot)$. In Section 3 a number of conditions is stated under which the latter property indeed holds. The sufficiency of these conditions is a direct consequence of the proofs given; their necessity, on the other hand, cannot be established unambiguously, although it can be argued that they are "almost necessary". This fuzzy conjecture results from the fact that they coincide with the three types of conditions which imply constraint-qualification in Karush-Kuhn-Tucker points, viz. linearity, convexity combined with Slater's condition, and the Mangasarian-Fromovitz regularity condition (cf. [Mangasarian & Fromovitz, 1967], [Minoux, 1986]). Therefore, it is not to be expected that similar statements concerning the continuity of the value-function hold under significantly weaker conditions, because these three types of conditions have been proven to play a too crucial role in Mathematical Programming. Part I is concluded by a summary.

The main contributions of Part I are the following. First of all, we already argued that considering right-hand-side perturbations is not so restrictive as one might be led to believe at first sight. Secondly, all results are presented within one single framework, being the continuity of the feasible set map. Some earlier results on the subject are thereby put into a unifying framework. Thirdly, it is proven that the continuity of the feasible set map is not only sufficient, but in some sense also necessary for the continuity of the value-function. Fourthly, it is demonstrated that the value-function is continuous on its domain of finiteness if all constraints are strictly quasi-convex and of the "less-than-or-equal" type. This result is a unification and an extension of two known results (cf. [Evans & Gould, 1970] and [Hogan, 1973[a]]). Finally, two new results are stated that combine some of the conditions under which the value-function was already shown to be continuous.

Before concluding this section, let us relate Part I to some previous work on the subject. In [Dantzig et al., 1967] the continuity of the value-function is studied through the behaviour of the *optimal set map*; this is a point-to-set map which identifies with each choice of parameter values, the set of optimal solutions of the resulting mathematical programme. The authors present conditions which ensure continuity of the value-function in case $\alpha(\cdot)$ is described by means of *parameterized affine functions*. As such, our result in the linear case (Theorem 3.1) is included in their work. Their proof, however, is significantly more complicated because they embedded the result in a considerably more general setting (general perturbations versus right-hand-side perturbations). B Results concerning the continuity of value-functions in Linear Programming can be found in [Böhm, 1975], [Martin, 1975], [Bereanu, 1976], [Robinson, 1977], [Mangasarian & Meyer, 1979] and [Wets, 1985].

In [Evans & Gould, 1970] the case of *right-hand-side perturbations* of *inequality constraints* is taken into consideration. First, the authors derive an abstract necessary and sufficient condition for the continuity of the value-function at a point. Then they show that in case all constraint functions are strictly quasi-convex, their condition holds and, as a result, the value-function is continuous on a certain domain. This result is a more restrictive version of our result for the strictly quasi-convex case (Theorem 3.3).

The results in [Greenberg & Pierskalla, 1972] are an extension of the ones in [Evans & Gould, 1970]. The former two authors prove that under the same conditions as derived by the latter two, the value-function also varies continuously with perturbations in the constraint functions. They also extend the work of Evans and Gould by allowing for *equality* constraints in the description of the feasible set map. This result, however, is comprised by our Theorem 3.3. In [Greenberg & Pierskalla, 1975] the results are also proven to hold in case there are an infinite number of constraints.

In [Dirickx et al., 1972] attention is also restricted to *right-hand-side perturbations*. The authors come up with a rather abstract sufficient condition for the continuity of the value-function by imposing conditions on the value-function itself; they show that this function is continuous on its domain of finiteness if the objective and constraint functions are convex and if all directional derivatives of the value-function exist and are bounded from below.

In [Daniel, 1973[b]] the continuity of the curve of (unique) optimal solutions in Definite Quadratic Programmes is analysed; in [Guddat, 1976] Convex Quadratic Programmes are studied. In [Bank & Hansel, 1984] the continuity of the value-function in Mixed-Integer Quadratic Programmes is dealt with.

Besides [Meyer, 1970], it was also [Hogan, 1973[b]] that served as an example to us. With identical notions and proof-techniques, these authors derive similar results as in our Section 2, but for the more general problem setting (1.1). For the case in which the map $\alpha(\cdot)$ is supposed to be described by means of inequalities, Hogan derives three interesting results. First, a more general version of our result for the convex case (Theorem 3.2) is proven to hold: Hogan relaxes the requirement of right-hand-side perturbations. Secondly, a generalization of the abstract result in [Evans & Gould, 1970] is presented. This result is not so interesting from an operational point of view because it states conditions on the map $\alpha(\cdot)$ itself, rather than on the constraint functions which describe $\alpha(\cdot)$. Finally, the case in which the constraints are assumed to be a mixture of equalities and inequalities is considered. Though resembling our Theorem 3.5, the two results are, in fact, incomparable; Hogan's requirements are weaker than ours, but so is his result when applied to our case.

[Brosowski, 1984] and [Zencke & Hettich, 1987] consider the case in which $\alpha(\cdot)$ is described by means of an (in)finite number of affine inequality constraints. Although the latter paper largely concentrates on the existence of directional derivatives, they do have one interesting result in common: the value-function is proven to be continuous if a Slater-type condition is assumed to hold. However, if their framework is applied to the more restrictive cases we consider, their result is comprised by ours.

Besides continuity, other important properties of perturbed mathematical programmes have been studied. A large body of literature exists on the Lipschitzean behaviour and the existence of all kinds of derivatives of the value-function, the curve of (local) minimizers and the curve of Karush-Kuhn-Tucker points; cf. [Aubin, 1984], [Auslander, 1979,1984], [Cornet & Vial, 1986], [Demyanov & Zabrodin, 1986], [Dontchev & Jongen, 1986], [Fiacco, 1976], [Fujiwara, 1985], [Gauvin, 1979], [Gauvin & Dubeau, 1982], [Gollan, 1984[a],1984[b]], [Hogan, 1973[c]], [Janin, 1984], [Jittorntrum, 1984], [Kojima & Hirabayashi, 1984], [Mangasarian & Shiau, 1987], [Penot, 1984,1988], [Robinson, 1974,1982,1987], [Rockafellar, 1982,1984], [Seeger, 1988], [Shapiro, 1985[a],1985[b],1988], [Spingarn, 1980] and [Stern & Topkis, 1976]. A notion of stability which is not stated in terms of optimization problems, is discussed by [Daniel, 1973[a],1975] and [Robinson, 1975,1976,1980]. Finally, all kinds of convexity and concavity properties of the value-function have been studied by [Kyparisis & Fiacco, 1987]. Although the list of references is already comprehensive, the reader should be aware that it is not yet complete. The references which are selected here, are a representative subset of, and a complete introduction to all the papers which are dealing with one of these related subjects. For extensive surveys on sensitivity and stability in nonlinear programming we refer to [Fiacco & Hutzler, 1982], [Bank et al., 1983] and [Fiacco, 1983].

Section 2

Some basic definitions and results

The approach we pursue in Section 2 is largely based on [Meyer, 1970] and [Hogan, 1973^b]. The results in this section can easily be generalized to the more general problem statement (1.1) (see e.g. [Debreu, 1959], [Berge, 1963] and the two references cited above). Theorem 2.2, which states that the continuity of the feasible set map is, in some sense, a necessary condition for the continuity of the value-function, seems to be new.

Consider the following family of mathematical programming problems:

$$\begin{aligned} \mathcal{P}(r): \quad & \max_{x} \quad f(x) \\ & \text{s.t.} \quad G(x) \diamond r \\ & \qquad\; x \in X \end{aligned} \tag{2.1}$$

where $f(\cdot)$ and $G(\cdot)$ are functions from $D \subseteq \mathbb{R}^n$ to the $\mathbb{R} \cup \{\pm\infty\}$ and $(\mathbb{R} \cup \{\pm\infty\})^m$ respectively, X is a subset of D and $\diamond \in \{\leq, =\}^m$ $(m, n \in \mathbb{N})$. As already mentioned, $G(x) \diamond r$ is a shorter notation for

$G_i(x) \diamond_i r_i$, where $G_i(\cdot)$, $\diamond_i$, and r_i are the i-th components of $G(\cdot)$, $\diamond$ and r respectively ($i = 1, \ldots, m$). More formally:

$$\begin{array}{lcl} X \subseteq D \subseteq \mathbb{R}^n & ; & \diamond \in \{\leq, =\}^m \\ f(\cdot) : D \rightarrow \mathbb{R} \cup \{\pm\infty\} & ; & G(\cdot) : D \rightarrow (\mathbb{R} \cup \{\pm\infty\})^m \end{array} \tag{2.2}$$

For each right-hand-side $r \in \mathbb{R}^m$ the feasible set of $\mathcal{P}(r)$ is defined by the following point-to-set map

$$\alpha(\cdot) : \mathbb{R}^m \rightarrow\rightarrow \mathbb{R}^n \text{ with } \alpha(r) = \{x \in X \mid G(x) \diamond r\} \tag{2.3}$$

The map $\alpha(\cdot)$ is called the *feasible set map*. The right-hand-sides for which the corresponding feasible set is non-empty, form the *set of feasible right-hand-sides*; it is defined as

$$RHS = \{r \in \mathbb{R}^m \mid \alpha(r) \neq \emptyset\}. \tag{2.4}$$

The *value-function* is defined as

$$\begin{array}{l} v(\cdot) : \mathbb{R}^m \rightarrow \mathbb{R} \cup \{\pm\infty\}, \text{ with} \\ v(r) = \begin{cases} \sup\limits_x \{f(x) \mid x \in \alpha(r)\} & \text{if } r \in RHS \\ -\infty & \text{otherwise} \end{cases} \end{array} \tag{2.5}$$

The subset of RHS on which $v(\cdot)$ is real-valued, is denoted by V; hence

$$V = \{r \in \mathbb{R}^m \mid v(r) \in \mathbb{R}\}. \tag{2.6}$$

Note that $V \subseteq RHS$. Our goal is to come up with conditions on $f(\cdot)$, $G(\cdot)$, X and $\diamond$, which enforce certain continuity properties on the value-function $v(\cdot)$. Unfortunately, the classical definition of continuity cannot be used, because this notion might be meaningless when applied to points which lie on the boundary of RHS or on the boundary of V. To resolve this difficulty, we use a slightly different notion of continuity.

Definition 2.1 *Let $W \subseteq \mathbb{R}^m$. A sequence $(y^k)_{\mathbb{N}}$ is said to be* in W *if* $\forall k \in \mathbb{N} : y^k \in W$.

Definition 2.2 ((Semi-)continuity with respect to) *Let $W \subseteq \mathbb{R}^m$. In addition, let $w(\cdot) : W \to \mathbb{R} \cup \{\pm\infty\}$ be an (extended) real-valued function on W. If for all sequences $(y^k)_{\mathbb{N}}$ in W with $\lim y^k = y \in W$,*

- $\limsup w(y^k) \leq w(y)$, *then $w(\cdot)$ is called* upper semi-continuous at y w.r.t. W;
- $\liminf w(y^k) \geq w(y)$, *then $w(\cdot)$ is called* lower semi-continuous at y w.r.t. W;
- $\lim w(y^k) = w(y)$, *then $w(\cdot)$ is called* continuous at y w.r.t. W.

From this definition it immediately follows that a function is continuous at y w.r.t. W if and only if it is both upper and lower semi-continuous at y w.r.t. W. Furthermore, a function is said to be (semi-)continuous w.r.t. W if and only if it has the same property at every point $y \in W$.

For points y that lie in the interior of W, the classical definition of continuity is reobtained. Therefore, Definition 2.2 can be considered to be an extension of the classical one in defining a notion of continuity at boundary points. In fact, Definition 2.2 boils down to the classical one if, instead of the natural topology on $\mathbb{R}^m$, the topology induced by

W is considered. Note that every function $w(\cdot)$ is continuous w.r.t. W if W is a singleton.

Before going into technical details, let us first try to develop some intuition. Continuity of the value-function means that for slight changes in the right-hand-side r, the optimal solution value of $\mathcal{P}(r)$ also changes only moderately. For a continuous objective function $f(\cdot)$ this seems to imply that the feasible set in (2.1) should not change drastically in case only minor changes in the right-hand-side are carried through. In other words, the feasible set map $\alpha(\cdot)$ should change continuously with changes in its arguments. In order to make this notion of continuity for point-to-set maps more precise, we adopt the terminology of [Hogan, 1973[b]]; in [Meyer, 1970] the same definitions occur under the names l.s.c. and u.s.c. respectively.

Definition 2.3 (Openness with respect to) *Let $W \subseteq \mathbb{R}^m$. In addition, let $\gamma(\cdot) : W \to\to \mathbb{R}^n$ be a point-to-set map. Then $\gamma(\cdot)$ is said to be* open at $y \in W$ w.r.t. W *if from*

- $(y^k)_{\mathbb{N}}$ *is a sequence in W with* $\lim y^k = y$, *and*
- $z \in \gamma(y)$,

it follows that there is a sequence $(z^k)_{\mathbb{N}}$ with $z^k \in \gamma(y^k)\ \forall k \in \mathbb{N}$ and $\lim z^k = z$.

Definition 2.4 (Closedness with respect to) *Let $W \subseteq \mathbb{R}^m$. Additionally, let $\gamma(\cdot) : W \to\to \mathbb{R}^n$ be a point-to-set map. Then $\gamma(\cdot)$ is said to be* closed at $y \in W$ w.r.t. W *if from*

- $(y^k)_{\mathbb{N}}$ *is a sequence in W with* $\lim y^k = y$, *and*

- $(z^k)_{\mathbb{N}}$ *is a sequence with* $z^k \in \gamma(y^k)\ \forall k \in \mathbb{N}$ *and* $\lim z^k = z$,

it follows that $z \in \gamma(y)$.

Definition 2.5 (Continuity with respect to) *Let* $W \subseteq \mathbb{R}^m$. *In addition, let* $\gamma(\cdot) : W \rightarrow\rightarrow \mathbb{R}^n$ *be a point-to-set map. Then* $\gamma(\cdot)$ *is said to be* continuous at $y \in W$ w.r.t. W *if* $\gamma(\cdot)$ *is both open and closed at* y *w.r.t.* W.

A point-to-set map is said to be open/closed/continuous w.r.t. W if it has the same property at every point $y \in W$ w.r.t. W. The same difference that exists between Definition 2.2 and the classical notion of (semi-)continuity for functions is encountered here too: Definitions 2.3–2.5 coincide with the "classical" ones if the natural topology on $\mathbb{R}^m$ is replaced by the topology induced by W. Moreover, every point-to-set map $\gamma(\cdot)$ is open w.r.t. W if W is a singleton. Note that this statement does not apply to the notion of closedness, unless $\gamma(W)$ is a topologically closed set. Also note that the closedness of $\gamma(\cdot)$ w.r.t. W implies the topological closedness of every set $\gamma(y)$ where $y \in W$. Definitions 2.3 and 2.4 reveal that there is a difference between the topological openness (closedness) of a set and the openness (closedness) of a point-to-set map; from the context it will always be clear which specific notion is referred to.

Definitions 2.3 and 2.4 precisely fulfil the purpose they are designed for. If the feasible set map is open, then it does not suffer from discontinuous expansions, because for *any* sequence $(r^k)_{\mathbb{N}}$ in W with $\lim r^k = r \in W$, *every* point $x \in \alpha(r)$ can be obtained as a limiting point of a sequence $(x^k)_{\mathbb{N}}$ with $x^k \in \alpha(r^k)\ \forall k \in \mathbb{N}$. Hence, at infinity, points will not suddenly "crop up". In such a case, situations like the one in Figure 2.1 will not occur.

As far as closedness is concerned, the situation is kind of reversed. If the feasible set map is closed, then it does not suffer from discontinuous contractions because for *any* sequence $(r^k)_{\mathbb{N}}$ in W with $\lim r^k = r \in W$,

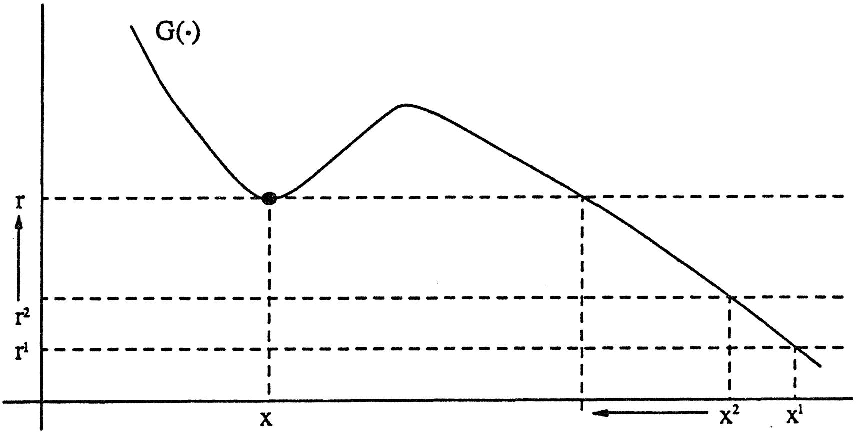

Figure 2.1: An example of a feasible set map $\alpha(r) = \{x \in \mathbb{R} \mid G(x) \leq r\}$ which is not open.

the limiting point x of *every* converging sequence $(x^k)_{\mathbb{N}}$ with $x^k \in \alpha(r^k)\ \forall k \in \mathbb{N}$, is an element of $\alpha(r)$. Hence, at infinity, points will not suddenly "drop out". In such a case, situations like the one in Figure 2.2 will not occur.

Now, if a feasible set map does not suffer from discontinuous expansions, the optimal solution value of a continuous objective function cannot increase discontinuously. To say it in other words, the *lower semi-continuity* of $v(\cdot)$ is likely to follow from the *openness* of $\alpha(\cdot)$. Similarly, the *upper semi-continuity* of $v(\cdot)$ might follow from the *closedness* of $\alpha(\cdot)$. Below, these assertions will be made more precise.

Definition 2.6 (Essentially boundedness with respect to) *Let $W \subseteq \mathbb{R}^m$. In addition, let $\gamma(\cdot) : W \to\!\!\to \mathbb{R}^n$ be a point-to-set map. Then $\gamma(\cdot)$ is said to be* essentially bounded at $y \in W$ w.r.t. W *if from*

- *$(y^k)_{\mathbb{N}}$ is a sequence in W with* $\lim y^k = y$ *and*
- *$(z^k)_{\mathbb{N}}$ is a sequence with $z^k \in \gamma(y^k)\ \forall k \in \mathbb{N}$,*

it follows that $(z^k)_{\mathbb{N}}$ is bounded.

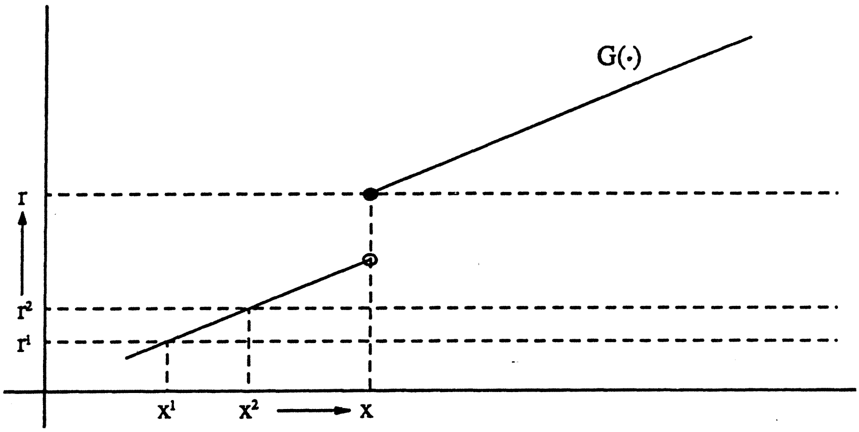

Figure 2.2: An example of a feasible set map $\alpha(r) = \{x \in \mathbb{R} \mid G(x) \leq r\}$ which is not closed.

This notion resembles Hogan's "uniformly compactness" (see [Hogan, 1973[b]]); it is slightly weaker but it has been introduced for the same purpose. Again, a point-to-set map is said to be essentially bounded w.r.t. W if it is essentially bounded at every point $y \in W$ w.r.t. W. Like [Hogan, 1973[b]] we can now derive the following result; note that we use $\cup_W \alpha(r)$ as a shorter notation for $\cup_{r \in W} \alpha(r)$.

Theorem 2.1 ((Semi-)continuity of $v(\cdot)$ – sufficient conditions) *Let (2.1)–(2.6) be given. In addition, let*

1. *$W \subseteq V$ and suppose that*

   - *$f(\cdot)$ is lower semi-continuous w.r.t. $\cup_W \alpha(r)$ and*
   - *$\alpha(\cdot)$ is open w.r.t. W*

   Then $v(\cdot)$ is lower semi-continuous w.r.t. W.

2. $W \subseteq V$ and suppose that

- *$f(\cdot)$ is upper semi-continuous w.r.t. $\cup_W \alpha(r)$*
- *$\alpha(\cdot)$ is essentially bounded w.r.t. W and*
- *$\alpha(\cdot)$ is closed w.r.t. W*

Then $v(\cdot)$ is upper semi-continuous w.r.t. W.

3. $W \subseteq V$ and suppose that

- *$f(\cdot)$ is continuous w.r.t. $\cup_W \alpha(r)$*
- *$\alpha(\cdot)$ is essentially bounded w.r.t. W and*
- *$\alpha(\cdot)$ is continuous w.r.t. W*

Then $v(\cdot)$ is continuous w.r.t. W.

Proof

1. Take any arbitrary sequence $(r^k)_{\mathbb{N}}$ in W with $\lim r^k = r^\infty \in W$. We have to prove that $\liminf v(r^k) \geq v(r^\infty)$.

   Let $\epsilon > 0$ be arbitrarily chosen. $W \subseteq V$ implies that

   $$\exists x^\infty \in \alpha(r^\infty) : \; f(x^\infty) > v(r^\infty) - \epsilon$$

   Furthermore, $\alpha(\cdot)$ is open w.r.t. W, so

   $$\exists (x^k)_{\mathbb{N}} \text{ with } x^k \in \alpha(r^k) \; \forall k \in \mathbb{N} \text{ and } \lim x^k = x^\infty$$

   From the lower semi-continuity of $f(\cdot)$ w.r.t. $\cup_W \alpha(r)$ it follows that

   $$v(r^\infty) < f(x^\infty) + \epsilon \leq \liminf f(x^k) + \epsilon \leq \liminf v(r^k) + \epsilon$$

As this relation holds for every $\epsilon > 0$, it follows that $\liminf v(r^k) \geq v(r^\infty)$, which proves the theorem.

2. Take any arbitrary sequence $(r^k)_{\mathbb{N}}$ in W with $\lim r^k = r^\infty \in W$. We have to prove that $\limsup v(r^k) \leq v(r^\infty)$. $W \subseteq V$ implies that

$$\forall k \in \mathbb{N}\ \exists x^k \in \alpha(r^k) : \ v(r^k) - \frac{1}{k} < f(x^k) \leq v(r^k)$$

As a result we get

$$\limsup v(r^k) = \limsup f(x^k)$$

Let $(x^{p(k)})_{\mathbb{N}}$ be a subsequence of $(x^k)_{\mathbb{N}}$, such that

$$\lim f(x^{p(k)}) = \limsup f(x^k)$$

From the essentially boundedness it follows that $(x^{p(k)})_{\mathbb{N}}$ has an accumulation point x^∞. Let $(x^{q(k)})_{\mathbb{N}}$ be a subsequence of $(x^{p(k)})_{\mathbb{N}}$ such that $\lim x^{q(k)} = x^\infty$. From the closedness of $\alpha(\cdot)$ it follows that $x^\infty \in \alpha(r^\infty)$, so

$$f(x^\infty) \leq v(r^\infty)$$

The upper semi-continuity of $f(\cdot)$ then, implies that

$$v(r^\infty) \geq f(x^\infty) \geq \limsup f(x^{q(k)}) = \lim f(x^{q(k)}) = \\ \lim f(x^{p(k)}) = \limsup f(x^k) = \limsup v(r^k)$$

3. According to the previous two results, $v(\cdot)$ is both lower and upper semi-continuous w.r.t. W, hence continuous w.r.t. W.

□

One of the implications of this result is the following

Corollary 2.1 (Closedness of the optimal set map)
Let (2.1)–(2.6) be given. In addition, let $\xi(\cdot)$ denote the optimal set map, *defined as*

$$\xi(\cdot) : \mathbb{R}^m \to\to \mathbb{R}^n \text{ with } \xi(r) = \{x \in \alpha(r) \mid f(x) = v(r)\}$$

If the conditions of Theorem 2.1 sub 3 are satisfied, then $\xi(\cdot)$ is closed w.r.t. W.

Proof This is left to the reader.

□

According to Theorem 2.1, continuity of the feasible set map (besides continuity of the objective function and some boundedness condition on the feasible set map) is a sufficient condition for the continuity of the value-function. What about its necessity? Of course, specific cases may exist in which the value-function is continuous whereas the feasible set map is not. However, if the value-function is to be continuous for *any* continuous objective function $f(\cdot)$, then the feasible set map must be both open and closed. In this respect, continuity of the feasible set map is a necessary condition as well.

Theorem 2.2 ((Semi-)continuity of $v(\cdot)$ – necessary conditions)
Let (2.1)–(2.5) be given. In addition, let

1. *$W \subseteq RHS$ and suppose that*

   - *for* any *objective function $f(\cdot)$ which is continuous w.r.t. D and for which the resulting value-function is real-valued on W, this value-function is lower semi-continuous w.r.t. W*

   Then $\alpha(\cdot)$ is open w.r.t. W.

2. $W \subseteq RHS$ and suppose that

- *$\forall r \in W : \alpha(r)$ is a closed set and*
- *for* any *objective function $f(\cdot)$ which is continuous w.r.t. D and for which the resulting value-function is real-valued on W, this value-function is upper semi-continuous w.r.t. W*

Then $\alpha(\cdot)$ is closed w.r.t. W.

Proof

1. Let $(r^k)_{\mathbb{N}}$ be any sequence in W, converging to $r^\infty \in W$. Let $x^\infty \in \alpha(r^\infty)$ be arbitrarily chosen. We have to prove that there is a sequence $(x^k)_{\mathbb{N}}$ with $\forall k \in \mathbb{N} : x^k \in \alpha(r^k)$ and $\lim x^k = x^\infty$.

   Let the objective function $f(\cdot)$ be chosen as $f(x) = -d(x, x^\infty)$, where $d(\cdot\,,\cdot)$ is any distance function. The corresponding value-function $\delta(\cdot)$ is assumed to be lower semi-continuous, so

   $$\liminf \delta(r^k) \geq \delta(r^\infty) = 0$$

   Furthermore, $W \subseteq RHS$, so

   $$\forall k \in \mathbb{N}\ \exists x^k \in \alpha(r^k) :\ \delta(r^k) - \frac{1}{k} < -d(x^k, x^\infty)$$

   From this relation it follows that

   $$\limsup d(x^k, x^\infty) \leq -\liminf \delta(r^k) \leq -\delta(r^\infty) = 0,$$

   which implies that $\lim d(x^k, x^\infty) = 0$, so $\lim x^k = x^\infty$.

2. Let $(r^k)_{\mathbb{N}}$ be any sequence in W, converging to $r^\infty \in W$. Furthermore, let $(x^k)_{\mathbb{N}}$ be a sequence, such that $\forall k \in \mathbb{N} : x^k \in \alpha(r^k)$ and $\lim x^k = x^\infty$. We have to prove that $x^\infty \in \alpha(r^\infty)$.

   Let the objective function $f(\cdot)$ be chosen as $f(x) = -d(x, x^\infty)$. Let $\delta(\cdot)$ denote the resulting value-function. Obviously, $\forall r \in \mathbb{R}^m : \delta(r) \leq 0$. From the definition of $\delta(\cdot)$ and the fact that $\forall k \in \mathbb{N} : x^k \in \alpha(r^k)$ and $\lim x^k = x^\infty$, it follows that

$$0 = \lim \delta(r^k)$$

Furthermore, $\delta(\cdot)$ is supposed to be upper semi-continuous, so

$$0 = \limsup \delta(r^k) \leq \delta(r^\infty) \leq 0$$

Consequently, $\delta(r^\infty) = 0$, which, combined with the closedness condition on the set $\alpha(r^\infty)$, implies that $x^\infty \in \alpha(r^\infty)$.

□

Theorem 2.2 suggests that continuity of the feasible set map is a necessary as well as sufficient condition, if for a sufficiently large class of objective functions (in any case, including certain distance functions) all resulting value-functions are to be continuous. The result does not suggest anything concerning the necessity of the continuity of the feasible set map as soon as only one specific case (i.e. one specific (type of) objective function) is taken into consideration. In any case, Theorem 2.1 and 2.2 imply that for general continuous objective functions, the continuity of the value-function may equally well be analysed through the continuity of the feasible set map.

In the sequel we will try to come up with conditions on the problem data $G(\cdot), X$ and $\diamond$, such that the resulting feasible set map $\alpha(\cdot)$ is both open and closed. Let us start with closedness. Closedness can be enforced under very weak conditions, as can be seen by the following theorem (see also [Hogan, 1973[b]]); recall that the topological closure-operator on sets is denoted by $cl(\cdot)$.

Theorem 2.3 (Closedness) *Let (2.1)–(2.5) be given. In addition, let $W \subseteq RHS$ and X be closed. Moreover, suppose that $G_i(\cdot)$ is lower semi-continuous w.r.t. $cl(\cup_W \alpha(r))$ if $\diamond_i \in \{\leq\}$ and that $G_i(\cdot)$ is continuous w.r.t. $cl(\cup_W \alpha(r))$ if $\diamond_i \in \{=\}$. Then $\alpha(\cdot)$ is closed w.r.t. W.*

Proof Let $(r^k)_{\mathbb{N}}$ be any sequence in W, converging to $r^\infty \in W$. Let $(x^k)_{\mathbb{N}}$ be a sequence, such that $\forall k \in \mathbb{N} : x^k \in \alpha(r^k)$ and $\lim x^k = x^\infty$. We have to prove that $x^\infty \in \alpha(r^\infty)$.

First of all, $\forall k \in \mathbb{N} : x^k \in X$, so from the closedness of X it follows that

$$x^\infty \in X$$

Furthermore, if $\diamond_i \in \{\leq\}$, then the lower semi-continuity of $G_i(\cdot)$ implies that

$$G_i(x^\infty) \leq \liminf G_i(x^k) \leq \liminf r_i^k = r_i^\infty$$

Finally, if $\diamond_i \in \{=\}$, then the continuity of $G_i(\cdot)$ implies that

$$G_i(x^\infty) = \lim G_i(x^k) = \lim r_i^k = r_i^\infty$$

As a result, $x^\infty \in \alpha(r^\infty)$.

□

The conditions mentioned in Theorem 2.3 are weak. This explains why the function $G(\cdot)$ which is depicted in Figure 2.2, had to be chosen so nasty. It also reveals that the upper semi-continuity of the value-function can be enforced by imposing only some weak conditions on the problem data.

Unfortunately, openness does not seem to be of the same simplicity. This is already revealed by the fact that even a "well-behaved" function $G(\cdot)$ like the one depicted in Figure 2.1, does not enforce openness of the feasible set map. Intuitively, this can be explained by the fact that this function $G(\cdot)$ has a local minimum which is not global as well. In fact, if one tries to draw (one-dimensional) functions $G(\cdot)$ such that the resulting feasible set map is, indeed, open, it seems impossible to succeed if $G(\cdot)$ has the aforementioned property. As we shall see in the next section, similar observations apply to the multi-dimensional case too. In any case, it is the lower semi-continuity of the value-function which seems difficult to assure.

Section 3

Continuity of feasible set maps

In this section we will study some specific cases in which the feasible set map $\alpha(\cdot)$, as defined as in (2.3), is both open and closed. In fact, our main concern will be openness, because closedness can easily be obtained under very weak conditions (Theorem 2.3). Let us first state a lemma which appears to be useful in proving the openness of a point-to-set map.

Lemma 3.1 *Let (2.1)–(2.5) be given. In addition, let $W \subseteq RHS$ and suppose that $(r^k)_{\mathbb{N}}$ is a sequence in W, converging to $r^\infty \in W$. Furthermore, let $x^\infty \in \alpha(r^\infty)$ be arbitrarily chosen, and suppose that*

$$\forall \epsilon > 0 \; \exists k_0 \in \mathbb{N} \; \forall k > k_0 \; \exists x^k \in \alpha(r^k) : \| x^k - x^\infty \| < \epsilon$$

Then a sequence $(y^k)_{\mathbb{N}}$ exists with $y^k \in \alpha(r^k) \; \forall k \in \mathbb{N}$ and $\lim y^k = x^\infty$.

Proof This is left to the reader.

3.1 The affine case

In this subsection, the feasible sets of the underlying mathematical programming problems will be assumed to be polyhedral.

Theorem 3.1 (The affine case) *Let (2.1)–(2.5) be given. Suppose that X is a closed polyhedral set and $G(x) = Ax - b$ where A is a matrix of order $m \times n$, and b is an m-vector. Then $\alpha(\cdot)$ is continuous w.r.t. RHS.*

Proof The closedness of $\alpha(\cdot)$ w.r.t. RHS immediately follows from Theorem 2.3. In order to prove the openness, let us suppose that $(r^k)_{\mathbb{N}}$ is a sequence in RHS, converging to $r^\infty \in RHS$. Let $x^\infty \in \alpha(r^\infty)$. We will construct a sequence $(x^k)_{\mathbb{N}}$ such that $\forall k \in \mathbb{N}$: $x^k \in \alpha(r^k)$ and $\lim x^k = x^\infty$. First we will consider the case where $X = \mathbb{R}^n$, then X will be assumed to be a general closed polyhedral set.

Case 1: $X = \mathbb{R}^n$

Consider for each $k \in \mathbb{N}$ the following mathematical programme

$$\begin{array}{ll} \min\limits_{x} & \| x - x^\infty \| \\ \text{s.t.} & Ax - b \diamond r^k \end{array} \tag{3.1}$$

where $\| \cdot \|$ denotes the max-norm. An equivalent formulation reads

$$\begin{array}{lrcll} \min\limits_{x,z} & z & & & \\ \text{s.t.} & -x_j + z & \geq & -x_j^\infty & \forall j \in \{1, \ldots, n\} \\ & x_j + z & \geq & x_j^\infty & \forall j \in \{1, \ldots, n\} \\ & Ax & \diamond & b + r^k & \end{array} \tag{3.2}$$

Programme (3.2) is a linear programme; its dual is

$$\begin{aligned}\max_{\lambda,\mu,\sigma} \quad & -\lambda^T x^\infty + \mu^T x^\infty - \sigma^T(b + r^k) \\ \text{s.t.} \quad & -\lambda + \mu - A^T\sigma = 0 \\ & \sum_{j=1}^n \lambda_j + \sum_{j=1}^n \mu_j = 1 \\ & \lambda, \mu \geq 0 \\ & \sigma_j \geq 0 \ \text{ if } \diamond_j \in \{\leq\}\end{aligned} \tag{3.3}$$

Because of the fact that $\forall k \in \mathbb{N}: \; r^k \in RHS$, it follows that (3.1) has an optimal solution, say x^k. Let the optimal solutions of (3.2) and (3.3) be denoted by (x^k, z^k) and $(\lambda^k, \mu^k, \sigma^k)$ respectively, where, without loss of generality, the latter point can be chosen to be an *extreme point* of the polyhedron which is defined by the feasible set of (3.3). Note that this polyhedron is independent of $k \in \mathbb{N}$, so $(\lambda^k, \mu^k, \sigma^k)$ belongs to some *finite* set. This implies that

$$\exists s \in \mathbb{R}_+ \; \forall k \in \mathbb{N} : \| \sigma^k \| \leq s$$

Furthermore, from primal and dual feasibility it follows that

$$\begin{aligned}&(-\lambda^k)^T x^\infty + (\mu^k)^T x^\infty - (\sigma^k)^T(b + r^k) = \\ &(-\lambda^k + \mu^k)^T x^\infty - (\sigma^k)^T(b + r^k) = \\ &(\sigma^k)^T(Ax^\infty) - (\sigma^k)^T(b + r^k) = \\ &(\sigma^k)^T(Ax^\infty - b - r^k) \leq \\ &(\sigma^k)^T(r^\infty - r^k)\end{aligned}$$

As a result,

$$\begin{aligned}&\| x^k - x^\infty \| = z^k = \\ &(-\lambda^k)^T x^\infty + (\mu^k)^T x^\infty - (\sigma^k)^T(b + r^k) \leq \\ &(\sigma^k)^T(r^\infty - r^k) \leq |(\sigma^k)^T(r^\infty - r^k)| \leq \\ &m \| \sigma^k \| \cdot \| r^k - r^\infty \| \leq ms \| r^k - r^\infty \|\end{aligned}$$

Now, let $\epsilon > 0$ be given. From $\lim r^k = r^\infty$ it follows that

$$\exists k_0 \in \mathbb{N}\ \forall k > k_0 : \| r^k - r^\infty \| < \frac{\epsilon}{ms}$$

Hence,

$$\forall k > k_0 : \| x^k - x^\infty \| \leq ms \| r^k - r^\infty \| < ms \frac{\epsilon}{ms} = \epsilon$$

By construction, $x^k \in \alpha(r^k)$ and $\lim x^k = x^\infty$, which proves the theorem for the case $X = \mathbb{R}^n$.

Case 2: X is a closed polyhedral set

Let $X = \{x \in \mathbb{R}^n \mid Ex - e \leq 0\}$ where E is a matrix of order $m' \times n$, and e is an m'-vector. Define the following point-to-set map

$$\beta(\cdot) : \mathbb{R}^{m+m'} \to\!\!\to \mathbb{R}^n, \text{ with}$$
$$\beta(r,t) = \{x \in \mathbb{R}^n \mid Ax - b \diamond r,\ Ex - e \leq t\}$$

Note that

$$\forall r \in \mathbb{R}^m : \ \alpha(r) = \beta(r, 0) \tag{3.4}$$

As a result,

$$\forall k \in \mathbb{N} : \ \beta(r^k, 0) \neq \emptyset \quad ; \ \beta(r^\infty, 0) \neq \emptyset$$
$$\lim(r^k, 0) = (r^\infty, 0) \qquad ; \ x^\infty \in \beta(r^\infty, 0)$$

So, according to Case 1, there is a sequence $(x^k)_{\mathbb{N}}$ with $x^k \in \beta(r^k, 0)\ \forall k \in \mathbb{N}$ and $\lim x^k = x^\infty$. This, however, proves the theorem, because from (3.4) it follows that $x^k \in \alpha(r^k)\ \forall k \in \mathbb{N}$ as well.

□

Similar results (with different kinds of proofs) can be found in [Hoffman, 1952], [Böhm, 1975], [Robinson, 1975], [Wets, 1985] and [Mangasarian & Shiau, 1987]. This result is also covered by Corollary II.3.1 of [Dantzig et al., 1967]. In that paper, however, a much more complicated proof is given, due to the fact that the authors embedded the result in a much more general setting; they consider affine constraints which are parameterized in both the constraint coefficients and the right-hand-sides. In [Zencke & Hettich, 1987] parameterized semi-infinite linear programmes are dealt with. On the one hand, this problem setting is more general than ours because a (possibly) *infinite* number of affine constraints which are parameterized in both the constraint coefficients and the right-hand-sides is taken into consideration. On the other hand, the problem setting is more restrictive because only *linear* objectives and the *interior* of the set RHS is taken into account.

3.2 The (quasi-)convex case

In this subsection, two results will be presented. In either of them it will be assumed that there are no equality constraints explicitly stated in the description of problem (2.1). This implies that $\diamond \in \{\leq\}^m$. Furthermore, in both results the function $G(\cdot)$ will, among other things, be assumed to be (strictly quasi-)convex. Note that the former assumption alone is not restrictive — any equality constraint can always be replaced by two inequalities — but in combination with the latter assumption, it is.

Definition 3.1 (Quasi-convexity) *Let W be a convex subset of some vector space, and let $H(\cdot) : W \to (\mathbb{R} \cup \{\pm\infty\})^m$ be some (extended) vector-valued function on W. Then $H(\cdot)$ is called*

- quasi-convex w.r.t. W *if for every component-function $H_i(\cdot)$ of $H(\cdot)$ it is true that $\forall y^1, y^2 \in W\ \forall\lambda \in [0,1]:\ H_i(\lambda y^1 + (1-\lambda)y^2) \leq \max\{H_i(y^1), H_i(y^2)\}$ $(i = 1, \ldots m)$.*

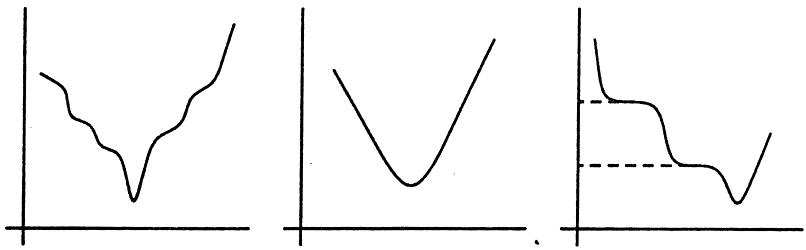

Figure 3.1: Differences between convexity and quasi-convexity

- strictly quasi-convex w.r.t. *W if for every component-function* $H_i(\cdot)$ *of* $H(\cdot)$ *it is true that* $\forall y^1 \in W\ \forall y^2 \in W \setminus \{y^1\}\ \forall \lambda \in (0,1)$: $H_i(\lambda y^1 + (1-\lambda)y^2) < \max\{H_i(y^1), H_i(y^2)\}$ *(i = 1, . . . m).*

The notion of quasi-convexity is a true generalization of convexity; every function which is (strictly) convex is (strictly) quasi-convex. The reverse statements are not necessarily true, as is shown by the examples depicted in Figure 3.1, where the first two functions are strictly quasi-convex and the third one is quasi-convex, whereas only the second one is (strictly) convex. In fact, under quasi-convexity a function cannot have a strict local minimum which is not global as well, and under strict quasi-convexity a function cannot have a local minimum which is not unique and global as well. And as already suggested by Figure 2.1, strict local minima and openness seem incoherent phenomena.

Theorem 3.2 (The convex case) *Let (2.1)-(2.5) be given. Let* $W \subseteq RHS$ *be open. Furthermore, let* X *be a closed convex set,* $\diamond \in \{\leq\}^m$ *and* $G(\cdot)$ *be convex on* $\alpha(r)\ \forall r \in W$ *and lower semi-continuous w.r.t.* $cl(\cup_W \alpha(r))$. *Then* $\alpha(\cdot)$ *is continuous w.r.t.* W.

Proof The closedness of $\alpha(\cdot)$ w.r.t. W immediately follows from Theorem 2.3, which, together with the convexity assumption on $G(\cdot)$, implies that $\forall r \in W$: $\alpha(r)$ is a closed convex set. For the openness, let

$(r^k)_{\mathbb{N}}$ be a sequence in W, converging to $r^\infty \in W$. Let $x^\infty \in \alpha(r^\infty)$ be arbitrarily chosen. We have to construct a sequence $(x^k)_{\mathbb{N}}$ such that $x^k \in \alpha(r^k)\ \forall k \in \mathbb{N}$ and $\lim x^k = x^\infty$.

W is an open set, so $\exists \bar{r} \in W : \bar{r} < r^\infty$. As a result,

$$\exists \bar{x} \in X : \ G(\bar{x}) \leq \bar{r} < r^\infty$$

Let $\epsilon > 0$ be arbitrarily chosen. Let $\lambda \in (0,1)$ be chosen such that

$$\lambda \parallel \bar{x} - x^\infty \parallel < \epsilon$$

Define $x^\lambda = \lambda\bar{x} + (1-\lambda)x^\infty$. From the convexity of $G(\cdot)$ on $\alpha(r^\infty)$ it follows that

$$G(x^\lambda) \leq \lambda G(\bar{x}) + (1-\lambda)G(x^\infty) < r^\infty$$

Now, $\lim r^k = r^\infty$, so

$$\exists k_0 \in \mathbb{N}\ \forall k > k_0 : \ G(x^\lambda) < r^k$$

Furthermore, X is convex, so $x^\lambda \in X$. Moreover,

$$\parallel x^\lambda - x^\infty \parallel = \lambda \parallel \bar{x} - x^\infty \parallel < \epsilon$$

From Lemma 3.1 the result follows, because $\forall k > k_0 : \ x^\lambda \in \alpha(r^k)$ and $\parallel x^\lambda - x^\infty \parallel < \epsilon$.

□

If besides the conditions stated in this theorem, $G(\cdot)$ is assumed to be convex on X and lower semi-continuous w.r.t. X, then $\alpha(\cdot)$ is continuous w.r.t. int(RHS), the interior of RHS. Unfortunately, the openness of $\alpha(\cdot)$ cannot be extended to the boundary of RHS without imposing additional assumptions on X and/or $G(\cdot)$. This negative result is demonstrated by means of the following example, which is due to [Hogan, 1973[a]].

$$X = \{x \in \mathbb{R}^2 \mid \| x \|_1 \leq 2\} \quad ; \quad \diamond \in \{\leq\}^2$$

$$G(\cdot) : \mathbb{R}^2 \to \mathbb{R}^2 \text{ with}$$

$$G(x) = \left(x_1, \| x - \begin{pmatrix} 0 \\ 1 \end{pmatrix} \|_2 + \| x - \begin{pmatrix} 0 \\ -1 \end{pmatrix} \|_2 - 2\right)^T$$

Here, $\| \cdot \|_1$ and $\| \cdot \|_2$ denote the sum-norm and the Euclidean norm respectively. The corresponding feasible set map is not open w.r.t. RHS, because

$$\forall k \in \mathbb{N} : r^k = \left(-\tfrac{1}{k}, 2\sqrt{1 + \left(\tfrac{1}{k}\right)^2} - 2\right)^T \in RHS$$

$$\alpha(r^k) = \left\{\left(-\tfrac{1}{k}, 0\right)^T\right\}$$

$$\lim r^k = r^\infty = (0,0)^T$$

However, $(0,-1)^T \in \alpha(r^\infty)$ and this point cannot be obtained as the limit of a converging sequence $(x^k)_{\mathbb{N}}$ with $x^k \in \alpha(r^k)\ \forall k \in \mathbb{N}$. Note that, because of Theorem 3.2, $(0,0)^T$ must be a boundary point of RHS. Under slightly different conditions on X and $G(\cdot)$, the boundary of RHS can, indeed, be included in the result.

Theorem 3.3 (The strictly quasi-convex case) *Let (2.1)-(2.5) be given. Let $W \subseteq RHS$. Furthermore, let X be a closed convex set, $\diamond \in \{\leq\}^m$, $G(\cdot)$ be strictly quasi-convex on $\alpha(r)$ $\forall r \in W$ and lower* *semi-continuous w.r.t. $cl(\cup_W \alpha(r))$, and $\alpha(\cdot)$ be essentially bounded w.r.t. W. Then $\alpha(\cdot)$ is continuous w.r.t. W.*

Proof The closedness of $\alpha(\cdot)$ w.r.t. W immediately follows from Theorem 2.3, which, together with the strict quasi-convexity assumption on $G(\cdot)$, implies that $\forall r \in W : \alpha(r)$ is a closed convex set. For the openness, let $(r^k)_{\mathbb{N}}$ be a sequence in W, converging to $r^\infty \in W$. Let $x^\infty \in \alpha(r^\infty)$ be arbitrarily chosen. We have to construct a sequence $(x^k)_{\mathbb{N}}$ such that $x^k \in \alpha(r^k)\ \forall k \in \mathbb{N}$ and $\lim x^k = x^\infty$.

