

# Decomposition in general mathematical programming

Olaf E. Flippo\* and Alexander H.G. Rinnooy Kan\*\*

*Econometric Institute, Erasmus University Rotterdam, Netherlands*

Received 17 April 1990

Revised manuscript received April 1992

In this paper a unifying framework is presented for the generalization of the decomposition methods originally developed by Benders (1962) and Dantzig and Wolfe (1960). These generalizations, called *Variable Decomposition* and *Constraint Decomposition* respectively, are based on the general duality theory developed by Tind and Wolsey. The framework presented is of a general nature since there are no restrictive conditions imposed on problem structure; moreover, inaccuracies and duality gaps that are encountered during computations are accounted for. The two decomposition methods are proven not to cycle if certain (fairly general) conditions are met. Furthermore, finite convergence can be ensured under the traditional finiteness conditions and asymptotic convergence can be guaranteed once certain continuity conditions are met. The obvious symmetry between both types of decomposition methods is explained by establishing a duality relation between the two, which extends a similar result in Linear Programming. A remaining asymmetry in the asymptotic convergence results is argued to be a direct consequence of a fundamental asymmetry that resides in the Tind–Wolsey duality theory. It can be shown that in case the latter asymmetry disappears, the former does too. Other decomposition techniques, such as Lagrangean Decomposition and Cross Decomposition, turn out to be captured by the general framework presented here as well.

*AMS Subject Classification:* 90C99.

*Key words:* Asymptotic convergence, decomposition, duality theory, mathematical programming.

## 1. Introduction

Decomposition has been recognized as a fundamental technique in optimization ever since the seminal papers of Benders (1962) and Dantzig and Wolfe (1960). The literature abounds with theoretical extensions of these two basic approaches, as well as with reports of successful applications (for references, see Flippo, 1991). In this paper, *Variable* and *Constraint Decomposition* will be introduced as proper generalizations of Benders Decomposition and Dantzig–Wolfe Decomposition respectively. In generalizing the latter two approaches, a duality theory is required

*Correspondence to:* Prof. O.E. Flippo, Faculty of Economics, University of Limburg, P.O. Box 616, 6200 MD Maastricht, Netherlands.

\* This study was supported by the Netherlands Foundation for Mathematics (SMC) with financial aid from the Netherlands Organization for Scientific Research (NWO).

\*\* Part of this work was done while on leave at the Wharton School of the University of Pennsylvania.

that extends the one in Linear Programming; a general duality theory that fits this purpose and that is used in this paper is due to Tind and Wolsey (1981).

Under minimal assumptions on problem structure, Variable and Constraint Decomposition turn out to include essentially all previous generalizations as special cases. In addition, our approach can cope with suboptimality in the primal and dual programs that are encountered during computations, as well as with the occurrence of duality gaps. Within our framework it is possible to establish for both decomposition techniques that intermediate primal and dual solutions will not be generated more than once during the solution process, and that finite convergence is assured under appropriate finiteness conditions. Moreover, asymptotic convergence (as well as asymptotic feasibility) for the Variable Decomposition Procedure can be established under fairly general continuity and stability conditions; for Constraint Decomposition a similar observation applies if, additionally, a finite parametrizability condition for the dual programs is met. This asymmetry between Variable and Constraint Decomposition is noteworthy in view of the strong relation between the two procedures; they will in fact be shown to be dual approaches. We shall argue that this asymmetry is indeed a direct consequence of a fundamental asymmetry that resides in the Tind-Wolsey duality theory on which our approach is based. Finally, it will be shown that several other decomposition techniques that have appeared in the literature, such as Lagrangean Decomposition and Cross Decomposition, are captured by the general framework presented here as well. Other relations to existing literature will be dealt with in the course of the text; for a generalization and unification of Variable and Constraint Decomposition in a max-min setting, we refer to Burkard et al. (1985) and Tind (1988, 1990).

## 2. Duality theory

In this section we briefly recall the main concepts and results from the Tind-Wolsey duality theory (1981) that will be required below. Consider the following *primal program*:

$$P: \quad \max_x f(x) \tag{1}$$

$$\text{s.t.} \quad G(x) \diamond 0, \tag{2}$$

$$x \in X, \tag{3}$$

with  $X \subseteq \mathbb{R}^n$ ,  $\diamond \in \{\leq, =\}^m$ ,  $f: X \rightarrow \mathbb{R} \cup \{\pm\infty\}$  and  $G: X \rightarrow \mathbb{R}^m$ . Here,  $G(x) \diamond 0$  is a shorthand notation for  $G_i(x) \diamond_i 0$  ( $i = 1, \dots, m$ ), with  $\diamond_i$  denoting either an equality (=) or an inequality sign of the "less-than-or-equal" type ( $\leq$ ). The reason for introducing such notation is to account for the simultaneous occurrence of equalities and inequalities in a compact way. The *value* (or *perturbation*) *function*  $v: \mathbb{R}^m \rightarrow \mathbb{R} \cup$

$\{\pm\infty\}$  is defined as

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

with  $\text{RHS} = \{r \in \mathbb{R}^m \mid \exists x \in X: G(x) \diamond r\}$  denoting the *set of feasible right-hand-sides*. If P is *regular*, i.e. if its *value*  $v(0)$  is not equal to  $\pm\infty$ , and if  $f(x) \geq v(0) - \varepsilon$  for some feasible solution  $x$  and some  $\varepsilon \geq 0$ , then  $x$  is called an