Case 1: $\{x^\infty\} = \alpha(r^\infty)$

Let $\forall k \in \mathbb{N} : x^k \in \alpha(r^k)$ be arbitrarily chosen. Because of the fact that $\alpha(\cdot)$ is essentially bounded w.r.t. W, it follows that $(x^k)_{\mathbb{N}}$ has an accumulation point, say $\tilde{x}$. Let $(x^{p(k)})_{\mathbb{N}}$ be a subsequence of $(x^k)_{\mathbb{N}}$ which converges to $\tilde{x}$. We then have

$$\forall k \in \mathbb{N} : r^{p(k)} \in W$$

$$\lim r^{p(k)} = r^\infty \in W$$

$$\forall k \in \mathbb{N} : x^{p(k)} \in \alpha(r^{p(k)})$$

$$\lim x^{p(k)} = \tilde{x}$$

So, from the closedness of $\alpha(\cdot)$ it follows that $\tilde{x} \in \alpha(r^\infty) = \{x^\infty\}$. Therefore, $(x^k)_{\mathbb{N}}$ has only one accumulation point, which is equal to x^∞. Hence, $\lim x^k = x^\infty$.

Case 2: $\{x^\infty\} \neq \alpha(r^\infty)$

Let $\bar{x} \in \alpha(r^\infty) \setminus \{x^\infty\}$. Let $\epsilon > 0$ be arbitrarily chosen and let $\lambda \in (0,1)$ be such that

$$\lambda \| \bar{x} - x^\infty \| < \epsilon$$

Define $x^\lambda = \lambda\bar{x} + (1-\lambda)x^\infty$. From the strict quasi-convexity of $G(\cdot)$ on $\alpha(r^\infty)$ it follows that

$$G(x^\lambda) < \max\{G(\bar{x}), G(x^\infty)\} \leq r^\infty$$

Now, $\lim r^k = r^\infty$, so

$$\exists k_0 \in \mathbb{N}\ \forall k > k_0 : \ G(x^\lambda) < r^k$$

Furthermore, X is convex, so $x^\lambda \in X$. Moreover,

$$\| x^\lambda - x^\infty \| = \lambda \| \bar{x} - x^\infty \| < \epsilon$$

From Lemma 3.1 the result follows, because $\forall k > k_0 : \ x^\lambda \in \alpha(r^k)$ and $\| x^\lambda - x^\infty \| < \epsilon$.

□

A similar result as our Theorem 3.2 can be found in [Hogan, 1973[b]]. Weaker versions of Theorem 3.3 can be found in [Evans & Gould, 1970], where W is required to be a subset of the *interior* of RHS, and in [Hogan, 1973[a]], where strict convexity on the constraint functions $G(\cdot)$ is imposed. Finally, a slightly weaker result than our Theorem 3.3 (continuity of $v(\cdot)$ w.r.t. the relative interior of RHS) under slightly weaker assumptions (quasi-convexity of the parameterized constraint functions $G_i(x,r)$ as a function of (x,r)) can be found in [Hogan, 1973[b]].

3.3 The MF-regular case

The conditions which will be considered here, resemble the famous regularity conditions which enforce constraint-qualification in Karush-Kuhn-Tucker points (cf. [Mangasarian & Fromovitz, 1967]).

Let (2.1)–(2.5) be given. Suppose $X = \{x \in \mathbb{R}^n \mid H(x) \circ 0\}$ where $H(\cdot)$ is a vector-valued function from $\mathbb{R}^n$ to $\mathbb{R}^\ell$ ($\ell \in \mathbb{N} \cup \{0\}$) and $\circ$ is an ℓ-vector of which all entries belong to $\{\leq, =\}$. Let $E_\diamond$ ($E_\circ$) and $I_\diamond$ ($I_\circ$) denote the index sets of equality and inequality constrained component-functions $G_i(\cdot)$ ($H_i(\cdot)$) of $G(\cdot)$ ($H(\cdot)$) respectively, hence

$$E_\diamond = \{i \in \{1,\dots,m\} \mid \diamond_i \in \{=\}\} \quad ; \quad I_\diamond = \{1,\dots,m\} \setminus E_\diamond$$
$$E_\circ = \{i \in \{1,\dots,\ell\} \mid \circ_i \in \{=\}\} \quad ; \quad I_\circ = \{1,\dots,\ell\} \setminus E_\circ$$

Furthermore, let D be open, and let $G(\cdot), H(\cdot)$ be once continuously differentiable on D, so

$$G(\cdot), H(\cdot) \in \mathcal{C}^1(D)$$

Finally, we define the Jacobians of the equalities and saturated inequalities as follows.

$$\begin{aligned}
J_G^{\leq}(x,r) &= \left(\frac{\partial G_i(x)}{\partial x_j}\right)_{i\in\{i\in I_\diamond | G_i(x)=r_i\}\,;\, j\in\{1,\dots,n\}} \\
J_G^{=}(x) &= \left(\frac{\partial G_i(x)}{\partial x_j}\right)_{i\in E_\diamond\,;\, j\in\{1,\dots,n\}} \\
J_H^{\leq}(x) &= \left(\frac{\partial H_i(x)}{\partial x_j}\right)_{i\in\{i\in I_\circ | H_i(x)=0\}\,;\, j\in\{1,\dots,n\}} \\
J_H^{=}(x) &= \left(\frac{\partial H_i(x)}{\partial x_j}\right)_{i\in E_\circ\,;\, j\in\{1,\dots,n\}}
\end{aligned}$$

Definition 3.2 (MF-regularity with respect to) *$x \in \alpha(r)$ is called* MF-regular w.r.t. $G(\cdot)$ and $H(\cdot)$ *if the following two conditions are satisfied.*

- $\begin{pmatrix} J_G^{=}(x) \\ J_H^{=}(x) \end{pmatrix}$ *has full row rank*

- *There is a direction $y \in \mathbb{R}^n$ such that*

$$J_G^{\leq}(x,r)y < 0 \quad ; \quad J_H^{\leq}(x)y < 0$$
$$J_G^{=}(x)y = 0 \quad ; \quad J_H^{=}(x)y = 0$$

We can now state the following result.

Theorem 3.4 (The MF-regular case) *Let (2.1)–(2.5) be given. Suppose $W \subseteq RHS$. Let $X = \{x \in \mathbb{R}^n \mid H(x) \circ 0\}$, where $H(\cdot)$ is a vector-valued function from $\mathbb{R}^n$ to $\mathbb{R}^\ell$ ($\ell \in \mathbb{N} \cup \{0\}$) and $\circ$ is an ℓ-vector of which all entries belong to $\{\leq, =\}$. Furthermore, let $D \supseteq cl(X)$ be open, $G(\cdot)$ be vector-valued, and $G(\cdot), H(\cdot) \in C^1(D)$. Finally, assume that $\forall x \in \cup_W \alpha(r)$: x is MF-regular w.r.t. $G(\cdot)$ and $H(\cdot)$. Then $\alpha(\cdot)$ is continuous w.r.t. W.*

Proof The closedness of $\alpha(\cdot)$ w.r.t. W immediately follows from the continuity of $G(\cdot)$ and $H(\cdot)$ w.r.t. D. In order to prove the openness, let us suppose that $(r^k)_{\mathbb{N}}$ is a sequence in W, converging to $r^\infty \in W$. Let $x^\infty \in \alpha(r^\infty)$. We have to construct a sequence $(x^k)_{\mathbb{N}}$ such that $\forall k \in \mathbb{N} : x^k \in \alpha(r^k)$ and $\lim x^k = x^\infty$.

First we observe that for each direction $\bar{\rho} \in \mathbb{R}^m$ in the parameter space, there is an open neighbourhood $\Lambda \subseteq \mathbb{R}$ of 0, an open neighbourhood $\Omega \subseteq \mathbb{R}^m$ of $\bar{\rho}$ and a continuously differentiable function $\psi(\cdot) : \Lambda \times \Omega \to \mathbb{R}^n$ such that

$$\begin{aligned} &\forall \rho \in \Omega : \ \psi(0, \rho) = 0, \text{ and} \\ &\forall 0 \leq \lambda \in \Lambda \ \forall \rho \in \Omega : \ x^\infty + \psi(\lambda, \rho) \in \alpha(r^\infty + \lambda\rho) \end{aligned} \tag{3.5}$$

Intuitively, this result implies that small enough steps in the parameter space can be reacted upon by small enough (nonlinear) steps in the solution space, such that the resulting solution remains feasible for the perturbed system of constraints. The proof of this statement relies upon the MF-regularity assumption and the Implicit Function Theorem; details can be found in [Gauvin & Dubeau, 1982]. Let us define

$$x^k \in \operatorname{argmin}\{\| x - x^\infty \| \mid x \in \alpha(r^k)\}$$

Due to the fact that $\alpha(r^k)$ is non-empty and closed, such a choice is always possible. Suppose $\lim x^k \neq x^\infty$, then

$$\exists \epsilon > 0 \; \forall k_0 \in \mathbb{N} \; \exists k > k_0 : \| x^k - x^\infty \| \geq \epsilon$$

Hence, there is a subsequence $(x^{p(k)})_{\mathbb{N}}$ such that

$$\forall k \in \mathbb{N} : \| x^{p(k)} - x^\infty \| \geq \epsilon$$

Let $\lambda^k = \| r^k - r^\infty \|$ and let ρ^k be chosen such that $\| \rho^k \| = 1$ and $\lambda^k \rho^k = r^k - r^\infty$. Evidently, there is a subsequence $(q(k))_{\mathbb{N}}$ of $(p(k))_{\mathbb{N}}$ such that $\lim \rho^{q(k)}$ exists, and equals, say ρ^∞.

As a result, there is an open neighbourhood $\Lambda \subseteq \mathbb{R}$ of 0, an open neighbourhood $\Omega \subseteq \mathbb{R}^m$ of ρ^∞, and a continuously differentiable function $\psi(\cdot) : \Lambda \times \Omega \rightarrow \mathbb{R}^n$ such that (3.5) is satisfied.

Let $k_0 \in \mathbb{N}$ be chosen such that $\forall k > k_0 : \; (\lambda^{q(k)}, \rho^{q(k)}) \in \Lambda \times \Omega$. From (3.5) it follows that

$$\forall k > k_0 : \; x^\infty + \psi(\lambda^{q(k)}, \rho^{q(k)}) \in \alpha(r^{q(k)})$$

Note that $\psi(\cdot) \in C^1(\Lambda \times \Omega)$, so

$$\lim \psi(\lambda^{q(k)}, \rho^{q(k)}) = \psi(0, \rho^\infty) = 0$$

By definition of x^k, we have $\| x^\infty - x^{q(k)} \| \leq \| x^\infty - (x^\infty + \psi(\lambda^{q(k)}, \rho^{q(k)})) \|$ for each $k > k_0$. Consequently, $\lim \| x^\infty - x^{q(k)} \| = 0$. This result, however, contradicts the fact that $(x^{q(k)})_{\mathbb{N}}$ is a subsequence of $(x^{p(k)})_{\mathbb{N}}$. Therefore, $\lim x^k = x^\infty$.

□

Note that the MF-regularity assumption implies that W can only be a subset of the *interior* of RHS (cf. (3.5)). Also note, that for $X = \mathbb{R}$ and a one-dimensional real-valued inequality constrained function $G(\cdot) \in C^1(\mathbb{R})$, the MF-regularity condition implies that the derivative of $G(\cdot)$ is fixed in sign. This excludes, for instance, the situation of Figure 2.1, where $G(\cdot)$ has a stationary point. Finally, in [Gauvin & Dubeau, 1982] it is proven that even in a more general problem setting (general versus right-hand-side perturbations), the same result holds. What we have done is adopted their approach and put it into the framework of Section 2.

3.4 Mixtures

In this section we will unify some of the previous results. Theorem 3.5 is a combination of Theorem 3.1–3.3. More precisely, Theorem 3.1 and some slightly weaker versions of Theorem 3.2 and 3.3 can be reobtained from Theorem 3.5 as a special case. The same reasoning applies to Theorem 3.6 as a combination of Theorem 3.1, 3.3 and 3.4. Both results seem to be new. First, the following description of the feasible set map $\alpha(\cdot)$ will be considered.

$$\alpha(r,s,t) = \{x \in X \mid G^1(x) \leq r,\ G^2(x) \leq s,\ Ax - b \diamond t\} \tag{3.6}$$

Here, $G^1(\cdot)$ and $G^2(\cdot)$ are (extended) vector-valued functions which are defined on D, and $\diamond$ is a vector all entries of which are elements of $\{\leq, =\}$.

Theorem 3.5 (Mixed case 1) *Let (2.1)–(2.5) be given. Let* $W = \{(r,s,t) \mid \exists \bar{s} < s :\ \alpha(r,\bar{s},t) \neq \emptyset\}$. *In addition, suppose that*

- *X is closed and convex*

- *X is polyhedral if the matrix A is non-vacuous*
- *$G^1(\cdot)$ is strictly quasi-convex on X and continuous w.r.t. X if $G^1(\cdot)$-type of constraints occur explicitly in the description of $\alpha(\cdot)$*
- *$G^2(\cdot)$ is convex on X and continuous w.r.t. X*
- *$\alpha(\cdot)$ is essentially bounded w.r.t. W if $G^1(\cdot)$-type of constraints occur explicitly in the description of $\alpha(\cdot)$*

Then $\alpha(\cdot)$ is continous w.r.t. W.

Proof The closedness of $\alpha(\cdot)$ w.r.t. W immediately follows from Theorem 2.3. In order to prove the openness, consider a sequence $(r^k, s^k, t^k)_{\mathbb{N}}$ in W which converges to $(r^\infty, s^\infty, t^\infty) \in W$. Let $x^\infty \in \alpha(r^\infty, s^\infty, t^\infty)$ be arbitrarily chosen. We have to construct a sequence $(x^k)_{\mathbb{N}}$ with $x^k \in \alpha(r^k, s^k, t^k)$ $\forall k \in \mathbb{N}$ and $\lim x^k = x^\infty$. For this purpose we will distinguish between two cases.

Case 1: $G^1(\cdot)$-type of constraints do not occur explicitly in (3.6)

From $(s^\infty, t^\infty) \in W$ and the definition of W, it follows that there is an $\bar{x} \in \alpha(s^\infty, t^\infty)$ such that $G^2(\bar{x}) < s^\infty$. Let $\epsilon > 0$ be arbitrarily chosen, and let $\lambda \in (0,1)$ be chosen such that

$$\lambda \parallel \bar{x} - x^\infty \parallel < \frac{\epsilon}{2}$$

Define $x^\lambda = \lambda\bar{x} + (1-\lambda)x^\infty$. The following statements are easily verified

$$x^\lambda \in \alpha(s^\infty, t^\infty)\ ;\ G^2(x^\lambda) < s^\infty\ ;\ \parallel x^\lambda - x^\infty \parallel < \frac{\epsilon}{2}$$

According to Theorem 3.1 there is a sequence $(x^k)_{\mathbb{N}}$ such that

$$x^k \in X \; ; \; Ax^k - b \diamond t^k \; ; \; \lim x^k = x^\lambda$$

(If there are no affine constraints explicitly stated in the description of $\alpha(\cdot)$, then choose $x^k = x^\lambda \; \forall k \in \mathbb{N}$.) From the above two statements and the continuity of $G^2(\cdot)$ w.r.t. X it follows that

$$\exists k_1 \in \mathbb{N} \; \forall k > k_1 : \| x^k - x^\lambda \| < \frac{\epsilon}{2}$$

$$\exists k_2 \in \mathbb{N} \; \forall k > k_2 : \; G^2(x^k) < s^k$$

Hence, $\forall k > \max\{k_1, k_2\} : \; x^k \in \alpha(s^k, t^k)$ and $\| x^k - x^\infty \| \leq \| x^k - x^\lambda \| + \| x^\lambda - x^\infty \| < \epsilon$. Applying Lemma 3.1 yields the desired result.

Case 2: $G^1(\cdot)$-type of constraints do occur explicitly in (3.6)

Let $\bar{x} \in \alpha(r^\infty, s^\infty, t^\infty) \setminus \{x^\infty\}$ and $\epsilon > 0$ be arbitrarily chosen. (If such an $\bar{x}$ does not exist then we can use the same argument as in the proof of Theorem 3.3 to show that *any* sequence $(x^k)_{\mathbb{N}}$ with $x^k \in \alpha(r^k, s^k, t^k)$ converges to x^∞.) Let $\lambda \in (0,1)$ be chosen such that

$$\lambda \| \bar{x} - x^\infty \| < \frac{\epsilon}{2}$$

Define $x^\lambda = \lambda \bar{x} + (1 - \lambda) x^\infty$. We then have

$$x^\lambda \in \alpha(r^\infty, s^\infty, t^\infty) \; ; \; G^1(x^\lambda) < r^\infty \; ; \; \| x^\lambda - x^\infty \| < \frac{\epsilon}{2}$$

According to Case 1 there is a sequence $(x^k)_{\mathbb{N}}$ with

$$x^k \in X \; ; \; G^2(x^k) \leq s^k \; ; \; Ax^k - b \diamond t^k \; ; \; \lim x^k = x^\lambda$$

From the above two statements and the continuity of $G^1(\cdot)$ w.r.t. X it follows that

$$\exists k_1 \in \mathbb{N} \; \forall k > k_1 : \| x^k - x^\lambda \| < \frac{\epsilon}{2}$$

$$\exists k_2 \in \mathbb{N} \; \forall k > k_2 : \; G^1(x^k) < r^k$$

Hence, $\forall k > \max\{k_1, k_2\}$: $x^k \in \alpha(r^k, s^k, t^k)$ and $\| x^k - x^\infty \| \leq \| x^k - x^\lambda \| + \| x^\lambda - x^\infty \| < \epsilon$. Again, the result follows from Lemma 3.1.

□

The condition imposed on W is nothing but a Slater-type of condition with respect to the non-linear constraints which are convex but not strictly quasi-convex. This immediately follows from the fact that $(r, s, t) \in W$ if and only if $\exists \bar{x} \in \alpha(r, s, t) : G^2(\bar{x}) < s$. Introducing such a condition for the strictly quasi-convex functions would be superfluous because strict quasi-convexity implies Slater's condition as soon as $\alpha(r, s, t)$ contains at least two points.

Although the set W does not coincide with the entire set of feasible right-hand-sides, it fully contains the relative interior of the latter set. Loosely speaking, W equals the set of feasible right-hand-sides, except for that part of the relative boundary which is induced by the non-linear constraints which are convex but not strictly quasi-convex. Compared to the result in [Hogan, 1973^b] our result is stronger, because it may include more than just the relative interior of RHS, but it has been derived under stronger assumptions as well.

Let us now consider the following description of $\alpha(\cdot)$.

$$\begin{aligned} \alpha(r, s, t) &= \{x \in X \mid G^1(x) \leq r,\ G^2(x) \diamond s,\ Ax - b \circ t\}, \text{ with} \\ X &= \{x \in \mathbb{R}^n \mid H^1(x) \leq 0,\ H^2(x) \diamond' 0,\ Cx - d \circ' 0\} \end{aligned} \tag{3.7}$$

Here, $G^1(\cdot), G^2(\cdot), H^1(\cdot)$ and $H^2(\cdot)$ are vector-valued functions defined on $\mathbb{R}^n$, and $\diamond, \diamond', \circ$ and $\circ'$ are vectors of which all entries belong to $\{\leq, =\}$.

Theorem 3.6 (Mixed case 2) *Let (2.1)–(2.5) be given. Let $W \subseteq RHS$. Suppose that*

- *$D \supseteq cl(X)$ and D is open*
- *$G^2(\cdot)$ and $H^2(\cdot)$ are vector-valued and once continuously differentiable functions on D, which are also continuous w.r.t. $\mathbb{R}^n$*
- *$\forall (r,s,t) \in W \; \forall x \in \alpha(r,s,t)$: x is MF-regular w.r.t. $G^2(\cdot), H^2(\cdot)$ and the affine constraint functions*
- *If $G^1(\cdot)$- or $H^1(\cdot)$-type of constraints are explicitly used in the description of $\alpha(\cdot)$, then*
  - *$G^1(\cdot), H^1(\cdot)$ are strictly quasi-convex on X and D, and continuous w.r.t. X and D respectively*
  - *$G^2(\cdot), H^2(\cdot)$ are quasi-convex on X and D respectively*
  - *$G_i^2(\cdot)$ is affine on X if $\diamond_i \in \{=\}$*
  - *$H_i^2(\cdot)$ is affine on D if $\diamond_i' \in \{=\}$*
  - *$\alpha(\cdot)$ is essentially bounded w.r.t. W*

Then $\alpha(\cdot)$ is continous w.r.t. W.

Proof The closedness of $\alpha(\cdot)$ w.r.t. W immediately follows from Theorem 2.3. In order to prove the openness, consider a sequence $(r^k, s^k, t^k)_{\mathbb{N}}$ in W which converges to $(r^\infty, s^\infty, t^\infty) \in W$. Let $x^\infty \in \alpha(r^\infty, s^\infty, t^\infty)$ be arbitrarily chosen. We have to construct a sequence $(x^k)_{\mathbb{N}}$ with $x^k \in \alpha(r^k, s^k, t^k) \; \forall k \in \mathbb{N}$ and $\lim x^k = x^\infty$. For this purpose we will distinguish between two cases.

Case 1: Neither $G^1(\cdot)$- nor $H^1(\cdot)$-type of constraints occur in (3.7)

The existence of such a sequence $(x^k)_{\mathbb{N}}$ immediately follows from Theorem 3.4.

Case 2: $G^1(\cdot)$- or $H^1(\cdot)$-type of constraints do occur in (3.7)

In this case the proof is essentially similar to the second part of the proof of Theorem 3.5; the construction of a sequence $(x^k)_{\mathbb{N}}$ with $G^2(x^k) \diamond s^k$, $Ax^k - b \circ t^k$, $H^2(x^k) \diamond' 0$, $Cx^k - d \circ' 0$ and $\lim x^k = x^\lambda$ can be done as in the proof of Theorem 3.4.

□

Section 4

Summary

In part I we reported on conditions under which the optimal solution value of a perturbed mathematical programme varies *continuously* with changes in the parameter reflecting perturbation. In fact, we only considered the case of *right-hand-side perturbations*, and we gave two reasons for doing so. First, the case of right-hand-side perturbations is general enough in view of the stability results that are needed in Parts II and III. Secondly, the limitation which results from the assumptions we adopted, does not so much stem from the fact that perturbations were assumed to appear only in the right-hand-side, but more from the fact that the map $\alpha(\cdot)$ was assumed to be described by means of a finite number of constraint functions, and that its domain as well as its codomain were supposed to be subsets of finitely dimensional Euclidean spaces. The results obtained can be summarized as follows.

Theorem 4.1 (Continuity of the value-function) *Let (2.1)–(2.6) be given. Suppose $W \subseteq V$ and assume that*

- $f(\cdot)$ *is continuous w.r.t.* $\cup_W \alpha(r)$
- $\alpha(\cdot)$ *is essentially bounded w.r.t.* W

- *at least one of the sets of conditions mentioned in Theorem 3.1–3.6 applies*

Then $v(\cdot)$ is continuous w.r.t. W.

Proof This is an immediate consequence of Theorem 2.1 and 3.1–3.6. □

As already mentioned in Section 2, the continuity of the feasible set map $\alpha(\cdot)$ (besides some boundedness condition on $\alpha(\cdot)$ and continuity of the objective function $f(\cdot)$) is not only a sufficient condition to ensure continuity of the value-function $v(\cdot)$, but is in some sense also necessary. This fact is revealed by Theorem 2.2 which states that the continuity of $\alpha(\cdot)$ is a necessary condition if for *any* continuous objective function, the resulting value-function is required to be continuous. So, in the absence of additional conditions on the objective function, it is not restrictive to analyse the continuity of the value-function by means of the continuity of the feasible set map.

Unfortunately, the same sort of reasoning does not apply to the conditions which are mentioned in Theorem 3.1–3.6; these conditions are sufficient, but their necessity cannot be established that unambiguously. However, if one is willing to accept some less rigorous argumentation, it seems possible to plead that, except for pathological cases, the conditions are "almost necessary". Recall that the conditions under which the value-function is proven to be continuous, can basically be classified into three groups. The first group consists of those mathematical programming problems in which only affine constraints occur (Theorem 3.1). The second group consists of programming problems on which a convexity and a Slater-type of condition are imposed; the latter condition can be imposed either directly by restricting oneself to the interior of the set of feasible right-hand-sides (Theorem 3.2), or indirectly, by requiring that the constraint functions are strictly quasi-convex (Theorem 3.3). Finally, the third group contains those programming problems for which the Mangasarian-Fromovitz regularity condition holds (Theorem 3.4). As is well-known, these three types of conditions are precisely the three classes of sufficient conditions which imply constraint-qualification in Karush-Kuhn-Tucker points. Many attempts have been made to loosen these conditions, but a significant step forward has not yet been made. Apparently, these three types of conditions are too intimately related with the property of a mathematical programme to be "well-behaved". This point of view can now be used as a justification for the conjecture that significantly weaker conditions which enforce continuity of the value-function will not be easily obtained.

The contributions of the preceding analysis are the following.

- The limitations which result from considering right-hand-side perturbations, do not originate from the fact that perturbations are assumed to appear only through the right-hand-side, but from the fact that the feasible set map is assumed to be described by means of a finite number of constraint functions, and that its domain as well as its codomain are supposed to be subsets of finitely dimensional Euclidean spaces.

- All results are presented within one single framework, viz. the continuity of the feasible set map.

- The continuity of the feasible set map is, in some sense, necessary and sufficient for the continuity of the value-function. Therefore, in the absence of additional conditions on the objective function, it is not restrictive to enforce the former in pursuing the latter.

- Our result in the strictly quasi-convex case is new; it is a unification and an extension of two earlier results (cf. [Evans & Gould, 1970] and [Hogan, 1973^a]).

- The formal statements and proofs of the mixed cases as described by Theorem 3.5 and 3.6 are new as well.

Part II

Duality theory in General Mathematical Programming

Section 1

Introduction

In Linear Programming the importance of duality theory is beyond dispute. Its usefulness not only originates from the fact that feasible dual solutions provide *lower bounds* on the quality of feasible primal solutions, an observation which is used in proving (near-)optimality, but also from the (economic) interpretation of the dual variables as *sensitivity measures*. Another, equally important feature of duality theory in Linear Programming, is that of *symmetry*, which means that the dual of a linear programme is, again, a linear programme. As such, the dual of the dual is well-defined, and, in fact, easily shown to equal the primal. From an analytical point of view, the primal and dual problems are therefore equally hard to solve, and usually an optimal solution of one of the problems can easily be obtained as a by-product of solving the other. This symmetry also allows for *reoptimization* when, after a problem has been solved, (some of) its data are perturbed.

Naturally, there have been many attempts to develop similar and equally powerful theories for more general mathematical programmes. Well-known developments include the ones in Convex Quadratic Programming ([Dorn, 1960], [Cottle, 1963]), in Convex Programming ([Geoffrion, 1972]), in General Non-Linear Programming ([Rockafellar, 1974]), in Integer Linear Programming ([Burdet & Johnson, 1977], [Wolsey, 1981]) and in General Mathematical Programming ([Gould, 1969, 1972], [Tind & Wolsey, 1981]). Unfortunately, it seems as if each gain in generality had to be paid for by a loss in applicability; although neither the capability of generating (tight) upper bounds, nor the interpretation of the dual variables as sensitivity measures has been affected, it is, except for the convex quadratic case, the symmetry that has. This is already revealed when Lagrangean duality in Convex Programming is considered; unlike the convex quadratic and linear cases, the primal and dual programmes are no longer equally hard to solve, and in the absence of additional properties on the primal, an optimal primal solution is no longer readily obtainable from an optimal dual solution. A more serious loss in symmetry occurs in the integer linear and the general cases; whereas the Lagrangean dual still is a finitely dimensional optimization problem, the decision variables in the integer linear and the general dual problems are (extended) real-valued functions, rather than (finitely dimensional) vectors of scalars. The conclusion seems justified that from a *theoretical* point of view, extensions of Linear Programming duality have succesfully been developed, but that from a *computational* point of view, this has not been the case; see also [Ponstein, 1983].

In this part a duality theory for *general* mathematical programmes is discussed. Such a programme, as well as its dual, will be introduced in Section 2. It will also be argued that this dual programme is in some sense the only natural one, if the primal programme is considered to be embedded in a family of parameterized problems which differ in their *right-hand-sides* only. In Section 3 the basic duality results are presented. These results concern *weak* and *strong duality*, as well as the *Farkas property*. In Section 4 some well known special cases will be considered, viz. Lagrangean duality (which, in its turn, comprises Convex Quadratic and Linear Programming duality), augmented Lagrangean duality, and Integer Linear Programming duality. Besides the fact that the dual programme as defined as in Section 2, is no longer an optimization problem in finite dimensions, another inconvenient property exists; the one-to-one correspondence between primal constraints and dual variables, as occurring in Lagrangean and Linear Programming duality, is lacking in the general case. In Section 5 we will prove that under certain conditions, this one-to-one correspondence can be restored without affecting weak and strong duality. In fact, three different cases will be distinguished. First, it is shown that with ***hardly any conditions imposed*** on the primal programme, this one-to-one correspondence can be restored. A less attractive feature of this result is that all dual solutions are assumed to have a codomain including the "value" $+\infty$. Secondly, it is demonstrated that this weak feature can be removed, provided a ***stability*** condition on the primal programme is met. This result elucidates a limitation of some well-known augmented Lagrangean methods; it is argued that some of these methods can only produce tight bounds in the case that the aforementioned stability condition is met. Thirdly, it is proven that if the primal programme is an ***integer (non-linear) programme***, a ***boundedness condition on the primal feasible set*** is sufficient to restore the one-to-one correspondence. Part II will be concluded by a summary. Sections 1–4 are largely based on [Tind & Wolsey, 1981]. The results of Section 5 are new, except for Theorem 5.1, which can also be found in [Gould, 1969].

Section 2

The primal and dual programme

Consider the following *primal programme*

$$
\begin{aligned}
\mathcal{P}: \quad & \max_{x} \quad f(x) \\
& \text{s.t.} \quad G(x) \diamond 0 \\
& \qquad\;\; x \in X
\end{aligned}
\tag{2.1}
$$

Here, $f(\cdot)$ and $G(\cdot)$ are functions from a set $D \subseteq \mathbb{R}^n$ to $\mathbb{R} \cup \{\pm\infty\}$ and $\mathbb{R}^m$ respectively, X is a subset of D and $\diamond \in \{\leq, =\}^m$ $(m, n \in \mathbb{N})$. More formally:

$$
\begin{aligned}
& X \subseteq D \subseteq \mathbb{R}^n && ; \quad \diamond \in \{\leq, =\}^m \\
& f(\cdot) : D \rightarrow \mathbb{R} \cup \{\pm\infty\} && ; \quad G(\cdot) : D \rightarrow \mathbb{R}^m
\end{aligned}
\tag{2.2}
$$

As in Part I, $G(x) \diamond r$ is a shorter notation for $G_i(x) \diamond_i r_i$, where $G_i(\cdot)$, r_i and $\diamond_i$ are the i-th components of $G(\cdot)$, r and $\diamond$ respectively

$(i = 1, \ldots, m)$. Furthermore, the *set of feasible right-hand-sides* is denoted by

$$RHS = \{r \in \mathbb{R}^m \mid \exists x \in X : \; G(x) \diamond r\} \tag{2.3}$$

The *value-function* is defined as

$$v(\cdot) : \mathbb{R}^m \to \mathbb{R} \cup \{\pm\infty\}, \text{ with}$$

$$v(r) = \begin{cases} \sup\limits_x \{f(x) \mid G(x) \diamond r, \; x \in X\} & \text{if } r \in RHS \\ -\infty & \text{otherwise} \end{cases} \tag{2.4}$$

Problem $\mathcal{P}$ is called *infeasible*, *unbounded* or *regular* according to whether $v(0)$ equals $-\infty$, $+\infty$ or a real number respectively. Moreover, $\mathcal{P}$ is called *solvable* if $f(x) = v(0)$ for some $x \in X$ with $G(x) \diamond 0$. If $\mathcal{P}$ is regular and $f(x) \geq v(0) - \epsilon$ for some non-negative ϵ and some feasible solution x, then x is called an *ϵ-optimal solution* for $\mathcal{P}$. A zero-optimal solution is called an *optimal* solution. The reason for introducing the notion of ϵ-optimality is that all regular programmes have ϵ-optimal solutions for every $\epsilon > 0$; a similar statement does not hold for $\epsilon = 0$.

Associated with $\mathcal{P}$ the following set of extended vector-valued functions is defined

$$\begin{aligned} \mathcal{F} \;=\; & \{g(\cdot) : \mathbb{R}^m \to \mathbb{R} \cup \{\pm\infty\} \mid \\ & g(r) \leq g(r') \;\; \forall r, r' \in RHS : \; r \diamond r'\} \end{aligned} \tag{2.5}$$

Finally, the *dual programme* of (2.1) is defined as

$$\begin{aligned} \mathcal{D} : \; & \min_{g(\cdot)} \;\; g(0) \\ & \text{s.t.} \;\; g(G(x)) \geq f(x) \;\; \forall x \in X \\ & \qquad g(\cdot) \in \mathcal{F} \end{aligned} \tag{2.6}$$

All notions which have been defined with respect to the primal programme can be defined similarly with respect to the dual. For instance, $\mathcal{D}$ is called *infeasible*, *unbounded* or *regular* depending on whether its optimal solution value equals $+\infty$, $-\infty$ or a real number respectively. Note that the decision variable in the dual programme is a *function*, rather than a (finitely dimensional) vector of *scalars*. It should also be noted that, by definition, the argument in the objective of $\mathcal{D}$ (i.c. 0) equals the right-hand-side of $\mathcal{P}$; if this right-hand-side were different, the argument in the objective of $\mathcal{D}$ would have to be modified accordingly.

The justification of calling (2.6) the dual of (2.1) stems from the fact that, as in Linear Programming, the dual programme provides *upper bounds* for the optimal solution value of the primal programme, *whatever* the value of the right-hand-side. This assertion is stated more rigorously in the following lemma.

Lemma 2.1 *Let (2.1)–(2.6) be given. Suppose $g(\cdot) \in \mathcal{F}$. Then the following two statements are equivalent.*

1. $\forall r \in \mathbb{R}^m : \; g(r) \geq v(r)$

2. *$g(\cdot)$ is a feasible solution for (2.6)*

Proof

1. $\Rightarrow$ 2.

Let $x \in X$ be arbitrarily chosen. From 1. it follows that

$$g(G(x)) \geq v(G(x))$$

By definition of $v(\cdot)$ we also have

$$v(G(x)) \geq f(x)$$

These two inequalities imply that

$$g(G(x)) \geq f(x) \ \ \forall x \in X$$

This result, combined with the fact that $g(\cdot) \in \mathcal{F}$, proves the first part of this lemma.

2. $\Rightarrow$ 1.

Let $r \in \mathbb{R}^m$ be arbitrarily chosen. If $r \notin RHS$, then $v(r) = -\infty$ and the inequality in 1. obviously holds. Suppose $r \in RHS$. Let $x \in X$ be such that $G(x) \diamond r$. From the feasibility of $g(\cdot)$ it follows that

$$g(G(x)) \geq f(x)$$

whereas $g(\cdot) \in \mathcal{F}$ and $G(x) \diamond r$ imply that

$$g(r) \geq g(G(x))$$

As a result, we have

$$g(r) \geq f(x) \ \ \forall x \in X : \ G(x) \diamond r$$

which proves the second part of this lemma.

□

Corollary 2.1 *Let (2.1)–(2.6) be given. If* $v(r) = +\infty$ *then* $g(r) = +\infty$ *for all feasible dual solutions* $g(\cdot)$.

Lemma 2.1 states that if a dual programme is to generate upper bounds for the optimal solution value of its primal, *irrespective of* the specific choice of the right-hand-sides in the primal, its feasible solutions should correspond to the feasible solutions of (2.6). Moreover, due to the minimization in (2.6), the *tightest* upper bound for (2.1) is come up with. These observations show that only (2.6) is to be considered a natural candidate as a dual programme, in case the primal programme is considered to be embedded in a family of parameterized programmes that differ in their right-hand-sides only. This observation also reveals that the duality theory which is discussed here, only "dualizes" right-hand-side perturbations; it is not designed to deal with perturbations in either the objective or the constraint functions, because in such cases, the set of feasible dual solutions is also subject to perturbation. Before continuing with duality results, let us first prove a simple statement concerning $\mathcal{D}$.

Lemma 2.2 *Let (2.1)–(2.6) be given. In addition, suppose that $\mathcal{P}$ is such that $v(G(x)) > -\infty$ and $f(x) < +\infty$ for all $x \in X$. Let $\mathcal{D}$ be a regular programme. Then every ϵ-optimal dual solution $\bar{g}(\cdot)$ satisfies* $-\epsilon \leq \sup_x\{f(x) - \bar{g}(G(x)) \mid x \in X\} \leq 0$.

Proof Let $\epsilon \geq 0$ be given, and suppose $\bar{g}(\cdot)$ is an ϵ-optimal solution for $\mathcal{D}$. From the feasibility of $\bar{g}(\cdot)$ it follows that (cf. Lemma 2.1)

$$\forall x \in X: \; \bar{g}(G(x)) \geq v(G(x)) \geq f(x)$$

This, combined with the assumptions on $\mathcal{P}$, implies that $\forall x \in X$: $f(x) - \bar{g}(G(x))$ is well-defined and non-positive. Therefore, $\sup_x\{f(x) - \bar{g}(G(x)) \mid x \in X\}$ is well-defined and non-positive. Let us denote this supremal value by s. Obviously, $s \leq 0$. Suppose $s < -\epsilon$. Then a $\sigma \in \mathbb{R}$ exists, with $s \leq \sigma < -\epsilon$. Now consider the function $g(\cdot) = \bar{g}(\cdot) + \sigma$. Then $g(\cdot)$ is a feasible solution for $\mathcal{D}$, so

$$\varphi(\mathcal{D}) \leq g(0) = \bar{g}(0) + \sigma \leq \varphi(\mathcal{D}) + \epsilon + \sigma < \varphi(\mathcal{D}),$$

where $\varphi(\mathcal{D})$ is the optimal solution value of $\mathcal{D}$. This, however, is an obvious contradiction. Consequently, $s \geq -\epsilon$.

□

Section 3

Basic duality results

In this section three elementary duality results will be discussed, viz. *weak duality*, *strong duality*, and the *Farkas property*. The primal and dual programmes that will be considered here, are the ones which are defined in Section 2. Recall that $\varphi(\cdot)$ denotes the optimal solution value of a given optimization problem.

Theorem 3.1 (Weak duality) *Let (2.1)–(2.6) be given. Suppose x and $g(\cdot)$ are feasible solutions for $\mathcal{P}$ and $\mathcal{D}$ respectively. Then $f(x) \leq g(0)$.*

Proof Obviously, $f(x) \leq v(0)$. From Lemma 2.1 it follows that $f(x) \leq v(0) \leq g(0)$, which proves the theorem.

□

Corollary 3.1 $v(0) \leq g(0)$ *for all feasible dual solutions* $g(\cdot)$

Corollary 3.2 $v(0) \leq \varphi(\mathcal{D})$

Weak duality merely states that each feasible dual solution provides an upper bound for the optimal solution value of the primal programme.

Theorem 3.2 (Strong duality) *Let (2.1)–(2.6) be given. Then the optimal solution values of the primal and dual programme coincide. Hence,* $v(0) = \varphi(\mathcal{D})$.

Proof $v(\cdot)$ is a feasible solution for $\mathcal{D}$, so $\varphi(\mathcal{D}) \leq v(0)$. From weak duality it follows that $v(0) \leq \varphi(\mathcal{D})$.

□

Corollary 3.3 *$\mathcal{D}$ has an optimal solution $g^*(\cdot)$ with $g^*(0) = v(0)$.*

Proof Define $g^*(\cdot) = v(\cdot)$. The result follows from the proof of Theorem 3.2.

□

According to strong duality, the discrepancy between the optimal primal and dual solution values, also known as the *duality gap*, is equal to zero.

Theorem 3.3 (The Farkas property) *Let (2.1)–(2.6) be given. The following two statements are then equivalent.*

1. $\exists x \in X : \; G(x) \diamond 0$

2. $h(0) \geq 0 \;\; \forall h(\cdot) \in \mathcal{F} : \; (h(G(x)) \geq 0 \; \forall x \in X)$

Proof

1. $\Rightarrow$ 2.

Consider a primal programme like $\mathcal{P}$ that has an objective function which equals zero at all points $x \in \mathbb{R}^n$. The result follows from applying weak duality to this primal programme and its dual.

2. $\Rightarrow$ 1.

Apply strong duality to the pair of primal and dual programmes which have been defined in the first part of this proof.

□

In Theorem 3.3 a necessary and sufficient condition is given for the feasible set of $\mathcal{P}$ to be empty or not. In this respect it can be regarded as an extension of the well-known Farkas Lemma for polyhedra (e.g. [Rockafellar, 1970]).

One final remark is in order. If the conditions mentioned in Lemma 2.2 are satisfied, then $\mathcal{D}$ is strongly related to the following programme.

$$\begin{aligned}\mathcal{D}': \quad \min_{g(\cdot)} \quad & \sup_x\{f(x) + g(0) - g(G(x)) \mid x \in X\} \\ \text{s.t} \quad & g(G(x)) > -\infty \;\; \forall x \in X \\ & g(0) \in \mathbb{R} \\ & g(\cdot) \in \mathcal{F}\end{aligned} \tag{3.1}$$

In this problem too, an optimization takes place over functions $g(\cdot) \in \mathcal{F}$, and weak and strong duality apply to the pair of programmes $\mathcal{P}$ and $\mathcal{D}$'. The latter property follows from the fact that $g(0) \geq \sup_x\{f(x) + g(0) - g(G(x)) \mid x \in X\}$ for any solution $g(\cdot)$ that is feasible for both $\mathcal{D}$ and $\mathcal{D}$'. This observation, combined with weak duality, implies that the optimal solution value of $\mathcal{D}'$ is a number between the optimal solution values of $\mathcal{P}$ and $\mathcal{D}$. Moreover, it also follows that any ϵ-optimal solution for $\mathcal{D}$ is also an ϵ-optimal solution for $\mathcal{D}'$ (cf. Lemma 2.2). Note that $\mathcal{D}'$ looks quite familiar, because it resembles the well-known Lagrangean dual. In fact, the latter is obtained from the former by considering only *affine functions* $g(\cdot)$ in (3.1). Similarly, the inner-optimization in $\mathcal{D}'$ can be regarded as an extension of the problem which results from applying *Lagrangean relaxation* to $\mathcal{P}$. In Section 4.1 it will be demonstrated that the dual programme $\mathcal{D}$ too, comprises the Lagrangean dual as a special case.

Section 4

Some well-known special cases

The results in the previous section generalize the corresponding results in Linear Programming. A major drawback of the general duality theory, however, is its limited usefulness, due to the fact that the dual space $\mathcal{F}$ is immense. Moreover, a complete description of the value-function, an a priori known optimal dual solution, is more difficult to obtain than solving the given mathematical programme $\mathcal{P}$. Even worse, it may well be impossible to give a complete description of the value-function by means of either an explicit formula or a finite table of function values. Therefore, one usually must confine oneself to dual solutions $g(\cdot)$ which have ***special structure***. Mathematically, this idea implies that the dual solution space $\mathcal{F}$ in (2.6) is replaced by a ***subset*** of $\mathcal{F}$. Obviously, this will not affect the property of weak duality. The property of strong duality, on the other hand, will generally cease to hold, unless, of course, the value-function is a member of the subset under consideration, or unless the primal programme has the appropriate special structure. For the Farkas property, a similar remark applies. In this section we will show that under the appropriate reduction of the dual space, some well-known dual programmes are recovered from the one that has been defined in Section 2. In Section 5 a new reduction of $\mathcal{F}$ will be introduced, and strong duality will be proven to hold if the primal programme meets some additional conditions. In what follows, the subset of $\mathcal{F}$ under consideration, will be denoted by (a generic) Γ. So, for various choices of $\Gamma \subseteq \mathcal{F}$, the following optimization problem will be considered to be the dual of (2.1).

$$\begin{aligned} \min_{g(\cdot)} \quad & g(0) \\ \text{s.t.} \quad & g(G(x)) \geq f(x) \;\; \forall x \in X \\ & g(\cdot) \in \Gamma \end{aligned} \tag{4.1}$$

4.1 Lagrangean duality

Let Γ be the set of *affine functions* in $\mathcal{F}$, so

$$\Gamma = \{g(\cdot) \in \mathcal{F} \mid \exists(\mu, \theta) \in \mathbb{R}^{m+1} \; \forall r \in \mathbb{R}^m : g(r) = \mu^T r + \theta\} \tag{4.2}$$

In that case, (4.1) boils down to

$$\begin{aligned} \min_{\mu,\theta} \quad & \theta \\ \text{s.t.} \quad & \mu^T G(x) + \theta \geq f(x) \;\; \forall x \in X \\ & \mu_i \geq 0 \;\; \forall i \in \{1, \ldots, m\} : \diamond_i \in \{\leq\} \end{aligned} \tag{4.3}$$

The non-negativity requirement in (4.3) follows from the monotonicity condition in $\mathcal{F}$. Note that the property of *weak duality* applies to the pair of programmes (2.1) and (4.3).

The property of *strong duality* can be proven to hold if $\mathcal{P}$ is a convex programme (i.e. X closed and convex, $f(\cdot)$ concave on X, $G(\cdot)$ convex on X and lower semi-continuous w.r.t. X and $\diamond \in \{\leq\}^m$) and, additionally, $v(0) \in \mathbb{R}$ and some Lipschitz condition on $v(\cdot)$ with respect to the right-hand-side of (2.1) (i.c. the origin) is verified (see

[Geoffrion, 1972]). This can be explained intuitively by observing that, under these conditions, $v(\cdot)$ is concave and not infinitely steep at the origin. As a consequence, $v(\cdot)$ is supported at the origin by some affine function $g^*(\cdot)$. As a result, $g^*(r) \geq v(r)\ \forall r \in \mathbb{R}^m$ and $g^*(0) = v(0)$. Furthermore, $g^*(\cdot) \in \mathcal{F}$, because the converse would imply the existence of an index $i \in \{1, \ldots, m\}$ for which $\mu_i < 0$. Choosing r a scalar multiple of the i-th unit-vector with the scalar approaching $+\infty$, the function value of $g^*(\cdot)$ would tend to $-\infty$, whereas the function value of $v(\cdot)$ would never fall below $v(0) \in \mathbb{R}$; an obvious contradiction. As a result, $g^*(\cdot)$ is a feasible solution for (4.3) with an objective value equal to the optimal solution value of the primal programme, which proves strong duality for this case.

Under the same assumptions, the *Farkas property* too, remains valid, because it boils down to a separating hyperplane theorem with respect to the closed convex set RHS. Note that the assumption of concavity on the primal objective function is superfluous here.

Finally, let us consider a trivial reformulation of (4.3). For each choice of μ, the optimal value for θ is equal to $\sup_x\{f(x) - \mu^T G(x) \mid x \in X\}$. Therefore, (4.3) is equivalent to the following programme, which is nothing but the well-known *Lagrangean dual*.

$$\begin{aligned} \min_{\mu} \quad & \sup_{x}\{f(x) - \mu^T G(x) \mid x \in X\} \\ \text{s.t.} \quad & \mu_i \geq 0 \ \ \forall i \in \{1, \ldots, m\} : \diamond_i \in \{\leq\} \end{aligned} \tag{4.4}$$

This proves that if only affine dual functions are considered in (2.6), the *Lagrangean dual* results. In this respect, Lagrangean duality can be considered to be a special case of the general duality theory that is discussed in Sections 2 and 3. More details on Lagrangean duality can be found in the outstanding paper by [Geoffrion, 1972], where it is also shown that in a similar way, Linear Programming duality and Convex Quadratic Programming duality (cf. [Dorn, 1960], [Cottle, 1963]) are special cases of Lagrangean duality. As a result, (2.6) comprises the latter two dual programmes as special cases as well.

4.2 Augmented Lagrangean duality

Let Γ be the set of *real-valued, finitely representable* dual solutions. So,

$$\Gamma = \{g(\cdot) \in \mathcal{F} \mid \exists(\lambda, \theta) \in \Lambda \times \mathbb{R} : \ g(\cdot) = h(\lambda, \cdot) + \theta\} \tag{4.5}$$

where $\Lambda \subseteq \mathbb{R}^k$ ($k \in \mathbb{N}$) is a *given* set, and $h(\cdot\,,\cdot) : \Lambda \times \mathbb{R}^m \to \mathbb{R}$ is a *given* function with $h(\lambda, 0) = 0$. The reason why we call such dual solutions *finitely representable* is obvious: each function $g(\cdot) \in \Gamma$ is completele described by means of a *finite* number of scalars, viz. λ_i ($i = 1, \ldots, k$) and θ. Note that the requirement $h(\lambda, 0) = 0$ is not restrictive; if it is not met, then it can be enforced by a redefinition of $h(\cdot\,,\cdot)$. Under this definition of Γ, (4.1) can be reformulated as

$$\begin{aligned} \min_{\lambda,\theta} \quad & \theta \\ \text{s.t.} \quad & h(\lambda, G(x)) + \theta \geq f(x) \ \ \forall x \in X \\ & h(\lambda, \cdot) \in \mathcal{F} \\ & \lambda \in \Lambda \end{aligned} \tag{4.6}$$

The variable θ can easily be eliminated from (4.6), which results in the following mathematical programme

$$\begin{aligned} \min_{\lambda} \quad & \sup_{x}\{f(x) - h(\lambda, G(x)) \mid x \in X\} \\ \text{s.t.} \quad & h(\lambda, \cdot) \in \mathcal{F} \\ & \lambda \in \Lambda \end{aligned} \tag{4.7}$$

The inner optimization in (4.7) covers all approaches based on *augmented Lagrangeans*, in which the idea of penalizing constraint violations is merged into a Lagrangean approach. For instance, the *separable quadratic augmented Lagrangean approach* for equality constrained primal programmes ($\diamond \in \{=\}^m$), is recovered by taking

$$h(\cdot) : \mathbb{R}^{m+1+m} \to \mathbb{R}, \text{ as}$$
$$h(\mu, \sigma, r) = \mu^T r + \sigma r^T r \tag{4.8}$$

Substituting (4.8) into (4.7) renders

$$\begin{aligned} \min_{\mu,\sigma} \quad & \sup_x \{f(x) - \mu^T G(x) - \sigma G(x)^T G(x) \mid x \in X\} \\ \text{s.t.} \quad & (\mu, \sigma) \in \mathbb{R}^{m+1} \end{aligned} \tag{4.9}$$

Due to the fact that (4.9) is a minimization problem, every negative choice of σ is dominated by its positive counterpart $-\sigma$. Therefore, we may add to (4.9) a non-negativity requirement on σ. The resulting inner optimization is then consistent with the separable quadratic augmented Lagrangean that can be found in the literature. Other augmented Lagrangeans (see e.g. [Rockafellar, 1974]) can just as well be cast into the framework of the general duality theory. For a comprehensive survey on augmented Lagrangean methods, or *Lagrange multiplier methods* as they are also called, the reader is referred to the excellent monograph by [Bertsekas, 1982]; for an early reference on this topic, see [Everett, 1963]. Dual programmes like (4.7) could be termed *augmented Lagrangean duals* or *Lagrange multiplier duals*. In view of their relation with general duality theory, however, we prefer to call them *finitely dimensional dual programmes.*

In Section 5 it will be argued that for a specific class of augmented Lagrangeans (i.c. the separable ones), strong duality is unlikely to hold, unless some severe conditions on the primal programme are met. This apparently new observation reveals some of the limitations of separable augmented Lagrangeans and justifies the use of (more general) non-separable ones (see e.g. [Bertsekas, 1982, pp. 207,223,229]). The conditions we will present in Section 5 mainly concern either *discreteness* of the primal solution space or *stability* of the primal programme with respect to right-hand-side perturbations. In fact, it will be shown that strong duality is unlikely to hold in any separable dual programme, not just in finitely dimensional ones.

4.3 Integer Linear Programming duality

Suppose (2.1) is an *integer linear programme*. Hence, $f(x) = c^T x$, $G(x) = Ax - b$ and $X = \{x \in \mathbb{Z}^n \mid x \geq 0\}$ for suitable choices of c, A and b, where c and b are vectors and A is a matrix of appropiate dimensions. Let (4.1) be the dual programme, where Γ is chosen to consist of only those functions in $\mathcal{F}$, which are ***superadditive with respect to the vector*** $-b$. To be more specific,

$$\begin{aligned}\Gamma = \{g(\cdot) \in \mathcal{F} \mid g(-b) = 0,\ g(r) > -\infty\ \forall r \in RHS,\\ \forall r, r' \in RHS:\ g(r + r' + b) \geq g(r) + g(r')\}\end{aligned} \tag{4.10}$$

How does such a definition of Γ affect the basic duality results? Again, ***weak duality*** continues to hold because $\Gamma \subseteq \mathcal{F}$. As far as ***strong duality*** is concerned, note that $v(-b) > 0$ if and only if $\forall r \in RHS : v(r) = +\infty$. With this observation in mind, it is not difficult to prove that $v(\cdot) \in \Gamma$ if and only if $\exists r \in RHS : v(r) < +\infty$. The latter condition is therefore a sufficient condition for strong duality. Recall that in Theorem 3.3 the ***Farkas property*** was proven by first replacing the objective function in the primal programme by an objective function which equals zero at all points $x \in \mathbb{R}^n$, followed by applying strong duality to this primal and its dual. In the integer linear case we are considering here, the primal programme thus obtained, obviously satisfies the aforementioned sufficient condition for strong duality. As a result, the ***Farkas property***, with $\mathcal{F}$ replaced by Γ, also holds in Integer Linear Programming duality.

The reason why a set Γ as defined as in (4.10) is taken into consideration, is explained by the fact that it is now possible to get rid of the primal variables in the dual programme. For all $g(\cdot) \in \Gamma$ we have

$$\forall x \in X :\ g(Ax - b) \geq c^T x \Leftrightarrow \forall j \in \{1, \ldots, n\} :\ g(a_j - b) \geq c_j$$

Here, a_j is the j-th column of the matrix A and c_j is the j-th component of the vector c. Consequently, the dual programme can be formulated as

$$\begin{aligned} \min_{g(\cdot)} \quad & g(0) \\ \text{s.t.} \quad & g(a_j - b) \geq c_j \;\; \forall j \in \{1, \ldots, n\} \\ & g(\cdot) \in \Gamma \end{aligned}$$

with Γ as defined as in (4.10). For more information concerning *Integer Linear Programming duality* we refer to [Burdet & Johnson, 1977] and [Wolsey, 1981]. A slightly different but basically similar theory applies to the case of *Mixed-Integer Linear Programming*; see [Jeroslow, 1979].

Section 5

Additively separable dual solutions

As already mentioned in Section 1, the usefulness of Linear Programming duality is mainly due to a property we called *symmetry*, meaning that the dual of a linear programme is, again, a linear programme. In the general case, symmetry does not exist; as revealed by (2.1) and (2.6), the dual of a general optimization problem in finite dimensions is not a problem in finite dimensions. Unfortunately, the lack of symmetry goes even further than that. In Linear Programming, each primal variable corresponds to exactly one dual constraint and, due to symmetry, each dual variable is associated with exactly one primal constraint. In the general case, this phenomenon does not occur; although each point x in primal space X induces a unique dual constraint, it is no longer true that each primal constraint $G_i(x) \diamond_i 0$ corresponds to a single dual variable. In this section it will be shown that the one-to-one correspondence between primal constraints and dual variables can be restored by considering only *additively separable dual solutions*; by definition, such solutions are *the sum of* an appropriate number of *one-dimensional functions*. As far as *strong duality* is concerned, three results will be presented. First it will be shown that with hardly any additional conditions imposed, the property of strong duality continues to hold in case the dual space is restricted to the set of additively separable, extended real-valued functions. This result implies that the dual programme which results from just considering additively separable dual solutions, is a true alternative for (2.6). Secondly, it will be proven that, provided a severe condition on the primal programme is met, strong duality also holds if, in addition to additive separability, the dual solutions are required to be real-valued and continuous as well. The severe condition mainly concerns *stability* of the primal feasible set with respect to right-hand-side perturbations, and, as a result, *linearity* of the primal equality constraints. It should be noted, by the way, that the continuity requirement with respect to the additively separable dual solutions is more or less imposed by the requirement of real-valuedness. Finally, it is demonstrated that, in case the dual space is restricted to the set of real-valued additively separable functions, strong duality also holds if the primal programme is a *bounded discrete non-linear programme*. Theorem 5.1 can also be found in [Gould, 1969], where a simple and constructive proof is given; the remaining two theorems are new.

Let (2.1)–(2.5) be given. Let (4.1) be the dual programme of (2.1), where $\Gamma \subseteq \mathcal{F}$ is yet to be defined.

Definition 5.1 (Additive separability) *Let* $g(\cdot) : \mathbb{R}^m \to T$ *be a function, where* $T \subseteq \mathbb{R} \cup \{+\infty\}$ *or* $T \subseteq \mathbb{R} \cup \{-\infty\}$. *Then* $g(\cdot)$ *is called* additively separable *if there are functions* $g_1(\cdot), \ldots, g_m(\cdot) : \mathbb{R} \to T$ *such that* $\forall r \in \mathbb{R}^m : \; g(r) = \sum_{i=1}^{m} g_i(r_i)$.

The reason why the set T is introduced, is to avoid the situation in which $+\infty$ and $-\infty$ have to be added up. An additively separable function is thus the sum of a number of one-dimensional functions. It should be noted though, that, *mutatis mutandis*, all results in this section also apply to additively separable dual solutions which are the sum of an appropriate number of lower dimensional functions (not necessarily of dimension one).

In case the dual solution space is restricted to additively separable functions only, problem (4.1) is easily reformulated into

$$\begin{aligned} \min_{g_1(\cdot),\ldots,g_m(\cdot)} \quad & \sum_{i=1}^{m} g_i(0) \\ \text{s.t.} \quad & \sum_{i=1}^{m} g_i(G_i(x)) \geq f(x) \ \forall x \in X \\ & g_i(\cdot) \in \Gamma_i, \ \ i = 1, \ldots, m \end{aligned} \tag{5.1}$$

Here, $\Gamma_1, \ldots, \Gamma_m$ are sets of one-dimensional functions, such that $\Gamma_1 + \cdots + \Gamma_m \subseteq \Gamma$ is well-defined. The major distinction between (4.1) and (5.1) is that in the latter, each dual variable corresponds to a unique primal constraint, whereas such a relation does not exist in the former. In other words, the one-to-one correspondence between primal constraints and dual variables is restored if in the dual programme, only additively separable solutions are taken into consideration.

Before discussing the main results concerning strong duality between (2.1) and (5.1), let us first state a lemma on the codomain of dual solutions. As far as notation is concerned, the vector in which each entry i equals the maximum (minimum) of the i-th component of a vector a and a vector b, will be denoted by $\max\{a, b\}$ ($\min\{a, b\}$). A similar notation applies to scalars. Finally, the index set of equality constraints in (2.1) will be denoted by $E_\diamond$ (cf. Section 3.3 of Part I).

Lemma 5.1 *Let (2.1)–(2.5) be given. Let (4.1) be the dual programme of (2.1).*

1. *If* $\Gamma = \{g(\cdot) \in \mathcal{F} \mid \forall r \in \mathbb{R}^m : \ g(r) > -\infty\}$ *then (4.1) has an optimal solution* $g^*(\cdot) \in \Gamma$ *with* $g^*(0) = v(0)$ *if and only if* $v(0) > -\infty$.

2. *If* $\Gamma = \{g(\cdot) \in \mathcal{F} \mid \forall r \in \mathbb{R}^m : \ g(r) < +\infty\}$ *then (4.1) has an optimal solution* $g^*(\cdot) \in \Gamma$ *with* $g^*(0) = v(0)$ *if and only if* $\forall r \in \mathbb{R}^m : \ v(r) < +\infty$.

3. *If* $\Gamma = \{g(\cdot) \in \mathcal{F} \mid \forall r \in \mathbb{R}^m : g(r) \in \mathbb{R}\}$ *then (4.1) has an optimal solution* $g^*(\cdot) \in \Gamma$ *with* $g^*(0) = v(0)$*, provided that* $v(0) > -\infty$ *and*

$$w(s,t) = \sup_x \{f(x) \mid x \in X,\ G(x) \leq s, \\ -G_i(x) \leq t_i \ \forall i \in E_\diamond\} \tag{5.2}$$

is less than $+\infty$ *for all values of* (s,t) *in the non-negative orthant* $\mathbb{R}_+^{m+|E_\diamond|}$.

Proof

1. The only-if part is trivially obtained, since $g^*(0) = v(0) = -\infty$ would imply $g^*(\cdot) \notin \Gamma$; an obvious contradiction. In order to prove the if-part, define the function $g^*(\cdot) : \mathbb{R}^m \to \mathbb{R} \cup \{\pm\infty\}$ where $g^*(r)$ equals $w(s,t)$ for $s = \max\{0, r\}$ and $t_i = -\min\{0, r_i\}$ $\forall i \in E_\diamond$ (cf. (5.2)). It is now easy to show that

$$g^*(r) \geq v(r) \ \ \forall r \in \mathbb{R}^m$$
$$g^*(r) \geq v(0) > -\infty \ \ \forall r \in \mathbb{R}^m$$
$$g^*(r) \leq g(r') \ \ \forall r, r' \in \mathbb{R}^m : \ r \diamond r'$$
$$g^*(0) = v(0)$$

As a result, $g^*(\cdot) \in \Gamma$. The result now follows from Lemma 2.1.

2. The only-if part immediately follows from Corollary 2.1. To prove the if-part, let $g^*(\cdot)$ be chosen equal to $v(\cdot)$. The result now follows from the feasibility, hence optimality, of the latter function.

3. Let $g^*(\cdot)$ be chosen as in 1. The result follows from a similar argument as in 1. and from the fact that $g^*(\cdot) \in \Gamma$.

□

Corollary 5.1 *Let the conditions of Lemma 5.1 sub 3 be satisfied. In addition, suppose that $w(s,t)$ as defined as in (5.2), is upper (lower) semi-continuous with respect to the non-negative orthant $\mathbb{R}_+^{m+|E_\diamond|}$. Then (4.1) has an optimal solution $g^*(\cdot) \in \Gamma$ with $g^*(0) = v(0)$ which is upper (lower) semi-continuous on $\mathbb{R}^m$.*

Proof The optimal dual solution that is defined in the proof of Lemma 5.1 sub 1. is easily shown to be upper (lower) semi-continuous on $\mathbb{R}^m$.

□

In the following three theorems, results on additively separable dual solutions will be stated. These theorems are all proven by mathematical induction to m, the number of constraints in the primal which are "dualized". Only the first theorem will be proven extensively; the remaining ones are similar, and therefore left to the reader to some extent.

The first theorem in this sequence states that, basically, dual solutions can always be required to be additively separable without invalidating weak and strong duality. A less attractive feature of this theorem is the fact that the codomain of all dual solutions must include $+\infty$.

Theorem 5.1 (Separability – general version) *Let (2.1)–(2.5) be given. Let (4.1) be the dual of (2.1), where $\Gamma \subseteq \mathcal{F}$ is defined as*

$$\begin{aligned} \Gamma &= \Gamma_1 + \cdots + \Gamma_m, \text{ with} \\ \Gamma_i &= \{g(\cdot) : \mathbb{R} \to \mathbb{R} \cup \{+\infty\} \mid \\ &\qquad g(r_i) \leq g(r_i') \;\; \forall r_i, r_i' \in \mathbb{R} : \; r_i \diamond_i r_i'\}, \; i = 1, \ldots, m \end{aligned} \tag{5.3}$$

As far as the primal programme is concerned, suppose that $\forall x \in X$: $f(x) < +\infty$ and $v(0) \in \mathbb{R}$. Then (4.1) has an optimal solution $g^*(\cdot) \in \Gamma$ with $g^*(0) = v(0)$.

Proof The proof is by mathematical induction to m. Note by the way, that of all sets Γ_i, at most two are distinct.

Basis of induction: $m = 1$

For $m = 1$ the result immediately follows from Lemma 5.1 sub 1.

Induction hypothesis: $m = \ell - 1$ ($\ell \in \mathbb{N} \setminus \{1\}$)

Suppose the statement is true for any primal–dual pair like (2.1) and (4.1), where $m = \ell - 1$ ($\ell \in \mathbb{N} \setminus \{1\}$ arbitrarily chosen), where Γ is defined as in (5.3), where the optimal solution value of the primal is a real number and where $\forall x \in X : \; f(x) < +\infty$.

Induction step: $m = \ell$

Let (2.1)–(2.4) be given. Suppose that $m = \ell$, that $f(x) < +\infty$ $\forall x \in X$ and that $v(0) \in \mathbb{R}$. Define

$$\widetilde{X} = \{x \in X \mid G_i(x) \diamond_i 0, \;\; i = 1, \ldots, \ell - 1\} \tag{5.4}$$

and consider the following reformulation of (2.1)

$$\begin{aligned} \max_x \;\; & f(x) \\ \text{s.t.} \;\; & G_\ell(x) \diamond_\ell 0 \\ & x \in \widetilde{X} \end{aligned} \tag{5.5}$$

As a primal programme, (5.5) satisfies the conditions that are mentioned in the theorem, so from the basis of induction it follows that

$$\exists g_\ell^*(\cdot) \in \Gamma_\ell : \begin{cases} g_\ell^*(G_\ell(x)) \geq f(x) \;\; \forall x \in \widetilde{X} \\ g_\ell^*(0) = v(0) \end{cases} \tag{5.6}$$

Let us now consider the following primal programme

$$\begin{aligned} \max_{x} \quad & f(x) - g_\ell^*(G_\ell(x)) \\ \text{s.t.} \quad & G_i(x) \diamond_i 0, \;\; i = 1, \ldots, \ell - 1 \\ & x \in X \end{aligned} \tag{5.7}$$

Note that the objective function in (5.7) is well-defined, because

$$\forall x \in X: \; f(x) < +\infty \wedge g_\ell^*(G_\ell(x)) > -\infty \tag{5.8}$$

Furthermore, the optimal solution value of (5.7) is real-valued and non-positive, because

1. $v(0) > -\infty \Rightarrow \exists \tilde{x} \in \widetilde{X}: \; G_\ell(\tilde{x}) \diamond_\ell 0 \wedge f(\tilde{x}) > -\infty \Rightarrow$
   $+\infty > v(0) = g_\ell^*(0) \geq g_\ell^*(G_\ell(\tilde{x})) \geq f(\tilde{x}) > -\infty \Rightarrow$
   $\exists \tilde{x} \in \widetilde{X}: \; f(\tilde{x}) - g_\ell^*(G_\ell(\tilde{x})) > -\infty$
2. $\forall x \in \widetilde{X}: \; f(x) - g_\ell^*(G_\ell(x)) \leq 0 < +\infty$

So, according to the induction hypothesis

$$\exists \left(g_1^*(\cdot), \ldots, g_{\ell-1}^*(\cdot)\right) \in \Gamma_1 \times \cdots \times \Gamma_{\ell-1} \tag{5.9}$$

satisfying

$$\begin{aligned} & \sum_{i=1}^{\ell-1} g_i^*(G_i(x)) \geq f(x) - g_\ell^*(G_\ell(x)) \;\; \forall x \in X \\ & \sum_{i=1}^{\ell-1} g_i^*(0) \leq 0 \end{aligned} \tag{5.10}$$

Now consider $\left(g_1^*(\cdot), \ldots, g_\ell^*(\cdot)\right) \in \Gamma_1 \times \cdots \times \Gamma_\ell$. From (5.10) it follows that

$$\sum_{i=1}^{\ell} g_i^*(G_i(x)) \geq f(x) \;\; \forall x \in X \tag{5.11}$$

As a result, $\sum_{i=1}^{\ell} g_i^*(\cdot)$ is a feasible solution for (4.1). Consequently, weak duality implies that

$$\sum_{i=1}^{\ell} g_i^*(0) \geq v(0) \tag{5.12}$$

However, (5.10) also implies that

$$\sum_{i=1}^{\ell} g_i^*(0) \leq g_\ell^*(0) = v(0) \tag{5.13}$$

This implies that $\sum_{i=1}^{\ell} g_i^*(\cdot) \in \Gamma$ is an optimal solution for (4.1) which closes the duality gap.

□

In the following theorem the dual solutions are required to be both real-valued and additively separable. Under severe assumptions on the primal programme, the property of strong duality is still verified. In one of these assumptions, the notions of *continuity* and *essentially boundedness* for point-to-set map reappear (cf. Definitions 2.3–2.6 of Part I).

Theorem 5.2 (Separability – restricted version) *Let (2.1)–(2.5) be given. Let (4.1) be the dual of (2.1), where $\Gamma \subseteq \mathcal{F}$ is defined as*

$$\begin{aligned} \Gamma &= \Gamma_1 + \cdots + \Gamma_m, \textit{ with} \\ \Gamma_i &= \{g(\cdot) : \mathbb{R} \to \mathbb{R} \mid g(\cdot) \textit{ is continuous on } \mathbb{R} \textit{ and} \\ &\qquad g(r_i) \leq g(r_i') \ \forall r_i, r_i' \in \mathbb{R} : \ r_i \diamond_i r_i'\}, \ i = 1, \ldots, m \end{aligned} \tag{5.14}$$

As far as the primal programme is concerned, suppose that

- $cl(X) \subseteq D$
- $\diamond \in \{\leq\}^m$
- *$f(\cdot)$ and $G(\cdot)$ are continuous w.r.t. D*

- $\forall x \in D: \ f(x) < +\infty$

- $v(0) > -\infty$

- *the feasible set map* $\alpha(\cdot) : \mathbb{R}^m \to\!\!\to \mathbb{R}^n$*, which is defined as* $\alpha(r) = \{x \in X \mid G(x) \diamond r\}$*, is continuous w.r.t.* $\mathbb{R}^m_+$

- *there is an index* $j \in \{1, \ldots, m\}$ *such that the point-to-set map* $\beta(\cdot) : \mathbb{R} \to\!\!\to \mathbb{R}^n$*, which is defined as* $\beta(r_j) = \{x \in X \mid G_j(x) \diamond_j r_j\}$*, is essentially bounded w.r.t.* $\mathbb{R}_+$

Then (4.1) has an optimal solution $g^*(\cdot) \in \Gamma$ *with* $g^*(0) = v(0)$.

Proof Without loss of generality we may choose the index j equal to 1. The proof of the theorem is then similar to the previous one. Note that the assumption on $\diamond$ implies that $E_\diamond = \emptyset$, so all sets Γ_i are identical.

Basis of induction: $m = 1$

First we will prove by contradiction that $\forall r \in \mathbb{R}_+ : v(r) < +\infty$. Suppose $v(r) = +\infty$ for some $r \in \mathbb{R}_+$. Then there is a sequence $(x^k)_{\mathbb{N}}$ such that $x^k \in \alpha(r)\ \forall k \in \mathbb{N}$ and $\lim f(x^k) = +\infty$. In the case that $m = 1$ the point-to-set map $\alpha(\cdot)$ coincides with the point-to-set map $\beta(\cdot)$. The condition of essentially boundedness then implies that $(x^k)_{\mathbb{N}}$ is bounded, which means that $(x^k)_{\mathbb{N}}$ has an accumulation point, say $x^\infty \in cl(X) \subseteq D$. The continuity assumption on $f(\cdot)$ w.r.t. D then implies that $f(x^\infty) = +\infty$. However, $\forall x \in D : f(x) < +\infty$ by assumption. This contradiction proves that $\forall r \in \mathbb{R}_+ : v(r) < +\infty$.

Applying Theorem 2.1 sub 3. of Part I and Corollary 5.1 proves the theorem for the case that $m = 1$.

Induction hypothesis: $m = \ell - 1$ ($\ell \in \mathbb{N} \setminus \{1\}$)

Suppose the statement is true for any primal–dual pair like (2.1) and (4.1), where $m = \ell - 1$ ($\ell \in \mathbb{N} \setminus \{1\}$ arbitrarily chosen), where the primal programme (2.1) satisfies the conditions mentioned in this theorem and where Γ is defined as in (5.14).

Induction step: $m = \ell$

Let (2.1)–(2.4) be given, where $m = \ell$ and where the primal programme satisfies the conditions mentioned in this theorem. As mentioned before, we may assume without loss of generality, that $\beta(r_1)$ is essentially bounded w.r.t. $\mathbb{R}_+$. The remaining part of this proof is along the same lines as the proof of Theorem 5.1; details are left to the reader. Note that checking the validity of the assumptions of (2.1) on the problems (5.5) and (5.7) is somewhat wearisome, but straightforward.

□

The ultimate goal of Theorem 5.2 was to state conditions under which the property of strong duality with real-valued and additively separable dual solutions could be verified. So why do we impose continuity on both the dual solutions $g_i^*(\cdot)$ and the feasible set map $\alpha(\cdot)$ as well? In proving the induction step, programmes (5.5) and (5.7) should have real-valued, and (5.7) even additively separable, optimal dual solutions. According to Lemma 5.1 sub 2 then, the value-functions of these programmes should not equal $+\infty$ anywhere. A natural way to enforce this, is to impose upper semi-continuity and real-valuedness on the objective functions, and some boundedness condition on (the closure of) the feasible sets. For (5.7) this implies that $g_\ell^*(\cdot)$ should be *lower* semi-continuous, which in general, does not result without the lower semi-continuity of the value-function of (5.5). On the other hand, the value-function of (5.5) and $g_\ell^*(\cdot)$ are upper semi-continuous under very weak assumptions (cf. Section 2 of Part I). This argumentation leads to the conclusion that the value-function of (5.5), and for the induction argument to work, also the value-function of (5.7), should be continuous, which in a sense is equivalent to the continuity of the feasible set map $\alpha(\cdot)$ (cf. Theorem 2.1 and 2.2 of Part I). The reasonableness of the continuity requirements (and the essentially boundedness condition on $\beta(\cdot)$) is hereby established.

What about the assumption on $\diamond$? In case (2.1) has equality constraints ($E_\diamond \neq \emptyset$), the framework of Theorem 5.2 is still applicable when each equality is replaced by two inequalities. However, if we take a closer look at the stability results in Section 3 of Part I, it seems as if such a reformulation satisfies stability, only if the equality constrained functions are *affine*. So, if attention is restricted to real-valued and additively separable dual solutions, and strong duality is still expected to hold, then all equality constraints had better be affine! If non-linear equality constraints are to be dealt with in Theorem 5.2, an improved version of Corollary 5.1 seems to be needed.

In the final theorem of this section, discrete programmes will be discussed. It will be proven that under some finiteness condition on the feasible set of the primal programme, strong duality with real-valued and additively separable dual solutions can be verified.

Theorem 5.3 (Separability - the discrete case) *Let (2.1)-(2.5) be given. Let (4.1) be the dual of (2.1), where* $\Gamma \subseteq \mathcal{F}$ *is defined as*

$$\begin{aligned} \Gamma &= \Gamma_1 + \cdots + \Gamma_m, \text{ with} \\ \Gamma_i &= \{g(\cdot) : \mathbb{R} \to \mathbb{R} \mid \qquad\qquad (5.15) \\ &\qquad g(r_i) \leq g(r_i') \ \forall r_i, r_i' \in \mathbb{R} : \ r_i \diamond_i r_i'\}, \ i = 1, \ldots, m \end{aligned}$$

As far as the primal programme is concerned, suppose that $v(0) > -\infty$, $\forall x \in X : \ f(x) < +\infty$ *and* $\{x \in X \mid (G_j(x) \leq s) \wedge (-G_j(x) \leq t \text{ if } \diamond_j \in \{=\})\}$ *is finite for all* $(s,t) \in \mathbb{R}^2_+$, *for at least one* $j \in \{1, \ldots, m\}$. *Then (4.1) has an optimal solution* $g^*(\cdot) \in \Gamma$ *with* $g^*(0) = v(0)$.

Proof Without loss of generality we may choose the index j equal to 1. The proof of the theorem is then, again, similar to the one of

Theorem 5.1. Note that of all sets Γ_i, at most two are distinct. The basis of induction follows from Lemma 5.1 sub 3. Remaining details are left to the reader.

□

From the preceding three theorems it follows that in the dual programme, separability per se can easily be obtained, but separability, combined with real-valuedness, can only be established under severe conditions on the primal programme. These conditions mainly concern either *finiteness*, or *stability* of the primal feasible set. In view of the results of Part I, the latter condition also seems to imply that in case non-linear equality constraints are present in the primal programme, stability, and hence strong duality, will generally not hold. These observations elucidate the limited applicability of some augmented Lagrangean methods. Basically, these methods are directed towards solving a dual programme with real-valued and finitely representable dual solutions. So, in the case that *separable* augmented Lagrangeans are used, a duality gap will generally exist, unless additional assumptions are met.

Section 6

Summary

In Part II a duality theory for general mathematical programmes is discussed. This general theory appears to be a true generalization of Linear Programming duality (and some other well-known special cases), in the sense that weak and strong duality, as well as the Farkas property, are also valid in the general case. A tremendous distinction between the two, however, is caused by the fact that the property of symmetry holds for the latter, whereas it does not hold for the former. More specifically, the dual programme of a general optimization problem in finite dimensions, is not an optimization problem in finite dimensions, and, contrary to the situation in Linear Programming, the one-to-one correspondence between primal constraints and dual variables does not exist. Indeed, the lack of applicability of the general theory seems to originate from this lack of symmetry.

If the dual solution space is reduced to contain only real-valued, finitely representable dual solutions, then the dual programme becomes an optimization problem in finite dimensions. This class of dual solutions seems to be the only interesting one from a computational point of view. In fact, considering only this class of dual solutions leads to what could be called *augmented Lagrangean duality*, and methods based on this concept are considered to be among the most efficient ones nowadays available, to solve constrained non-linear programming problems in practice.

In Part II research aimed at restoring the one-to-one correspondence between primal constraints and dual variables. This correspondence can be established by considering only additively separable functions in the dual solution space. In Section 5 conditions are stated, which imply strong duality between a primal and its "separable" dual. The contributions of Part II can be summarized as follows.

- In case the dual solution space is reduced to additively separable functions, which are also required to be real-valued and continuous, then strong duality holds provided some severe condition is met. This condition mainly concerns stability of the primal feasible set under right-hand-side perturbations. An important observation is that the continuity of the dual solutions is more or less imposed by the property of real-valuedness.
- Strong duality holds if in the dual programme attention is restriced to real-valued, addititively separable functions, and if the primal programme is in some sense, a bounded, discrete optimization problem.

The algorithmic implication of the first result for applied constrained optimization theory is considerable. In case separable augmented Lagrangeans are used to solve a given mathematical programme, then strong duality cannot be guaranteed to hold, unless the mathematical programme satisfies a severe (stability) assumption. If one would like to apply such methods to problems which go beyond these conditions, a non-separable augmented Lagrangean should be used if a duality gap is to be avoided.

Part III

Decomposition methods in General Mathematical Programming

Section 1

Introduction

In the early sixties, G.B. Dantzig, P. Wolfe and J.F. Benders devised two clever approaches for solving specially structured mathematical programming problems (cf. [Dantzig & Wolfe, 1960], [Benders, 1962]). Both methods aim at solving a given mathematical programme by means of alternately solving a parameterized descendant of this problem, usually referred to as the *subprogramme*, and adjusting the value of the parameter concerned. The latter is done by partially solving a mathematical programme, usually referred to as the *master programme*, which is nothing but a reformulation of the original programme. The part of the master programme which is solved during an iteration, is extended over iterations, which implies that it resembles more and more the master programme, hence the original programme, as iterations go by. In the subprogramme too, only a part of the given mathematical programme is taken into consideration, the remaining part being partially accounted for through the parameter. In the approach of Dantzig and Wolfe, for instance, the subprogramme is obtained from the original linear programme, by moving a weighted sum of some of the constraints to the objective function, the parameter being the vector of weights. In the approach of Benders, the subprogramme is obtained from the original problem by parametrically fixing some of the decision variables to specific values.

The concepts of decomposition have been the subject of extensive research ever since their introduction. To a certain extent, research aimed at adjusting the original ideas to develop decomposition schemes which subsumed the existing ones. Generalizations of Dantzig-Wolfe Decomposition have been discussed in [Dantzig, 1963] for the convex case, in [Sweeney & Murphy, 1979], [Holm & Tind, 1985] and [Tind & Holm, 1986] for the integer linear case, and in [Burkard et al., 1985] for the general case. Generalizations of Benders Decomposition have been developed in [Balas, 1969] and in [Lazimy, 1982,1985] for the convex quadratic case, in [Geoffrion, 1972[c]] and in [Duran & Grossmann, 1986] for two partially convex cases, in [Wolsey, 1981] for the general case, and in [Burkard et al., 1985] for the separable case. Other decomposition schemes, which all seem to have descended from the original procedures, have been developed as well. The most famous one in this respect is Lagrangean Relaxation, which is strongly related to Dantzig-Wolfe Decomposition; see e.g. [Geoffrion, 1974], [Fisher et al., 1975], [Goffin, 1977], [Fisher, 1981] and [Dyer & Walker, 1982].

The usefulness of decomposition is illustrated by the vast literature on applications, both theoretical and practical. For instance, Dantzig-Wolfe Decomposition type of methods have succesfully been applied to economic lot sizing problems ([Dzielinski & Gomory, 1965]), cutting stock problems ([Gilmore & Gomory, 1961,1963]), multi-commodity flow problems ([Assad, 1978]), routing problems ([Desrosiers et al., 1984]) and crew scheduling problems ([Crainic & Rousseau, 1987]). Benders Decomposition type of methods have succesfully been applied to variable factor programming problems ([Wilson, 1966]), location problems ([Geoffrion & Graves, 1974]), mixed-integer linear programming problems ([McDaniel & Devine, 1977]), multi-commodity flow problems ([Assad, 1978]) and quadratic assignment problems ([Bazaraa & Sherali, 1980]), as well as to practical problems dealing with scheduling of nursing personnel in a hospital ([Warner & Prawda, 1972]), planning of electric power generation ([Noonan & Giglio, 1977], [Bloom, 1983], [Rouhani et al., 1985]), investment and allocation decisions in water planning ([Armstrong & Willis, 1977]) and the determination of chemical equilibria ([Clasen, 1984]). For applications of Lagrangean Relaxation the reader is referred to e.g. [Geoffrion, 1974], [Fisher et al., 1975], [Goffin, 1977], [Fisher, 1981] and [Dyer & Walker, 1982]. For two monographs on decomposition methods in the linear and the (partially) convex case, the reader is referred to [Dirickx & Jennergren, 1979] and [Holmberg, 1985].

In Part III we present a unifying framework for extending the two aforementioned decomposition methods. Several names exist for these methods. Benders Decomposition is also known as *Primal Decomposition* and *Resource Directive Decomposition*, while Dantzig-Wolfe Decomposition is also referred to as *Column Generation*, *Generalized Linear Programming*, *Dual Decomposition* and *Price Directive Decomposition*. We denote our generalizations by *Variable Decomposition* and *Constraint Decomposition* respectively. The reason for adding these two names to the list of already existing ones, is to facilitate terminology. The names seem to be well-chosen because in these methods, decomposition takes place with respect to the set of variables and constraints respectively.

In Section 2 *Variable Decomposition* is discussed. The presentation is subdivided into a number of subsections. In *Subsection 2.1 and 2.2* the variable decomposition procedure is introduced and some complementary observations are made. The most crucial observation concerns an essential feature of the dual subprogrammes, without which the applicability of the procedure becomes questionable. It is shown that the formulation of the primal subprogrammes which is considered in Subsection 2.1, leads to dual subprogrammes with the desired property. It is also argued that in former extensions of Benders Decomposition (except for the one in [Wolsey, 1981]), additional conditions on the primal programme were needed, just because of the use of inappropriate formulations of the primal and dual subprogrammes. In *Subsection 2.3* the variable decomposition procedure is related to existing literature. The main difference with former generalizations of Benders Decomposition is twofold. First of all, we do not have to impose any additional assumptions on problem structure. Our development therefore comprises the ones in [Balas, 1969] and [Lazimy, 1982,1985] for the convex quadratic case, the ones in [Geoffrion, 1972] and [Duran & Grossmann, 1986] for two partially convex cases, and the one in [Burkard et al., 1985] for the separable case. One of the implications of our development is that (some of) the additional assumptions in [Lazimy, 1982,1985] and in [Burkard et al., 1985] are proven to be superfluous. A second difference between our approach and former generalizations is that the primal and dual solutions which are generated during the iterative process, are allowed to be inaccurate. Additionally, duality gaps between the underlying primal and dual subprogrammes are also allowed to occur. These two features of Variable Decomposition are appealing, especially as far as applications are concerned; exact solutions may not exist, and even if they do, it may be computationally burdensome, prohibitive, or even impossible, to identify them. Because of these features, Variable Decomposition also comprises the approach in [Wolsey, 1981], which is an extension of Benders Decomposition to the general case as well. Another advantage of our discussion is that we do not need additional notions that cloud the apparent similarity between the original Benders Decomposition and its extension. *Subsection 2.4* deals with the question which conditions ensure a non-cyclic behaviour of the procedure, and which conditions imply convergence within a finite number of steps. In *Subsection 2.5* the question of asymptotic convergence is addressed. Asymptotic convergence means that each accumulation point of a sequence of intermediate solutions, is (near-)optimal. It is proven that, apart from some minor conditions on problem data, *stability* of the primal subprogrammes and *closedness* of the point-to-set map which largely describes the construction of the sequence of intermediate primal solutions, are sufficient conditions for asymptotic convergence. It is also argued that, apart from pathological cases, both conditions are necessary as well. In *Subsection 2.6* a new application of Variable Decomposition is discussed. It turns out that if the approach is applied to mixed-integer non-linear programmes with an underlying convex structure, the mixed-integer part is separated from the non-linearities; in that case, the procedure amounts to alternately solving *purely convex programmes* and *mixed-integer linear programmes*.

A similar discussion with respect to *Constraint Decomposition* follows in Section 3. After its introduction and some minor observations in *Subsections 3.1 and 3.2*, the procedure is related to existing literature in *Subsection 3.3*. After that, its non-cyclic behaviour and finite convergence is established in *Subsection 3.4*. In *Subsection 3.5* the asymptotic convergence of the procedure is analysed. It is proven that under much more restrictive conditions than in the case of Variable Decomposition, Constraint Decomposition also converges asymptotically. The conditions mainly concern the *finite representability* and *continuity* of the intermediate dual solutions. The necessity of these sufficient conditions is also discussed.

As a direct consequence of the discussions in Section 2 and 3, it follows that Variable and Constraint Decomposition can be considered to be dual methods. More specifically, in Section 4 it will be shown that under extremely mild conditions, the latter approach applied to a mathematical programme is equivalent to the former applied to a well-chosen dual. The dual programme which has to be considered for this purpose, is the *additively separable dual programme* which is discussed in Section 5 of Part II.

In Section 5 extensions of some related decomposition schemes are presented. In fact, three related methods will be discussed, viz. *Lagrangean Decomposition*, or *Variable Splitting* as it is also sometimes referred to (cf. [Jörnsten et al., 1985], [Minoux, 1986], [Guignard & Kim, 1987]), *Cross Decomposition* (cf. [Van Roy, 1980,1983]) and *Kornai-Lipták Decomposition* (cf. [Kornai & Lipták, 1965], [Aardal & Ari, 1990]). The former is nothing but a Lagrangean Relaxation approach. The latter two can be regarded as being in between Variable and Constraint Decomposition; from both decomposition methods some parts are adopted, while at the same time, other parts are ignored. Part III is concluded by a summary.

The main contributions of Part III arise from the generality of the discussion. In deriving the variable and constraint decomposition procedures, no restrictive assumptions on the primal programme are made. Furthermore, the (intermediate) primal and dual solutions, which are generated during the iterative process, are allowed to be inaccurate and duality gaps between the underlying primal and dual programmes are allowed to occur. As a result of these features, Variable and Constraint Decomposition properly include former generalizations of Benders and Dantzig-Wolfe Decomposition. At first glance, the approach in [Burkard et al., 1985] seems to contradict this allegation. The mathematical programmes considered there, are formulated in the context of *algebraic optimization*. In such problems the underlying structure is described by totally ordered commutative (semi-)groups, which comprises the ordered group $(\mathbb{R}, +, \leq)$ we consider. However, except for the asymptotic convergence results, the approach in Part III can be extended to account for these abstract fields of optimization as well, which means that our approach has not really lost in generality. Moreover, we accounted for inaccuracies in intermediate primal and dual solutions as well as for duality gaps between the underlying primal and dual programmes. This fact, combined with the observation that in [Burkard et al., 1985] separability assumptions on the primal objective and constraint functions were needed in deriving their extension of Benders Decomposition, validates the above claim. The non-cyclic and finite convergence results in Subsection 2.4 and 3.4 are, due to their generality, new, although similar results were already mentioned in [Wolsey, 1981] and [Burkard et al., 1985]. The asymptotic convergence results are also new; as far as Variable Decomposition is concerned, a similar result for the partially convex case can be found in [Hogan, 1973], and as far as Constraint Decomposition is concerned, a weaker result which takes Lagrangean duality into account and which concentrates on the convergence of dual solutions, can be found in [Magnanti et al., 1976]. Finally, the procedure which results from applying Variable Decomposition to mixed-integer non-linear programmes with an underlying convex structure, also contributes to the variety of known applications. The duality relation between Variable and Constraint Decomposition for the general case is an extension of a similar result for the the linear case ([Lasdon, 1970]) and the separable case ([Burkard et al., 1985]). Finally, the extensions of Lagrangean Decomposition, Cross Decomposition and Kornai-Lipták Decomposition are obtained straightforwardly from the preceding discussions.

Section 2

Variable Decomposition

In this section the variable decomposition procedure is discussed. Basically, it is a generalization of Benders Decomposition to general mathematical programmes. The idea underlying this approach is also known as (Generalized) Benders Decomposition, Primal Decomposition and Resource Directive Decomposition. In Subsection 2.1 and 2.2 the procedure is explained and some remarks are made. In Subsection 2.3 the procedure is related to existing literature. Non-cyclic behaviour, finite convergence and asymptotic convergence are the topics of Subsections 2.4 and 2.5. Section 2 is concluded by an application of the procedure to mixed-integer non-linear programmes with an underlying convex structure.

2.1 Problem manipulations and solution strategy

Consider the following *primal programme*

$$\begin{aligned} \mathcal{P}: \quad & \max_{x,y} \quad f(x,y) \\ & \text{s.t.} \quad G(x,y) \diamond 0 \\ & \qquad\; (x,y) \in U \cap (X \times Y) \end{aligned} \tag{2.1}$$

where $f(\cdot)$ and $G(\cdot)$ are functions from $D \subseteq \mathbb{R}^n$ to $\mathbb{R} \cup \{\pm\infty\}$ and $\mathbb{R}^m$ respectively, $U \cap (X \times Y) \subseteq D$ and $\diamond \in \{\leq, =\}^m$ ($m, n \in \mathbb{N}$). More formally,

$$\begin{aligned} & U \cap (X \times Y) \subseteq D \subseteq \mathbb{R}^n \quad ; \quad \diamond \in \{\leq, =\}^m \\ & f(\cdot) : D \to \mathbb{R} \cup \{\pm\infty\} \quad ; \quad G(\cdot) : D \to \mathbb{R}^m \end{aligned} \tag{2.2}$$

As in Part I and II, $G(x,y) \diamond r$ is a shorter notation for $G_i(x,y) \diamond_i r_i$, $i = 1, \ldots, m$, where $G_i(\cdot)$, $\diamond_i$ and r_i are the i-th components of $G(\cdot)$, $\diamond$ and r respectively. Furthermore, it will be assumed that the following condition is satisfied.

> If (x,y) is a feasible solution for $\mathcal{P}$, then a $y' \in Y$ exists, such that (x,y') is a feasible solution for $\mathcal{P}$ with $f(x,y') > -\infty$. (2.3)

This assumption is not a severe one; it is already met if the codomain of the objective function $f(\cdot)$ does not include $-\infty$. The reason why we persist in taking *extended* objective functions $f(\cdot)$ into consideration, is to allow for the possibility of $f(\cdot)$ being the result of an optimization problem itself. The necessity of introducing assumption (2.3) will be explained in this subsection, after the variable decomposition procedure has been described.

In (2.1) *two types of constraints* occur, one of which is defined by means of *a constraint function* $G(\cdot)$, and the other by means of *a set* $U \cap (X \times Y)$. This fact does not imply that these two types are necessarily distinct. The set $U \cap (X \times Y)$ may partially, or even completely, be described by means of (in)equalities. The reason for keeping the two types separate, is to have the opportunity to treat one differently from the other. The distinction between the sets U and $X \times Y$ stems from the fact that the former allows for interdependence between the variables x and y, whereas the latter does not.

The reason for identifying *two types of variables* in (2.1), is to examine the possibility of splitting up the joint optimization in $\mathcal{P}$ into two separate ones, viz. one optimization with respect to the variables of one type and one optimization with respect to the variables of the other. Obviously, it would be illusory to think that a complete independence between the two optimizations could be achieved. Therefore, we seek to decompose the joint optimization in (2.1) into, on the one hand, an optimization with respect to the x-variables and, on the other hand, an optimization with respect to the y-variables *conditionally on* the x-variables. The latter optimization is easily obtained from $\mathcal{P}$ by fixing the x-variables to a certain value $\bar{x} \in X$. By doing so, the following family of parameterized *primal subprogrammes* is obtained

$$\begin{aligned} \mathcal{P}(\bar{x}): \quad \max_{x,y} \quad & f(x,y) \\ \text{s.t.} \quad & G(x,y) \diamond 0 \\ & x = \bar{x} \\ & (x,y) \in U \cap (\widetilde{X} \times Y) \end{aligned} \tag{2.4}$$

where $\widetilde{X}$ may be any superset of X. If we let $\varphi(\cdot)$ denote the optimal objective value of a given optimization problem (being defined as $-\infty$ ($+\infty$) if the feasible set of the maximization (minimization) problem in question is empty), then the optimization with respect to the x-variables looks like

$$\begin{aligned} \max_{\bar{x}} \quad & \varphi(\mathcal{P}(\bar{x})) \\ \text{s.t.} \quad & \varphi(\mathcal{P}^0(\bar{x})) \geq 0 \\ & \bar{x} \in X \end{aligned} \tag{2.5}$$

where $\mathcal{P}^0(\bar{x})$ is a mathematical programme obtained from $\mathcal{P}(\bar{x})$ by replacing the objective function $f(\cdot)$ with an objective function which equals zero at all points $(x, y) \in U \cap (X \times Y)$. As a result, the constraint $\varphi(\mathcal{P}^0(\bar{x})) \geq 0$ guarantees that in (2.5) only those values for $\bar{x} \in X$ are taken into account that can be supplemented with a value $\bar{y}$, such that $(\bar{x}, \bar{y})$ is a feasible solution for $\mathcal{P}$. It is a trivial observation that (2.5) is equivalent to the original optimization problem $\mathcal{P}$ in the following sense.

Theorem 2.1 *Let (2.1)–(2.5) be given. Then the optimal objective function values of $\mathcal{P}$ and (2.5) coincide. In addition, if $\mathcal{P}$ is regular, then (x^*, y^*) is an ϵ-optimal solution for $\mathcal{P}$ if and only if $\exists \epsilon' \in [0, \epsilon]$: x^* is an ϵ'-optimal solution for (2.5) and (x^*, y^*) is an $(\epsilon - \epsilon')$-optimal solution for $\mathcal{P}(x^*)$.*

Proof This is left to the reader. For the notions of infeasibility, unboundedness, regularity and ϵ-optimality we refer to Section 2 of Part II.

□

The nested optimization in (2.5) is the core of Variable Decomposition: the optimization with respect to, on the one hand, the x-variables and, on the other hand, the y-variables is done by two distinct, albeit related programmes. Unfortunately, (2.5) itself is usually not suitable to be dealt with directly, due to the complexity of both the objective function and the constraints. Therefore, some additional problem manipulations have to be carried out in order to end up with a formulation which is more amenable to solution. The key problem manipulations involved are called *projection*, *dualization* and *outer approximation*, while the key solution strategy is referred to as *relaxation*.[1] In fact, *projection* has already been carried out, being the transition from (2.1) to (2.5). For the second key problem manipulation, we need the *dual subprogrammes*

[1]This terminology has been adopted from [Geoffrion, 1972[a]]. We changed his notion of *outer linearization* into *outer approximation* because linear approximations are too restrictive for the general case we are dealing with.

$\mathcal{D}(\bar{x})$, which are the dual programmes of $\mathcal{P}(\bar{x})$. These programmes are defined as (cf. Section 2 of Part II)

$$\begin{aligned} \mathcal{D}(\bar{x}): \quad & \min_{g(\cdot)} \; g(0,\bar{x}) \\ & \text{s.t.} \quad g(G(x,y),x) \geq f(x,y) \; \forall (x,y) \in U \cap (\widetilde{X} \times Y) \\ & \qquad\; g(\cdot) \in \mathcal{F} \end{aligned} \tag{2.6}$$

As in Part II the dual solution space $\mathcal{F}$ equals

$$\begin{aligned} \mathcal{F} \; = \; & \{ g(\cdot) : \mathbb{R}^{m+n_1} \rightarrow \mathbb{R} \cup \{\pm\infty\} \mid \\ & g(r,\bar{x}) \leq g(r',\bar{x}) \;\; \forall (r,\bar{x}),(r',\bar{x}) \in RHS : \; r \diamond r' \} \end{aligned} \tag{2.7}$$

where n_1 is the dimension of the x-vector, and RHS is the *set of feasible right-hand-sides*, which is defined as

$$\begin{aligned} RHS = \{ & (r,\bar{x}) \in \mathbb{R}^{m+n_1} \mid \\ & \exists (x,y) \in U \cap (\widetilde{X} \times Y) : \; G(x,y) \diamond r, \; x = \bar{x} \} \end{aligned} \tag{2.8}$$

The dual programmes $\mathcal{D}^0(\bar{x})$ of $\mathcal{P}^0(\bar{x})$ are defined similarly; they are obtained from (2.6) by replacing $f(x,y)$ with 0.

Dualization can now be carried out by replacing the two primal programmes in (2.5) by their dual counterparts. Due to *strong duality*, an equivalent mathematical programme is obtained.

$$\begin{aligned} \max_{x} \quad & \varphi(\mathcal{D}(x)) \\ \text{s.t.} \quad & \varphi(\mathcal{D}^0(x)) \geq 0 \\ & x \in X \end{aligned} \tag{2.9}$$

Note that we skipped the bar from the x-variables. In order to unify the way in which the objective $\varphi(\mathcal{D}(x))$ and the constraint $\varphi(\mathcal{D}^0(x)) \geq 0$ can be handled, we restate (2.9) as

$$\begin{aligned} \max_{x,\theta} \quad & \theta \\ \text{s.t.} \quad & -\varphi(\mathcal{D}(x)) + \theta \leq 0 \\ & -\varphi(\mathcal{D}^0(x)) \leq 0 \\ & (x,\theta) \in X \times \mathbb{R} \end{aligned} \tag{2.10}$$

Obviously, $\mathcal{P}$ and (2.10) are equivalent problems, in the sense that one is infeasible, unbounded or regular if and only if the other is. Moreover, from (2.3) it follows that feasible values for the x-variables in the two programmes coincide. In addition, if one of the two programmes is regular, then (near-)optimal solution values for the x-variables in the two programmes coincide as well. *Outer approximation* is based on the fact that a given scalar is less than or equal to an infimal value of a set if and only if this scalar is less than or equal to every element of this set. To be more specific, if the feasible sets of $\mathcal{D}(x)$ and $\mathcal{D}^0(x)$ are denoted by Δ and Δ^0 respectively, hence

$$\begin{aligned} \Delta &= \{g(\cdot) \in \mathcal{F} \mid g(G(x,y),x) \geq f(x,y)\ \forall (x,y) \in U \cap (\widetilde{X} \times Y)\} \\ \Delta^0 &= \{h(\cdot) \in \mathcal{F} \mid h(G(x,y),x) \geq 0\ \forall (x,y) \in U \cap (\widetilde{X} \times Y)\} \end{aligned}$$

then the following statements hold

$$\begin{aligned} \theta \leq \varphi(\mathcal{D}(x)) \quad &\Leftrightarrow \quad \theta \leq g(0,x)\ \ \forall g(\cdot) \in \Delta \\ 0 \leq \varphi(\mathcal{D}^0(x)) \quad &\Leftrightarrow \quad 0 \leq h(0,x)\ \ \forall h(\cdot) \in \Delta^0 \end{aligned}$$

These observations ultimately lead to the final restatement of $\mathcal{P}$.

$$\begin{array}{lll} \mathcal{VD}(\Delta, \Delta^0): & \max_{x,\theta} & \theta \\ & \text{s.t.} & -g(0,x) + \theta \leq 0 \;\; \forall g(\cdot) \in \Delta \\ & & -h(0,x) \qquad \leq 0 \;\; \forall h(\cdot) \in \Delta^0 \\ & & (x,\theta) \in X \times \mathbb{R} \end{array} \tag{2.11}$$

$\mathcal{VD}(\Delta, \Delta^0)$ is what is usually referred to as the *master programme*. The master programme is easily proven to be equivalent to the original programme $\mathcal{P}$, in the sense that the former is infeasible, unbounded or regular if and only if the latter is. Moreover, feasible as well as (near-) optimal solutions for the x-variables in the two programmes coincide. Note that the so-called *cut sets* Δ and Δ^0 are *independent* of x, which appears to be a crucial property for the applicability of the variable decomposition approach. In fact, the reason why in former generalizations of Benders Decomposition additional conditions had to be imposed on problem structure, is explained by a violation of this property (cf. [Lazimy, 1982, 1985], [Geoffrion, 1972[c]], [Burkard et al., 1985]). This will be discussed in more detail in the following two subsections.

The transition from (2.9) to (2.11) is called outer approximation because both the objective function and the feasible set in (2.9) are described by means of the *intersection of cuts*. The constraints which are defined by Δ are called *value cuts* because their intersection determines the objective value in (2.9). For a similar reason, the cuts described by the set Δ^0 are referred to as *feasibility cuts*.

Regarding the number of variables, the master programme is to be preferred to the original one. Unfortunately, this preference changes as soon as the number of constraints is taken into consideration. In general, the master programme contains a huge, if not infinite number of constraints, even if all dominated ones would be identified and eliminated in advance. Therefore, *relaxation* is the only reasonable key solution strategy that can be applied to solve $\mathcal{VD}(\Delta, \Delta^0)$. Here, relaxation just means that some (or most) of the constraints which describe the feasible set, are ignored. More formally, if $\overline{\Delta} \subseteq \Delta$ and $\overline{\Delta}^0 \subseteq \Delta^0$, then the corresponding relaxation of $\mathcal{VD}(\Delta, \Delta^0)$ reads

$$\mathcal{VD}(\overline{\Delta}, \overline{\Delta}^0): \quad \max_{x,\theta} \quad \theta$$
$$\text{s.t.} \quad -g(0,x) + \theta \leq 0 \;\; \forall g(\cdot) \in \overline{\Delta}$$
$$-h(0,x) \qquad \leq 0 \;\; \forall h(\cdot) \in \overline{\Delta}^0 \tag{2.12}$$
$$(x, \theta) \in X \times \mathbb{R}$$

$\mathcal{VD}(\overline{\Delta}, \overline{\Delta}^0)$ is what is usually referred to as the *relaxed master programme*. Optimal solutions for this programme supply *upper bounds* for the master programme, and thus for the original programme $\mathcal{P}$, because every solution (x, θ) which is feasible for $\mathcal{VD}(\Delta, \Delta^0)$ is also feasible for $\mathcal{VD}(\overline{\Delta}, \overline{\Delta}^0)$ with the same objective value. The converse of this statement is generally not true. Therefore, we have

Theorem 2.2 (Upper bounds) *Let (2.1)–(2.12) be given.*

1. $\forall \overline{\Delta} \subseteq \Delta \; \forall \overline{\Delta}^0 \subseteq \Delta^0 : \; \varphi\left(\mathcal{VD}(\overline{\Delta}, \overline{\Delta}^0)\right) \geq \varphi(\mathcal{P})$

2. *If $\mathcal{P}$ is regular then* $\exists \overline{\Delta} \subseteq \Delta \; \exists \overline{\Delta}^0 \subseteq \Delta^0 \; \forall \epsilon \geq 0 : \; \varphi\left(\mathcal{VD}(\overline{\Delta}, \overline{\Delta}^0)\right) \leq \varphi(\mathcal{P}) + \epsilon$

Proof This is left to the reader.

□

The situation is reversed when $\mathcal{P}(\bar{x})$ is considered. The feasible set of a primal subprogramme can be thought of as being a cross-section of the feasible set of the original programme, viz. consisting of those solutions of $\mathcal{P}$, which have a predetermined value for the x-variables. As a result, optimal solutions for $\mathcal{P}(\bar{x})$ supply *lower bounds* for $\mathcal{P}$.

Theorem 2.3 (Lower bounds) *Let (2.1)–(2.12) be given.*

1. $\forall \bar{x} \in X : \; \varphi(\mathcal{P}(\bar{x})) \leq \varphi(\mathcal{P})$

2. *If $\mathcal{P}$ is regular then* $\forall \epsilon > 0 \; \exists \bar{x} \in X : \; \varphi(\mathcal{P}(\bar{x})) \geq \varphi(\mathcal{P}) - \epsilon$

Proof This is left to the reader.

□

In the two aforementioned theorems the second result indicates that the bounds can be chosen as tight as desired. The question remains which subsets $\overline{\Delta} \subseteq \Delta$ and $\overline{\Delta}^0 \subseteq \Delta^0$ should be taken into consideration. In fact, two questions are involved here, viz. which subsets should be taken initially and how should these initial choices be modified in order to end up with a (near-)optimal solution for $\mathcal{P}$. Note that the constraints in the relaxed master programmes originate from feasible solutions of the dual programmes $\mathcal{D}(\bar{x})$ and $\mathcal{D}^0(\bar{x})$. These latter programmes in their turn, require predetermined values for the x-variables, which can be obtained from any relaxed master programme. This observation suggests an iterative procedure involving subprogrammes and relaxed master programmes. The following result appears to be useful, in the sense that it reveals what the variable decomposition procedure is going to look like.

Theorem 2.4 *Let (2.1)–(2.12) be given. Suppose* $\overline{\Delta} \subseteq \Delta$, $\overline{\Delta}^0 \subseteq \Delta^0$ *and* $\varphi\left(\mathcal{VD}(\overline{\Delta}, \overline{\Delta}^0)\right) < +\infty$.

1. *If* $\mathcal{VD}(\overline{\Delta}, \overline{\Delta}^0)$ *is infeasible, so is* $\mathcal{P}$

2. *Let* $(\bar{x}, \bar{\theta})$ *be any feasible solution for* $\mathcal{VD}(\overline{\Delta}, \overline{\Delta}^0)$, *then*

   - $\varphi(\mathcal{P}(\bar{x})) < \bar{\theta} \Leftrightarrow \exists \bar{g}(\cdot) \in \Delta \setminus \overline{\Delta} : \; -\bar{g}(0, \bar{x}) + \bar{\theta} > 0$
   - $\mathcal{P}(\bar{x})$ *is infeasible* $\Leftrightarrow \exists \bar{h}(\cdot) \in \Delta^0 \setminus \overline{\Delta}^0 : \; -\bar{h}(0, \bar{x}) > 0$

3. *Let* $(\bar{x}, \bar{\theta})$ *be an* ϵ_1*-optimal solution for* $\mathcal{VD}(\overline{\Delta}, \overline{\Delta}^0)$. *Suppose* $\bar{y}$ *is an* ϵ_2*-optimal solution for* $\mathcal{P}(\bar{x})$, *and assume that there is a* $\bar{g}(\cdot) \in \overline{\Delta}$ *such that* $\bar{g}(\cdot)$ *is an* ϵ_3*-optimal solution for* $\mathcal{D}(\bar{x})$, *then* $(\bar{x}, \bar{y})$ *is an* $(\epsilon_1 + \epsilon_2 + \epsilon_3)$*-optimal solution for* $\mathcal{P}$.

Proof

1. Trivial (cf. Theorem 2.2).

2. From strong duality it follows that $\varphi(\mathcal{P}(\bar{x})) < \bar{\theta} \Leftrightarrow \varphi(\mathcal{D}(\bar{x})) < \bar{\theta}$. Furthermore, $\varphi(\mathcal{D}(\bar{x})) < \bar{\theta} \Leftrightarrow \exists \bar{g}(\cdot) \in \Delta : \bar{g}(0, \bar{x}) < \bar{\theta}$. Finally, $\forall g(\cdot) \in \overline{\Delta} : g(0, \bar{x}) \geq \bar{\theta}$, hence $\bar{g}(\cdot) \in \Delta \setminus \overline{\Delta}$.

   The second assertion immediately follows from the Farkas property. Again, $\bar{h}(\cdot) \in \Delta^0 \setminus \overline{\Delta}^0$ because $h(0, \bar{x}) \geq 0 \;\; \forall h(\cdot) \in \overline{\Delta}^0$.

3. $\varphi(\mathcal{P}) \leq \varphi\left(\mathcal{VD}(\overline{\Delta}, \overline{\Delta}^0)\right) \leq \bar{\theta} + \epsilon_1 \leq \bar{g}(0, \bar{x}) + \epsilon_1 \leq$
   $\varphi(\mathcal{D}(\bar{x})) + \epsilon_1 + \epsilon_3 = \varphi(\mathcal{P}(\bar{x})) + (\epsilon_1 + \epsilon_3) \leq$
   $f(\bar{x}, \bar{y}) + (\epsilon_1 + \epsilon_2 + \epsilon_3)$

□

In defining the variable decomposition procedure we will need the following definitions; a superscript k is used to denote the value of a variable during iteration k.

UB^k = best upper bound for $\varphi(\mathcal{P})$ over the first k iterations

LB^k = best lower bound for $\varphi(\mathcal{P})$ over the first k iterations

$(x^{inc,k}, y^{inc,k})$ = best solution for $\mathcal{P}$ over the first k iterations[2]

[2]The superscript "inc" comes from "current incumbent"; see [Geoffrion & Marsten, 1972].

VARIABLE DECOMPOSITION PROCEDURE – START

0. INITIATION PHASE

BEGIN

0.1 **LET** $k := 1$;

0.2 **LET** $\Delta^k :\subseteq \Delta$, $\Delta^{0,k} :\subseteq \Delta^0$ **BE SUCH THAT** $\varphi(\mathcal{VD}(\Delta^k, \Delta^{0,k})) < +\infty$;

0.3 **IF** impossible **THEN STOP** because $\varphi(\mathcal{P}) = +\infty$

0.4 **ELSE BEGIN LET** $UB^0 := +\infty$; **LET** $LB^0 := -\infty$ **END**;

1. MASTER PROGRAMME PHASE

1.1 **SOLVE** $\mathcal{VD}(\Delta^k, \Delta^{0,k})$;

1.2 **IF** $\varphi(\mathcal{VD}(\Delta^k, \Delta^{0,k})) = -\infty$ **THEN STOP** because $\varphi(\mathcal{P}) = -\infty$;

ELSE BEGIN

1.3 **LET** $(\epsilon_1^k, x^k, \theta^k)$ **BE SUCH THAT** (x^k, θ^k) is an ϵ_1^k-optimal solution for $\mathcal{VD}(\Delta^k, \Delta^{0,k})$;

1.4 **LET** $UB^k := \min\{UB^{k-1}, \theta^k + \epsilon_1^k\}$

END;

2. SUBPROGRAMME PHASE

2.1 **SOLVE** $\mathcal{P}(x^k)$; **SOLVE** $\mathcal{D}(x^k)$;

2.2 **IF** $\varphi(\mathcal{P}(x^k)) = -\infty$ **THEN BEGIN**

LET $\epsilon_2^k := \epsilon_3^k := 0$; **LET** $LB^k := LB^{k-1}$;

2.3 **LET** $h^k(\cdot) \in \Delta^0$ **BE SUCH THAT** $h^k(0, x^k) < 0$;

2.4 **LET** $\Delta^k \subseteq: \Delta^{k+1} :\subseteq \Delta$; **LET** $\Delta^{0,k} \cup \{h^k(\cdot)\} \subseteq: \Delta^{0,k+1} :\subseteq \Delta^0$

END

ELSE BEGIN

2.5 **LET** (ϵ_2^k, y^k) **BE SUCH THAT** y^k is an ϵ_2^k-optimal solution for $\mathcal{P}(x^k)$;

2.6 **LET** $(\epsilon_3^k, g^k(\cdot))$ **BE SUCH THAT** $g^k(\cdot)$ is an ϵ_3^k-optimal solution for $\mathcal{D}(x^k)$;

2.7 **LET** $\Delta^k \cup \{g^k(\cdot)\} \subseteq: \Delta^{k+1} :\subseteq \Delta$; **LET** $\Delta^{0,k} \subseteq: \Delta^{0,k+1} :\subseteq \Delta^0$;

2.8 **LET** $LB^k := \max\{LB^{k-1}, f(x^k, y^k)\}$;

2.9 **IF** $LB^k > LB^{k-1}$ **THEN LET** $(x^{inc,k}, y^{inc,k}) := (x^k, y^k)$

ELSE LET $(x^{inc,k}, y^{inc,k}) := (x^{inc,k-1}, y^{inc,k-1})$

END;

| *3. OPTIMALITY CHECK* |

3.1 **LET** ϵ_0^k **BE SUCH THAT** $\epsilon_0^k \geq \epsilon_1^k + \epsilon_2^k + \epsilon_3^k$;

3.2 **IF** $UB^k - LB^k \leq \epsilon_0^k$ **THEN STOP** because $(x^{inc,k}, y^{inc,k})$ is a $(UB^k - LB^k)$-optimal solution for $\mathcal{P}$, which meets the required accuracy of ϵ_0^k

3.3 **ELSE BEGIN LET** $k := k + 1$; **GOTO** 1 **END**

END.

| **VARIABLE DECOMPOSITION PROCEDURE – END** |

In this procedure many questions are left unanswered, such as how to initiate it (step 0.2), how to solve the relaxed master programmes (steps 1.1 and 1.3) and the primal subprogrammes (steps 2.1 and 2.5), how to generate dual solutions (steps 0.2, 2.1, 2.3, 2.4, 2.6 and 2.7) and how to specify the inaccuracy parameters (steps 1.3, 2.5, 2.6 and 3.1). The reason for this is twofold. First of all, general answers do not exist to these questions; the unboundedness at initial iterations, for instance, is a notorious phenomenon, even if $\mathcal{P}$ is an ordinary linear programme. Secondly, it may well be possible that, depending on the specific problem characteristics involved, a choice between alternatives can be made. A similar remark applies to the extensions of the set Δ^k

$(\Delta^{0,k})$ to Δ^{k+1} $(\Delta^{0,k+1})$ in steps 2.4 and 2.7. For the generation of dual solutions the situation is not much different. Usually, these solutions can be obtained as a by-product once the primal programmes have been solved, but it may be possible to do better. To summarize, we have been discussing a *framework* here; applying the procedure to a specific problem(-class) still requires a significant amount of fine-tuning. We will close this subsection with some additional explanations.

Step 1.4 $\theta^k + \epsilon_1^k$ is an upper bound on $\varphi\left(\mathcal{VD}(\Delta^k, \Delta^{0,k})\right)$, hence on $\varphi(\mathcal{P})$.

Step 2.3 From (2.3) it follows that $\varphi(\mathcal{P}(x^k)) = -\infty$ if and only if $\varphi(\mathcal{P}^0(x^k)) = -\infty$. This observation explains the necessity of (2.3), because in this way, the executability of step 2.3 is guaranteed.

Step 2.4 $h^k(0, x^k) < 0$ implies that x^k will not be generated in subsequent iterations if the dual solution $h^k(\cdot) \in \Delta^0$ is added to $\Delta^{0,k+1}$. Note that $h^k(\cdot) \notin \Delta^{0,k}$, because $-h(0, x^k) \leq 0 \ \forall h(\cdot) \in \Delta^{0,k}$.

Step 2.5 $\mathcal{P}(x^k)$ is a regular programme, because $-\infty < \varphi(\mathcal{P}(x^k)) \leq \varphi(\mathcal{P}) \leq \varphi\left(\mathcal{VD}(\Delta^k, \Delta^{0,k})\right) \leq \varphi\left(\mathcal{VD}(\Delta^1, \Delta^{0,1})\right) < +\infty$.

Step 2.7 From $UB^k \leq \theta^k + \epsilon_1^k$ and $g^k(0, x^k) - \epsilon_2^k - \epsilon_3^k \leq f(x^k, y^k) \leq LB^k$, it follows that $\left(\theta^k + \epsilon_1^k\right) - \left(g^k(0, x^k) - \epsilon_2^k - \epsilon_3^k\right) \geq UB^k - LB^k$. Combining this inequality with $\epsilon_0^k \geq \epsilon_1^k + \epsilon_2^k + \epsilon_3^k$ and $UB^k - LB^k > \epsilon_0^k$ yields that $-g^k(0, x^k) + \theta^k > 0$. As a result, (x^k, θ^k) is cut off in subsequent iterations by adding the dual solution $g^k(\cdot) \in \Delta$ to Δ^{k+1}. Note that $g^k(\cdot) \notin \Delta^k$, because $-g(0, x^k) + \theta^k \leq 0$ $\forall g(\cdot) \in \Delta^k$.

Step 3.1 ϵ_0^k is a non-negative parameter, bounding the overall inaccuracy which is allowed in the final solution of $\mathcal{P}$. Note that all four inaccuracy parameters are allowed to change among iterations.

Step 3.2 $0 \leq \varphi(\mathcal{P}) - f(x^{inc,k}, y^{inc,k}) \leq UB^k - LB^k$. Consequently, $(x^{inc,k}, y^{inc,k})$ is a $(UB^k - LB^k)$-optimal solution for $\mathcal{P}$.

2.2 A major and some minor remarks

In this subsection some additional remarks concerning the variable decomposition procedure will be made. The first and by far most crucial observation concerns an essential feature of the family of dual subprogrammes $\mathcal{D}(\cdot)$, without which the applicability of the variable decomposition procedure is seriously called into question. The second observation stresses the fact that the procedure as presented in the previous subsection allows for inaccuracies, and hence duality gaps, during the iterative process and comments on the choice of the inaccuracy parameters involved. The third remark relates to the monotonicity of the upper bounds in case optimal solutions are generated in the master programme phase throughout the execution of the procedure. In the fourth and final remark, the consequences of ignoring feasibility cuts in the (relaxed) master programme(s) is touched upon.

The feasible set of $\mathcal{D}(\cdot)$

As one can see, all dual subprogrammes $\mathcal{D}(x)$ share a common feasible set; a similar remark applies to the dual programmes $\mathcal{D}^0(x)$. If this were not the case, relaxations of the master programme would look like

$$\begin{array}{lll} \max\limits_{x,\theta} & \theta & \\ \text{s.t.} & -g(0,x)+\theta \leq 0 & \forall g(\cdot) \in \overline{\Delta}(x) \\ & -h(0,x) \leq 0 & \forall h(\cdot) \in \overline{\Delta}^0(x) \\ & (x,\theta) \in X \times \mathbb{R} & \end{array} \tag{2.13}$$

where $\Delta(x)$ ($\Delta^0(x)$), the feasible set of $\mathcal{D}(x)$ ($\mathcal{D}^0(x)$), now depends on x, and where $\overline{\Delta}(x)$ ($\overline{\Delta}^0(x)$) is a subset of $\Delta(x)$ ($\Delta^0(x)$). In general, (2.13) is an intractable programme, because its constraints are of a very complicated nature. To illustrate, suppose that (2.13) has to be solved and that $-\bar{g}(0,x)+\theta \leq 0$ is one of its constraints. In addition, suppose that the feasibility of a trial solution $(\bar{x},\bar{\theta})$ has to be checked.

If $\bar{g}(\cdot) \in \overline{\Delta}(\bar{x})$, then $(\bar{x}, \bar{\theta})$ has to satisfy the constraint $-\bar{g}(0, x) + \theta \leq 0$. On the other hand, if $\bar{g}(\cdot) \notin \overline{\Delta}(\bar{x})$, then the constraint $-\bar{g}(0, x) + \theta \leq 0$ does not have to be met. Loosely speaking, each $x \in X$ defines its own set of constraints and it is exactly this fact that makes (2.13) highly intractable. Therefore, it seems crucial for the feasible sets of $\mathcal{D}(\bar{x})$ and $\mathcal{D}^0(\bar{x})$ to be independent of $\bar{x}$. What does this observation mean for the primal programmes $\mathcal{P}(\bar{x})$ and $\mathcal{P}^0(\bar{x})$? On the one hand, these programmes should only depend on $\bar{x}$ through their right-hand-sides; on the other hand, the primal constraints which are parameterized by $\bar{x}$ should belong to the set of constraints which are dualized. This explains why in (2.4) the constraints $x = \bar{x}$ are formulated in an explicit way. Former generalizations of Benders Decomposition have not made use of this elementary but crucial observation. For instance, the way in which a dual feasible set is obtained in [Lazimy, 1982,1985] which is independent of $\bar{x}$, is in view of our argument not only unnecessarily complicated, but also unnecessarily restrictive, because additional requirements on the original primal programme had to be introduced. The reason why in [Geoffrion, 1972^c] a constraint-qualification is needed and why in [Burkard et al., 1985] only separable problems are dealt with (like the one mentioned below), can also be explained by our argument. For more detailed information on this, the reader is referred to the next subsection.

Of course, when dealing with special cases, the primal subprogrammes may be formulated differently, as long as dual subprogrammes are obtained which satisfy the aforementioned requirement. For instance, if $U = \mathbb{R}^n$ and if $f(\cdot)$ and $G(\cdot)$ are additively separable with respect to x and y, i.e. $\forall (x, y) \in X \times Y$,

$$f(x, y) = f_1(x) + f_2(y), \; G(x, y) = G_1(x) + G_2(y) \tag{2.14}$$

then the primal subprogrammes may be formulated as

$$\begin{aligned} f_1(\bar{x}) \quad + \quad \max_y \quad & f_2(y) \\ \text{s.t.} \quad & G_2(y) \diamond -G_1(\bar{x}) \\ & y \in Y \end{aligned} \tag{2.15}$$

The feasible set of the resulting dual subprogramme will, indeed, be independent of $\bar{x}$. It is in fact this approach that has been pursued in [Benders, 1962] and in [Burkard et al., 1985].

Inaccuracies and duality gaps

As is revealed by the superscript k, all four inaccuracy parameters ϵ_i^k ($i = 0, \ldots, 3$) are allowed to vary between iterations. The overall inaccuracy ϵ_0^k, though, has to be chosen at least as large as the sum of the inaccuracies which are allowed in the blocks of which the procedure is composed; if ϵ_0^k has not been assigned in this way and if the optimality condition has not yet been satisfied, then the (near-)optimal solution of the current relaxed master programme is not necessarily cut off, implying that this solution may be generated in *all* subsequent iterations (see the additional explanation concerning step 2.7 in Subsection 2.1). Of course, if $UB^k - LB^k \leq \epsilon_0^k$, then $(x^{inc,k}, y^{inc,k})$ is a (near-)optimal solution for $\mathcal{P}$ which meets the required accuracy of ϵ_0^k, irrespective of whether ϵ_0^k outweighs the sum of the other three. But if the procedure is to be guaranteed not to exhibit cyclic behaviour, ϵ_0^k has to be chosen as indicated. For more information, the reader is referred to Subsection 2.4.

From a computational point of view it may be undesirable, or even prohibitive, to optimize the dual subprogrammes over the entire dual solution space Δ. In performing steps 0.2, 2.1, 2.3, 2.4, 2.6 and 2.7, one may therefore be tempted, or even be forced, to restrict attention to specially structured dual solutions only. Obviously, this may introduce duality gaps between primal and dual subprogrammes. The variable decomposition procedure as presented in the previous subsection, allows for such gaps, because these values are fully absorbed by the values of ϵ_3^k. One has to realize, however, that the larger the gap, the larger the overall inaccuracy ϵ_0^k one has to accept. An example of such an approach is given in [Rouhani et al., 1985]. In this paper, attention is restricted to affine dual solutions only (Lagrangean duality; see [Geoffrion, 1972^b] and Subsection 4.1 of Part II), although all kinds of convexity conditions may not have been met.

It should also be noted that the values for ϵ_2^k and ϵ_3^k need not be specified separately. If in steps 2.5 and 2.6 a primal and a dual solution y^k and $g^k(\cdot)$ have been obtained for which $f(x^k, y^k)$ and $g^k(0, x^k)$ are sufficiently close, then $\epsilon_2^k + \epsilon_3^k$ may be chosen equal to the difference of these two values. In this way, the individual values for ϵ_2^k and ϵ_3^k are unknown, but also unimportant.

Monotonous upper bounds

It is a trivial observation that, by construction, $(UB^k)_{\mathbb{N}}$ and $(LB^k)_{\mathbb{N}}$ constitute a monotonically non-increasing and non-decreasing sequence of upper and lower bounds respectively. If at each iteration, the relaxed master programme is solved to optimality, i.e. $\epsilon_1^k = 0 \ \forall k \in \mathbb{N}$, then the values for θ^k which are generated during the master programme phase, constitute a monotonically non-increasing sequence of upper bounds. In that case, step 1.4 can be modified into: "**LET** $UB^k := \theta^k$." A similar statement does *not* hold for the lower bounds.

Feasibility cuts

From a theoretical point of view, the constraint $\varphi(\mathcal{P}^0(\bar{x})) \geq 0$ in (2.5) may be considered to be redundant, because any $\bar{x} \in X$ that violates it yields an objective value of $-\infty$, which is a most undesirable objective value as far as maximization is concerned. The consequences of leaving out this constraint are only minor. Both the problem manipulations and the solution strategy are straightforwardly adapted and the feasibility cuts will no longer occur in the (relaxed) master programme(s); *mutatis mutandis*, the results in Subsections 2.4–2.6 will also remain valid. For more detailed information on this, the reader is referred to [Flippo et al., 1987].

2.3 Relations to existing literature

In this subsection Variable Decomposition will be related to existing literature. More specifically, it will be argued that Benders Decomposition and former extensions are subsumed by the approach of Subsection 2.1. The approaches that will be discussed here concern the ones in [Benders, 1962] for the partially linear case, [Balas, 1969] and [Lazimy, 1982,1985] for the convex quadratic case, [Geoffrion, 1972^c] and [Duran & Grossmann, 1986] for two partially convex cases, [Wolsey, 1981] for the general case, and [Burkard et al., 1985] for the separable case.

[Benders, 1962]

Benders can be regarded as the founding father of the procedure. In his paper *partially linear programmes* are considered, which can be obtained from (2.1) by specifying $f(x,y) = t(x) + qy$, $G(x,y) = T(x) + Ey - b$, $U = \mathbb{R}^n$ and $Y = \mathbb{R}^{n_2}_+$ $(0 < n_2 < n)$. Here, $t(\cdot)$ and $T(\cdot)$ are functions and q, E and b are vectors and matrices of appropriate dimensions. In this case, the primal subprogrammes $\mathcal{P}(\cdot)$ are linear programmes, which can be formulated as (2.15). As a result, the dual programmes $\mathcal{D}(\cdot)$ and $\mathcal{D}^0(\cdot)$ can be chosen to be the Linear Programming duals. If, in addition, all inaccuracies are chosen equal to zero, Benders's approach is recovered. Finite convergence is implied by Theorem 2.9 (cf. Subsection 2.4).

[Balas, 1969]

In Balas's paper *mixed-integer quadratic programmes with a convex structure* are considered. Such programmes can be obtained from (2.1) by specifying $f(x,y) = p^T x + q^T y + x^T C_{11} x + 2x^T C_{12} y + y^T C_{22} y$, $G(x,y) = Ax + Ey - b$, $U = \mathbb{R}^n$, $X = \{x \in \mathbb{Z}^{n_1} \mid x \geq 0\}$ and $Y = \{y \in \mathbb{R}^{n_2} \mid y \geq 0\}$ $(n_1 + n_2 = n)$. Here, p, q, and b are vectors, and C_{11}, C_{12}, C_{22}, A, and E are matrices of appropriate dimensions. Furthermore, it is assumed that

$$C = \begin{pmatrix} C_{11} & C_{12} \\ C'_{12} & C_{22} \end{pmatrix}$$

is negative semi-definite. As a result, the dual programmes $\mathcal{D}(\cdot)$ and $\mathcal{D}^0(\cdot)$ can be chosen to be the Convex Quadratic Programming duals (cf. [Dorn, 1960], [Cottle, 1963]), because strong duality and the Farkas property hold in this case. If, in addition, all inaccuracies are chosen equal to zero, Balas's approach results, although Balas does not solve relaxed master programmes to (near-)optimality, but attacks them by implicit enumeration instead. As an unfavourable consequence of this, useful information is lost because upper bounds are no longer obtained from these relaxations. Furthermore, contrary to what Balas claims, finite convergence is not implied by Theorem 2.9, but follows from Theorem 2.8 if, additionally, X is assumed to be bounded (cf. Subsection 2.4).

[Lazimy, 1982, 1985]

Lazimy considers *the same type of problems as Balas does*. First, the author explains that straightforwardly generalizing Benders's approach to the quadratic case yields an intractable master programme, due to the fact that its constraints are quadratic in x. However, the fact that each $x \in X$ in the master programme defines its own set of constraints is a far more serious problem (see Subsection 2.2), which remains unmentioned. To circumvent the quadratic constraints, Lazimy proposes a transformation of variables. In this way primal subprogrammes $\mathcal{P}(\cdot)$ are obtained, which depend on $\bar{x}$ through their right-hand-sides only. As we have explained in Subsection 2.2 this is crucial for the applicability of the variable decomposition procedure. In order to make the transformation of variables work, a constraint-qualification on the matrix C is imposed, which excludes Mixed-Integer Linear Programming as a special case. The fact that, in the end, the same procedure is obtained as the one described in Subsection 2.1 when applied to the quadratic case using Convex Quadratic Programming duality (cf. [Dorn, 1960], [Cottle, 1963]), inevitably leads to the conclusion that Lazimy's approach is both unnecessarily complicated and restrictive;

see also [Flippo & Rinnooy Kan, 1990].

[Geoffrion, 1972c]

In [Geoffrion, 1972c] mathematical programming problems are considered which may be referred to as *partially convex programmes*. Such programmes are obtained from (2.1) if $U = \mathbb{R}^n$ and $\diamond \in \{\leq\}^m$. In addition, it is assumed that for each $\bar{x} \in X$ the primal subprogramme

$$\begin{aligned} \mathcal{P}(\bar{x}):\quad \max_{y} \quad & f(\bar{x}, y) \\ \text{s.t.} \quad & G(\bar{x}, y) \leq 0 \\ & y \in Y \end{aligned}$$

is a convex programme. Finally, some minor regularity conditions are supposed to hold such that the feasible set of the dual subprogrammes $\mathcal{D}(\bar{x})$ may be restricted to the set of monotonically non-decreasing affine functions, without invalidating strong duality and the Farkas property (Lagrangean duality; see [Geoffrion, 1972b] and Subsection 4.1 of Part II). Under the aforementioned conditions, the dual programmes of $\mathcal{P}(\bar{x})$ can be defined as

$$\begin{aligned} \mathcal{D}(\bar{x}):\quad \min_{u,\zeta} \quad & \zeta \\ \text{s.t.} \quad & u^T G(\bar{x}, y) + \zeta \geq f(\bar{x}, y) \ \ \forall y \in Y \\ & u \geq 0 \end{aligned}$$

The dual programmes of $\mathcal{P}^0(\bar{x})$ are similarly defined

$$\begin{aligned} \mathcal{D}^0(\bar{x}):\quad \min_{v,\xi} \quad & \xi \\ \text{s.t.} \quad & v^T G(\bar{x}, y) + \xi \geq 0 \ \ \forall y \in Y \\ & v \geq 0 \end{aligned}$$

The use of these dual programmes leads to the following formulation of the master programme

$$\begin{array}{ll} \max\limits_{\bar{x},\theta} & \theta \\ \text{s.t.} & \theta \le \zeta \quad \forall u \ge 0 \, \forall \zeta : \ (\zeta \ge f(\bar{x},y) - u^T G(\bar{x},y) \ \forall y \in Y) \\ & 0 \le \xi \quad \forall v \ge 0 \, \forall \xi : \ (\xi \ge -v^T G(\bar{x},y) \ \forall y \in Y) \\ & (\bar{x},\theta) \in X \times \mathbb{R} \end{array}$$

For any $u \ge 0$ fixed, the only value cut that is non-redundant is $\theta \le \sup_y\{f(\bar{x},y) - u^T G(\bar{x},y) \mid y \in Y\}$; a similar statement holds for the feasibility cuts. The master programme is therefore reformulated as

$$\begin{array}{ll} \max\limits_{\bar{x},\theta} & \theta \\ \text{s.t.} & \theta \le \sup\limits_{y}\{f(\bar{x},y) - u^T G(\bar{x},y) \mid y \in Y\} \quad \forall u \ge 0 \\ & 0 \le \sup\limits_{y}\{-v^T G(\bar{x},y) \mid y \in Y\} \quad \forall v \ge 0 \\ & (\bar{x},\theta) \in X \times \mathbb{R} \end{array}$$

Geoffrion conjectures that relaxations of this master programme are only apt to solution if for every $u, v \ge 0$ both suprema are "essentially independent" of $\bar{x}$, in the sense that they can be evaluated for all $\bar{x} \in X$ with little or no more effort than is required to evaluate them for a single $\bar{x} \in X$. This is what Geoffrion calls *Property P*. In view of Subsection 2.2 it is not surprising that such an additional property is needed, due to the fact that the feasible sets of $\mathcal{D}(\bar{x})$ and $\mathcal{D}^0(\bar{x})$ depend on $\bar{x}$. Obviously, if these programmes can be reformulated in such a way that their feasible sets are independent of $\bar{x}$, then *Property P* holds. The converse of this statement, however, is not easily verified, because *Property P* has not been rigorously defined.

[Wolsey, 1981]

Suppose there are no constraint functions explicitly stated in (2.1) (i.e. $m = 0$), then the following dual subprogrammes are obtained

$$
\begin{aligned}
\mathcal{D}(\bar{x}): \ & \min_{\sigma(\cdot)} \ \sigma(\bar{x}) \\
& \text{s.t.} \ \ \sigma(x) \geq f(x,y) \ \ \forall (x,y) \in U \cap (X \times Y) \\
& \qquad \sigma(\cdot): \mathbb{R}^{n_1} \rightarrow \mathbb{R} \cup \{\pm\infty\}
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{D}^0(\bar{x}): \ & \min_{\tau(\cdot)} \ \tau(\bar{x}) \\
& \text{s.t.} \ \ \tau(x) \geq 0 \ \ \forall (x,y) \in U \cap (X \times Y) \\
& \qquad \tau(\cdot): \mathbb{R}^{n_1} \rightarrow \mathbb{R} \cup \{\pm\infty\}
\end{aligned}
$$

The resulting master programme then reads

$$
\begin{aligned}
\max_{x,\theta} \ & \theta \\
\text{s.t.} \ & \theta \leq \sigma(x) \ \ \forall \sigma(\cdot) \text{ feasible in } \mathcal{D}(\cdot) \\
& 0 \leq \tau(x) \ \ \forall \tau(\cdot) \text{ feasible in } \mathcal{D}^0(\cdot) \\
& (x,\theta) \in X \times \mathbb{R}
\end{aligned}
\tag{2.16}
$$

Most of the constraints in (2.16) are redundant. Let $X_1 = \{x \in X \mid \varphi(\mathcal{P}(x)) > -\infty\}$. Suppose $x^* \in X_1$, then $\theta^* > \sigma(x^*)$ for some $\sigma(\cdot)$ feasible for $\mathcal{D}(\cdot)$ if and only if $\theta^* > \sigma_{x^*}(x^*)$ for some $\sigma_{x^*}(\cdot)$ which is optimal for $\mathcal{D}(x^*)$. Furthermore, under (2.3) we have that $x^* \in X \setminus X_1$ if and only if $0 > \tau_{x^*}(x^*)$ for some $\tau_{x^*}(\cdot)$ which is feasible for $\mathcal{D}^0(x^*)$. As a result, (2.16) is equivalent to

$$\begin{aligned}
\max_{x,\theta}\quad & \theta \\
\text{s.t.}\quad & \theta \le \sigma_{x^*}(x) \quad \forall x^* \in X_1 \\
& 0 \le \tau_{x^*}(x) \quad \forall x^* \in X \setminus X_1 \\
& (x,\theta) \in X \times \mathbb{R}
\end{aligned} \tag{2.17}$$

Wolsey's master programme is similar to (2.17). In his approach, $\sigma_{x^*}(\cdot)$ is a *support function* with respect to $x^* \in X_1$ and $\tau_{x^*}(\cdot)$ is a *cut function* with respect to $x^* \in X \setminus X_1$. In our terminology, these notions are defined as follows

Definition 2.1 (Support function) *$\sigma_{x^*}(\cdot)$ is a* support function *with respect to x^* if $x^* \in X_1$, $\sigma_{x^*}(x) \geq \varphi(\mathcal{P}(x))$ $\forall x \in X_1$ and $\sigma_{x^*}(x^*) = \varphi(\mathcal{P}(x^*))$.*

Definition 2.2 (Cut function) *$\tau_{x^*}(\cdot)$ is a* cut function *with respect to x^* if $x^* \in X \setminus X_1$, $\tau_{x^*}(x) \geq 0$ $\forall x \in X_1$ and $\tau_{x^*}(x^*) < 0$.*

From Lemma 2.1 of Part II it follows that $\sigma_{x^*}(\cdot)$ is a support function if and only if $x^* \in X_1$ and $\sigma_{x^*}(\cdot)$ is optimal for $\mathcal{D}(x^*)$. It is also a trivial observation that $\tau_{x^*}(\cdot)$ is a cut function if and only if $x^* \in X \setminus X_1$, $\tau_{x^*}(\cdot)$ is feasible for $\mathcal{D}^0(x^*)$ and $\tau_{x^*}(x^*) < 0$. From these observations it follows that Wolsey's master programme is not only similar, but even identical to (2.17), which justifies the conclusion that Wolsey's framework fits into ours. The main difference is that Wolsey ignores (most of the) redundant constraints in the master programme. However, such constraints may not be redundant in *relaxations* of the master programme. Furthermore, in Wolsey's approach it is not allowed to dualize constraints which are formulated explicitly in the original programme $\mathcal{P}$ (i.c. $G(x,y) \diamond 0$).

[Burkard et al., 1985]

The mathematical programming problems considered by Burkard c.s. are formulated in the context of *algebraic optimization*. In such problems, the underlying structure is described by totally ordered commutative (semi-)groups. Such general structures include the ordered group $(\mathbb{R}, +, \leq)$ we considered. However, the approach in Subsection 2.1 can be amplified to account for these abstract fields of optimization as well, which means that our approach is, in fact, not overly restrictive. On the other hand, Burkard c.s. assume that both the objective $f(\cdot)$ and the constraint function $G(\cdot)$ are *separable* in x and y. This has been done in order to end up with primal subprogrammes like the ones in (2.15), so with dual programmes $\mathcal{D}(\bar{x})$ and $\mathcal{D}^0(\bar{x})$ in which the feasible sets are independent of $\bar{x}$. As explained in Subsection 2.2 this is crucial for the applicability of the variable decomposition procedure. The generality of Subsection 2.1, however, reveals that it is unnecessarily restrictive to consider only separable objective and constraint functions in the original problem formulation $\mathcal{P}$. Furthermore, the case in which inaccuracies and/or duality gaps are present during the iterative process, is not covered by their approach, due to the absence of a metric.

[Duran & Grossmann, 1986]

Duran and Grossmann consider a class of *additively separable, partially convex mixed-integer programming problems*, which follows from (2.1) by specifying $U = \mathbb{R}^n$, X as the intersection of a given polyhedron with $\mathbb{Z}^{n_1}$, Y as a polyhedron, $f(x,y) = c^T x + r(y)$ $\forall(x,y) \in X \times Y$ for some vector c and some real-valued function $r(\cdot)$, $G(x,y) = Bx + S(y)$ $\forall(x,y) \in X \times Y$ for some matrix B and some vector-valued function $S(\cdot)$ and $\diamond \in \{\leq\}^m$. Furthermore, it is assumed that X is a finite set, Y is a bounded polyhedron and $-r(\cdot)$ and $S(\cdot)$ are convex and continuously differentiable functions with respect to Y. Finally, Slater's condition is assumed to hold, i.e. $\exists(x,y) \in X \times Y$: $Bx + S(y) < 0$. Let us now apply the following version of Variable Decomposition. To obtain the dual subprogrammes, only the projecting constraints $x = \bar{x}$ are dualized. This leads to the following master programme.

$$\begin{aligned}\max_{x,\theta}\quad & \theta \\ \text{s.t.}\quad & \theta \le g(x) \ \ \forall g(\cdot) \in \Delta \\ & 0 \le h(x) \ \ \forall h(\cdot) \in \Delta^0 \\ & (x,\theta) \in X \times \mathbb{R}\end{aligned} \tag{2.18}$$

Here, Δ and Δ^0 are the feasible sets of the dual subprogrammes $\mathcal{D}(\bar{x})$ and $\mathcal{D}^0(\bar{x})$ respectively; hence

$$\begin{aligned}\Delta &= \{g(\cdot) : \mathbb{R}^{n_1} \to \mathbb{R} \cup \{\pm\infty\} \mid g(x) \ge c^T x + r(y) \\ &\qquad \forall (x,y) \in \widetilde{X} \times Y : \ Bx + S(y) \le 0\} \\ \Delta^0 &= \{h(\cdot) : \mathbb{R}^{n_1} \to \mathbb{R} \cup \{\pm\infty\} \mid h(x) \ge 0 \\ &\qquad \forall (x,y) \in \widetilde{X} \times Y : \ Bx + S(y) \le 0\}\end{aligned}$$

where $\widetilde{X}$ is any superset of X. Due to the convexity and continuous differentiability conditions on $-r(\cdot)$ and $S(\cdot)$, the master programme (2.18) can be reformulated into

$$\begin{aligned}\max_{x,y,\theta}\quad & \theta \\ \text{s.t.}\quad & \theta \le c^T x + r(\tilde{y}) + \nabla r(\tilde{y})(y - \tilde{y}) \quad \forall \tilde{y} \in Y \\ & 0 \le -Bx - S(\tilde{y}) - \nabla S(\tilde{y})(y - \tilde{y}) \quad \forall \tilde{y} \in Y \\ & (x,y,\theta) \in X \times Y \times \mathbb{R}\end{aligned} \tag{2.19}$$

where ∇ denotes the Jacobian of the function involved. The equivalence between the two master programmes is easily established. More specifically, we can derive the following result.

Theorem 2.5 *If $(\bar{x},\bar{\theta})$ is feasible for (2.18), then there is a $\bar{y} \in Y$ such that $(\bar{x},\bar{y},\bar{\theta})$ is feasible for (2.19). Vice versa, if $(\bar{x},\bar{y},\bar{\theta})$ is feasible for (2.19), then $(\bar{x},\bar{\theta})$ is feasible for (2.18).*

Proof Let $(\bar{x}, \bar{\theta})$ be feasible for (2.18). It follows by definition that $0 \leq h(\bar{x})\ \forall h(\cdot) \in \Delta^0$, so

$$0 \leq \varphi(\mathcal{D}^0(\bar{x})) = \varphi(\mathcal{P}^0(\bar{x}))$$

This inequality, and the conditions on $r(\cdot)$ and Y, imply that $\mathcal{P}(\bar{x})$ has an optimal solution, say $\bar{y} \in Y$. From the convexity assumptions it follows that

$$\forall \tilde{y} \in Y: \ B\bar{x} + S(\tilde{y}) + \nabla S(\tilde{y})(\bar{y} - \tilde{y}) \leq B\bar{x} + S(\bar{y}) \leq 0$$

Furthermore, $\bar{\theta} \leq g(\bar{x})\ \forall g(\cdot) \in \Delta$. This implies that

$$\forall \tilde{y} \in Y: \ \bar{\theta} \leq \varphi(\mathcal{D}(\bar{x})) = \varphi(\mathcal{P}(\bar{x})) =$$
$$c^T\bar{x} + r(\bar{y}) \leq c^T\bar{x} + r(\tilde{y}) + \nabla r(\tilde{y})(\bar{y} - \tilde{y})$$

which proves that $(\bar{x}, \bar{y}, \bar{\theta})$ is feasible for (2.19). In order to prove the second part of the theorem, let $(\bar{x}, \bar{y}, \bar{\theta})$ be feasible for (2.19). Consider $\tilde{y} = \bar{y}$. It follows that $0 \leq -B\bar{x} - S(\bar{y})$. This implies that $\varphi(\mathcal{P}^0(\bar{x})) = 0$. Applying strong duality yields

$$0 \leq h(\bar{x}) \ \ \forall h(\cdot) \in \Delta^0$$

A similar argument yields that $\bar{\theta} \leq c^T\bar{x} + r(\bar{y})$. Strong duality then implies that

$$\bar{\theta} \leq \varphi(\mathcal{P}(\bar{x})) = \varphi(\mathcal{D}(\bar{x})) \leq g(\bar{x}) \ \ \forall g(\cdot) \in \Delta$$

Hence, $(\bar{x}, \bar{\theta})$ is feasible for (2.18). This completes the proof.

□

As a result, the outer approximation approach of Duran and Grossmann can be cast into the framework of Variable Decomposition, in which, additionally, a reformulation of the master programme has been carried out by expanding its dimension. Note, however, that although the procedure can be regarded as a special instance of Variable Decomposition from a *conceptual* point of view, it may substantially differ from a straightforward implementation of Variable Decomposition from a *computational* point of view, due to the different ways in which intermediate primal solutions in *relaxations* of these two master programmes are cut off. One negative effect of this is that in the approach of Duran and Grossmann, infeasible values for the integer variables x can only be cut off from all subsequent iterations through an "ad-hoc" strategy (see [Duran & Grossmann, 1986, pp. 315,316,320]).

2.4 Non-cyclic behaviour and finite convergence

In this section it will be demonstrated that the variable decomposition procedure as explained in Subsection 2.1, does not exhibit cyclic behaviour. In addition, conditions will be stated under which the procedure is guaranteed to terminate within a finite number of steps.

Theorem 2.6 (Non-cyclic behaviour of complete solutions)
Let (2.1)-(2.12) be given.

1. *In step 1.3 no solution (x^k, θ^k) will be generated more than once.*

2. *In step 2.3 no solution $h^k(\cdot)$ will be generated more than once.*

3. As soon as in step 2.6 a solution $g^k(\cdot)$ is generated for the second time, the algorithm will terminate.

Proof

1. This has already been proven in Subsection 2.1 (see the additional explanation concerning steps 2.4 and 2.7)

2. Assume that at iteration k, $h^k(\cdot)$ is generated for the second time. This implies that $h^k(\cdot) \in \Delta^{0,k}$. From step 1.3 it then follows that $h^k(0, x^k) \geq 0$, whereas from step 2.3 we know that $h^k(0, x^k) < 0$. This contradiction proves the second part of the theorem.

3. Suppose that at iteration k, $g^k(\cdot)$ is generated for the second time. Consequently, $g^k(\cdot) \in \Delta^k$. From step 1.3 it then follows that $-g^k(0, x^k) + \theta^k \leq 0$. According to Subsection 2.1 (see the additional explanation concerning step 2.7), $UB^k - LB^k > \epsilon_0^k$ would imply that $-g^k(0, x^k) + \theta^k > 0$. As a result, $UB^k - LB^k \leq \epsilon_0^k$, hence the procedure terminates at iteration k.

□

If the algorithm which is used to solve the relaxed master programmes is of such a nature that its outcome (x^k, θ^k) is completely determined by the cut sets $\Delta^k \subseteq \Delta$ and $\Delta^{0,k} \subseteq \Delta^0$, then the second and third part of the proof of the theorem could be simplified; if a dual solution $g^k(\cdot)$ $(h^k(\cdot))$ would be generated more than once, then the same cut sets would appear more than once, implying that the same solution (x^k, θ^k) would be obtained more than once, which contradicts the first result of the theorem. However, assuming such a deterministic kind of algorithm in step 1.3 is unnecessarily restrictive. On the one hand, the algorithm may be stochastic, in the sense that all kinds of decisions which have to be taken during the course of the algorithm, are randomized. On the other hand, the inaccuracy parameter ϵ_1^k may vary between iterations, possibly leading to different solutions as well. For these two reasons the proof in Theorem 2.6 is more general than the one which is discussed in this paragraph.

The requirement $\epsilon_0^k \geq \epsilon_1^k + \epsilon_2^k + \epsilon_3^k$ appears to be crucial if the procedure is to be safeguarded against cycling. Loosely speaking, it means that, at any time, the overall inaccuracy parameter ϵ_0^k should not be chosen less than the sum of the inaccuracies which are allowed in the building blocks of the procedure.

Although it is already a great comfort to know that the variable decomposition procedure does not exhibit cyclic behaviour with respect to the "full" solutions (x^k, θ^k), $g^k(\cdot)$ and $h^k(\cdot)$, it should still be considered a most undesirable situation if the procedure could generate cycles with respect to the "partial" solutions x^k in step 1.3. In such a case, the same (finite) sequence of primal subprogrammes would repeatedly have to be solved in step 2.1. The result of such cyclic behaviour would be that both the upper and the lower bound would not be improved upon by more than the maximal reduction possible in the inaccuracy parameters ϵ_1^k and ϵ_2^k. Fortunately, this kind of cyclic behaviour can be avoided quite easily, as is revealed by the following theorem. For this purpose we define $\max\{\epsilon_2^k + \epsilon_3^k\}$ to be the largest value of $\epsilon_2^k + \epsilon_3^k$ that has been used so far.

Theorem 2.7 (Non-cyclic behaviour of partial solutions)
Let (2.1)–(2.12) be given. Suppose that at each iteration k that passes through steps 2.5–2.9, ϵ_0^k is chosen such that $\epsilon_0^k \geq \epsilon_1^k + \max\{\epsilon_2^k + \epsilon_3^k\}$, then the procedure terminates as soon as a solution x^k is generated for the second time.

Proof Suppose $x^\ell = x^k$ for some $k > \ell$. From the additional explanation in Subsection 2.1 concerning step 2.4, it follows that $\varphi(\mathcal{P}(x^\ell)) = \varphi(\mathcal{P}(x^k)) > -\infty$. From step 1.3 it then follows that $-g^\ell(0, x^k) + \theta^k \leq 0$, implying that

$$\theta^k \leq g^\ell(0, x^k) = g^\ell(0, x^\ell) \leq g^\ell(0, x^\ell) + \epsilon_0^k - \epsilon_1^k - (\epsilon_2^\ell + \epsilon_3^\ell)$$

From steps 2.5 and 2.6 it also follows that

$$f(x^\ell, y^\ell) + \epsilon_2^\ell \geq \varphi(\mathcal{P}(x^\ell)) = \varphi(\mathcal{D}(x^\ell)) \geq g^\ell(0, x^\ell) - \epsilon_3^\ell$$

These two inequalities imply that

$$UB^k - LB^k \leq UB^k - LB^\ell \leq (\theta^k + \epsilon_1^k) - f(x^\ell, y^\ell) \leq$$
$$(\theta^k + \epsilon_1^k) - (g^\ell(0, x^\ell) - \epsilon_2^\ell - \epsilon_3^\ell) \leq \epsilon_0^k$$

Consequently, the algorithm terminates at iteration k.

□

So, the procedure will not exhibit cyclic behaviour with respect to the partial solutions x^k either. As revealed by Theorem 2.7, the inaccuracies in steps 2.5–2.9 are the most critical ones, in the sense that there is no gain in demanding a specific value for $\epsilon_2^k + \epsilon_3^k$, once a larger value for this quantity has ever been allowed. Therefore, the sequence $(\epsilon_2^k + \epsilon_3^k)_{\mathbb{N}}$ could just as well be chosen monotonically non-decreasing, in which case the condition boils down to the one stated in step 3.1.

From the previous two theorems it follows that, even in its full generality, Variable Decomposition will not exhibit cyclic behaviour. To enforce finite convergence of the procedure, severe conditions have to be imposed on problem structure. Below we will formulate two such conditions.

Theorem 2.8 (Finite convergence – primal version)
Let (2.1)–(2.12) be given. Suppose that at each iteration k that passes through steps 2.5–2.9, ϵ_0^k is chosen such that $\epsilon_0^k \geq \epsilon_1^k + \max\{\epsilon_2^k + \epsilon_3^k\}$. If, in addition, all $x^k \in X$ which are generated in step 1.3, belong to some finite subset *of X, then the procedure terminates within a finite number of steps.*

Proof The result immediately follows from the fact that under the aforementioned condition on x^k, the procedure cannot generate infinitely many different values for x^k. As a consequence, a previously generated value for x^k will be reobtained within a finite number of steps. From Theorem 2.7 it then follows that the procedure will terminate within a finite number of steps as well.

□

This theorem proves its usefulness, for instance, in the common situation where the set X is a bounded set of vectors with integer-valued components.

Theorem 2.9 (Finite convergence – dual version)
Let (2.1)–(2.12) be given. Suppose that all $g^k(\cdot) \in \Delta$ and $h^k(\cdot) \in \Delta^0$ which are generated in steps 2.6 and 2.3 respectively, belong to finite subsets *of Δ and Δ^0 respectively, then the procedure terminates within a finite number of steps.*

Proof This proof is similar to the previous one, except that all occurrences of x^k should be replaced by $g^k(\cdot)$ and $h^k(\cdot)$, and Theorem 2.7 should be replaced by Theorem 2.6.

□

This theorem applies, for instance, if the primal subprogrammes are *linear programmes*. In that case attention can be restricted to dual solutions $g^k(\cdot)$ and $h^k(\cdot)$ in steps 2.6 and 2.3 respectively, which correspond to the extreme points and rays of a certain polyhedron.

2.5 Asymptotic convergence

In this subsection, the asymptotic convergence properties of the variable decomposition procedure will be discussed. In Theorem 2.10 sufficient conditions will be stated which imply asymptotic convergence. Apart from some mathematical technicalities, these conditions boil down to *closedness* of the point-to-set map which largely describes the construction of the sequence of intermediate primal solutions x^k in the master programme phase, and *stability* of the family of primal subprogrammes $\mathcal{P}(\cdot)$ (cf. Part I). After Theorem 2.10, the necessity of the two conditions is brought into question. Although neither condition is proven to be necessary from a strictly mathematical point of view, it is argued that, except for some pathological cases maybe, asymptotic convergence is unlikely to occur without them. More specifically, it will be argued that the stability condition is necessary, in that it is inherent to the notion of asymptotic convergence in Variable Decomposition, whereas the closedness condition is necessary in that is inherent to the asymptotic convergence property in general. Unfortunately, the property of *stability* significantly narrows down the class of problems for which a convergent variable decomposition scheme can be guaranteed. As revealed in Part I, stability cannot be expected, unless severe conditions are imposed on the original problem $\mathcal{P}$. The property of *closedness* is further analysed in Theorem 2.11. It is demonstrated that in all computationally relevant cases (i.e. the dual space being restricted to *finitely representable* dual solutions; see Subsection 4.2 of Part II), closedness is preserved if these functions satisfy some *continuity* conditions, and if they can be represented by elements of some (finitely dimensional) *compact* set. This subsection is concluded by Theorem 2.12, which is a result on the asymptotic behaviour of the variable decomposition procedure, in case the procedure does not terminate within a finite number of steps and is unable to come up with an infinite sequence of *feasible* solutions (x^k, y^k). In such a case, the procedure will not generate value cuts from a certain iteration onwards. Under similar conditions on the (common) feasible set of the dual subprogrammes as mentioned in Theorem 2.11, the procedure is proven to at least converge to a feasible value for the x-variables.

It should be noted that (parts of) the results in this subsection hold under slightly weaker or slightly modified conditions as well. However, we resisted the temptation to bury ourselves into further technicalities. On the contrary, we aimed at coherent and useful results. The results in this subsection are new. For the special case of partially convex programmes, Theorem 2.10 boils down to Hogan's Theorem 3.1.1 (cf. [Hogan, 1973]).

By definition, an asymptotic analysis requires the occurrence of *infinite* sequences of solutions. However, the procedure may terminate within a *finite* number of steps. Therefore, if reference is made to asymptotic behaviour, it is tacitly assumed that the algorithm does not terminate prematurely. Let FS_X denote the set of feasible x-values, hence

$$FS_X = \{x \in X \mid \varphi(\mathcal{P}(x)) > -\infty\} \tag{2.20}$$

Note that (2.3) implies that $x \in FS_X$ if and only if there is a $y \in Y$ such that $(x, y) \in U$ and $G(x, y) \diamond 0$. Furthermore, the index set of iterations that pass through steps 2.5–2.9 is denoted by I. So,

$$I = \{k \in \mathbb{N} \mid x^k \in FS_X\} \tag{2.21}$$

Finally, some point-to-set maps are defined which largely describe the construction of the sequence of intermediate primal solutions x^k in the master programme phase. On close inspection, there are two of them. The first map describes the generation of x^{k+1} from x^k, in case a value cut is added to the current relaxed master programme. This process can be defined by means of the composed point-to-set map $\alpha(\beta(\cdot))$, where

$$\begin{aligned}
&\alpha(\cdot) : \Delta \to\to X \times \mathbb{R}, \text{ with} \\
&\alpha(g(\cdot)) = \{(x, \theta) \in X \times \mathbb{R} \mid -g(0, x) + \theta \leq 0\} \\
&\beta(\cdot) : FS_X \times \mathbb{R}_+ \to\to \Delta, \text{ with} \\
&\beta(x, \epsilon_3) = \{g(\cdot) \in \Delta \mid g(\cdot) \text{ is } \epsilon_3\text{-optimal for } \mathcal{D}(x)\}
\end{aligned} \tag{2.22}$$

The second map is defined similarly. It describes the generation of x^{k+1} from x^k, in case a feasibility cut is added to the current relaxed master programme. This process can be defined by means of the composed point-to-set map $\alpha^0(\beta^0(\cdot))$, where

$$\begin{aligned} &\alpha^0(\cdot) : \Delta^0 \to\!\!\to X, \text{ with} \\ &\alpha^0(h(\cdot)) = \{x \in X \mid -h(0,x) \le 0\} \\ \\ &\beta^0(\cdot) : X \setminus FS_X \to\!\!\to \Delta^0, \text{ with} \\ &\beta^0(x) = \{h(\cdot) \in \Delta^0 \mid -h(0,x) > 0\} \end{aligned} \tag{2.23}$$

In studying the asymptotic convergence of the variable decomposition procedure, we will distinguish between two cases, viz. $|I| = +\infty$ on the one hand, and $|I| < +\infty$ on the other. Theorem 2.10 and 2.11 apply to the former case, Theorem 2.12 to the latter.

Case 1. $|I| = +\infty$

It should be noted that in the following two theorems, all sequences and all lim sup's are no longer taken with respect to the index set $\mathbb{N}$, but with respect to the index set I instead. Note that the assumption on I only makes sense if the initiation phase has been passed through succesfully, and if the **STOP**-statement in step 1.2 will never be executed. Therefore, we will assume that $\mathcal{P}$ is a regular programme.

Theorem 2.10 (Asymptotic convergence) *Let (2.1)–(2.12) and (2.20)–(2.23) be given. Suppose that $\varphi(\mathcal{P}) \in \mathbb{R}$, that $U \cap (X \times Y)$ is a compact set, and that $f(\cdot)$ is upper semi-continuous with respect to $U \cap (X \times Y)$. Furthermore, suppose that $G_i(\cdot)$ is lower semi-continuous with respect to $U \cap (X \times Y)$ in case $\diamond_i \in \{\le\}$ and continuous with respect to $U \cap (X \times Y)$ in case $\diamond_i \in \{=\}$, and that $\varphi(\mathcal{P}(\cdot))$ is lower semi-continuous with respect to FS_X. Moreover, let for all $k \in I$, ϵ_i^k be chosen from a compact set $E_i \subseteq \mathbb{R}_+$ $(i = 1, \ldots, 3)$. Finally, assume that the point-to-set map $\alpha(\beta(\cdot))$ is closed with respect to $FS_X \times E_3$. Then the variable decomposition procedure converges, in the sense that*

1. *If $|I| = +\infty$, then every accumulation point of the sequence $(x^k)_I$ is a* $\limsup(\epsilon_1^k + \epsilon_3^k)$*-optimal solution for (2.5).*

2. *If* $|I| = +\infty$, *then every accumulation point of* $(x^k, y^k)_I$ *is a* $\limsup(\epsilon_1^k + \epsilon_2^k + \epsilon_3^k)$*-optimal solution for* $\mathcal{P}$.

3. *If* $|I| = +\infty$, *then every accumulation point of* $(x^{inc,k})_I$ *is a* $\limsup(\epsilon_1^k + \epsilon_2^k + \epsilon_3^k)$*-optimal solution for (2.5), and every accumulation point of* $(x^{inc,k}, y^{inc,k})_I$ *is a* $\limsup(\epsilon_1^k + \epsilon_2^k + \epsilon_3^k)$*-optimal solution for* $\mathcal{P}$.

4. *If* $|I| = +\infty$, *then* $0 \leq \lim(UB^k - LB^k) \leq \limsup(\epsilon_1^k + \epsilon_2^k + \epsilon_3^k)$.

5. *Let* $\max\{k \in I \mid \varphi(\mathcal{P}(x^k)) = -\infty\} \leq \ell_0 \in \mathbb{N}$. *Furthermore, suppose that there is an* $\eta > 0$ *such that* $\forall k \geq \ell_0 : \ \epsilon_0^k \geq \epsilon_1^k + \epsilon_2^k + \epsilon_3^k + \eta$, *then the algorithm terminates within a finite number of steps, say* k_0, *and* $(x^{inc,k_0}, y^{inc,k_0})$ *is a* $(UB^{k_0} - LB^{k_0})$*-optimal solution for* $\mathcal{P}$, *which meets the required accuracy of* $\epsilon_0^{k_0}$.

Proof

1. Let $x^\infty \in FS_X$ be any accumulation point of $(x^k)_I$. Note that such a point exists, and is necessarily an element of FS_X, because $\{x^k \mid k \in I\} \subseteq FS_X$ and FS_X is a compact set. From Theorem 2.2 we know that

$$\forall k \in I: \ \theta^k + \epsilon_1^k \geq \varphi(\mathcal{VD}(\Delta^k, \Delta^{0,k})) \geq \varphi(\mathcal{P}), \text{ so}$$
$$\forall k \in I: \ \theta^k \geq \varphi(\mathcal{P}) - \epsilon_1^k \geq \varphi(\mathcal{P}) - \sup(E_1)$$

On the other hand, if ℓ is the smallest index for which $UB^\ell = UB^k$, then we also know that

$$\forall k \in I: \ \theta^k \leq \varphi(\mathcal{VD}(\Delta^k, \Delta^{0,k})) \leq \varphi(\mathcal{VD}(\Delta^\ell, \Delta^{0,\ell})) \leq$$
$$\theta^\ell + \epsilon_1^\ell = UB^\ell \leq UB^1$$

Due to the fact that $UB^1 \in \mathbb{R}$, $\varphi(\mathcal{P}) \in \mathbb{R}$ and E_1 is a compact set, it follows that $\forall k \in I: \theta^k$ is chosen from some compact set, say Θ. Consequently, there is a subsequence $(p(k))_I$ of I such that

$$\lim(x^{p(k)}, \theta^{p(k)}, \epsilon_1^{p(k)}, \epsilon_3^{p(k)}) = (x^\infty, \theta^\infty, \epsilon_1^\infty, \epsilon_3^\infty)$$

for some $(\theta^\infty, \epsilon_1^\infty, \epsilon_3^\infty) \in \mathbb{R} \times E_1 \times E_3$. We now have that

$$\lim(x^{p(k)}, \theta^{p(k)}) = \lim(x^{p(k+1)}, \theta^{p(k+1)}) = (x^\infty, \theta^\infty), \text{ and}$$
$$(x^{p(k+1)}, \theta^{p(k+1)}) \in \alpha(\beta(x^{p(k)}, \epsilon_3^{p(k)}))$$

These two relations, combined with the closedness of $\alpha(\beta(\cdot))$ with respect to $FS_X \times E_3$, imply that

$$(x^\infty, \theta^\infty) \in \alpha(\beta(x^\infty, \epsilon_3^\infty))$$

Consequently, there is an ϵ_3^∞-optimal solution $g^\infty(\cdot)$ of $\mathcal{D}(x^\infty)$ for which

$$\begin{aligned} &\theta^\infty - \epsilon_3^\infty \le g^\infty(0, x^\infty) - \epsilon_3^\infty \le \\ &\varphi(\mathcal{D}(x^\infty)) = \varphi(\mathcal{P}(x^\infty)) \le \varphi(\mathcal{P}) \end{aligned} \tag{2.24}$$

On the other hand, it is obvious that

$$\forall k \in I: \theta^{p(k)} + \epsilon_1^{p(k)} \ge \varphi(\mathcal{P}) \Rightarrow \theta^\infty + \epsilon_1^\infty \ge \varphi(\mathcal{P})$$

The latter two relations imply that

$$\varphi(\mathcal{P}) - \epsilon_1^\infty - \epsilon_3^\infty \le \theta^\infty - \epsilon_3^\infty \le \varphi(\mathcal{P}(x^\infty)) \le \varphi(\mathcal{P}) \tag{2.25}$$

This fact proves that x^∞ is a $\limsup(\epsilon_1^k + \epsilon_3^k)$-optimal solution for (2.5), because $x^\infty \in FS_X$ and $\epsilon_1^\infty + \epsilon_3^\infty \le \limsup(\epsilon_1^k + \epsilon_3^k)$.

2. Let (x^∞, y^∞) be any accumulation point of $(x^k, y^k)_I$. Note that such a point exists, and is necessarily a feasible solution for $\mathcal{P}$, because $\{(x^k, y^k) \mid k \in I\}$ is a subset of the feasible set of $\mathcal{P}$, which is compact. Just as in 1., there is a subsequence $(p(k))_I$ of I such that

$$\lim(x^{p(k)}, y^{p(k)}, \theta^{p(k)}, \epsilon_1^{p(k)}, \epsilon_2^{p(k)}, \epsilon_3^{p(k)}) = (x^\infty, y^\infty, \theta^\infty, \epsilon_1^\infty, \epsilon_2^\infty, \epsilon_3^\infty)$$

for some $(\theta^\infty, \epsilon_1^\infty, \epsilon_2^\infty, \epsilon_3^\infty) \in \mathbb{R} \times E_1 \times E_2 \times E_3$. From the upper semi-continuity of $f(\cdot)$, the lower semi-continuity of $\varphi(\mathcal{P}(\cdot))$ and (2.25) it follows that

$$\begin{aligned}
&f(x^\infty, y^\infty) \geq \limsup f(x^{p(k)}, y^{p(k)}) \geq \\
&\limsup \left(\varphi(\mathcal{P}(x^{p(k)})) - \epsilon_2^{p(k)}\right) \geq \\
&\liminf \varphi(\mathcal{P}(x^{p(k)})) - \epsilon_2^\infty \geq \\
&\varphi(\mathcal{P}(x^\infty)) - \epsilon_2^\infty \geq \varphi(\mathcal{P}) - \epsilon_1^\infty - \epsilon_2^\infty - \epsilon_3^\infty \geq \\
&\varphi(\mathcal{P}) - \limsup(\epsilon_1^k + \epsilon_2^k + \epsilon_3^k)
\end{aligned} \tag{2.26}$$

which proves the theorem.

3. Let $(x^{inc,\infty}, y^{inc,\infty})$ be any accumulation point of $(x^{inc,k}, y^{inc,k})_I$. Note again, that such a point exists, and is necessarily a feasible solution for $\mathcal{P}$, because $\{(x^{inc,k}, y^{inc,k}) \mid k \in I\}$ is a subset of the feasible set of $\mathcal{P}$, which is compact. Just as in 1., there is a subsequence $(p(k))_I$ of I such that

$$\begin{aligned}
&\lim(x^{inc,p(k)}, y^{inc,p(k)}, x^{p(k)}, y^{p(k)}, \theta^{p(k)}, \epsilon_1^{p(k)}, \epsilon_2^{p(k)}, \epsilon_3^{p(k)}) = \\
&(x^{inc,\infty}, y^{inc,\infty}, x^\infty, y^\infty, \theta^\infty, \epsilon_1^\infty, \epsilon_2^\infty, \epsilon_3^\infty)
\end{aligned}$$

for some $(x^\infty, y^\infty, \theta^\infty, \epsilon_1^\infty, \epsilon_2^\infty, \epsilon_3^\infty) \in X \times Y \times \mathbb{R} \times E_1 \times E_2 \times E_3$. From (2.26) and the upper semi-continuity of $f(\cdot)$, it follows that

$$\begin{aligned}
&\varphi(\mathcal{P}(x^{inc,\infty})) \geq f(x^{inc,\infty}, y^{inc,\infty}) \geq \\
&\limsup f(x^{inc,p(k)}, y^{inc,p(k)}) \geq \limsup f(x^{p(k)}, y^{p(k)}) \geq \\
&\varphi(\mathcal{P}) - \limsup(\epsilon_1^k + \epsilon_2^k + \epsilon_3^k)
\end{aligned}$$

So, it follows that $x^{inc,\infty}$ is a $\limsup(\epsilon_1^k + \epsilon_2^k + \epsilon_3^k)$-optimal solution for (2.5) as well as that $(x^{inc,\infty}, y^{inc,\infty})$ is a $\limsup(\epsilon_1^k + \epsilon_2^k + \epsilon_3^k)$-optimal solution for $\mathcal{P}$.

4. Let us consider a subsequence as defined as in 3. By construction, $(LB^k)_{\mathbb{N}}$ $((UB^k)_{\mathbb{N}})$ is a monotonically non-decreasing (non-increasing) sequence of lower (upper) bounds, which is bounded from above (below) by $\varphi(\mathcal{P})$. As a result, the sequence of lower (upper) bounds is convergent. For the lower bounds it follows that

$$\begin{aligned}&LB^\infty = \liminf LB^{p(k)} = \liminf f(x^{inc,p(k)}, y^{inc,p(k)}) \geq \\ &\liminf f(x^{p(k)}, y^{p(k)}) \geq \liminf \left(\varphi(\mathcal{P}(x^{p(k)})) - \epsilon_2^{p(k)}\right) = \\ &\liminf \varphi(\mathcal{P}(x^{p(k)})) - \epsilon_2^\infty \geq \varphi(\mathcal{P}(x^\infty)) - \epsilon_2^\infty\end{aligned}$$

For the upper bounds we have (cf. (2.24))

$$\begin{aligned}&UB^\infty = \lim UB^{p(k)} \leq \lim(\theta^{p(k)} + \epsilon_1^{p(k)}) = \\ &\theta^\infty + \epsilon_1^\infty \leq \varphi(\mathcal{P}(x^\infty)) + \epsilon_1^\infty + \epsilon_3^\infty\end{aligned}$$

Combining both results yields

$$\begin{aligned}&0 \leq \lim(UB^k - LB^k) = UB^\infty - LB^\infty \leq \\ &\varphi(\mathcal{P}(x^\infty)) + \epsilon_1^\infty + \epsilon_3^\infty - \varphi(\mathcal{P}(x^\infty)) + \epsilon_2^\infty \leq \\ &\limsup(\epsilon_1^k + \epsilon_2^k + \epsilon_3^k)\end{aligned}$$

5. Suppose the variable decomposition procedure has not terminated before iteration ℓ_0. Because of 4. we know that

$$\exists k_0 \geq \ell_0 : \; UB^{k_0} - LB^{k_0} < \epsilon_1^{k_0} + \epsilon_2^{k_0} + \epsilon_3^{k_0} + \eta \leq \epsilon_0^{k_0}$$

Hence, the variable decomposition procedure terminates after k_0 iterations, and $(x^{inc,k_0}, y^{inc,k_0})$ is recognized as a $(UB^{k_0} - LB^{k_0})$-optimal solution for $\mathcal{P}$, which meets the required accuracy of $\epsilon_0^{k_0}$ □

What about the necessity of these conditions? As far as the *closedness* of $\alpha(\beta(\cdot))$ is concerned, we can be brief. Ever since Zangwill introduced his general framework for analysing the asymptotic behaviour of iterative procedures (cf. [Zangwill, 1969]), closedness has generally been accepted as *the* fundamental concept for asymptotic convergence. In order to show that *stability* becomes an essential requirement if the asymptotic convergence analysis is specifically applied to a variable decomposition scheme, it is crucial to understand that it is, in fact, the lower semi-continuity of the objective function which is an essential prerequisite. To justify this assertion, suppose that the function $z(\cdot)$ which is defined as

$$z(x) = \begin{cases} x & \text{if } x \in [0, \frac{1}{2}) \\ 1\frac{1}{2} - x & \text{if } x \in [\frac{1}{2}, 1] \end{cases}$$

has to be maximized over the interval $[0, 1]$. Suppose we have an algorithm that generates a sequence of solutions $(x^k)_{\mathbb{N}}$, where $x^k = \frac{1}{2} - \frac{1}{k+1}$. Obviously, this algorithm converges asymptotically to the optimal solution $x^* = \frac{1}{2}$. However, the algorithm does not converge with respect to function values, because $\lim f(x^k) = \frac{1}{2} < 1 = f(x^*)$. The implication of this phenomenon is twofold. First, any solution which is generated by the algorithm, is valued at least 50% less than the optimal solution is, and secondly, any stopping rule of the type "upper bound − lower bound $\leq \epsilon$" (as in step 3.2) would actually be unable to ever stop the iterative procedure if ϵ is chosen too small. This example clearly shows that the notion of asymptotic convergence is meaningless if the objective function is allowed to suddenly jump upwards in the transition from $\lim z(x^k)$ to $z(x^*)$. A condition that precisely excludes such behaviour, is that of lower semi-continuity of $z(\cdot)$.

Let us now return to the asymptotic convergence analysis of the variable decomposition procedure. On close inspection, Variable Decomposition aims at solving (2.5) instead of $\mathcal{P}$. As a result of the previous discussion, an asymptotic convergence analysis for the variable decomposition procedure will only be meaningful, if the objective function of (2.5) is lower semi-continuous. This explains the stability condition on $\mathcal{P}(\cdot)$ in Theorem 2.10. Note that, actually, stability means continuity of $\varphi(\mathcal{P}(\cdot))$, whereas we only require lower semi-continuity. However, the difference between the two notions is only minor; in Section 2 of Part I it is shown that the upper semi-continuity of the objective function in (2.5) can be enforced by imposing only weak conditions on the original problem and, in fact, the upper semi-continuity of $\varphi(\mathcal{P}(\cdot))$ here, follows from the assumptions that are mentioned in the theorem. Therefore, the notions of lower semi-continuity and stability coincide in this case. Finally, the compactness conditions on the feasible set of $\mathcal{P}$ and of (2.5) are essential to make sure that any infinite sequence in either set has at least one accumulation point, which necessarily belongs to the same set. To conlude, it seems that, although the conditions in Theorem 2.10 are not proven to be necessary from a strictly mathematical point of view, similar results cannot be expected to hold without them.

The fact that in 3. $x^{inc,\infty}$ can only be proven to be $\limsup(\epsilon_1^k + \epsilon_2^k + \epsilon_3^k)$-optimal (instead of $\limsup(\epsilon_1^k + \epsilon_3^k)$-optimal), is due to the fact that $(x^{inc,k})_{\mathbb{N}}$ depends on $(x^k, y^k)_{\mathbb{N}}$, hence on $(\epsilon_2^k)_{\mathbb{N}}$.

Finally, the condition in 5. concerning the choice of the inaccuracy parameters, is also intuitively justifiable. It states that the required bound on the inaccuracy of the overall solution during a certain iteration, should always be strictly greater than the sum of the inaccuracies which have already been allowed during the same iteration. In addition, this difference is not allowed to gradually vanish over iterations.

This is as far as Theorem 2.10 is concerned. One important special case, and in more practical applications the only meaningful one, is when (near-)optimal solutions of the dual subprogrammes can be described by means of a *finite* number of parameters (cf. Subsection 4.2 of Part II). For instance, if the primal subprogrammes $\mathcal{P}(\cdot)$ satisfy certain convexity conditions, then the dual solution space may be restricted to the finitely representable affine functions (Lagrangean duality; see [Geoffrion, 1972^b] and Subsection 4.1 of Part II). In the next theorem we will show that in such "finitely dimensional cases", the closedness condition on the point-to-set map $\alpha(\beta(\cdot))$ is satisfied, if, additionally, the dual solutions involved satisfy some *continuity* conditions and can be represented by means of elements of some (finitely dimensional) *compact* set.

Theorem 2.11 (Closedness of $\alpha(\beta(\cdot))$) *Let (2.1)–(2.12), (2.20)–(2.23) be given. Suppose that X is closed and $\varphi(\mathcal{P}(\cdot))$ is upper semi-continuous with respect to FS_X. In addition, let there be a non-empty and compact set $T \subseteq \mathbb{R}^\tau$ ($\tau \in \mathbb{N}$) and a function $w(\cdot) : T \times \mathbb{R}^{m+n_1} \to \mathbb{R} \cup \{\pm\infty\}$ that is continuous with respect to $T \times \{(r,x)\}$ $\forall (r,x) \in RHS$ as well as with respect to $T \times \{0\} \times FS_X$ as well as upper semi-continuous with respect to $T \times \{0\} \times X$, and for which*

$$\begin{aligned} &\forall (x,\epsilon_3) \in FS_X \times E_3 : \ [(z,\theta) \in \alpha(\beta(x,\epsilon_3)) \Rightarrow \\ &\exists t \in T : \ w(t,\cdot,\cdot) \in \beta(x,\epsilon_3) \wedge (z,\theta) \in \alpha(w(t,\cdot,\cdot))] \end{aligned} \tag{2.27}$$

Then $\alpha(\beta(\cdot))$ is closed with respect to $FS_X \times E_3$.

Proof Let $(x^k, \epsilon_3^k)_{\mathbb{N}}$ be a sequence in $FS_X \times E_3$ which converges to $(x^\infty, \epsilon_3^\infty) \in FS_X \times E_3$, and let $(z^k, \theta^k)_{\mathbb{N}}$ be a sequence converging to $(z^\infty, \theta^\infty)$, where

$$\forall k \in \mathbb{N} : \ (z^k, \theta^k) \in \alpha(\beta(x^k, \epsilon_3^k)) \tag{2.28}$$

We have to prove that $(z^\infty, \theta^\infty) \in \alpha(\beta(x^\infty, \epsilon_3^\infty))$. From (2.27) and (2.28) it follows that

$$\forall k \in \mathbb{N} \ \exists t^k \in T : \ -w(t^k, 0, z^k) + \theta^k \leq 0$$

where $w(t^k, \cdot, \cdot)$ is an ϵ_3^k-optimal solution for $\mathcal{D}(x^k)$. T is compact, so without loss of generality we may assume that $\lim t^k = t^\infty \in T$. From the upper semi-continuity of $w(\cdot)$ with respect to $T \times \{0\} \times X$ it follows that

$$-w(t^\infty, 0, z^\infty) + \theta^\infty \leq 0 \tag{2.29}$$

and from the continuity of $w(\cdot)$ with respect to $T \times \{(r,x)\}$ $\forall (r,x) \in RHS$ it follows that

$$w(t^{\infty}, \cdot, \cdot) \in \Delta \tag{2.30}$$

Recall that $w(t^k, \cdot, \cdot)$ is an ϵ_3^k-optimal solution for $\mathcal{D}(x^k)$, so

$$w(t^k, 0, x^k) \leq \varphi(\mathcal{D}(x^k)) + \epsilon_3^k = \varphi(\mathcal{P}(x^k)) + \epsilon_3^k$$

In addition, $w(\cdot)$ is continuous with respect to $T \times \{0\} \times FS_X$ and $\varphi(\mathcal{P}(\cdot))$ is upper semi-continuous with respect to FS_X, so

$$\begin{aligned} & w(t^{\infty}, 0, x^{\infty}) = \lim w(t^k, 0, x^k) \leq \\ & \limsup \left(\varphi(\mathcal{P}(x^k)) + \epsilon_3^k\right) \leq \varphi(\mathcal{P}(x^{\infty})) + \epsilon_3^{\infty} = \\ & \varphi(\mathcal{D}(x^{\infty})) + \epsilon_3^{\infty} \end{aligned} \tag{2.31}$$

From (2.30) and (2.31) it follows that $w(t^{\infty}, \cdot, \cdot)$ is an ϵ_3^{∞}-optimal solution for $\mathcal{D}(x^{\infty})$. This, on its turn, combined with (2.29), implies that $(z^{\infty}, \theta^{\infty}) \in \alpha(\beta(x^{\infty}, \epsilon_3^{\infty}))$.

□

Note that (2.27) is a condition on the problem $\mathcal{P}$, rather than a condition on the solution procedure. It states that the *existence* of finitely representable dual solutions is the issue. It does not state that it is necessary to actually *generate* such solutions during the iterative process, although in any practical situation one will.

Case 2. $|I| < +\infty$

As we are only interested in the limiting behaviour of the variable decomposition procedure, we will assume that the procedure will not terminate within a finite number of steps. This assumption, combined with the supposition that $|I| < +\infty$, implies that from a certain iteration onwards, the variable decomposition procedure will no longer generate values for the x-variables that can be supplemented by values for the y-variables, such that feasible solutions for $\mathcal{P}$ result. Another implication is, that from the same iteration onwards, the procedure will no longer be able to improve on the lower bound LB^k.

In the next theorem, we will state a convergence result which seems to complement a synthesis of the previous two. It will be shown that under appropriate conditions, any accumulation point of the sequence of solutions $(x^k)_{\mathbb{N}\setminus I}$ is *admissible* for $\mathcal{P}$, in the sense that each one of them can be supplemented by a value for the y-variables, such that a feasible solution for $\mathcal{P}$ results. Loosely speaking, this result means that, although inadmissible solutions might be generated throughout the algorithm, an admissible solution is at least approached as closely as desired. As in Theorem 2.11, we have to assume that finitely representable, continuous dual solutions *exist*. Unlike Theorem 2.11, however, we have to assume that such dual solutions are actually *generated* in step 2.3 of the algorithm.

Theorem 2.12 (Asymptotic feasibility) *Let (2.1)-(2.12), (2.20)-(2.23) be given. Suppose that X is compact and $\varphi(\mathcal{P}) \in \mathbb{R}$. Furthermore, suppose that $|I| < +\infty$ and assume that the variable decomposition procedure does not terminate within a finite number of steps. In addition, let there be a non-empty and compact set $T \subseteq \mathbb{R}^\tau$ ($\tau \in \mathbb{N}$) and a function $w(\cdot) : T \times \mathbb{R}^{m+n_1} \to \mathbb{R} \cup \{\pm\infty\}$ that is continuous with respect to $T \times \{0\} \times X$, and for which*

$$\begin{aligned} &\forall x \in X \setminus FS_X : \ [z \in \alpha^0(\beta^0(x)) \Rightarrow \\ &\exists t \in T : \ w(t,\cdot,\cdot) \in \beta^0(x) \wedge z \in \alpha^0(w(t,\cdot,\cdot))] \end{aligned} \tag{2.32}$$

Finally, assume that $(\epsilon^k)_{\mathbb{N}\setminus I}$ is a sequence of non-negative values, converging to 0, and that at each iteration $k \in \mathbb{N} \setminus I$, a function $w(t^k,\cdot,\cdot) \in \beta^0(x^k)$ is generated in step 2.3, where t^k is an ϵ^k-optimal solution for

$$\begin{aligned} \min_t \quad & w(t, 0, x^k) \\ s.t. \quad & t \in T \end{aligned} \tag{2.33}$$

in case (2.33) is a regular programme, and where t^k is an optimal solution for (2.33) in case of unboundedness. Then each accumulation point x^∞ of $(x^k)_{\mathbb{N}\setminus I}$ – and there is at least one – is an admissible solution for $\mathcal{P}$, i.e. $x^\infty \in FS_X$.

Proof Let $x^\infty \in X$ be any accumulation point of $(x^k)_{\mathbb{N}\setminus I}$. Note that there is at least one such point and that all such points are in X, because $(x^k)_{\mathbb{N}\setminus I}$ is an infinite sequence of points belonging to the compact set X. The compactness of T then implies that there is a $t^\infty \in T$, and a subsequence $(p(k))_{\mathbb{N}\setminus I}$ of $\mathbb{N} \setminus I$, such that

$$\lim(t^{p(k)}, x^{p(k)}, \epsilon^{p(k)}) = (t^\infty, x^\infty, 0)$$

From step 2.3 it follows that

$$x^{p(k+1)} \in \alpha^0(w(t^k, \cdot, \cdot))$$

which, together with the continuity assumption on $w(\cdot)$, implies that

$$w(t^\infty, 0, x^\infty) = \lim w(t^{p(k)}, 0, x^{p(k+1)}) \geq 0$$

Now, let $t \in T$ be arbitrarily chosen. We know that $t^{p(k)}$ is an $\epsilon^{p(k)}$-optimal solution for (2.33) in case (2.33) is regular, and that $t^{p(k)}$ is optimal in case of unboundedness. Therefore,

$$\begin{aligned} & w(t, 0, x^\infty) = \lim w(t, 0, x^{p(k)}) \geq \\ & \lim \left(w(t^{p(k)}, 0, x^{p(k)}) - \epsilon^{p(k)} \right) = w(t^\infty, 0, x^\infty) \geq 0 \end{aligned}$$

This result, combined with (2.32), implies that $x^\infty \in FS_X$.

□

One application of Theorem 2.12 is when the original programme $\mathcal{P}$ is a *convex programme*. In that case, the feasibility cuts $-h(0,x) \leq 0$ can always be chosen to be *affine*, i.e. of the form $t^T x + t_0 \leq 0$, if some mild conditions on the problem data X, Y, U and $G(\cdot)$ are met. In that case, the feasibility cuts are actually *separating hyperplanes*. Such cuts obviously obey the continuity requirement that is mentioned in the theorem; the compactness condition on T, however, still has to be enforced. In the special case of *Linear Programming*, this latter condition too, is met. In that case, the only functions $h(\cdot)$ in Δ^0 which have to be considered, are the ones which relate to the extreme rays of a certain polyhedral cone. These extreme rays, in their turn, can be normalized, thereby satisfying the compactness condition on T; (2.33) then states that the "(almost) most violated" normalized extreme ray should be generated. In the case of Linear Programming, however, a much stronger property holds, viz. *finite* convergence (cf. Theorem 2.9).

The reader may have observed a slight discrepancy between the conditions mentioned in Theorems 2.11 and 2.12. In the former theorem the *existence* of finitely representable dual solutions is required, whereas in the latter theorem, such solutions actually have to be *generated*. An explication for this phenomenon possibly stems from the fact that in the latter theorem, the notion of closedness seems to be useless. Closedness of $\alpha^0(\beta^0(\cdot))$ would necessarily imply that only a *finite* number of feasibility cuts would be generated during the iterative process, implying that there is nothing left to prove in Theorem 2.12. To justify this assertion, let us assume that the sequences $(x^k)_{\mathbb{N}\setminus I}$ and $(h^k(\cdot))_{\mathbb{N}\setminus I}$ are infinite. Consider a subsequence $(x^{p(k)})_{\mathbb{N}\setminus I}$ of $(x^k)_{\mathbb{N}\setminus I}$ which converges to $x^\infty \in X$. Note that the existence of such a converging subsequence is guaranteed by the compactness condition on X. Consequently, $\lim x^{p(k)} = \lim x^{p(k+1)} = x^\infty$ and $x^{p(k+1)} \in \alpha^0(\beta^0(x^{p(k)}))$. Closedness of this point-to-set map would imply that a dual solution $h^\infty(\cdot) \in \beta^0(x^\infty)$ existed, for which $-h^\infty(0,x^\infty) \leq 0$, or, to say it in other words, for which $h^\infty(\cdot) \notin \beta^0(x^\infty)$. This is an obvious contradiction, implying that the aforementioned sequences are actually finite.

To summarize, in the absence of a useful notion of closedness, we have to rely on what actually happens during the execution of the iterative procedure. This may explain why in Theorem 2.12 we cannot confine ourselves to the *existence* of finitely representable, continuous dual solutions, but why we actually have to *generate* them.

2.6 The special case of Mixed-Integer Non-Linear Programming with an underlying convex structure

In this subsection it will be demonstrated that the variable decomposition procedure applied to a mixed-integer non-linear programme with an underlying convex structure, amounts to alternately solving mixed-integer *linear* programmes and *ordinary* non-linear *convex* programmes; in other words, the integrality requirements are separated from the non-linearities.

Consider such a mixed-integer non-linear programme, which, apart from the integrality requirements, has a convex structure. In terms of (2.1) this means that X contains integrality requirements on its variables, that Y and $U \cap (\widetilde{X} \times Y)$ are convex sets for some convex set $\widetilde{X} \supseteq X$, that $-f(\cdot)$ and $G(\cdot)$ are convex functions on $U \cap (\widetilde{X} \times Y)$ and that $\diamond \in \{\leq\}^m$. In that case, the primal subprogrammes (2.4) are *convex programming problems*. Under some mild constraint qualifications then, the (common) feasible set of their duals can be restricted to the set of affine functions without invalidating strong duality and the Farkas property (Lagrangean duality; cf. [Geoffrion, 1972^b] and Subsection 4.1 of Part II). In doing so, the following dual subprogrammes are obtained

$$\begin{array}{ll} \min\limits_{u,v,w} & v^T\bar{x} + w \\ \text{s.t.} & u^TG(x,y) + v^Tx + w \geq f(x,y) \quad \forall (x,y) \in U \cap (\widetilde{X} \times Y) \\ & u \geq 0 \end{array}$$

The resulting master programme reads

$$\begin{array}{llll} \max\limits_{\bar{x},\theta} & \theta & & \\ \text{s.t.} & -v^T\bar{x} + \theta & \leq w & \forall (u,v,w) \text{ satisfying } 1.\ u \geq 0, \text{ and} \\ & & & 2.\ u^TG(x,y) + v^Tx + w \geq f(x,y) \\ & & & \forall (x,y) \in U \cap (\widetilde{X} \times Y) \\ & -s^T\bar{x} & \leq t & \forall (r,s,t) \text{ satisfying } 1.\ r \geq 0, \text{ and} \\ & & & 2.\ r^TG(x,y) + s^Tx + t \geq 0 \\ & & & \forall (x,y) \in U \cap (\widetilde{X} \times Y) \\ & \bar{x} \in X & & \end{array} \tag{2.34}$$

This master programme contains a huge number of dominated constraints. Many of them can be eliminated by using the equivalence of the following two statements

- $-v^T\bar{x} + \theta \leq w \quad \forall (u,v,w)$ satisfying 1. $u \geq 0$, and
  2. $u^TG(x,y) + v^Tx + w \geq f(x,y)$
  $\forall (x,y) \in U \cap (\widetilde{X} \times Y)$

- $-v^T\bar{x} + \theta \leq \sup\{f(x,y) - u^TG(x,y) - v^Tx \mid (x,y) \in U \cap (\widetilde{X} \times Y)\} \quad \forall (u,v) : \ u \geq 0$

A similar equivalence holds for the feasibility cuts in (2.34). As a result, (2.34) can be replaced by

$$
\begin{aligned}
\max_{\bar{x},\theta} \quad & \theta \\
\text{s.t.} \quad & -v^T\bar{x} + \theta \le \sup\{f(x,y) - u^T G(x,y) - v^T x \mid \\
& \qquad (x,y) \in U \cap (\widetilde{X} \times Y)\} \;\; \forall (u,v) : u \ge 0 \\
& -s^T\bar{x} \le \sup\{-r^T G(x,y) - s^T x \mid \\
& \qquad (x,y) \in U \cap (\widetilde{X} \times Y)\} \;\; \forall (r,s) : r \ge 0 \\
& \bar{x} \in X
\end{aligned}
\tag{2.35}
$$

Note that both (2.34) and (2.35) are *mixed-integer linear programming problems*. As a result, applying Variable Decomposition to mixed-integer non-linear programming problems with an underlying convex structure, amounts to alternately solving mixed-integer *linear* relaxed master programmes and *ordinary* non-linear *convex* primal subprogrammes. To say it differently, the integrality requirements and the non-linearities are dealt with separately.

Section 3

Constraint Decomposition

In this section, Constraint Decomposition is discussed. Basically, it is a generalization of Dantzig-Wolfe Decomposition (cf. [Dantzig & Wolfe, 1960]) to general mathematical programmes. The idea underlying this approach is also known as Column Generation, Generalized Linear Programming, (Generalized) Dantzig-Wolfe Decomposition, Dual Decomposition and Price Directive Decomposition. In Subsection 3.1 the procedure will be explained. Some comments are made in Subsection 3.2 and the procedure is related to existing literature in Subsection 3.3. Non-cyclic behaviour and finite convergence are the topics of Subsection 3.4. Our discussion is concluded by Subsection 3.5, in which the asymptotic convergence of Constraint Decomposition is analysed.

3.1 Problem manipulations and solution strategy

Consider the following *primal programme*

$$\begin{aligned} \mathcal{P}: \quad \max_{x} \quad & f(x) \\ \text{s.t.} \quad & G(x) \diamond 0 \\ & x \in X \end{aligned} \tag{3.1}$$

Here, $f(\cdot)$ and $G(\cdot)$ are functions from $D \subseteq \mathbb{R}^n$ to $\mathbb{R} \cup \{\pm\infty\}$ and $\mathbb{R}^m$ respectively, $X \subseteq D$ and $\diamond \in \{\leq, =\}^m$ ($m, n \in \mathbb{N}$). More formally,

$$\begin{aligned} & X \subseteq D \subseteq \mathbb{R}^n && ; \quad \diamond \in \{\leq, =\}^m \\ & f(\cdot) : D \to \mathbb{R} \cup \{\pm\infty\} && ; \quad G(\cdot) : D \to \mathbb{R}^m \end{aligned} \tag{3.2}$$

As in Part I and II, $G(x) \diamond r$ is a shorter notation for $G_i(x) \diamond_i r_i$, $i = 1, \ldots, m$, where $G_i(\cdot)$, $\diamond_i$ and r_i are the i-th components of $G(\cdot)$, $\diamond$ and r respectively. For reasons to be explained in Theorem 3.1, we will assume that

$$\forall x \in X : \; f(x) < +\infty \tag{3.3}$$

Note that (3.3) is not a severe condition. The reason why we persist in taking *extended* real-valued functions $f(\cdot)$ into consideration, is to include the possibility of $f(\cdot)$ being the result of an optimization problem itself. As one can see, two types of constraints are distinguished in (3.1), viz. the (in)equality constraints $G(x) \diamond 0$, and the (possibly) more general constraints $x \in X$. The reason for such a distinction originates from the fact that the former type will be treated differently from the latter. Consider for all $\overline{X} \subseteq X$ the following dual pair of programmes:

$$\begin{aligned} \mathcal{P}(\overline{X}): \quad \max_{x} \quad & f(x) \\ \text{s.t.} \quad & G(x) \diamond 0 \\ & x \in \overline{X} \end{aligned} \tag{3.4}$$

$$\mathcal{D}(\overline{X}): \min_{g(\cdot)} \; g(0)$$
$$\text{s.t.} \quad g(G(x)) \geq f(x) \;\; \forall x \in \overline{X} \tag{3.5}$$
$$g(\cdot) \in \Gamma$$

Here, the dual solution space Γ equals

$$\Gamma = \{g(\cdot) : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\} \mid g(r) \leq g(r') \;\; \forall r, r' \in RHS : \; r \diamond r'\} \tag{3.6}$$

where the set of feasible right-hand-sides RHS is defined as

$$RHS = \{r \in \mathbb{R}^m \mid \exists x \in X : \; G(x) \diamond r\} \tag{3.7}$$

From Lemma 5.1 sub 1 of Part II, it follows that strong duality holds between the pair of programmes $\mathcal{P}(\overline{X})$ and $\mathcal{D}(\overline{X})$ if and only if $\varphi(\mathcal{P}(\overline{X})) > -\infty$; here, $\varphi(\cdot)$ denotes the optimal solution value of a given optimization problem, being defined as $-\infty(+\infty)$ if the feasible set of the maximization (minimization) problem in question is empty.

$\mathcal{P}(X)$ and $\mathcal{D}(X)$ are usually referred to as the *primal and dual master programme*. By definition, $\mathcal{P}(X)$ is equivalent to the original programme $\mathcal{P}$. For any $\overline{X} \subseteq X$, $\mathcal{P}(\overline{X})$ is called a *restricted primal master programme*. Considering such programmes only makes sense if one is able to distinguish whether a given subset $\overline{X} \subseteq X$ takes into account the relevant part of the feasible set of $\mathcal{P}$, and if not, one is able to modify $\overline{X}$ in some sensible way. Fortunately, this can be done through the *relaxed dual master programmes* $\mathcal{D}(\overline{X})$; due to the fact that $\mathcal{D}(\overline{X})$ is a *relaxation* of $\mathcal{D}(X)$, it follows that $g(\cdot)$ is a (near-)optimal solution for the latter problem if it is both (near-)optimal for the former and feasible for the latter. As a result, testing optimality can be done through testing dual feasibility. Let us now elaborate on these intuitive ideas.

Theorem 3.1 *Let (3.1)–(3.7) be given.*

1. *Suppose that* $g(\cdot) \in \Gamma$ *and* $\overline{X} \subseteq X$, *then* $g(\cdot)$ *is feasible for* $\mathcal{D}(\overline{X})$ *if and only if* $\sup_x\{f(x) - g(G(x)) \mid x \in \overline{X}\} \leq 0$

2. *Suppose that* $g(\cdot) \in \Gamma$ *and* $g(0) \in \mathbb{R}$, *then* $g(\cdot)$ *is feasible for* $\mathcal{D}(X)$ *if and only if* $\varphi\left(\mathcal{CD}(g(\cdot))\right) \leq g(0)$. *Here,*

$$\begin{aligned} \mathcal{CD}(g(\cdot)):\quad & \max_x \quad f(x) + g(0) - g(G(x)) \\ & s.t. \quad x \in X \end{aligned} \tag{3.8}$$

Proof

1. $g(\cdot)$ is feasible for $\mathcal{D}(\overline{X})$ if and only if $g(\cdot) \in \Gamma$ and $\forall x \in \overline{X}$: $f(x) \leq g(G(x))$. From (3.3) and (3.6) it follows that

$$\forall x \in \overline{X} : \ (f(x) < +\infty \wedge g(G(x)) > -\infty)$$

As a result, $f(x) - g(G(x))$ is well-defined for all $x \in \overline{X}$. This observation implies that $g(\cdot) \in \Gamma$ is feasible for $\mathcal{D}(\overline{X})$ if and only if $\forall x \in \overline{X} : \ f(x) - g(G(x)) \leq 0$. The latter statement is equivalent to $\sup_x\{f(x) - g(G(x)) \mid x \in \overline{X}\} \leq 0$.

2. $g(\cdot) \in \Gamma$ and $g(0) \in \mathbb{R}$ imply that $g(\cdot)$ is feasible for $\mathcal{D}(X)$ if and only if $\sup_x\{f(x)+g(0)-g(G(x)) \mid x \in X\} = g(0)+\sup_x\{f(x) - g(G(x)) \mid x \in X\} \leq g(0)$ (cf. 1.).

□

Problem (3.8) looks familiar; it can be regarded as a "*generalized*" *Lagrangean relaxation* of $\mathcal{P}$, in the sense that functions $g(\cdot)$ other than affine ones, may be used to relax the constraints $G(x)\diamond 0$. As is revealed by the following theorem, these programmes supply *upper bounds* for the original programme $\mathcal{P}$, and because of strong duality, these upper bounds can be as tight as desired.

Theorem 3.2 (Upper bounds) *Let (3.1)–(3.8) be given.*

1. $\varphi(\mathcal{CD}(g(\cdot))) \geq \varphi(\mathcal{P})$ *for every* $g(\cdot) \in \Gamma$ *with* $g(0) \in \mathbb{R}$

2. *If* $\mathcal{P}$ *is regular then* $\varphi(\mathcal{CD}(g(\cdot))) \leq \varphi(\mathcal{P}) + \epsilon$ *for every* ϵ*-optimal solution* $g(\cdot) \in \Gamma$ *of* $\mathcal{D}(X)$

Proof

1. Let $g(\cdot) \in \Gamma$ be such that $g(0) \in \mathbb{R}$. As in the proof of Theorem 3.1, $f(x) + g(0) - g(G(x))$ is well-defined for all $x \in X$, so $\varphi(\mathcal{CD}(g(\cdot)))$ is well-defined as well. If $\varphi(\mathcal{P}) = -\infty$ then the statement obviously holds, so assume that $\varphi(\mathcal{P}) > -\infty$. From (3.6) it follows that $g(G(x)) \leq g(0)$ for all $x \in X$ with $G(x) \diamond 0$. Consequently,

$$\varphi(\mathcal{CD}(g(\cdot))) = \sup_x \{f(x) + g(0) - g(G(x)) \mid x \in X\} \geq$$
$$\sup_x \{f(x) + g(0) - g(G(x)) \mid G(x) \diamond 0,\ x \in X\} \geq$$
$$\sup_x \{f(x) \mid G(x) \diamond 0,\ x \in X\} = \varphi(\mathcal{P})$$

2. Let $\epsilon \geq 0$ be given and suppose that $g(\cdot) \in \Gamma$ is an ϵ-optimal solution for $\mathcal{D}(X)$; the existence of such a solution follows from the regularity assumption on $\mathcal{P}$ and from Lemma 5.1 sub 1 of Part II. Applying Theorem 3.1 sub 2 yields $\varphi(\mathcal{CD}(g(\cdot))) \leq g(0)$ $\leq \varphi(\mathcal{D}(X)) + \epsilon = \varphi(\mathcal{P}) + \epsilon$.

□

Of course, restricted primal master programmes supply *lower bounds* for the original programme $\mathcal{P}$, and these lower bounds can also be as tight as desired. More specifically,

Theorem 3.3 (Lower bounds) *Let (3.1)–(3.8) be given.*

1. $\forall \overline{X} \subseteq X: \ \varphi(\mathcal{P}(\overline{X})) \leq \varphi(\mathcal{P})$

2. *If $\mathcal{P}$ is regular then* $\exists \overline{X} \subseteq X \ \forall \epsilon \geq 0: \ \varphi(\mathcal{P}(\overline{X})) \geq \varphi(\mathcal{P}) - \epsilon$

Proof This is left to the reader.

□

The two aforementioned theorems indicate that the upper and lower bounds can be chosen as tight as desired. The question remains, however, which subset $\overline{X} \subseteq X$ should be taken into consideration. In fact, two questions are involved here, viz. which subset should be taken initially and how should this initial choice be modified in order to end up with a (near-)optimal solution for $\mathcal{P}$. In defining the constraint decomposition procedure, we will need the following definitions; a superscript k is used to denote the value of a variable during iteration k.

UB^k = best upper bound for $\varphi(\mathcal{P})$ over the first k iterations
LB^k = best lower bound for $\varphi(\mathcal{P})$ over the first k iterations
$x^{inc,k}$ = best solution for $\mathcal{P}$ over the first k iterations[1]

CONSTRAINT DECOMPOSITION PROCEDURE – START

0. INITIATION PHASE

BEGIN

0.1 **LET** $k := 1$;

0.2 **LET** $X^k :\subseteq X$ **BE SUCH THAT** $\varphi(\mathcal{P}(X^k)) > -\infty$;

[1]The superscript "inc" comes from "current incumbent"; see [Geoffrion & Marsten, 1972].

0.3 **IF** impossible **THEN STOP** because $\varphi(\mathcal{P}) = -\infty$

0.4 **ELSE BEGIN LET** $UB^0 := +\infty$; **LET** $LB^0 := -\infty$ **END**;

1. MASTER PROGRAMME PHASE

1.1 **SOLVE** $\mathcal{P}(X^k)$; **SOLVE** $\mathcal{D}(X^k)$;

1.2 **IF** $\varphi(\mathcal{P}(X^k)) = +\infty$ **THEN STOP** because $\varphi(\mathcal{P}) = +\infty$

ELSE BEGIN

1.3 **LET** (x^k, ϵ_1^k) **BE SUCH THAT** x^k is an ϵ_1^k-optimal solution for $\mathcal{P}(X^k)$;

1.4 **LET** $(g^k(\cdot), \epsilon_2^k)$ **BE SUCH THAT** $g^k(\cdot)$ is an ϵ_2^k-optimal solution for $\mathcal{D}(X^k)$;

1.5 **LET** $LB^k := \max\{LB^{k-1}, f(x^k)\}$;

1.6 **IF** $LB^k > LB^{k-1}$ **THEN LET** $x^{inc,k} := x^k$

ELSE LET $x^{inc,k} := x^{inc,k-1}$

END;

2. SUBPROGRAMME PHASE

2.1 **SOLVE** $\mathcal{CD}(g^k(\cdot))$;

2.2 **IF** $\varphi(\mathcal{CD}(g^k(\cdot))) = +\infty$ **THEN BEGIN LET** $\epsilon_3^k := 0$;

2.3 **LET** $y^k \in X$ **BE SUCH THAT** $f(y^k) + g^k(0) - g^k(G(y^k)) > g^k(0)$;

2.4 **LET** $X^k \cup \{y^k\} \subseteq: X^{k+1} :\subseteq X$; **LET** $UB^k := UB^{k-1}$

END ELSE BEGIN

2.5 **LET** (y^k, ϵ_3^k) **BE SUCH THAT** y^k is an ϵ_3^k-optimal solution for $\mathcal{CD}(g^k(\cdot))$;

2.6 **LET** $X^k \cup \{y^k\} \subseteq: X^{k+1} :\subseteq X$;

2.7 **LET** $UB^k := \min\{UB^{k-1}, f(y^k) + g^k(0) - g^k(G(y^k)) + \epsilon_3^k\}$

END;

3. OPTIMALITY CHECK

3.1 **LET** ϵ_0^k **BE SUCH THAT** $\epsilon_0^k \geq \epsilon_1^k + \epsilon_2^k + \epsilon_3^k$;

3.2 **IF** $UB^k - LB^k \leq \epsilon_0^k$ **THEN STOP** because $x^{inc,k}$ is a $(UB^k - LB^k)$-optimal solution for $\mathcal{P}$, which meets the required accuracy of ϵ_0^k

3.3 **ELSE BEGIN LET** $k := k + 1$; **GOTO** 1 **END**

END.

CONSTRAINT DECOMPOSITION PROCEDURE – END

As in the variable decomposition procedure of Subsection 2.1, many questions are left unanswered here, such as how to initiate the procedure (step 0.2), how to solve the restricted primal master programmes (steps 1.1 and 1.3), how to obtain dual solutions (steps 1.1 and 1.4), how to solve the subprogrammes (steps 2.1, 2.3 and 2.5), how to extend X^k to X^{k+1} (steps 2.4 and 2.6) and how to specify the inaccuracy parameters ϵ_i^k (steps 1.3, 1.4, 2.5 and 3.1). The reason for this is, again, twofold. On the one hand, general answers do not exist for these questions; the infeasibility at initial iterations, for instance, is a notorious phenomenon, even if $\mathcal{P}$ is an ordinary linear programme. On the other hand, it may well be possible that, depending on the specific problem characteristics involved, several alternatives exist. Dual solutions, for instance, can usually be obtained as a by-product of solving the primal, but it may be possible to obtain them otherwise. To summarize, we have been discussing a *framework* here; applying the procedure to a specific problem(-class) still requires a lot of fine-tuning, which highly determines its performance. Let us close this subsection with some additional remarks.

Step 1.2 $\varphi(\mathcal{P}) \geq \varphi(\mathcal{P}(X^k)) = +\infty$.

Step 1.3 $\mathcal{P}(X^k)$ is a regular programme.

Step 1.5 $f(x^k)$ is a lower bound for $\varphi(\mathcal{P}(X^k))$, hence for $\varphi(\mathcal{P})$.

Step 2.3 From the fact that $g^k(\cdot)$ is a feasible solution for $\mathcal{D}(X^k)$, it follows that $y^k \notin X^k$.

Step 2.4 Due to the fact that $f(y^k) - g^k(G(y^k)) > 0$, it follows that $g^k(\cdot)$ will not be generated in subsequent iterations if $y^k \in X \setminus X^k$ is added to X^{k+1}.

Step 2.5 $\mathcal{CD}(g^k(\cdot))$ is a regular programme, because $\varphi\left(\mathcal{CD}(g^k(\cdot))\right) \geq \varphi(\mathcal{P}) \geq \varphi(\mathcal{P}(X^1)) > -\infty$ for all $g^k(\cdot) \in \Gamma$ with $g(0) \in \mathbb{R}$ (cf. Theorem 3.2).

Step 2.7 $\left(f(y^k) + g^k(0) - g^k(G(y^k)) + \epsilon_3^k\right)$ is an upper bound for $\varphi(\mathcal{P})$, because it is one for $\varphi\left(\mathcal{CD}(g^k(\cdot))\right)$.

Suppose $\epsilon_0^k \geq \epsilon_1^k + \epsilon_2^k + \epsilon_3^k$ and $UB^k - LB^k > \epsilon_0^k$. Then it follows that

$$\begin{aligned}&\left(f(y^k) + g^k(0) - g^k(G(y^k)) + \epsilon_3^k\right) - \left(g^k(0) - \epsilon_1^k - \epsilon_2^k\right) \geq\\ &UB^k - \left(\varphi(\mathcal{D}(X^k)) - \epsilon_1^k\right) = UB^k - \left(\varphi(\mathcal{P}(X^k)) - \epsilon_1^k\right) \geq\\ &UB^k - f(x^k) \geq UB^k - LB^k > \epsilon_0^k\end{aligned}$$

hence $f(y^k) - g^k(G(y^k)) > \epsilon_0^k - \epsilon_1^k - \epsilon_2^k - \epsilon_3^k \geq 0$. The solution $g^k(\cdot)$ is feasible for $\mathcal{D}(X^k)$, so $y^k \notin X^k$. As a result, $g^k(\cdot)$ will not be generated in subsequent iterations since $y^k \in X \setminus X^k$ is added to X^{k+1}.

Step 3.1 ϵ_0^k is a non-negative parameter, bounding the overall inaccuracy which is allowed in the final solution of $\mathcal{P}$. Note that all four inaccuracy parameters are allowed to vary among iterations.

Step 3.2 $0 \leq \varphi(\mathcal{P}) - f(x^{inc,k}) \leq UB^k - LB^k$. Therefore, $x^{inc,k}$ is a $(UB^k - LB^k)$-optimal solution for $\mathcal{P}$, which meets the required accuracy of ϵ_0^k.

The test $UB^k - LB^k \leq \epsilon_0^k$ is essentially a test on dual optimality. This assertion was already made before the actual description of the procedure; its justification is based on the fact that at each iteration k, a dual feasible solution for $\mathcal{D}(X)$ is readily available. In addition, this dual feasible solution is $(UB^k - LB^k)$-optimal for

$\mathcal{D}(X)$. In proving this statement, consider an arbitrary iteration k. Let ℓ denote the smallest index for which $UB^\ell = UB^k$. Note that, by definition, $UB^\ell < UB^{\ell-1}$, so during iteration ℓ, step 2.7 was executed. Define the function $\tilde{g}(\cdot)$ as follows

$$\tilde{g}(\cdot) = g^\ell(\cdot) + f(y^\ell) - g^\ell(G(y^\ell)) + \epsilon_3^\ell$$

Note that $\tilde{g}(0) = UB^\ell = UB^k$. Obviously, $\tilde{g}(\cdot) \in \Gamma$ because $\tilde{g}(\cdot)$ is readily obtained from $g^\ell(\cdot)$ by a vertical shift. Furthermore, the ϵ_3^ℓ-optimality of y^ℓ in $\mathcal{CD}(g^\ell(\cdot))$ implies that

$$\begin{aligned}
&\forall x \in X: \; f(x) - \tilde{g}(G(x)) = \\
&f(x) - \left(g^\ell(G(x)) + f(y^\ell) - g^\ell(G(y^\ell)) + \epsilon_3^\ell\right) = \\
&(f(x) + g^\ell(0) - g^\ell(G(x))) - \\
&\left(f(y^\ell) + g^\ell(0) - g^\ell(G(y^\ell)) + \epsilon_3^\ell\right) \leq 0
\end{aligned}$$

Consequently, $\tilde{g}(\cdot)$ is a feasible solution for $\mathcal{D}(X)$. Moreover,

$$\tilde{g}(0) - \varphi(\mathcal{D}(X)) \leq UB^\ell - LB^k = UB^k - LB^k$$

As a result, $\tilde{g}(\cdot)$ is a $(UB^k - LB^k)$-optimal solution for $\mathcal{D}(X)$. These observations imply that the test in step 3.2 is, in fact, a test on dual optimality.

3.2 Some minor remarks

In this subsection, two additional remarks concerning the constraint decomposition procedure will be made. The first remark concerns the choice of the inaccuracy parameters ϵ_i^k, $(i = 0, \ldots, 3)$; the second remark relates to the monotonicity of the lower bounds, in case optimal primal solutions are generated in the master programme phase throughout the execution of the procedure.

Inaccuracies and duality gaps

As is revealed by the superscript k, all four inaccuracy parameters ϵ_i^k $(i = 0, \ldots, 3)$ are allowed to vary between iterations. The overall inaccuracy ϵ_0^k, though, has to be chosen at least as large as the sum of the inaccuracies which are allowed in the blocks of which the procedure is composed, in order to make sure that the current solution will be cut off in all subsequent iterations. In this way, the procedure is prevented from getting stuck in a cycle (see the additional explanation concerning steps 2.3–2.7 in Subsection 3.1). Of course, if $UB^k - LB^k \leq \epsilon_0^k$, then $x^{inc,k}$ is a (near-)optimal solution which fits the required accuracy ϵ_0^k, irrespective of whether $\epsilon_0^k \geq \epsilon_1^k + \epsilon_2^k + \epsilon_3^k$. However, if non-cyclic behaviour is to be guaranteed, then ϵ_0^k has to be chosen as indicated. For more detailed information, the reader is referred to Subsection 3.4.

If the restricted primal master programmes $\mathcal{P}(\cdot)$ are of a specific structure, attention may be restricted to specially structured solutions in the relaxed dual master programmes $\mathcal{D}(\cdot)$. For instance, if $\mathcal{P}(\overline{X})$ is a convex programme (i.e. having a concave objective function and a convex feasible set) then, under a mild regularity condition, attention can be restricted to affine solutions in $\mathcal{D}(\overline{X})$ (Lagrangean duality; cf. [Geoffrion, 1972^b] and Subsection 4.1 of Part II). From a computational point of view, however, it may be desirable, or even compulsory, to restrict attention to a (well-structured) *subset* of the feasible set of $\mathcal{D}(\overline{X})$, even if strong duality can thereby no longer be guaranteed. The constraint decomposition procedure as presented in the previous subsection allows for such an approach, because the resulting duality gaps are fully absorbed by the values for ϵ_2^k. One has to realize, however, that the larger the gap, the larger the overall inaccuracy ϵ_0^k one has to accept. It should also be noted that the values for ϵ_1^k and ϵ_2^k need not be specified separately. If in steps 1.3 and 1.4 a primal and a dual solution x^k and $g^k(\cdot)$ have been obtained for which $f(x^k)$ and $g^k(0)$ are sufficiently close, then $\epsilon_1^k + \epsilon_2^k$ may be chosen equal to the difference of these two values. In this way, the individual values for ϵ_1^k and ϵ_2^k are unknown, but also unimportant.

Monotonous lower bounds

If at each iteration $\epsilon_1^k = 0$, then the values $f(x^k)$ constitute a monotonically non-decreasing sequence of lower bounds. As a result, steps 1.5 and 1.6 can then be changed into: "LET $LB^k := f(x^k)$;" and "LET $x^{inc,k} := x^k$" respectively. A similar statement does *not* apply, however, to the sequence of upper bounds UB^k.

3.3 Relations to existing literature

In this subsection Constraint Decomposition will be related to existing literature. More specifically, it will be argued that Dantzig-Wolfe Decomposition and two former extensions are captured by the approach of Subsection 3.1. The procedures that will be discussed here concern the ones in [Dantzig & Wolfe, 1960] for the linear case, [Dantzig, 1963] for the convex case and [Burkard et al., 1985] for the general case.

[Dantzig & Wolfe, 1960]

Dantzig and Wolfe can be regarded as the founding fathers of the procedure. In their paper, *linear programmes* are considered, which can be obtained from (3.1) by choosing $f(x) = c^T x$, $G(x) = Ax - b$, and $X = \{x \in \mathbb{R}^n \mid Ex - h \leq 0,\ x \geq 0\}$; here, c, b and h are given vectors, and A and E are given matrices of appropriate dimensions. Let us denote the (finitely many) extreme points of the polyhedron X by x^j $(j \in J)$ and its (finitely many) extreme rays by r^ℓ $(\ell \in L)$. Let us consider the following restriction of the primal solution space X.

$$\begin{aligned}\overline{X} = \{x \in \mathbb{R}^n \mid\ & x \text{ can be written as a convex} \\ & \text{combination of } x^j\ (j \in \overline{J} \subseteq J) \text{ plus a} \\ & \text{non-negative combination of } r^\ell\ (\ell \in \overline{L} \subseteq L)\}\end{aligned} \tag{3.9}$$

The restricted primal master programmes thus obtained, satisfy the sufficient conditions which imply strong duality between these programmes and their Lagrangean duals (cf. [Geoffrion, 1972[b]] and Subsection 4.1 of Part II). A simple transformation of variables, however, causes each primal-dual pair of master programmes to be transformed into a pair of *linear programmes*, which are dual to one another in a Linear Programming sense.

In the subprogramme phase an extreme point x^j or an extreme ray $\{x^j + \lambda r^\ell \mid \lambda \geq 0\}$ of X is generated, depending on whether the subprogramme $\mathcal{CD}(g^k(\cdot))$ is bounded or not. In step 2.4 or 2.6 then, X^{k+1} is obtained from X^k by extending the index sets $\overline{J}$ and/or $\overline{L}$. As a result, an *uncountable* number of points is added to the current restriction of X.

Finally, if all inaccuracy parameters ϵ_i^k $(i = 0, \ldots, 3)$ are chosen equal to zero, Dantzig and Wolfe's original constraint decomposition procedure is obtained. Note that such a choice is possible, because all programmes involved are *linear* programmes. The finite convergence of the procedure follows from Theorem 3.6 (cf. Subsection 3.4).

[Dantzig, 1963]

In [Dantzig, 1963] pp. 471–478 a constraint decomposition procedure is discussed for the *convex* case. Such programmes can be obtained from (3.1) by assuming that X is a compact and convex set, $f(\cdot)$ is a real-valued and concave function on X, $G(\cdot)$ is a vector-valued and convex function on X, and $\diamond \in \{\leq\}^m$. Suppose that at the start of iteration k, a finite number of points $y^j \in X$ $(j \in \overline{J} \subseteq J)$ has already been generated. In the master programme phase of iteration k, the primal solution space is then restricted to the set $\{y^j \mid j \in \overline{J}\}$, leading to a restricted primal master programme of the following type.

$$\begin{aligned} \max_{x} \quad & f(x) \\ \text{s.t.} \quad & G(x) \leq 0 \\ & x \in \{y^j \mid j \in \overline{J}\} \end{aligned} \tag{3.10}$$

In case only affine functions are considered in the dual of (3.10), then the corresponding relaxed dual master programmes coincides with the Lagrangean dual of (3.10), viz.

$$\begin{aligned} \min_{u,\zeta} \quad & \zeta \\ \text{s.t.} \quad & u^T G(x) + \zeta \geq f(x) \ \ \forall x \in \{y^j \mid j \in \overline{J}\} \\ & u \geq 0 \end{aligned} \tag{3.11}$$

The subprogrammes are fully determined by (near-)optimal solutions of the relaxed dual master programmes. These subprogrammes therefore look like

$$\begin{aligned} \max_{x} \quad & f(x) - u^T G(x) \\ \text{s.t.} \quad & x \in X \end{aligned} \tag{3.12}$$

Unfortunately, the resulting algorithm does not seem to be a promising tool for solving mathematical programmes of the aforementioned type, for at least two reasons. On the one hand, there will usually be a large gap between the optimal solution values of the restricted primal and relaxed dual master programmes (3.10) and (3.11) respectively. This will usually lead to large values for $\epsilon_1^k + \epsilon_2^k$, implying that the values for the overall inaccuracies ϵ_0^k will usually have to be chosen large. As a result, it seems as if one has to content oneself with truly inaccurate solutions for the original programme $\mathcal{P}$. On the other hand, the restricted primal master programmes which are defined in (3.10) are unlikely to be able to constantly generate good lower bounds; the y^k's which are generated in the subprogramme phase are usually not feasible for $\mathcal{P}$. Consequently, the feasible set of $\mathcal{P}(X^k)$ will usually not be extended from iteration to iteration, implying that the lower bound will usually not be increased significantly over iterations. Fortunately, in the convex case we are dealing with, these two major drawbacks can easily be eliminated. The crucial observation in this respect is that (3.11) is a *linear programme*, of which the dual reads

$$\begin{aligned}
\max_{\lambda} \quad & \sum_{j \in \overline{J}} \lambda_j f(y^j) \\
\text{s.t.} \quad & \sum_{j \in \overline{J}} \lambda_j G(y^j) \leq 0 \\
& \sum_{j \in \overline{J}} \lambda_j = 1 \\
& \lambda_j \geq 0, \ j \in \overline{J}
\end{aligned} \tag{3.13}$$

From Linear Programming duality it follows that there is no duality gap between (3.13) and the Lagrangean dual of (3.10). As the Lagrangean dual generally closes the duality gap in case the primal programme is a convex optimization problem, (3.13) is sometimes referred to as the *convexification* of (3.10) (cf. [Magnanti et al., 1976]). Also note that any feasible solution λ for (3.13) leads to a feasible solution $x = \sum_{j \in \overline{J}} \lambda_j y^j$ for $\mathcal{P}$. Because of the latter property, one is tempted to take (3.13), instead of (3.10), as the restricted primal master programme. By doing so, both the aforementioned drawbacks are eliminated; on the one hand, there is no duality gap between (3.13) and (3.11), meaning that any positive value for $\epsilon_1^k + \epsilon_2^k$ is solely due to an unwillingness to solve the two linear programmes to optimality, and on the other hand, the feasible set of (3.13) will usually be extended from iteration to iteration, meaning that the lower bound will usually be improved upon as iterations go by. The question remains of course, whether it is actually allowed to replace (3.10) with (3.13). Fortunately, the answer is yes. It is easily verified that the only way in which the algorithm is affected, is through the determination of the lower bounds; all statements in Subsection 3.1 remain valid, though. The description of Dantzig's decomposition method for the convex case is hereby complete. It is an iterative method between the relaxed dual master programmes (3.11) and the subprogrammes (3.12), where the current incumbents $x^{inc,k}$ and the lower bounds LB^k are determined by the restricted primal master programmes (3.13).

[Burkard et al., 1985]

The generalization of the Dantzig-Wolfe decomposition procedure which is discussed in [Burkard et al., 1985] is essentially the same as ours. As in Variable Decomposition, the former discussion covers the field of *algebraic optimization*. The approach in Subsection 3.1, however, can be amplified to account for these abstract fields of optimization as well. Moreover, we allowed for inaccuracies in intermediate primal and dual solutions, as well as for duality gaps between the underlying primal and dual master programmes.

3.4 Non-cyclic behaviour and finite convergence

In this subsection it will be explained that the constraint decomposition procedure, like the variable decomposition procedure in Section 2, does not exhibit cyclic behaviour. In addition, conditions will be stated under which the procedure is guaranteed to terminate within a finite number of steps.

Theorem 3.4 (Non-cyclic behaviour of complete solutions)
Let (3.1)–(3.8) be given.

1. *In step 1.4 no solution $g^k(\cdot)$ will be generated more than once*

2. *As soon as in step 2.3 or 2.5 a solution y^k is generated for the second time, the algorithm will terminate*

Proof The first assertion immediately follows from the additional explanation concerning step 2.4. In order to prove the second one, let us suppose that $y^\ell = y^k$ for some $\ell < k$. From step 1.4 it follows that

$$f(y^k) = f(y^\ell) \leq g^k(G(y^\ell)) = g^k(G(y^k))$$

This fact, combined with the additional explanation concerning steps 2.4 and 2.7, yields that $UB^k - LB^k \leq \epsilon_0^k$. Hence, the algorithm terminates at iteration k.

□

Definition 3.1 (Essentially identical) *Let $g(\cdot)$ and $g'(\cdot)$ be two extended real-valued functions, defined on a common domain. Then $g(\cdot)$ and $g'(\cdot)$ are called* essentially identical *if and only if $\exists a \in \mathbb{R} : g(\cdot) = g'(\cdot) + a$.*

Two functions which are not essentially identical, are called *essentially different*. The reason for introducing such a notion is clear: if two functions are essentially identical, they yield the same subprogramme $\mathcal{CD}(\cdot)$, implying that the upper bound cannot be improved upon by more than the largest reduction in ϵ_3^k possible. Fortunately, the procedure can be safeguarded against generating essentially identical solutions $g^k(\cdot)$ quite easily. Let $\max\{\epsilon_3^k\}$ denote the largest value of ϵ_3^k that has been used so far.

Theorem 3.5 (Non-cyclic behaviour of partial solutions) *Let (3.1)–(3.8) be given. In addition, suppose that a real-valued upper bound UB^{k_0} has been obtained at iteration k_0, and that at each subsequent iteration the following conditions are satisfied.*

- *If $\varphi(\mathcal{CD}(g^k(\cdot))) = +\infty$ then y^k in step 2.3 is chosen in such a way that $f(y^k) + g^k(0) - g^k(G(y^k)) \geq UB^k$*

- *If $\varphi(\mathcal{CD}(g^k(\cdot))) < +\infty$ then ϵ_0^k in step 3.1 is chosen in such a way that $\epsilon_0^k \geq \epsilon_1^k + \epsilon_2^k + \max\{\epsilon_3^k\}$*

Then the procedure terminates as soon as in step 1.4 a solution $g^k(\cdot)$ is generated which is essentially identical to a previously generated solution $g^\ell(\cdot)$ for some $\ell \in \{k_0, \ldots, k-1\}$.

Proof Suppose $g^\ell(\cdot)$ is essentially identical to $g^k(\cdot)$ for some $\ell \in \{k_0, \ldots, k-1\}$. We know that $y^\ell \in X^k$, so $g^k(G(y^\ell)) \geq f(y^\ell)$. This implies that

$$f(y^\ell) + g^k(0) - g^k(G(y^\ell)) \leq g^k(0) \tag{3.14}$$

Furthermore, $g^\ell(\cdot)$ is essentially identical to $g^k(\cdot)$, hence

$$g^\ell(0) - g^\ell(G(y^\ell)) = g^k(0) - g^k(G(y^\ell)) \tag{3.15}$$

In case $\varphi(\mathcal{CD}(g^\ell(\cdot))) = \varphi(\mathcal{CD}(g^k(\cdot))) = +\infty$, (3.14) and (3.15) imply that

$$\begin{aligned} &UB^k - LB^k \leq UB^\ell - LB^k \leq UB^\ell - f(x^k) \leq \\ &\left(f(y^\ell) + g^\ell(0) - g^\ell(G(y^\ell))\right) - \left(g^k(0) - \epsilon_1^k - \epsilon_2^k\right) = \\ &\left(f(y^\ell) + g^k(0) - g^k(G(y^\ell))\right) - \left(g^k(0) - \epsilon_1^k - \epsilon_2^k\right) \leq \epsilon_0^k \end{aligned}$$

As a result, the procedure terminates at iteration k. On the other hand, if $\varphi(\mathcal{CD}(g^\ell(\cdot))) = \varphi(\mathcal{CD}(g^k(\cdot))) < +\infty$, then (3.14) and (3.15) imply that

$$\begin{aligned} &UB^k - LB^k \leq UB^\ell - LB^k \leq UB^\ell - f(x^k) \leq \\ &\left(f(y^\ell) + g^\ell(0) - g^\ell(G(y^\ell)) + \epsilon_3^\ell\right) - \left(g^k(0) - \epsilon_1^k - \epsilon_2^k\right) = \\ &\left(f(y^\ell) + g^k(0) - g^k(G(y^\ell)) + \epsilon_3^\ell\right) - \left(g^k(0) - \epsilon_1^k - \epsilon_2^k\right) \leq \epsilon_0^k \end{aligned}$$

Again, the procedure terminates at iteration k.

□

As one can see, Theorem 3.5 does not exclude the possibility of repeatedly generating essentially identical functions $g^k(\cdot)$ completely. However, once a solution $g^k(\cdot)$ is generated in step 1.4 with $\varphi(\mathcal{CD}(g^k(\cdot))) < +\infty$, a real-valued upper bound UB^k is obtained. From that iteration onwards, all solutions in step 1.4 can be guaranteed to be essentially different, unless termination occurs. As in the variable decomposition procedure, the inaccuracy in the subprogramme phase is the most critical one if repeatedly generating essentially identical functions $g^k(\cdot)$ is to be avoided. Apparently, there is no gain in requiring a specific accuracy ϵ_3^k, once a less accurate solution has ever been allowed in step 2.5. The sequence $(\epsilon_3^k)_{\mathbb{N}}$ could therefore just as well be chosen monotonically non-decreasing, in which case the condition stated in step 3.1 results.

From the previous two theorems it follows that even in its full generality, Constraint Decomposition does not exhibit cyclic behaviour. To enforce finite convergence, severe conditions have to be imposed on problem structure. Below we will formulate two such conditions.

Theorem 3.6 (Finite convergence – primal version)
Let (3.1)–(3.8) be given. In addition, suppose that every $y^k \in X$ that is generated in step 2.3 or 2.5 belongs to some finite subset of X. *Then the procedure is guaranteed to terminate within a finite number of steps.*

Theorem 3.7 (Finite convergence – dual version)
Let (3.1)–(3.8) be given. In addition, suppose that a real-valued upper bound UB^{k_0} has been obtained at iteration k_0 and that at each subsequent iteration, the following conditions are satisfied.

- *If $\varphi(\mathcal{CD}(g^k(\cdot))) = +\infty$ then y^k in step 2.3 is chosen in such a way that $f(y^k) + g^k(0) - g^k(G(y^k)) \geq UB^k$*

- *If $\varphi(\mathcal{CD}(g^k(\cdot))) < +\infty$ then ϵ_0^k in step 3.1 is chosen in such a way that $\epsilon_0^k \geq \epsilon_1^k + \epsilon_2^k + \max\{\epsilon_3^k\}$*

Finally, suppose that every $g^k(\cdot)$ which is generated in step 1.4, can be obtained by vertically shifting a member of some finite *set of functions. Then the procedure will terminate within a finite number of steps.*

Proof The results are a direct consequence of the fact that the procedure neither generates the same values for $y \in X$, nor essentially identical functions $g(\cdot) \in \Gamma$ more than twice.

□

These convergence results prove their usefulness, for instance, if the subprogrammes can be converted into linear programmes. In that case, all intermediate primal solutions y^k relate to a *finite* number of extreme points and/or extreme rays. More specifically, the finite convergence of the original Dantzig-Wolfe decomposition procedure is covered by these results.

3.5 Asymptotic convergence

In this subsection, the asymptotic convergence properties of the constraint decomposition procedure will be discussed. In Theorem 3.8 sufficient conditions will be stated which imply asymptotic convergence. These conditions concern, apart from some mild continuity and compactness requirements on problem data, the *existence of finitely representable dual solutions* for the relaxed dual master programmes (cf. Subsection 4.2 of Part II). Moreover, it is assumed that these dual solutions are *continuous*, and that they can be represented by means of elements of some (finitely dimensional) *compact* set. Finally, such dual solutions are supposed to be actually *generated* throughout the execution of the algorithm.

Theorem 3.8 (Asymptotic convergence) *Let (3.1)–(3.8) be given. Suppose that X is compact, $\varphi(\mathcal{P}) \in \mathbb{R}$ and that $f(\cdot)$ and $G(\cdot)$ are continuous functions with respect to X. Furthermore, suppose that $\forall k : \epsilon_i^k$ is chosen from some compact set E_i $(i = 1, \ldots, 3)$. Finally, suppose that there is a non-empty and compact set $T \subseteq \mathbb{R}^\tau$ $(\tau \in \mathbb{N})$ and a function $w(\cdot) : T \times \mathbb{R}^m \to \mathbb{R} \cup \{\pm\infty\}$ which is continuous and real-valued with respect to $T \times RHS$, such that $w(t, \cdot) \in \Gamma$ $\forall t \in T$ and*

$$\text{If } g^k(\cdot) \text{ is generated in step 1.4 during iteration } k, \text{ then } \exists t^k \in T : \; w(t^k, \cdot) = g^k(\cdot) \tag{3.16}$$

Then the algorithm converges asymptotically, in the sense that

1. *If the algorithm does not terminate within a finite number of steps, then for every accumulation point t^∞ of $(t^k)_{\mathbb{N}}$ it follows that $w(t^\infty, \cdot) + \epsilon_3^\infty$ is an $(\epsilon_2^\infty + \epsilon_3^\infty)$-optimal solution for $\mathcal{D}(X)$ for some accumulation point $(\epsilon_2^\infty, \epsilon_3^\infty)$ of $(\epsilon_2^k, \epsilon_3^k)_{\mathbb{N}}$.*

2. *If the algorithm does not terminate within a finite number of steps, then every accumulation point of $(x^k)_{\mathbb{N}}$ is a $\limsup(\epsilon_1^k + \epsilon_2^k + \epsilon_3^k)$-optimal solution for $\mathcal{P}$.*

3. *If the algorithm does not terminate within a finite number of steps, then every accumulation point of $(x^{inc,k})_{\mathbb{N}}$ is a $\limsup(\epsilon_1^k+\epsilon_2^k+\epsilon_3^k)$-optimal solution for $\mathcal{P}$.*

4. *If the algorithm does not terminate within a finite number of steps, then $0 \leq \lim(UB^k - LB^k) \leq \limsup(\epsilon_1^k + \epsilon_2^k + \epsilon_3^k)$.*

5. *Suppose there is an iteration ℓ_0 and an $\eta > 0$ such that $\forall k \geq \ell_0 :$ $\epsilon_0^k \geq \epsilon_1^k + \epsilon_2^k + \epsilon_3^k + \eta$, then the algorithm terminates within a finite number of steps, say k_0, and x^{inc,k_0} is a $(UB^{k_0} - LB^{k_0})$-optimal solution for $\mathcal{P}$, which meets the required accuracy of $\epsilon_0^{k_0}$.*

Proof First of all, it should be noted that under the aforementioned conditions, steps 2.2 and 2.4 of the algorithm will never be executed. This assertion is easily proven. The image of X under $f(\cdot)$ does not include $+\infty$, and the image of RHS under $w(t,\cdot)$ does not include $-\infty$ for any $t \in T$. This implies that the objective function of $\mathcal{CD}(w(t^k,\cdot))$ is less than $+\infty$ at all points $x \in X$ for any iteration k. As a consequence of the compactness of X and the continuity of $f(\cdot)$, $G(\cdot)$ and $w(t^k,\cdot)$, it follows that the optimal objective function value of $\mathcal{CD}(w(t^k,\cdot))$ is bounded from above for all iterations k. Secondly, it should be observed that due to $\varphi(\mathcal{P}) \in \mathbb{R}$, finite convergence of the algorithm necessarily implies that the optimality criterion in step 3.2 is met. Finally, it should be mentioned that all references to accumulation points only make sense in case such points exist. From the compactness conditions it follows that such points exist if and only if the algorithm does not terminate prematurely.

1. Let t^∞ be any accumulation point of $(t^k)_{\mathbb{N}}$. Note that the compactness condition on T implies that such a point exists, and is necessarily an element of T. Furthermore, the compactness conditions on E_2, E_3 and X imply that there is a subsequence $(p(k))_{\mathbb{N}}$ of $\mathbb{N}$ such that

$$\lim(t^{p(k)}, \epsilon_2^{p(k)}, \epsilon_3^{p(k)}, y^{p(k)}) = (t^\infty, \epsilon_2^\infty, \epsilon_3^\infty, y^\infty)$$

for some $(\epsilon_2^\infty, \epsilon_3^\infty, y^\infty) \in E_2 \times E_3 \times X$. From the fact that $y^{p(k)}$ is an $\epsilon_3^{p(k)}$-optimal solution of $\mathcal{CD}(w(t^{p(k)},\cdot))$, it follows that

$$\forall x \in X: \; f(x) - w(t^{p(k)}, G(x)) \leq$$
$$f(y^{p(k)}) - w(t^{p(k)}, G(y^{p(k)})) + \epsilon_3^{p(k)}$$

The continuity conditions on $f(\cdot)$, $G(\cdot)$ and $w(\cdot)$ imply that

$$\forall x \in X: \; f(x) - w(t^\infty, G(x)) \leq \quad f(y^\infty) - w(t^\infty, G(y^\infty)) + \epsilon_3^\infty \tag{3.17}$$

Note that y^∞ is hereby proven to be an ϵ_3^∞-optimal solution for $\mathcal{CD}(w(t^\infty, \cdot))$. From step 1.4 it also follows that

$$w(t^{p(k+1)}, G(y^{p(k)})) \geq f(y^{p(k)})$$

which, together with the continuity conditions on $f(\cdot)$, $G(\cdot)$ and $w(\cdot)$, implies that

$$w(t^\infty, G(y^\infty)) \geq f(y^\infty) \tag{3.18}$$

Combining (3.17) and (3.18) renders

$$\forall x \in X: \; w(t^\infty, G(x)) + \epsilon_3^\infty \geq f(x)$$

It is obvious that Γ is closed with respect to scalar addition, so

$$w(t^\infty, \cdot) + \epsilon_3^\infty \in \Gamma$$

The last two statements prove the feasibility of $w(t^\infty, \cdot) + \epsilon_3^\infty$ in $\mathcal{D}(X)$. Finally, it is also true that $w(t^{p(k)}, 0) - \epsilon_2^{p(k)} \leq \varphi(\mathcal{D}(X^{p(k)})) = \varphi(\mathcal{P}(X^{p(k)})) \leq \varphi(\mathcal{P}) = \varphi(\mathcal{D}(X))$. Again from the continuity condition on $w(\cdot)$ it follows that

$$w(t^\infty, 0) + \epsilon_3^\infty \leq \varphi(\mathcal{D}(X)) + \epsilon_2^\infty + \epsilon_3^\infty$$

As a result, $w(t^\infty, \cdot) + \epsilon_3^\infty$ is an $(\epsilon_2^\infty + \epsilon_3^\infty)$-optimal solution for $\mathcal{D}(X)$

2. Let x^∞ be any accumulation point of $(x^k)_{\mathbb{N}}$. Note that the compactness condition on X implies that such a point exists, and is necessarily an element of X. As above, we construct a subsequence $(p(k))_{\mathbb{N}}$ of $\mathbb{N}$, such that

$$\lim(t^{p(k)}, \epsilon_1^{p(k)}, \epsilon_2^{p(k)}, \epsilon_3^{p(k)}, x^{p(k)}) = (t^\infty, \epsilon_1^\infty, \epsilon_2^\infty, \epsilon_3^\infty, x^\infty)$$

for some $(t^\infty, \epsilon_1^\infty, \epsilon_2^\infty, \epsilon_3^\infty) \in T \times E_1 \times E_2 \times E_3$. The continuity of $G(\cdot)$ implies that x^∞ is a feasible solution for $\mathcal{P}$. Furthermore, from the continuity of $f(\cdot)$ it follows that

$$
\begin{aligned}
&f(x^\infty) = \lim f(x^{p(k)}) \geq \\
&\limsup \left(\varphi(\mathcal{P}(X^{p(k)})) - \epsilon_1^{p(k)}\right) = \\
&\limsup \left(\varphi(\mathcal{D}(X^{p(k)})) - \epsilon_1^{p(k)}\right) \geq \\
&\lim \left(w(t^{p(k)}, 0) - \epsilon_1^{p(k)} - \epsilon_2^{p(k)}\right) = \\
&w(t^\infty, 0) - \epsilon_1^\infty - \epsilon_2^\infty = \\
&\left(w(t^\infty, 0) + \epsilon_3^\infty\right) - \left(\epsilon_1^\infty + \epsilon_2^\infty + \epsilon_3^\infty\right) \geq \\
&\varphi(\mathcal{P}) - \left(\epsilon_1^\infty + \epsilon_2^\infty + \epsilon_3^\infty\right)
\end{aligned}
$$

The last inequality follows from the fact that $w(t^\infty, \cdot) + \epsilon_3^\infty$ is a feasible solution for $\mathcal{D}(X)$ (cf. 1.).

3. Let $x^{inc,\infty}$ be any accumulation point of $(x^{inc,k})_{\mathbb{N}}$. Note that the compactness condition on X implies that such a point exists, and is necessarily an element of X. As above, we construct a subsequence $(p(k))_{\mathbb{N}}$ of $\mathbb{N}$, such that

$$\lim(x^{inc,p(k)}, x^{p(k)}) = (x^{inc,\infty}, x^\infty)$$

for some $x^\infty \in X$. The continuity of $G(\cdot)$ implies that $x^{inc,\infty}$ is a feasible solution for $\mathcal{P}$. Furthermore, $f(x^{inc,k}) \geq f(x^k)$, so

$$f(x^{inc,\infty}) \geq f(x^\infty) \geq \varphi(\mathcal{P}) - \left(\epsilon_1^\infty + \epsilon_2^\infty + \epsilon_3^\infty\right)$$

4. From the compactness conditions on X, T, E_1, E_2 and E_3 it follows that there is a subsequence $(p(k))_{\mathbb{N}}$ of $\mathbb{N}$ such that

$$\lim(t^{p(k)}, \epsilon_1^{p(k)}, \epsilon_2^{p(k)}, \epsilon_3^{p(k)}, y^{p(k)}) = (t^\infty, \epsilon_1^\infty, \epsilon_2^\infty, \epsilon_3^\infty, y^\infty)$$

for some $(t^\infty, \epsilon_1^\infty, \epsilon_2^\infty, \epsilon_3^\infty, y^\infty) \in T \times E_1 \times E_2 \times E_3 \times X$. Note that $(UB^k)_{\mathbb{N}}$ $((LB^k)_{\mathbb{N}})$ is a monotonically non-increasing (non-decreasing) sequence of upper (lower) bounds, which is bounded from below (above) by $\varphi(\mathcal{P})$. Hence, both sequences converge to, say UB^∞ and LB^∞ respectively. On the one hand, (3.18) implies that

$$UB^\infty \leq$$
$$\lim\left(f(y^{p(k)}) + w(t^{p(k)},0) - w(t^{p(k)},G(y^{p(k)})) + \epsilon_3^{p(k)}\right) \leq$$
$$f(y^\infty) + w(t^\infty,0) - w(t^\infty,G(y^\infty)) + \epsilon_3^\infty \leq w(t^\infty,0) + \epsilon_3^\infty$$

On the other hand,

$$LB^\infty \geq \lim f(x^{p(k)}) \geq$$
$$\lim\left(w(t^{p(k)},0) - \epsilon_1^{p(k)} - \epsilon_2^{p(k)}\right) = w(t^\infty,0) - \epsilon_1^\infty - \epsilon_2^\infty$$

The above two relations imply that

$$\lim(UB^k - LB^k) = UB^\infty - LB^\infty \leq \epsilon_1^\infty + \epsilon_2^\infty + \epsilon_3^\infty$$

5. Suppose the algorithm does not terminate within a finite number of steps. We know that

$$\lim(UB^k - LB^k) \leq \lim\sup(\epsilon_1^k + \epsilon_2^k + \epsilon_3^k) \leq \lim\sup \epsilon_0^k - \eta$$

As a result,

$$\exists k: \; UB^k - LB^k \leq \epsilon_0^k$$

which would imply that the optimality criterion is met at iteration k; an obvious contradiction. Therefore, the algorithm is bound to terminate after a finite number of steps.

□

Corollary 3.1 (Asymptotic convergence – perturbed version) *Let the conditions of Theorem 3.8 be satisfied. Then every accumulation point y^∞ of $(y^k)_{\mathbb{N}}$ is a* $\lim\sup \epsilon_3^k$*-optimal solution for the mathematical programme which is obtained from $\mathcal{P}$ by changing the right-hand-side* 0 *into $G(y^\infty)$.*

Proof Let y^∞ be any accumulation point of $(y^k)_{\mathbb{N}}$. Note that the closedness condition on X implies that $y^\infty \in X$, so the feasibility of y^∞ with respect to the aforementioned perturbed primal programme is easily verified. As in the proof of Theorem 3.8, we construct a subsequence $(p(k))_{\mathbb{N}}$ of $\mathbb{N}$, such that

$$\lim(t^{p(k)}, \epsilon_2^{p(k)}, \epsilon_3^{p(k)}, y^{p(k)}) = (t^\infty, \epsilon_2^\infty, \epsilon_3^\infty, y^\infty)$$

for some $(t^\infty, \epsilon_2^\infty, \epsilon_3^\infty) \in T \times E_2 \times E_3$. From (3.17) we know that y^∞ is an ϵ_3^∞-optimal solution for $\mathcal{CD}(w(t^\infty, \cdot))$. This fact, combined with $w(t^\infty, \cdot) \in \Gamma$, renders

$$\begin{aligned}
&\sup_x\{f(x) \mid G(x) \diamond G(y^\infty),\ x \in X\} \leq \\
&\sup_x\{f(x) + w(t^\infty, G(y^\infty)) - w(t^\infty, G(x)) \mid \\
&\qquad\qquad\qquad G(x) \diamond G(y^\infty),\ x \in X\} \leq \\
&\sup_x\{f(x) + w(t^\infty, G(y^\infty)) - w(t^\infty, G(x)) \mid x \in X\} = \\
&\sup_x\{f(x) + w(t^\infty, 0) - w(t^\infty, G(x)) \mid x \in X\} + \\
&\qquad\qquad\qquad w(t^\infty, G(y^\infty)) - w(t^\infty, 0) \leq \\
&f(y^\infty) + w(t^\infty, 0) - w(t^\infty, G(y^\infty)) + \epsilon_3^\infty + \\
&\qquad\qquad\qquad w(t^\infty, G(y^\infty)) - w(t^\infty, 0) = \\
&f(y^\infty) + \epsilon_3^\infty
\end{aligned}$$

This proves the corollary.

□

Although there seems to be a lot of similarity between Variable Decomposition and Constraint Decomposition (in fact, these methods will be proven to be dual approaches in Section 4), there remains an annoying discrepancy with respect to the generality of the asymptotic convergence results. In Variable Decomposition, *we did not have to rely on finitely representable dual solutions* in order to prove convergence. In Subsection 2.5, convergence was proven under a closedness condition on $\alpha(\beta(\cdot))$, the point-to-set map which largely described the variable decomposition procedure. Furthermore, it was also proven that in case finitely representable, and in some sense continuous dual solutions *existed*, closedness of this point-to-set map followed.

In proving the convergence of Constraint Decomposition, *we assumed the existence of finitely representable and continuous dual solutions* $w(\cdot)$ *right from the start*; we even assumed that such solutions were actually *generated* throughout the execution of the procedure. As in Variable Decomposition, the former assumption, together with the continuity conditions on $f(\cdot)$, $G(\cdot)$ and $w(\cdot)$ and the compactness condition on X, implies the closedness of the point-to-set map which largely describes the constraint decomposition procedure. Here too, this map can be defined as the composed map $\alpha(\beta(\cdot))$, where

$$\begin{aligned}
&\alpha(\cdot) : X \rightarrow\rightarrow T, \text{ with}\\
&\alpha(x) = \{t \in T \mid w(t, G(x)) \geq f(x)\}\\
&\beta(\cdot) : T \times E_2 \rightarrow\rightarrow X, \text{ with}\\
&\beta(t, \epsilon_2) = \{x \in X \mid x \text{ is } \epsilon_2\text{-optimal for } \mathcal{CD}(w(t, \cdot))\}
\end{aligned} \tag{3.19}$$

Unfortunately, it seems impossible to generalize Theorem 3.8 to non-finitely representable dual solutions. In such a case, a closedness condition on the analogue of $\alpha(\beta(\cdot))$ would no longer be superfluous; it would be needed in deriving a similar expression as in (3.18), for instance. But such a general condition of closedness on $\alpha(\beta(\cdot))$ would require the notion of convergence of a sequence of functions. How could such a notion be defined? Pointwise convergence would not suffice, because then Theorem 3.8 would not follow as a special case. More specifically, $\lim_{k\to\infty} t^k = t^\infty$ would not necessarily imply that $\lim_{k\to\infty} w(t^k, \cdot)$ coincides with $w(t^\infty, \cdot)$ in case finitely representable functions are taken into consideration. The only way out seems to define a norm, so that by definition $\lim_{k\to\infty} g^k(\cdot) = g^\infty(\cdot)$ if and only if $\lim_{k\to\infty} \| g^k(\cdot) - g^\infty(\cdot) \| = 0$. Consequently, normed functional spaces would have to be taken into consideration. This is an enormous complication and beyond the scope of this monograph.

Section 4

Variable Decomposition and Constraint Decomposition as dual methods

Many striking similarities exist between the two decomposition procedures that are discussed in the previous two sections. For instance, in **Variable Decomposition**, *relaxations* of the original programme are obtained through the master programmes, while *restrictions* are obtained through the subprogrammes. Moreover, *primal* solutions of the relaxed master programmes are sent to the subprogrammes, which in their turn, send *dual* solutions back. The situation is reversed when **Constraint Decomposition** is considered. In that case, *restrictions* of the original programme are obtained through the master programmes, while *relaxations* are obtained through the subprogrammes. Furthermore, *dual* solutions of the restricted master programmes are sent to the subprogrammes, which in their turn, send *primal* solutions back. Moreover, the results concerning non-cyclic behaviour and finite and asymptotic convergence of the two methods also bear a great deal of resemblance. Finally, it is a well-known fact that in Linear Programming, Benders and Dantzig-Wolfe Decomposition are dual approaches, in the sense that the latter applied to a linear programme is equivalent to the former applied to its dual (cf. [Lasdon, 1970]). The question now arises whether Variable Decomposition and Constraint Decomposition can be considered to be dual methods in the general case as well. Consider the following primal programme

$$\begin{aligned} \mathcal{P}: \max_{x} \quad & f(x) \\ \text{s.t.} \quad & G(x) \diamond 0 \\ & H(x) \circ 0 \\ & x \in X \end{aligned} \tag{4.1}$$

where $f(\cdot), G(\cdot)$ and $H(\cdot)$ are functions from a set $D \subseteq \mathbb{R}^n$ to $\mathbb{R} \cup \{-\infty\}$, $\mathbb{R}^{m_1}$ and $\mathbb{R}^{m_2}$ respectively, X is a subset of D, $\diamond \in \{\leq, =\}^{m_1}$ and $\circ \in \{\leq, =\}^{m_2}$ $(m_1, m_2, n \in \mathbb{N})$. In the sequel we will also assume that

$$\varphi(\mathcal{P}) \in \mathbb{R} \tag{4.2}$$

From Theorem 5.1 of Part II it follows that the dual programme of (4.1) may be formulated as

$$\begin{aligned} \mathcal{D}: \min_{g(\cdot),h(\cdot)} \quad & g(0) + h(0) \\ \text{s.t.} \quad & g(G(x)) + h(H(x)) \geq f(x) \;\; \forall x \in X \\ & g(\cdot) \in \Gamma_G, \; h(\cdot) \in \Gamma_H \end{aligned} \tag{4.3}$$

where

$$\begin{aligned} \Gamma_G = \{ & g(\cdot) : \mathbb{R}^{m_1} \to \mathbb{R} \cup \{+\infty\} \mid \\ & g(r) \leq g(r') \;\; \forall r, r' \in \mathbb{R}^{m_1} : \; r \diamond r' \} \\ \Gamma_H = \{ & h(\cdot) : \mathbb{R}^{m_2} \to \mathbb{R} \cup \{+\infty\} \mid \\ & h(s) \leq h(s') \;\; \forall s, s' \in \mathbb{R}^{m_2} : \; s \circ s' \} \end{aligned} \tag{4.4}$$

Applying **Constraint Decomposition to (4.1)** leads to an iterative scheme involving the restricted primal master programmes

$$\begin{aligned} \mathcal{P}(\overline{\Pi}): \quad \max_{x} \quad & f(x) \\ \text{s.t.} \quad & G(x) \diamond 0 \\ & x \in \overline{\Pi} \end{aligned} \tag{4.5}$$

the relaxed dual master programmes

$$\begin{aligned} \mathcal{D}(\overline{\Pi}): \quad \min_{g(\cdot)} \quad & g(0) \\ \text{s.t.} \quad & g(G(x)) \geq f(x) \ \ \forall x \in \overline{\Pi} \\ & g(\cdot) \in \Gamma_G \end{aligned} \tag{4.6}$$

and the subprogrammes

$$\begin{aligned} \mathcal{CD}(g(\cdot)): \quad \max_{x} \quad & f(x) + g(0) - g(G(x)) \\ \text{s.t.} \quad & H(x) \circ 0 \\ & x \in X \end{aligned} \tag{4.7}$$

where $\overline{\Pi} \subseteq \Pi = \{x \in X \mid H(x) \circ 0\}$. Note that there is no duality gap between those restricted primal master programmes $\mathcal{P}(\overline{\Pi})$ and relaxed dual master programmes $\mathcal{D}(\overline{\Pi})$ for which $\varphi(\mathcal{P}(\overline{\Pi})) > -\infty$ (cf. Lemma 5.1 sub 1. of Part II). Now, if we would like to apply **Variable Decomposition to the dual of (4.1)**, i.c. (4.3), we would temporarily have to fix some of the decision variables, while optimization would take place over the remaining ones. In (4.3) the function $g(\cdot)$ can be temporarily held fixed. This is exactly the reason why we formulated the dual programme as we have. The following primal subprogrammes are hereby obtained.

$$\begin{aligned} \mathcal{D}(\bar{g}(\cdot)): \quad \min_{h(\cdot)} \quad & \bar{g}(0) + h(0) \\ \text{s.t.} \quad & h(H(x)) \geq f(x) - \bar{g}(G(x)) \;\; \forall x \in X \\ & h(\cdot) \in \Gamma_H \end{aligned} \tag{4.8}$$

Because $\mathcal{D}(\bar{g}(\cdot))$ is dual to $\mathcal{CD}(\bar{g}(\cdot))$ for those $\bar{g}(\cdot) \in \Gamma_G$ with $\bar{g}(0) \in \mathbb{R}$, we can reformulate (4.3) by applying the key problem manipulations of projection and dualization as indicated. This leads to

$$\begin{aligned} \min_{\bar{g}(\cdot)} \quad \max_{x} \quad & f(x) + \bar{g}(0) - \bar{g}(G(x)) \\ \text{s.t.} \quad & H(x) \circ 0 \\ & x \in X \\ \text{s.t.} \quad & \bar{g}(0) \in \mathbb{R}, \; \bar{g}(\cdot) \in \Gamma_G \end{aligned} \tag{4.9}$$

Introducing a dummy variable θ gives us the following equivalent formulation

$$\begin{aligned} \min_{\bar{g}(\cdot),\theta} \quad & \theta \\ \text{s.t.} \quad & \theta \geq f(x) + \bar{g}(0) - \bar{g}(G(x)) \;\; \forall x \in X : \; H(x) \circ 0 \\ & \bar{g}(0) \in \mathbb{R}, \; \bar{g}(\cdot) \in \Gamma_G, \; \theta \in \mathbb{R} \end{aligned} \tag{4.10}$$

Finally, a change of variables is carried out; the function $g(\cdot)$ is substituted for every occurrence of $\bar{g}(\cdot) - \bar{g}(0) + \theta$. As such, the following relaxed master programmes are finally obtained.

$$\begin{aligned} \mathcal{VD}(\overline{\Pi}): \quad \min_{g(\cdot)} \quad & g(0) \\ \text{s.t.} \quad & g(G(x)) \geq f(x) \;\; \forall x \in \overline{\Pi} \\ & g(0) \in \mathbb{R}, \; g(\cdot) \in \Gamma_G \end{aligned} \tag{4.11}$$

Note that $\mathcal{D}(\overline{\Pi})$ equals $\mathcal{VD}(\overline{\Pi})$ for those $\overline{\Pi} \subseteq \Pi$ with $\varphi(\mathcal{P}(\overline{\Pi})) > -\infty$! *To summarize, both Constraint Decomposition applied to the primal programme (4.1) and Variable Decomposition applied to its dual programme (4.3) amount to iterating between the master programmes (4.6) and the subprogrammes (4.7).* Loosely speaking, the two decomposition schemes can, even in the general case, be considered as dual methods. With this observation in mind, the many striking similarities between the two decomposition methods and their results are no longer surprising.

The observed dual relation between the two decomposition schemes also adds to the intuitive understanding of the two types of subprogrammes. The ones in Variable Decomposition reveal how the original *primal programme* is affected when some of the decision variables in the *primal space* are held fixed. In Constraint Decomposition, however, they show how the original *primal programme* is affected when some of the decision variables in the *dual space* are held fixed. Apparently, the latter effect is twofold; first, the primal feasible set is enlarged and, secondly, the primal objective function is modified.

Section 5

Extensions of some known variations

Many variations of the two decomposition schemes which are discussed in Sections 2 and 3 are possible. Here, we will discuss a few. Subsection 5.1 deals with some miscellaneous matters. In Subsection 5.2 Cross Decomposition is discussed, and a brief remark on Kornai-Lipták Decomposition follows in Subsection 5.3.

5.1 Miscellaneous

First of all, it should be noted that for a given problem formulation, many implementations are still possible for both the variable and the constraint decomposition procedure. In Variable Decomposition one first has to decide which decision variables will be dealt with through the master programme and in Constraint Decomposition one has to select those constraints which together form the set X (cf. (3.1)). Obviously, not all possibilities are equally desirable.

Although it was not mentioned explicitly in Section 2 and 3, it is very important to realize that both decomposition methods fully depend on the mathematical *formulation* of an optimization problem, more than on problem *characteristics*. This statement is not only true in case a completely different formulation is considered, but also if the original formulation is only slightly modified. In this respect, one could think of first duplicating existing variables and/or constraints, or first adding new variables and/or constraints which are in some sense redundant, before applying decomposition. For instance, if all variables are first duplicated and then Constraint Decomposition is applied in such a way that the set X consists of all but the duplicating constraints, then a procedure results which has been called *Variable Splitting* in [Jörnsten et al., 1985], and *Lagrangean Decomposition* in [Guignard & Kim, 1987]. As another example, suppose a linear programme is given in which activities share some common resources, and that only a limited amount of each of these resources is available. In that case, artificial variables may be introduced, indicating which amount of each type of resource is allocated to each activity. As a consequence, new constraints have to be added, enforcing that such an allocation does not exceed the amounts of resources available. The application of Variable Decomposition to such a reformulation, where the artificial variables are the ones that are temporarily held fixed in the primal subprogrammes, is known as Resource Directive Decomposition.

Other variations of Variable and Constraint Decomposition are obtained if the master programme is set aside in favour of some other device that is able to feed the subprogrammes. As far as Variable Decomposition is concerned, [Balas, 1969] may serve as an example, although Balas's device which generates integer values for the integer variables, is very much similar to the relaxed master programmes. As far as Constraint Decomposition is concerned, many references can be given. All Lagrangean Relaxation approaches in Integer (Linear) Programming, for instance, fall into this framework. In these approaches the master programmes are replaced by a (heuristic) update mechanism for the dual solutions, such as a subgradient optimization procedure; see e.g. [Geoffrion, 1974], [Fisher et al., 1975], [Goffin, 1977], [Fisher, 1981], [Dyer & Walker, 1982] and the references cited there. Cross

Decomposition and Kornai-Lipták Decomposition are examples of approaches that are variations on Variable Decomposition and Constraint Decomposition *simultaneously*; they will be discussed in the next two subsections.

5.2 Cross Decomposition

Basically, Cross Decomposition is a variation of Benders Decomposition and Dantzig-Wolfe Decomposition *simultaneously*, and as such, it can also be generalized to general mathematical programmes. Originally, it has been developed for programmes with a linear structure and it has been applied quite successfully to facility location problems (cf. [Van Roy, 1980,1983]); the results in this section are an extension of the ones developed by Van Roy. In [Burkard et al., 1985] too, a generalization of Cross Decomposition is presented. On the one hand, their treatment is more general than ours, because in their paper the field of *algebraic optimization* is being dealt with. The discussion in this subsection, however, can be amplified to account for such general structures as well. On the other hand, their results are more restrictive, because in the problems they consider, all primal and dual solutions are supposed to have a separable structure.

Suppose we are given a mathematical programme to which both Variable and Constraint Decomposition could be applied.

$$\begin{aligned} \mathcal{P}: \quad \max_{x,y} \quad & f(x,y) \\ \text{s.t.} \quad & G(x,y) \diamond 0 \\ & (x,y) \in U \cap (X \times Y) \end{aligned} \tag{5.1}$$

Here, $f(\cdot)$ and $G(\cdot)$ are functions from $D \subseteq \mathbb{R}^n$ to $\mathbb{R} \cup \{\pm\infty\}$ and $\mathbb{R}^m$ respectively, $U \cap (X \times Y) \subseteq D$, and $\diamond \in \{\leq, =\}^m$ $(m, n \in \mathbb{N})$. Furthermore, it will be assumed that

- $\forall(x,y) \in U \cap (X \times Y):\ f(x,y) < +\infty$
- If (x,y) is a feasible solution for $\mathcal{P}$, then a $y' \in Y$ exists, such that (x,y') is a feasible solution for $\mathcal{P}$ with $f(x,y') > -\infty$.

Let us consider the following implementation of **Variable Decomposition**, where iterations take place between, on the one hand, the primal subprogrammes

$$\begin{aligned} \mathcal{P}(\bar{x}): \quad \max_{x,y} \quad & f(x,y) \\ \text{s.t.} \quad & G(x,y) \diamond 0 \\ & x = \bar{x} \\ & (x,y) \in U \cap (\widetilde{X} \times Y) \end{aligned} \tag{5.2}$$

and the dual subprogrammes

$$\begin{aligned} \mathcal{D}(\bar{x}): \quad \min_{\sigma(\cdot)} \quad & \sigma(0,\bar{x}) \\ \text{s.t.} \quad & \sigma(G(x,y),x) \geq f(x,y) \\ & \qquad\qquad \forall(x,y) \in U \cap (\widetilde{X} \times Y) \\ & \sigma(\cdot): \mathbb{R}^{m+n_1} \rightarrow \mathbb{R} \cup \{+\infty\} \\ & \sigma(r,x) \leq \sigma(r',x) \ \ \forall r \diamond r' \ \forall x \in \widetilde{X} \end{aligned} \tag{5.3}$$

and, on the other hand, the relaxed master programmes

$$\begin{aligned} \mathcal{VD}(\overline{\Delta}, \overline{\Delta}^0): \quad \max_{x,\theta} \quad & \theta \\ \text{s.t.} \quad & -\sigma(0,x) + \theta \leq 0 \ \ \forall \sigma(\cdot) \in \overline{\Delta} \\ & -\tau(0,x) \qquad\ \leq 0 \ \ \forall \tau(\cdot) \in \overline{\Delta}^0 \\ & (x,\theta) \in X \times \mathbb{R} \end{aligned} \tag{5.4}$$

Here, $\widetilde{X}$ is any superset of X, independent of $\bar{x} \in X$, and $\overline{\Delta}$ ($\overline{\Delta}^0$) is a subset of Δ (Δ^0), where Δ is the common feasible set of the dual subprogrammes (5.3) and Δ^0 is obtained from the definition of Δ by changing the occurrence of $f(x,y)$ into 0 (cf. Subsection 2.1). **Constraint Decomposition** applied to (5.1), amounts to iterating between, on the one hand, the restricted primal master programmes

$$\mathcal{P}(\overline{\Pi}): \quad \max_{x,y} \quad f(x,y) \qquad \text{s.t.} \quad G(x,y) \diamond 0, \quad (x,y) \in \overline{\Pi} \tag{5.5}$$

and the relaxed dual master programmes

$$\begin{aligned} \mathcal{D}(\overline{\Pi}): \quad & \min_{g(\cdot)} \quad g(0) \\ \text{s.t.} \quad & g(G(x,y)) \geq f(x,y) \ \ \forall (x,y) \in \overline{\Pi} \\ & g(\cdot): \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\} \\ & g(r) \leq g(r') \ \ \forall r \diamond r' : \\ & \qquad (\exists (x,y) \in \Pi : \ G(x,y) \diamond r) \end{aligned} \tag{5.6}$$

and, on the other hand, the subprogrammes

$$\mathcal{CD}(\bar{g}(\cdot)): \quad \max_{x,y} \quad f(x,y) + \bar{g}(0) - \bar{g}(G(x,y)) \qquad \text{s.t.} \quad (x,y) \in U \cap (X \times Y) \tag{5.7}$$

Here, $\overline{\Pi} \subseteq \Pi = U \cap (X \times Y)$; see also Subsection 3.1. In most applications the master programmes $\mathcal{VD}(\overline{\Delta}, \overline{\Delta}^0)$, $\mathcal{P}(\overline{\Pi})$ and $\mathcal{D}(\overline{\Pi})$ are more difficult to solve than the corresponding subprogrammes $\mathcal{P}(\bar{x})$, $\mathcal{D}(\bar{x})$ and $\mathcal{CD}(\bar{g}(\cdot))$, which explains the incentive to ignore the master programmes as much as possible. The implication, of course, is that another device has to be chosen to feed the subprogrammes $\mathcal{P}(\bar{x})$, $\mathcal{D}(\bar{x})$ and $\mathcal{CD}(\bar{g}(\cdot))$, depending on whether Variable or Constraint Decomposition is applied. In this sense, Cross Decomposition is a variation on Variable Decomposition in that the relaxed master programmes $\mathcal{VD}(\overline{\Delta},\overline{\Delta}^0)$ in (5.4) are ignored in favour of the subprogrammes $\mathcal{CD}(\bar{g}(\cdot))$ in (5.7). More specifically, if a solution $\bar{\sigma}(\cdot)$ ($\bar{\tau}(\cdot)$) is obtained from the dual subprogramme $\mathcal{D}(\bar{x})$, then the subprogramme $\mathcal{CD}(\bar{g}(\cdot))$ is solved next, where $\bar{g}(\cdot) = \bar{\sigma}(\cdot,\bar{x})$ ($\bar{g}(\cdot) = \bar{\tau}(\cdot,\bar{x})$). This latter programme, in its turn, sends a primal solution $\bar{x}'$ back, so that in the next iteration $\mathcal{D}(\bar{x}')$ is going to be solved, etc. But if that is the way it is, then Cross Decomposition can just as well be considered to be a variation on Constraint Decomposition, in which the relaxed dual master programmes $\mathcal{D}(\overline{\Pi})$ in (5.6) are ignored in favour of the dual subprogrammes $\mathcal{D}(\bar{x})$ in (5.3)! To conclude, **Cross Decomposition** *is a variation on Variable Decomposition and Constraint Decomposition simultaneously.*

What about termination criteria? Obviously, all primal subprogrammes $\mathcal{P}(\bar{x})$ supply lower bounds for the optimal solution value of $\mathcal{P}$, and these lower bounds can be as tight as desired (cf. Theorem 2.3). On the other hand, the subprogrammes $\mathcal{CD}(\bar{g}(\cdot))$ with $\bar{g}(0) \in \mathbb{R}$, supply upper bounds for $\mathcal{P}$, and these upper bounds too, can be as tight as desired (cf. Theorem 3.2). Hence, if LB and UB denote the greatest lower bound and least upper bound which have been found so far, a natural termination criterion is $UB - LB \leq \epsilon_0$, where ϵ_0 is a (prescribed) tolerance level. In case this criterion is met, the current incumbent x^{inc} is an $(UB - LB)$-optimal solution for $\mathcal{P}$, which meets the accuracy required.

Although the incentive to ignore master programmes may be justified with good reasons, a price has to be paid; in the absence of master programmes, non-cyclic behaviour, let alone asymptotic convergence, can no longer be guaranteed. Intuitively, this can be explained by the fact that in Cross Decomposition only the most recently obtained information is taken into account. This is in sharp contrast with the use of master programmes in Variable and Constraint Decomposition, where an accumulation of information takes place.

The most natural way to avoid cycling is to rely on the variable or constraint decomposition procedures. Suppose that each time the cross decomposition procedure is in danger of generating a cycle, at least one full iteration of the variable (c.q. constraint) decomposition procedure is performed. In that case, all the nice properties of Variable (c.q. Constraint) Decomposition concerning non-cyclic behaviour pass on to this modified version of Cross Decomposition. It is even allowed to invoke the variable and constraint decomposition procedures alternately, as long as these procedures are called upon sufficiently often. Viewed in this way, the resulting procedure is a variation of Cross Decomposition, in the sense that the latter is equipped with a "spacer step" (see e.g. [Luenberger, 1984]). However, the resulting algorithm could just as well be considered as (a variation of) Variable (c.q. Constraint) Decomposition, where some extra time is spent in the subprogramme phase to build up the set Δ^{k+1} and/or $\Delta^{0,k+1}$ out of Δ^k and/or $\Delta^{0,k}$ (c.q. X^{k+1} out of X^k); see Subsections 2.1 and 3.1. The resulting procedure is therefore a ***hybrid*** approach. It can be visualized by Figure 5.1. Note that the diagram is a little bit deceptive because the entire picture can also be regarded as (a variation of) one of the three decomposition procedures individually. How to wander around in the diagram will heavily depend on the specific problem(-class) that has to be solved. A major concern remains, of course, how to safeguard the procedure against cycling. In [Van Roy, 1983] some anti-cycling strategies are given for the mixed-integer linear case, which are all based on the condition that the variable (c.q. constraint) decomposition procedure is called upon sufficiently often. For the general case we are dealing with, similar strategies can also be proven to prevent the procedure from cycling. For more detailed information on this, the reader is referred to [Flippo et al., 1987].

5.3 Kornai-Lipták Decomposition

As explained in the previous subsection, Cross Decomposition can be considered to be an iterative method in which at any moment, only

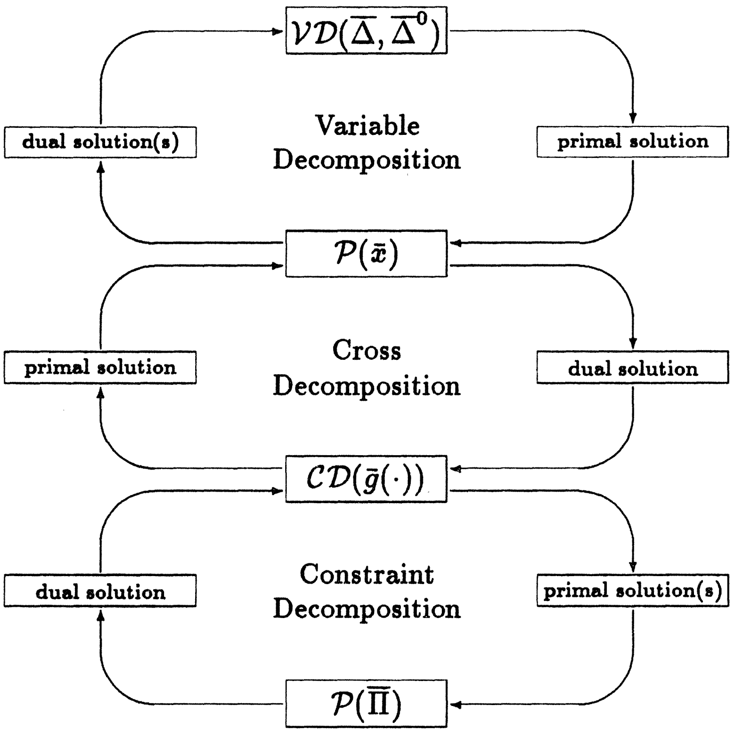

Figure 5.1: The three decomposition procedures related

the ***most recently obtained information*** is taken into account. An unfavourable consequence of this ignoring the past, is that cyclic behaviour can no longer be precluded. Strategies that are designed to prevent such behaviour do exist, but most of them, unfortunately, rely on the master programmes of the variable and/or constraint decomposition approach, in which the *entire past* has been accumulated. Kornai-Lipták Decomposition tries to meet these two extremes halfway; like Cross Decomposition it ignores the master programmes, but the information that is sent between the two subprogrammes $\mathcal{D}(\bar{x})$ and $\mathcal{CD}(\bar{g}(\cdot))$ partly takes the past into account. This is done by sending to the next programme to be solved, an *aggregation* of the past. Of course, such an approach would, from a memory-allocation point of view, only be an improvement compared to the variable and constraint decomposition procedures, if the aggregation formula can be updated from iteration to iteration, so that keeping track of the full past becomes unneccesary. For instance, as input for the subprogramme $\mathcal{D}(\bar{x})$, the arithmetic mean of all previously generated values for x could be used. A similar strategy for the dual information $\bar{g}(\cdot)$ is likely to be applicable, only if all generated dual solutions $g(\cdot)$ are finitely representable and if the aggregation formula is applied to the (finitely dimensional) parameter through which these dual solutions are characterized. For a specific class of linear programmes, asymptotic convergence of the Kornai-Lipták procedure has been established in case arithmetic means are used to aggregate both the intermediate primal and intermediate dual solutions; cf. [Kornai & Lipták, 1965], [Dirickx & Jennergren, 1979] and [Aardal & Ari, 1990].

Section 6

Summary

In part III we introduced and analysed extensions of two well-known decomposition methods, viz. Benders Decomposition and Dantzig-Wolfe Decomposition, to general mathematical programming problems. The former generalization is discussed in Section 2. The idea underlying this approach is also known as Generalized Benders Decomposition, Primal Decomposition and Resource Directive Decomposition. We prefer to call the approach *Variable Decomposition*, in order to reflect the partitioning of the set of variables into two mutually exclusive sets. In Section 2, the approach is also related to a part of existing literature, and an account on non-cyclic behaviour, asymptotic convergence and finite convergence is given as well. Finally, the approach is shown to separate integrality requirements from non-linearities, if it is applied to a mixed-integer non-linear programme with an underlying convex structure. A similar discussion on Dantzig-Wolfe Decomposition follows in Section 3. Generalizations of this approach are also known as Column Generation, Generalized Linear Programming, Dual Decomposition and Price Directive Decomposition. However, we chose to add the name *Constraint Decomposition*, to voice the partitioning of the set of constraints into two disjoint subsets. This approach too, is related to a part of existing literature, and non-cyclic behaviour, asymptotic convergence and finite convergence are also topics that are dealt with.

In Section 4, Variable and Constraint Decomposition are shown to be, in some sense, *dual* to one another, and finally, some *variations* are discussed in Section 5.

As already mentioned in Section 1, the main contributions of Part III arise from the generality of the discussion. On the one hand, *no restrictive assumptions on the original primal programme are made*, and on the other hand, *inaccuracies and duality gaps are allowed to occur*. As a result of this, former generalizations of Benders and Dantzig-Wolfe Decomposition are captured by ours. This statement also holds for the generalizations that are described in [Burkard et al., 1985], although a (straightforward) extension of our approach to the field of algebraic optimization remains to be carried out. More specifically, with respect to **Variable Decomposition** the following contributions have been made.

- For the applicability of the variable decomposition procedure, it is essential for the dual subprogrammes to have a *common feasible set*. This observation implies that the parameterization of the primal subprogrammes is to be modelled through their right-hand-sides, and that the parameterizing constraints are to be among the ones that are dualized. In fact, the need for constraint-qualifications in former generalizations of Benders Decomposition can be explained by a violation of this requirement (cf. Subsections 2.2 and 2.3).

- The results in Subsection 2.4 reveal the impact of inaccuracies and duality gaps during the iterative process, on possibly cyclic behaviour and finite convergence. The variable decomposition procedure can be *guaranteed not to cycle* if, quite naturally, at each iteration the overall accuracy required, *does not exceed the sum* of the inaccuracies that are already allowed in the blocks of which the procedure is composed. If, in addition, the primal or dual solution space is essentially *finite*, the finite convergence of the procedure is even guaranteed.

- Apart from some continuity and compactness conditions on problem data, *closedness* of the point-to-set map which largely describes the construction of the sequence of intermediate primal solutions x^k, and *stability* of the family of primal subprogrammes $\mathcal{P}(\cdot)$, turn out to be *sufficient conditions* for Variable Decomposition to converge asymptotically. Furthermore it is argued that, although neither of these two conditions is proven to be necessary from a strictly mathematical point of view, *asymptotic convergence can generally not be expected without them*; the notion of closedness is just too intimitely related to the notion of asymptotic convergence in general, whereas a meaningful convergence analysis for Variable Decomposition necessarily implies stability. After this result has been established, attention is restricted to the computationally relevant cases. This means that the solution space of the dual subprogrammes is restricted to *finitely representable functions* only. The aforementioned point-to-set map is then proven to be *closed*, if these finitely representable dual solutions are defined over a *compact* set and, additionally, meet some *continuity* requirements. Finally, it is demonstrated that *similar conditions* imply the *feasibility* of all the accumulation points of the sequence of primal solutions that are generated by the relaxed master programmes. A full account of these results can be found in Subsection 2.5.

- As is described in Subsection 2.6, an iterative scheme results between, on the one hand, a family of *(ordinary) convex subprogrammes*, and on the other hand, a family of *mixed-integer linear relaxed master programmes*, in case the variable decomposition procedure is applied to a *mixed-integer non-linear programme with an underlying convex structure*. In other words, in such programming problems, the integrality requirements can be separated from the non-linearities.

With respect to **Constraint Decomposition** similar results have been obtained. More specifically, the following contributions have been made.

- The results in Subsection 3.4 reveal the impact of inaccuracies and duality gaps during the iterative process, on possibly cyclic behaviour and finite convergence. The constraint decomposition procedure can be *guaranteed not to cycle* if, quite naturally, at each iteration the overall accuracy required, *does not exceed the sum* of the inaccuracies that are allowed in the blocks of which the procedure is composed. If, in addition, the primal or dual solution space is essentially *finite*, finite convergence of the procedure results.

- In Subsection 3.5 the constraint decomposition procedure is shown to *converge asymptotically* if, besides some compactness and continuity conditions on problem data, the solution spaces of the relaxed dual master programmes can be restricted to *finitely representable functions* only, which, additionally, are defined over a *compact* set and meet some *continuity* requirements.

- A straightforward extension of a known variation of Constraint Decomposition is presented in Section 5. This variation concerns *Lagrangean Decomposition*, or *Variable Splitting* as it is also sometimes referred to.

Results which relate to the **interaction between the two decomposition methods** are the following.

- The apparent similarity between Variable Decomposition and Constraint Decomposition is analysed in Section 4. There, the two methods are proven to be *dual approaches*, in that if the latter is applied to a mathematical programming problem, an identical iterative scheme results as if the former were applied to a well-chosen dual. The dual formulation that has to be considered for this purpose, is the *additively separable dual programme* which has been introduced in Section 5 of Part II.

- Straightforward extensions of two known variations on both methods *simultaneously* are presented in Section 5, viz. *Cross Decomposition* and *Kornai-Lipták Decomposition*.

As already mentioned in Subsection 3.5, there remains an annoying discrepancy between the generality of the asymptotic convergence results of Variable Decomposition on the one hand, and Constraint Decomposition on the other. In *Variable Decomposition*, we *did not* have to rely on *finitely representable* dual solutions in order to prove convergence; convergence was proven under a closedness condition on $\alpha(\beta(\cdot))$, which is the point-to-set map which largely describes the variable decomposition procedure. Furthermore, it was also proven that in case finitely representable and in some sense continuous dual solutions defined on a compact set existed, closedness of this point-to-set map followed. In proving the convergence of *Constraint Decomposition*, we *assumed* the existence of finitely representable and continuous dual solutions $w(\cdot)$ *right from the start*. This assumption, like in Variable Decomposition, implies closedness of the point-to-set map that largely describes the constraint decomposition procedure, in case some additional continuity and compactness conditions on problem data are met. Finally, it was also argued that it would probably be impossible to remove this discrepancy, unless maybe, the complex notion of normed functional spaces would be taken into consideration. How is it possible to be stuck in the seemingly conflicting situation where two methods, even in their full generality, can be proven to be dual to one another, while their asymptotic convergence results suffer from being equally general? It is our belief, that the asymptotic convergence results can only be expected to be "alike", if the primal and dual problems are. Recall that in general, the dual of an optimization problem in *finite* dimensions, is an optimization problem in *infinite* dimensions, which is of a completely different nature. Therefore, it seems that only in case the primal and dual programmes are both optimization problems in *finite* dimensions, the discrepancy in the asymptotic convergence results may vanish. And according to the results in Subsections 2.5 and 3.5, it does!

Epilogue

In the preceding discussion we analysed the notions of *Stability*, *Duality* and *Decomposition* in General Mathematical Programming. A full account of these notions is given in Parts I till III. For a list of new results we refer to the summaries of these parts. The preceding discussion also reveals the interrelations between the three notions. As already mentioned in the Prologue, duality theory appears to be an essential ingredient for both Variable and Constraint Decomposition. Furthermore, stability is proven to play a natural role in duality theory, if the one-to-one correspondence between constraints in the primal programme and variables in the dual is to be restored, and if the dual solutions are required to be real-valued. Stability is also proven to be an essential prerequisite for the asymptotic convergence of both decomposition methods.

In the Prologue we also mentioned the usual doubts about the use of theoretical research; in this respect, we argued that it seems impossible to encounter the limits of concepts and methods except by a process of abstraction. One might ask whether our analysis actually came up with any such limits. Fortunately, the answer is yes. First of all, we concluded in Section 4 of Part I that the three types of conditions which imply constraint-qualification in Karush-Kuhn-Tucker points (viz. linearity, convexity with Slater and MF-regularity) also seem to be essential in enforcing stability. Secondly, in Subsection 4.2 of Part II it was argued that the use of separable augmented Lagrangeans seems limited, unless at least one of the aforementioned three types of conditions is met. This observation, of course, strongly suggests the use of non-separable augmented Lagrangeans. Finally, the results in Subsection 2.5 of Part III indicate that stability is an essential prerequisite for the asymptotic convergence of Variable Decomposition. This implies that, again, one of the aforementioned three types of conditions should be met if Variable Decomposition is to converge. To conclude, the three types of conditions which enforce constraint-qualification in Karush-Kuhn-Tucker points, viz. linearity, convexity with Slater and MF-regularity, also play a major role in stability, duality and decomposition. In this respect, they seem to indicate whether a mathematical programme can be considered to be "well-behaved".

Of course, many things still remain to be explored. A natural research topic in the general duality theory we discussed in Part II, for instance, would be to look for necessary and/or sufficient conditions for the existence of *finitely dimensional dual programmes* that preserve *strong duality*. The importance of such a result follows from the close connections between finitely representable dual solutions, and the approaches based on *augmented Lagrangeans* (cf. Subsection 4.2 of Part II). As far as the decomposition methods in Part III are concerned, it would be worthwhile to explore under which conditions the asymptotic convergence results in Variable Decomposition remain unaffected, if constraints in the relaxed master programmes are allowed to be dropped; with respect to Constraint Decomposition a similar issue applies. Only with such a strategy one may hope to keep the problems in the master programme phase of reasonable size. The general theory on convergence of cutting plane algorithms, as developed in [Eaves & Zangwill, 1971], may thereby prove to be extremely useful. Finally, there are many implementation issues that are yet unresolved. To name only one, it will most likely benefit the performance of the variable decomposition procedure, if one chooses in an "intelligent" way from all the (near-) optimal dual solutions that qualify to be added as a cut to the relaxed master programme, the one that actually will. For an elaboration of this idea we refer to [Magnanti & Wong, 1981].

Author index

Subject index

References

Part I

- Aubin J.-P. (1984), "Lipschitz behavior of solutions to convex minimization problems", *Mathematics of Operations Research 9*, pp. 87–111.
- Auslander A. (1979), "Differentiable stability in Non Convex and Non Differentiable Programming", *Mathematical Programming Study 10*, pp. 29–41.
- Auslander A. (1984), "Stability in Mathematical Programming with nondifferentiable data", *SIAM Journal on Control and Optimization 22*, pp. 239–254.
- Bank B., J. Guddat, D. Klatte, B. Kummer & K. Tammer (1983), *Non-linear parametric optimization*, Birkhäuser Verlag, Basel-Boston-Stuttgart.
- Bank B. & R. Hansel (1984), "Stability of mixed-integer quadratic programming problems", *Mathematical Programming Study 21*, pp. 1–17.
- Bereanu B. (1976), "The continuity of the optimum in Parametric Programming and applications to Stochastic Programming", *Journal of Optimization Theory and Applications 18*, pp. 319–333.

- Berge C. (1963), *Topological spaces*, MacMillan, New York.

- Böhm V. (1975), "On the continuity of the optimal policy set for linear programs", *Journal of the Society for Industrial and Applied Mathematics 28*, pp. 303–306.

- Brosowski B. (1984), "Parametric Semi-Infinite Linear Programming I. Continuity of the feasible set and of the optimal value", *Mathematical Programming Study 21*, pp. 18–42.

- Cornet B. & J.-Ph. Vial (1986), "Lipschitzian solutions of perturbed nonlinear programming problems", *SIAM Journal on Control and Optimization 24*, pp. 1123–1137.

- Daniel J.W. (1973[a]), "On perturbations in systems of linear inequalities", *SIAM Journal on Numerical Analysis 10*, pp. 299–307.

- Daniel J.W. (1973[b]), "Stability of the solution of definite quadratic programs", *Mathematical Programming 5*, pp. 41–53.

- Daniel J.W. (1975), "Remarks on perturbations in systems of linear inequalities", *SIAM Journal on Numerical Analysis 12*, pp. 770–771.

- Dantzig G.B., J. Folkman & N. Shapiro (1967), "On the continuity of the minimum set of a continuous function", *Journal of Mathematical Analysis and Applications 17*, pp. 519–548.

- Debreu G. (1959), Theory of value, *Cowles Foundation Monograph 17*, Wiley, New York.

- Demyanov V.F. & I.S. Zabrodin (1986), "Directional differentiability of a continual maximum function of quasi-differentiable functions", *Mathematical Programming Study 29*, pp. 108–117.

- Dirickx Y.M.I., L.P. Jennergren & D.W. Peterson (1972), "A note on the continuity of a fundamental function in Mathematical Programming", *Preprint I/72-34*, International Institute of Management, D-1000 Berlin 33, Griegstraße 5, Germany.

- Dontchev A.L. & H.Th. Jongen (1986), "On the regularity of the Kuhn-Tucker curve", *SIAM Journal on Control and Optimization 24*, pp. 169–176.

- Evans J.P. & F.J. Gould (1970), "Stability in Nonlinear Programming", *Operations Research 18*, pp. 107–118.

- Fiacco A.V. (1976), "Sensitivity analysis for Nonlinear Programming using penalty methods", *Mathematical Programming 10*, pp. 287–311.

- Fiacco A.V. (1983), *Introduction to Sensitivity and Stability Analysis in Nonlinear Programming*, Academic Press, New York.

- Fiacco A.V. & W.P. Hutzler (1982), "Basic results in the development of sensitivity and stability analysis in Nonlinear Programming", *Computers & OR 9*, pp. 9–28.

- Fujiwara O. (1985), "A note on differentiability of global optimal values", *Mathematics of Operations Research 10*, pp. 612–618.

- Gauvin J. (1979), "The generalized gradient of a marginal function in Mathematical Programming", *Mathematics of Operations Research 4*, pp. 458–463.

- Gauvin J. & F. Dubeau (1982), "Differential properties of the marginal function", *Mathematical Programming Study 19*, pp. 101–119.

- Gollan B. (1984[a]), "Inner estimates for the generalized gradient of the optimal value function in Nonlinear Programming", *Mathematical Programming Study 22*, pp. 132–146.

- Gollan B. (1984[b]), "On the marginal function in Nonlinear Programming", *Mathematics of Operations Research 9*, pp. 208–221.

- Greenberg H.J. & W.P. Pierskalla (1972), "Extensions of the Evans-Gould stability theorems for mathematical programs", *Operations Research 20*, pp. 143–153.

- Greenberg H.J. & W.P. Pierskalla (1975), "Stability theorems for infinitely constrained mathematical programs", *Journal of Optimization Theory and Applications 16*, pp. 409–428.

- Guddat J. (1976), "Stability in Convex Quadratic Parametric Programming", *Mathematische Operationsforschung und Statistik 7*, pp. 223–245.

- Hoffman A.J. (1952), "On approximate solutions of systems of linear inequalities", *Journal of Research National Bureau of Standards 49*, pp. 263–265.

- Hogan W.W. (1973[a]), "The continuity of the perturbation function of a convex program", *Operations Research 21*, pp. 351–352.

- Hogan W.W. (1973[b]), "Point-to-set maps in Mathematical Programming", *SIAM Review 15*, pp. 591–603.

- Hogan W.W. (1973[c]), "Directional derivatives for extremal-value functions with applications to the completely convex case", *Operations Research 21*, pp. 188–209.

- Janin R. (1984), "Directional derivative of the marginal function in Non-Linear Programming", *Mathematical Programming Study 21*, pp. 110–126.

- Jittorntrum K. (1984), "Solution point differentiability without strict complementarity in Nonlinear Programming", *Mathematical Programming Study 21*, pp. 127–138.

- Kojima M. & R. Hirabayashi (1984), "Continuous deformation of nonlinear programs", *Mathematical Programming Study 21*, pp. 150–198.

- Kyparisis J. & A.V. Fiacco (1987), "Generalized convexity and concavity of the optimal value function in Nonlinear Programming", *Mathematical Programming 39*, pp. 285–304.

- Mangasarian O.L. & S. Fromovitz (1967), "The Fritz John necessary optimality conditions in the presence of equality and inequality constraints", *Journal of Mathematical Analysis and Applications 17*, pp. 37–47.

- Mangasarian O.L. & R.R. Meyer (1979), "Nonlinear perturbation of linear programs", *SIAM Journal on Control and Optimization 17*, pp. 745–752.

- Mangasarian O.L. & T.H. Shiau (1987), "Lipschitz continuity of solutions of linear inequalities, programs and complementarity problems", *SIAM Journal on Control and Optimization 25*, pp. 583–595.

- Martin D.H. (1975), "On the continuity of the maximum in parametric Linear Programming", *Journal of Optimization Theory and Applications 17*, 205–210.

- Meyer R. (1970), "The validity of a family of optimization methods", *SIAM Journal on Control 8*, pp. 41–54.

- Minoux M. (1986), *Mathematical Programming: theory and algorithms*, Wiley, New York.

- Penot J.-P. (1984), "Differentiability of relations and differential stability of perturbed optimization problems", *SIAM Journal on Control and Optimization 22*, 529–551.

- Penot J.-P. (1988), "Erratum: Differentiability of relations and differential stability of perturbed optimization problems", *SIAM Journal on Control and Optimization 26*, 997–998.

- Robinson S.M. (1974), "Perturbed Kuhn-Tucker points and rates of convergence for a class of nonlinear programming algorithms", *Mathematical Programming 7*, pp. 1–16.

- Robinson S.M. (1975), "Stability theory for systems of inequalities. Part I: Linear systems", *SIAM Journal on Numerical Analysis 12*, pp. 754–769.

- Robinson S.M. (1976), "Stability theory for systems of inequalities. Part II: Differentiable nonlinear systems", *SIAM Journal on Numerical Analysis*, pp. 497–513.

- Robinson S.M. (1977), "A characterization of stability in Linear Programming", *Operations Research 25*, pp. 435–447.

- Robinson S.M. (1980), "Strongly regular generalized equations", *Mathematics of Operations Research 5*, pp. 43–62.

- Robinson S.M. (1982), "Generalized equations and their solutions. Part II: Applications to Nonlinear Programming", *Mathematical Programming Study 19*, pp. 200–221.

- Robinson S.M. (1987), "Local structure of feasible sets in Nonlinear Programming. Part III: Stability and sensitivity", *Mathematical Programming Study 30*, pp. 45–66.

- Rockafellar R.T. (1982), "Lagrange multipliers and subderivatives of optimal value functions in Nonlinear Programming", *Mathematical Programming Study 17*, pp. 28–66.

- Rockafellar R.T. (1984), "Directional differentiability of the optimal value-function in a nonlinear programming problem", *Mathematical Programming Study 21*, pp. 213–226.

- Seeger A. (1988), "Second-order directional derivatives in parametric optimization", *Mathematics of Operations Research 13*, pp. 124–139.

- Shapiro A. (1985[a]), "Second-order sensitivity analysis and asymptotic theory of parameterized nonlinear programs, *Mathematical Programming 33*, pp. 280–289.

- Shapiro A. (1985[b]), "Second-order derivatives of extremal-value functions and optimality conditions for semi-infinite programs", *Mathematics of Operations Research 10*, pp. 207–219.

- Shapiro A. (1988), "Sensitivity analysis of nonlinear programs and differentiability properties of metric projections", *SIAM Journal on Control and Optimization 26*, pp. 628–645.

- Spingarn J.E. (1980), "Fixed and variable constraints in sensitivity analysis", *SIAM Journal on Control and Optimization 18*, pp. 297-310.

- Stern M.H. & D.M. Topkis (1976), "Rates of stability in Nonlinear Programming", *Operations Research 24*, pp. 462–476.

- Wets R.J.-B. (1985), "On the continuity of the value of a linear program and of related polyhedral-valued multifunctions", *Mathematical Programming Study 24*, pp. 14–29.

- Zencke P. & R. Hettich (1987), "Directional derivatives for the value-function in Semi-Infinite Programming", *Mathematical Programming 38*, pp. 323–340.

Part II

- Bertsekas D.P. (1982), "Constrained optimization and Lagrange multiplier methods", Academic Press, Series: Computer Science and Applied Mathematics, New York etc.

- Burdet C.A. & E.L. Johnson (1977), "A subadditive approach to solve linear integer programs", *Annals of Discrete Mathematics 1*, pp. 117–144.

- Cottle R.W. (1963), "Symmetric dual quadratic programs", *Quarterly of Applied Mathematics 21*, pp. 237–243.

- Dorn W.S. (1960), "Duality in Quadratic Programming", *Quarterly of Applied Mathematics 18*, pp. 155–161.

- Everett H. (1963), "Generalized Lagrange multiplier method for solving problems of optimum allocation of resources", *Operations Research 11*, pp. 399–417.

- Geoffrion A.M. (1972), "3. Duality in Nonlinear Programming: a simplified applications-oriented development", *Perspectives on Optimization: A Collection of Expository Articles*, A.M. Geoffrion (ed.), Addison-Wesley Publishing Company Inc., Reading Massachusetts.

- Gould F.J. (1969), "Extensions of Lagrange multipliers in nonlinear programming", *SIAM Journal on Applied Mathematics 17*, pp. 1280–1297.

- Gould F.J. (1972), "Nonlinear duality theorems", *Cahiers du Centre d'Etudes de Recherche Opérationnelle 14*, pp. 196–212.

- Jeroslow R.G. (1979), "Minimal inequalities", *Mathematical Programming 17*, pp. 1–15.

- Ponstein J. (1983), "Comments on the general duality survey by J. Tind and L.A. Wolsey", *Mathematical Programming 25*, pp. 240–244.

- Rockafellar R.T. (1970), *Convex analysis*, Princeton University Press, Princeton, New Jersey.

- Rockafellar R.T. (1974), "Augmented Lagrange multiplier functions and duality in Nonconvex Programming", *SIAM Journal on Control 12*, pp. 268–285.

- Tind J. & L.A. Wolsey (1981), "An elementary survey of general duality theory in Mathematical Programming", *Mathematical Programming 21*, pp. 241–261.

- Wolsey L.A. (1981), "Integer programming duality: price functions and sensitivity analysis", *Mathematical Programming 20*, pp. 173–195.

Part III

- Aardal K. & A. Ari (1990), "On the resemblance between the Kornai-Lipták and cross decomposition techniques for block-angular linear programs", *European Journal of Operational Research*, to appear.
- Armstrong R.D. & C.E. Willis (1977), "Simultaneous investment and allocation decisions applied to water planning", *Management Science 23*, pp. 1080–1088.
- Assad A.A. (1978), "Multicommodity network flows — a survey", *Networks 8*, pp. 37–91.
- Balas E. (1969), "Duality in Discrete Programming: II. The quadratic case", *Management Science 16*, pp. 14–32.
- Bazaraa M.S. & H.D. Sherali (1980), "Benders' partitioning scheme applied to a new formulation of the quadratic assignment problem", *Naval Research Logistics Quarterly 27*, pp. 29–41.
- Benders J.F. (1962), "Partitioning procedures for solving mixed variables programming problems", *Numerische Mathematik 4*, pp. 238–252.
- Bloom J.A. (1983), "Solving an electricity generating capacity expansion planning problem by generalized Benders' Decomposition", *Operations Research 31*, pp. 84–100.
- Burkard R.E., H.W. Hamacher & J. Tind (1985), "On general decompostion schemes in Mathematical Programming", *Mathematical Programming Study 24* (Special Issue), pp. 238–252.
- Clasen R.J. (1984), "The solution of the chemical equilibrium programming problem with generalized Benders Decomposition", *Operations Research 32*, pp. 70–79.
- Cottle R.W. (1963), "Symmetric dual quadratic programs", *Quarterly of Applied Mathematics 21*, pp. 237–243.

- Crainic T. & J.M. Rousseau (1987), "The column generation principle and the airline crew scheduling problem", *INFOR 25*, pp. 136–151.
- Dantzig G.B. (1963), *Linear Programming and extensions*, Princeton University Press, Princeton, New Jersey.
- Dantzig G.B. & P. Wolfe (1960), "Decomposition principle for Linear Programming", *Operations Research 8*, pp. 101–111.
- Desrosiers J., F. Soumis & M. Desrochers (1984), "Routing with time windows by column generation", *Networks 14*, pp. 545–565.
- Dirickx Y.M.I. & L.P. Jennergren (1979), *System analysis by multi-level methods with applications to economics and management*, Wiley, New York.
- Dorn W.S. (1960), "Duality in Quadratic Programming", *Quarterly of Applied Mathematics 18*, pp. 155–161.
- Duran M.A. & I.E. Grossmann (1986), "An outer-approximation algorithm for a class of mixed-integer nonlinear programs", *Mathematical Programming 36*, pp. 307–339.
- Dyer M.E. & J. Walker (1982), "Solving the subproblem in the Lagrangian dual of separable discrete programs with linear constraints", *Mathematical Programming 24*, pp. 107–112.
- Dzielinski B.P. & R.E. Gomory (1965), "Optimal programming of lot sizes, inventory and labor allocations", *Management Science 11*, pp. 874–890.
- Eaves B.C. & W.I. Zangwill (1971), "Generalized cutting plane algorithms", *SIAM Journal on Control 9*, pp. 529–542.
- Fisher M.L. (1981), "The Lagrangian Relaxation method for solving integer programming problems", *Management Science 27*, pp. 1–18.

- Fisher M.L., W.D. Northup & J.F. Shapiro (1975), "Using duality to solve discrete optimization problems: theory and computational experience", *Mathematical Programming Study 3*, pp. 56–94.

- Flippo O.E. & A.H.G. Rinnooy Kan (1990), "A note on Benders Decomposition in Mixed-Integer Quadratic Programming", *Operations Research Letters 9* (to appear).

- Flippo O.E., A.H.G. Rinnooy Kan & G. Van der Hoek (1987), "Duality and decomposition in General Mathematical Programming", *Report 8747/B*, Econometric Institute, Erasmus University Rotterdam, Rotterdam, The Netherlands.

- Geoffrion A.M. (1972[a]), "2. Elements of large-scale Mathematical Programming. Part I: Concepts. Part II: Synthesis of algorithms and bibliography", *Perspectives on Optimization: A Collection of Expository Articles*, A.M. Geoffrion (ed.), Addison-Wesley Publishing Company Inc., Reading Massachusetts.

- Geoffrion A.M. (1972[b]), "3. Duality in Nonlinear Programming: a simplified applications-oriented development", *Perspectives on Optimization: A Collection of Expository Articles*, A.M. Geoffrion (ed.), Addison-Wesley Publishing Company Inc., Reading Massachusetts.

- Geoffrion A.M. (1972[c]), "Generalized Benders Decomposition", *Journal of Optimization Theory and Application 10*, pp. 237–260.

- Geoffrion A.M. (1974), "Lagrangean Relaxation for Integer Programming", *Mathematical Programming Study 2*, pp. 82–114.

- Geoffrion A.M. & G.W. Graves (1974), "Multicommodity distribution system design by Benders' Decomposition", *Management Science 20*, pp. 822–844.

- Geoffrion A.M. & R.E. Marsten (1972), "5. Integer Programming algorithms: a framework and state-of-the art survey", *Perspectives on Optimization: A Collection of Expository Articles*, A.M. Geoffrion (ed.), Addison-Wesley Publishing Company Inc., Reading Massachusetts.

- Gilmore P.C. & R.E. Gomory (1961), "A linear programming approach to the cutting-stock problem", *Operations Research 9*, pp. 849–859.

- Gilmore P.C. & R.E. Gomory (1963), "A linear programming approach to the cutting-stock problem — Part II", *Operations Research 11*, pp. 863–888.

- Goffin J.L. (1977), "On convergence rates of subgradient optimization methods", *Mathematical Programming 13*, pp. 329–347.

- Guignard M. & S. Kim (1987), "Lagrangean Decomposition: a model yielding stronger Lagrangean bounds", *Mathematical Programming 39*, pp. 215–228.

- Hogan W.W. (1973), "Applications of a general convergence theory for outer-approximation algorithms", *Mathematical Programming 5*, pp. 151–168.

- Holm S. & J. Tind (1985), "Decomposition in Integer Programming by superadditive functions", *Working Paper 32-85-86*, Graduate School of Industrial Administration, Carnegie-Mellon University, Pittsburgh, USA.

- Holmberg K. (1985), *Decomposition in large scale Mathematical Programming*, Doctoral Dissertation, Department of Mathematics, Linköping University, Linköping, Sweden.

- Jörnsten K.O., M. Näsberg & P.A. Smeds (1985), "Variable Splitting — a new Lagrangean relaxation approach to some mathematical programming models", *Working Paper LiTh-MAT-R-85-04*, Linköping University, Linköping, Sweden.

- Kornai J. & Th. Lipták (1965), "Two-level planning", *Econometrica 33*, pp. 141–169.

- Lasdon L.S. (1970), *Optimization theory for large systems*, The MacMillan Company, London.

- Lazimy R. (1982), "Mixed-Integer Quadratic Programming", *Mathematical Programming 22*, pp. 332–349.
- Lazimy R. (1985), "Improved algorithm for mixed-integer quadratic programs and a computational study", *Mathematical Programming 32*, pp. 100–113.
- Luenberger D.G. (1984), *Linear and Nonlinear Programming*, 2^{nd} edition, Addison-Wesley Publishing Company Inc., Reading Massachusetts.
- Magnanti T.L., J.F. Shapiro & M.H. Wagner (1976), "Generalized Linear Programming solves the dual", *Management Science 22*, pp. 1195–1203.
- Magnanti T.L. & R.T. Wong (1981), "Accelerating Benders Decomposition: algorithmic enhancement and model selection criteria", *Operations Research 29*, pp. 464–484.
- McDaniel D. & M. Devine (1977), "A modified Benders' partitioning algorithm for Mixed Integer Programming", *Management Science 24*, pp. 312– 319.
- Minoux M. (1986), *Mathematical Programming: theory and algorithms*, Wiley, New York.
- Noonan F. & R.J. Giglio (1977), "Planning electric power generation: a nonlinear mixed integer model employing Benders Decomposition", *Management Science 23*, pp. 946–956.
- Rockafellar R.T. (1970), *Convex analysis*, Princeton University Press, Princeton, New Jersey.
- Rouhani R., L. Lasdon, W. Lebow & A.D. Waren (1985), "A generalized Benders decomposition approach to reactive source planning in power systems", *Mathematical Programming Study 25*, pp. 62–75.
- Sweeney D.J. & R.A. Murphy (1979), "A method of decomposition for integer programs", *Operations Research 27*, pp. 1128–1141.

- Tind J. & S. Holm (1986), "A unified approach for price-directive decomposition procedures in Integer Programming", *Working Paper 40-85-86*, Graduate School of Industrial Administration, Carnegie-Mellon University, Pittsburgh, USA.

- Van Roy T.J. (1980), *Cross Decomposition for large-scale Mixed Integer Linear Programming with applications to facility location on distribution networks*, Doctoral Dissertation, Applied Sciences, Katholieke Universiteit Leuven, Leuven, Belgium.

- Van Roy T.J. (1983), "Cross Decomposition for Mixed Integer Programming", *Mathematical Programming 25*, pp. 46–63.

- Warner D.M. & J. Prawda (1972), "A mathematical programming model for scheduling nursing personnel in a hospital", *Management Science 19*, pp. 411–422.

- Wilson R. (1966), "Programming variable factors", *Management Science 13*, pp. 144–151.

- Wolsey L.A. (1981), "A resource decomposition algorithm for general mathematical programs", *Mathematical Programming Study 14*, pp. 244–257.

- Zangwill W.I. (1969), *Nonlinear Programming: a unified approach*, Prentice-Hall Inc., Englewood Cliffs, New Jersey.

MATHEMATICAL CENTRE TRACTS

1 T. van der Walt. *Fixed and almost fixed points.* 1963.

2 A.R. Bloemena. *Sampling from a graph.* 1964.

3 G. de Leve. *Generalized Markovian decision processes, part I: model and method.* 1964.

4 G. de Leve. *Generalized Markovian decision processes, part II: probabilistic background.* 1964.

5 G. de Leve, H.C. Tijms, P.J. Weeda. *Generalized Markovian decision processes, applications.* 1970.

6 M.A. Maurice. *Compact ordered spaces.* 1964.

7 W.R. van Zwet. *Convex transformations of random variables.* 1964.

8 J.A. Zonneveld. *Automatic numerical integration.* 1964.

9 P.C. Baayen. *Universal morphisms.* 1964.

10 E.M. de Jager. *Applications of distributions in mathematical physics.* 1964.

11 A.B. Paalman-de Miranda. *Topological semigroups.* 1964.

12 J.A.Th.M. van Berckel, H. Brandt Corstius, R.J. Mokken, A. van Wijngaarden. *Formal properties of newspaper Dutch.* 1965.

13 H.A. Lauwerier. *Asymptotic expansions.* 1966, out of print; replaced by MCT 54.

14 H.A. Lauwerier. *Calculus of variations in mathematical physics.* 1966.

15 R. Doornbos. *Slippage tests.* 1966.

16 J.W. de Bakker. *Formal definition of programming languages with an application to the definition of ALGOL 60.* 1967.

17 R.P. van de Riet. *Formula manipulation in ALGOL 60, part 1.* 1968.

18 R.P. van de Riet. *Formula manipulation in ALGOL 60, part 2.* 1968.

19 J. van der Slot. *Some properties related to compactness.* 1968.

20 P.J. van der Houwen. *Finite difference methods for solving partial differential equations.* 1968.

21 E. Wattel. *The compactness operator in set theory and topology.* 1968.

22 T.J. Dekker. *ALGOL 60 procedures in numerical algebra, part 1.* 1968.

23 T.J. Dekker, W. Hoffmann. *ALGOL 60 procedures in numerical algebra, part 2.* 1968.

24 J.W. de Bakker. *Recursive procedures.* 1971.

25 E.R. Paërl. *Representations of the Lorentz group and projective geometry.* 1969.

26 European Meeting 1968. *Selected statistical papers, part I.* 1968.

27 European Meeting 1968. *Selected statistical papers, part II.* 1968.

28 J. Oosterhoff. *Combination of one-sided statistical tests.* 1969.

29 J. Verhoeff. *Error detecting decimal codes.* 1969.

30 H. Brandt Corstius. *Exercises in computational linguistics.* 1970.

31 W. Molenaar. *Approximations to the Poisson, binomial and hypergeometric distribution functions.* 1970.

32 L. de Haan. *On regular variation and its application to the weak convergence of sample extremes.* 1970.

33 F.W. Steutel. *Preservation of infinite divisibility under mixing and related topics.* 1970.

34 I. Juhász, A. Verbeek, N.S. Kroonenberg. *Cardinal functions in topology.* 1971.

35 M.H. van Emden. *An analysis of complexity.* 1971.

36 J. Grasman. *On the birth of boundary layers.* 1971.

37 J.W. de Bakker, G.A. Blaauw, A.J.W. Duijvestijn, E.W. Dijkstra, P.J. van der Houwen, G.A.M. Kamsteeg-Kemper, F.E.J. Kruseman Aretz, W.L. van der Poel, J.P. Schaap-Kruseman, M.V. Wilkes, G. Zoutendijk. *MC-25 Informatica Symposium.* 1971.

38 W.A. Verloren van Themaat. *Automatic analysis of Dutch compound words.* 1972.

39 H. Bavinck. *Jacobi series and approximation.* 1972.

40 H.C. Tijms. *Analysis of (s,S) inventory models.* 1972.

41 A. Verbeek. *Superextensions of topological spaces.* 1972.

42 W. Vervaat. *Success epochs in Bernoulli trials (with applications in number theory).* 1972.

43 F.H. Ruymgaart. *Asymptotic theory of rank tests for independence.* 1973.

44 H. Bart. *Meromorphic operator valued functions.* 1973.

45 A.A. Balkema. *Monotone transformations and limit laws.* 1973.

46 R.P. van de Riet. *ABC ALGOL, a portable language for formula manipulation systems, part 1: the language.* 1973.

47 R.P. van de Riet. *ABC ALGOL, a portable language for formula manipulation systems, part 2: the compiler.* 1973.

48 F.E.J. Kruseman Aretz, P.J.W. ten Hagen, H.L. Oudshoorn. *An ALGOL 60 compiler in ALGOL 60, text of the MC-compiler for the EL-X8.* 1973.

49 H. Kok. *Connected orderable spaces.* 1974.

50 A. van Wijngaarden, B.J. Mailloux, J.E.L. Peck, C.H.A. Koster, M. Sintzoff, C.H. Lindsey, L.G.L.T. Meertens, R.G. Fisker (eds.). *Revised report on the algorithmic language ALGOL 68.* 1976.

51 A. Hordijk. *Dynamic programming and Markov potential theory.* 1974.

52 P.C. Baayen (ed.). *Topological structures.* 1974.

53 M.J. Faber. *Metrizability in generalized ordered spaces.* 1974.

54 H.A. Lauwerier. *Asymptotic analysis, part 1.* 1974.

55 M. Hall, Jr., J.H. van Lint (eds.). *Combinatorics, part 1: theory of designs, finite geometry and coding theory.* 1974.

56 M. Hall, Jr., J.H. van Lint (eds.). *Combinatorics, part 2: graph theory, foundations, partitions and combinatorial geometry.* 1974.

57 M. Hall, Jr., J.H. van Lint (eds.). *Combinatorics, part 3: combinatorial group theory.* 1974.

58 W. Albers. *Asymptotic expansions and the deficiency concept in statistics.* 1975.

59 J.L. Mijnheer. *Sample path properties of stable processes.* 1975.

60 F. Göbel. *Queueing models involving buffers.* 1975.

63 J.W. de Bakker (ed.). *Foundations of computer science.* 1975.

64 W.J. de Schipper. *Symmetric closed categories.* 1975.

65 J. de Vries. *Topological transformation groups, 1: a categorical approach.* 1975.

66 H.G.J. Pijls. *Logically convex algebras in spectral theory and eigenfunction expansions.* 1976.

68 P.P.N. de Groen. *Singularly perturbed differential operators of second order.* 1976.

69 J.K. Lenstra. *Sequencing by enumerative methods.* 1977.

70 W.P. de Roever, Jr. *Recursive program schemes: semantics and proof theory.* 1976.

71 J.A.E.E. van Nunen. *Contracting Markov decision processes.* 1976.

72 J.K.M. Jansen. *Simple periodic and non-periodic Lamé functions and their applications in the theory of conical waveguides.* 1977.

73 D.M.R. Leivant. *Absoluteness of intuitionistic logic.* 1979.

74 H.J.J. te Riele. *A theoretical and computational study of generalized aliquot sequences.* 1976.

75 A.E. Brouwer. *Treelike spaces and related connected topological spaces.* 1977.

76 M. Rem. *Associons and the closure statement.* 1976.

77 W.C.M. Kallenberg. *Asymptotic optimality of likelihood ratio tests in exponential families.* 1978.

78 E. de Jonge, A.C.M. van Rooij. *Introduction to Riesz spaces.* 1977.

79 M.C.A. van Zuijlen. *Emperical distributions and rank statistics.* 1977.

80 P.W. Hemker. *A numerical study of stiff two-point boundary problems.* 1977.

81 K.R. Apt, J.W. de Bakker (eds.). *Foundations of computer science II, part 1.* 1976.

82 K.R. Apt, J.W. de Bakker (eds.). *Foundations of computer science II, part 2.* 1976.

83 L.S. van Benthem Jutting. *Checking Landau's "Grundlagen" in the AUTOMATH system.* 1979.

84 H.L.L. Busard. *The translation of the elements of Euclid from the Arabic into Latin by Hermann of Carinthia (?), books vii-xii.* 1977.

85 J. van Mill. *Supercompactness and Wallman spaces.* 1977.

86 S.G. van der Meulen, M. Veldhorst. *Torrix I, a programming system for operations on vectors and matrices over arbitrary fields and of variable size.* 1978.

88 A. Schrijver. *Matroids and linking systems.* 1977.

89 J.W. de Roever. *Complex Fourier transformation and analytic functionals with unbounded carriers.* 1978.

90 L.P.J. Groenewegen. *Characterization of optimal strategies in dynamic games.* 1981.

91 J.M. Geysel. *Transcendence in fields of positive characteristic.* 1979.

92 P.J. Weeda. *Finite generalized Markov programming.* 1979.

93 H.C. Tijms, J. Wessels (eds.). *Markov decision theory.* 1977.

94 A. Bijlsma. *Simultaneous approximations in transcendental number theory.* 1978.

95 K.M. van Hee. *Bayesian control of Markov chains.* 1978.

96 P.M.B. Vitányi. *Lindenmayer systems: structure, languages, and growth functions.* 1980.

97 A. Federgruen. *Markovian control problems; functional equations and algorithms.* 1984.

98 R. Geel. *Singular perturbations of hyperbolic type.* 1978.

99 J.K. Lenstra, A.H.G. Rinnooy Kan, P. van Emde Boas (eds.). *Interfaces between computer science and operations research.* 1978.

100 P.C. Baayen, D. van Dulst, J. Oosterhoff (eds.). *Proceedings bicentennial congress of the Wiskundig Genootschap, part 1.* 1979.

101 P.C. Baayen, D. van Dulst, J. Oosterhoff (eds.). *Proceedings bicentennial congress of the Wiskundig Genootschap, part 2.* 1979.

102 D. van Dulst. *Reflexive and superreflexive Banach spaces.* 1978.

103 K. van Harn. *Classifying infinitely divisible distributions by functional equations.* 1978.

104 J.M. van Wouwe. *Go-spaces and generalizations of metrizability.* 1979.

105 R. Helmers. *Edgeworth expansions for linear combinations of order statistics.* 1982.

106 A. Schrijver (ed.). *Packing and covering in combinatorics.* 1979.

107 C. den Heijer. *The numerical solution of nonlinear operator equations by imbedding methods.* 1979.

108 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science III, part 1.* 1979.

109 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science III, part 2.* 1979.

110 J.C. van Vliet. *ALGOL 68 transput, part I: historical review and discussion of the implementation model.* 1979.

111 J.C. van Vliet. *ALGOL 68 transput, part II: an implementation model.* 1979.

112 H.C.P. Berbee. *Random walks with stationary increments and renewal theory.* 1979.

113 T.A.B. Snijders. *Asymptotic optimality theory for testing problems with restricted alternatives.* 1979.

114 A.J.E.M. Janssen. *Application of the Wigner distribution to harmonic analysis of generalized stochastic processes.* 1979.

115 P.C. Baayen, J. van Mill (eds.). *Topological structures II, part 1.* 1979.

116 P.C. Baayen, J. van Mill (eds.). *Topological structures II, part 2.* 1979.

117 P.J.M. Kallenberg. *Branching processes with continuous state space.* 1979.

118 P. Groeneboom. *Large deviations and asymptotic efficiencies.* 1980.

119 F.J. Peters. *Sparse matrices and substructures, with a novel implementation of finite element algorithms.* 1980.

120 W.P.M. de Ruyter. *On the asymptotic analysis of large-scale ocean circulation.* 1980.

121 W.H. Haemers. *Eigenvalue techniques in design and graph theory.* 1980.

122 J.C.P. Bus. *Numerical solution of systems of nonlinear equations.* 1980.

123 I. Yuhász. *Cardinal functions in topology - ten years later.* 1980.

124 R.D. Gill. *Censoring and stochastic integrals.* 1980.

125 R. Eising. *2-D systems, an algebraic approach.* 1980.

126 G. van der Hoek. *Reduction methods in nonlinear programming.* 1980.

127 J.W. Klop. *Combinatory reduction systems.* 1980.

128 A.J.J. Talman. *Variable dimension fixed point algorithms and triangulations.* 1980.

129 G. van der Laan. *Simplicial fixed point algorithms.* 1980.

130 P.J.W. ten Hagen, T. Hagen, P. Klint, H. Noot, H.J. Sint, A.H. Veen. *ILP: intermediate language for pictures.* 1980.

131 R.J.R. Back. *Correctness preserving program refinements: proof theory and applications.* 1980.

132 H.M. Mulder. *The interval function of a graph.* 1980.

133 C.A.J. Klaassen. *Statistical performance of location estimators.* 1981.

134 J.C. van Vliet, H. Wupper (eds.). *Proceedings international conference on ALGOL 68.* 1981.

135 J.A.G. Groenendijk, T.M.V. Janssen, M.J.B. Stokhof (eds.). *Formal methods in the study of language, part I.* 1981.

136 J.A.G. Groenendijk, T.M.V. Janssen, M.J.B. Stokhof (eds.). *Formal methods in the study of language, part II.* 1981.

137 J. Telgen. *Redundancy and linear programs.* 1981.

138 H.A. Lauwerier. *Mathematical models of epidemics.* 1981.

139 J. van der Wal. *Stochastic dynamic programming, successive approximations and nearly optimal strategies for Markov decision processes and Markov games.* 1981.

140 J.H. van Geldrop. *A mathematical theory of pure exchange economies without the no-critical-point hypothesis.* 1981.

141 G.E. Welters. *Abel-Jacobi isogenies for certain types of Fano threefolds.* 1981.

142 H.R. Bennett, D.J. Lutzer (eds.). *Topology and order structures, part 1.* 1981.

143 J.M. Schumacher. *Dynamic feedback in finite- and infinite-dimensional linear systems.* 1981.

144 P. Eijgenraam. *The solution of initial value problems using interval arithmetic; formulation and analysis of an algorithm.* 1981.

145 A.J. Brentjes. *Multi-dimensional continued fraction algorithms.* 1981.

146 C.V.M. van der Mee. *Semigroup and factorization methods in transport theory.* 1981.

147 H.H. Tigelaar. *Identification and informative sample size.* 1982.

148 L.C.M. Kallenberg. *Linear programming and finite Markovian control problems.* 1983.

149 C.B. Huijsmans, M.A. Kaashoek, W.A.J. Luxemburg, W.K. Vietsch (eds.). *From A to Z, proceedings of a symposium in honour of A.C. Zaanen.* 1982.

150 M. Veldhorst. *An analysis of sparse matrix storage schemes.* 1982.

151 R.J.M.M. Does. *Higher order asymptotics for simple linear rank statistics.* 1982.

152 G.F. van der Hoeven. *Projections of lawless sequences.* 1982.

153 J.P.C. Blanc. *Application of the theory of boundary value problems in the analysis of a queueing model with paired services.* 1982.

154 H.W. Lenstra, Jr., R. Tijdeman (eds.). *Computational methods in number theory, part I.* 1982.

155 H.W. Lenstra, Jr., R. Tijdeman (eds.). *Computational methods in number theory, part II.* 1982.

156 P.M.G. Apers. *Query processing and data allocation in distributed database systems.* 1983.

157 H.A.W.M. Kneppers. *The covariant classification of two-dimensional smooth commutative formal groups over an algebraically closed field of positive characteristic.* 1983.

158 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science IV, distributed systems, part 1.* 1983.

159 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science IV, distributed systems, part 2.* 1983.

160 A. Rezus. *Abstract AUTOMATH.* 1983.

161 G.F. Helminck. *Eisenstein series on the metaplectic group, an algebraic approach.* 1983.

162 J.J. Dik. *Tests for preference.* 1983.

163 H. Schippers. *Multiple grid methods for equations of the second kind with applications in fluid mechanics.* 1983.

164 F.A. van der Duyn Schouten. *Markov decision processes with continuous time parameter.* 1983.

165 P.C.T. van der Hoeven. *On point processes.* 1983.

166 H.B.M. Jonkers. *Abstraction, specification and implementation techniques, with an application to garbage collection.* 1983.

167 W.H.M. Zijm. *Nonnegative matrices in dynamic programming.* 1983.

168 J.H. Evertse. *Upper bounds for the numbers of solutions of diophantine equations.* 1983.

169 H.R. Bennett, D.J. Lutzer (eds.). *Topology and order structures, part 2.* 1983.

CWI TRACTS

1 D.H.J. Epema. *Surfaces with canonical hyperplane sections.* 1984.

2 J.J. Dijkstra. *Fake topological Hilbert spaces and characterizations of dimension in terms of negligibility.* 1984.

3 A.J. van der Schaft. *System theoretic descriptions of physical systems.* 1984.

4 J. Koene. *Minimal cost flow in processing networks, a primal approach.* 1984.

5 B. Hoogenboom. *Intertwining functions on compact Lie groups.* 1984.

6 A.P.W. Böhm. *Dataflow computation.* 1984.

7 A. Blokhuis. *Few-distance sets.* 1984.

8 M.H. van Hoorn. *Algorithms and approximations for queueing systems.* 1984.

9 C.P.J. Koymans. *Models of the lambda calculus.* 1984.

10 C.G. van der Laan, N.M. Temme. *Calculation of special functions: the gamma function, the exponential integrals and error-like functions.* 1984.

11 N.M. van Dijk. *Controlled Markov processes; time-discretization.* 1984.

12 W.H. Hundsdorfer. *The numerical solution of nonlinear stiff initial value problems: an analysis of one step methods.* 1985.

13 D. Grune. *On the design of ALEPH.* 1985.

14 J.G.F. Thiemann. *Analytic spaces and dynamic programming: a measure theoretic approach.* 1985.

15 F.J. van der Linden. *Euclidean rings with two infinite primes.* 1985.

16 R.J.P. Groothuizen. *Mixed elliptic-hyperbolic partial differential operators: a case-study in Fourier integral operators.* 1985.

17 H.M.M. ten Eikelder. *Symmetries for dynamical and Hamiltonian systems.* 1985.

18 A.D.M. Kester. *Some large deviation results in statistics.* 1985.

19 T.M.V. Janssen. *Foundations and applications of Montague grammar, part 1: Philosophy, framework, computer science.* 1986.

20 B.F. Schriever. *Order dependence.* 1986.

21 D.P. van der Vecht. *Inequalities for stopped Brownian motion.* 1986.

22 J.C.S.P. van der Woude. *Topological dynamix.* 1986.

23 A.F. Monna. *Methods, concepts and ideas in mathematics: aspects of an evolution.* 1986.

24 J.C.M. Baeten. *Filters and ultrafilters over definable subsets of admissible ordinals.* 1986.

25 A.W.J. Kolen. *Tree network and planar rectilinear location theory.* 1986.

26 A.H. Veen. *The misconstrued semicolon: Reconciling imperative languages and dataflow machines.* 1986.

27 A.J.M. van Engelen. *Homogeneous zero-dimensional absolute Borel sets.* 1986.

28 T.M.V. Janssen. *Foundations and applications of Montague grammar, part 2: Applications to natural language.* 1986.

29 H.L. Trentelman. *Almost invariant subspaces and high gain feedback.* 1986.

30 A.G. de Kok. *Production-inventory control models: approximations and algorithms.* 1987.

31 E.E.M. van Berkum. *Optimal paired comparison designs for factorial experiments.* 1987.

32 J.H.J. Einmahl. *Multivariate empirical processes.* 1987.

33 O.J. Vrieze. *Stochastic games with finite state and action spaces.* 1987.

34 P.H.M. Kersten. *Infinitesimal symmetries: a computational approach.* 1987.

35 M.L. Eaton. *Lectures on topics in probability inequalities.* 1987.

36 A.H.P. van der Burgh, R.M.M. Mattheij (eds.). *Proceedings of the first international conference on industrial and applied mathematics (ICIAM 87).* 1987.

37 L. Stougie. *Design and analysis of algorithms for stochastic integer programming.* 1987.

38 J.B.G. Frenk. *On Banach algebras, renewal measures and regenerative processes.* 1987.

39 H.J.M. Peters, O.J. Vrieze (eds.). *Surveys in game theory and related topics.* 1987.

40 J.L. Geluk, L. de Haan. *Regular variation, extensions and Tauberian theorems.* 1987.

41 Sape J. Mullender (ed.). *The Amoeba distributed operating system: Selected papers 1984-1987.* 1987.

42 P.R.J. Asveld, A. Nijholt (eds.). *Essays on concepts, formalisms, and tools.* 1987.

43 H.L. Bodlaender. *Distributed computing: structure and complexity.* 1987.

44 A.W. van der Vaart. *Statistical estimation in large parameter spaces.* 1988.

45 S.A. van de Geer. *Regression analysis and empirical processes.* 1988.

46 S.P. Spekreijse. *Multigrid solution of the steady Euler equations.* 1988.

47 J.B. Dijkstra. *Analysis of means in some non-standard situations.* 1988.

48 F.C. Drost. *Asymptotics for generalized chi-square goodness-of-fit tests.* 1988.

49 F.W. Wubs. *Numerical solution of the shallow-water equations.* 1988.

50 F. de Kerf. *Asymptotic analysis of a class of perturbed Korteweg-de Vries initial value problems.* 1988.

51 P.J.M. van Laarhoven. *Theoretical and computational aspects of simulated annealing.* 1988.

52 P.M. van Loon. *Continuous decoupling transformations for linear boundary value problems.* 1988.

53 K.C.P. Machielsen. *Numerical solution of optimal control problems with state constraints by sequential quadratic programming in function space.* 1988.

54 L.C.R.J. Willenborg. *Computational aspects of survey data processing.* 1988.

55 G.J. van der Steen. *A program generator for recognition, parsing and transduction with syntactic patterns.* 1988.

56 J.C. Ebergen. *Translating programs into delay-insensitive circuits.* 1989.

57 S.M. Verduyn Lunel. *Exponential type calculus for linear delay equations.* 1989.

58 M.C.M. de Gunst. *A random model for plant cell population growth.* 1989.

59 D. van Dulst. *Characterizations of Banach spaces not containing l^1.* 1989.

60 H.E. de Swart. *Vacillation and predictability properties of low-order atmospheric spectral models.* 1989.

61 P. de Jong. *Central limit theorems for generalized multilinear forms.* 1989.

62 V.J. de Jong. *A specification system for statistical software.* 1989.

63 B. Hanzon. *Identifiability, recursive identification and spaces of linear dynamical systems, part I.* 1989.

64 B. Hanzon. *Identifiability, recursive identification and spaces of linear dynamical systems, part II.* 1989.

65 B.M.M. de Weger. *Algorithms for diophantine equations.* 1989.

66 A. Jung. *Cartesian closed categories of domains.* 1989.

67 J.W. Polderman. *Adaptive control & identification: Conflict or conflux?.* 1989.

68 H.J. Woerdeman. *Matrix and operator extensions.* 1989.

69 B.G. Hansen. *Monotonicity properties of infinitely divisible distributions.* 1989.

70 J.K. Lenstra, H.C. Tijms, A. Volgenant (eds.). *Twenty-five years of operations research in the Netherlands: Papers dedicated to Gijs de Leve.* 1990.

71 P.J.C. Spreij. *Counting process systems. Identification and stochastic realization.* 1990.

72 J.F. Kaashoek. *Modeling one dimensional pattern formation by anti-diffusion.* 1990.

73 A.M.H. Gerards. *Graphs and polyhedra. Binary spaces and cutting planes.* 1990.

74 B. Koren. *Multigrid and defect correction for the steady Navier-Stokes equations. Application to aerodynamics.* 1991.

75 M.W.P. Savelsbergh. *Computer aided routing.* 1991.

76 O.E. Flippo. *Stability, duality and decomposition in general mathematical programming.* 1991.

77 A.J. van Es. *Aspects of nonparametric density estimation.* 1991.

78 G.A.P. Kindervater. *Exercises in parallel combinatorial computing.* 1991.

79 J.J. Lodder. *Towards a symmetrical theory of generalized functions.* 1991.

80 S.A. Smulders. *Control of freeway traffic flow.* 1991.

81 P.H.M. America, J.J.M.M. Rutten. *A parallel object-oriented language: design and semantic foundations.* 1991.

82 F. Thuijsman. *Optimality and equilibria in stochastic games.* 1991.

83 R.J. Kooman. *Convergence properties of recurrence sequences.* 1991.

84 A.M. Cohen (ed.). *Computational aspects of Lie group representations and related topics. Proceedings of the 1990 Computational Algebra Seminar at CWI, Amsterdam.* 1991.

85 V. de Valk. *One-dependent processes.* 1991.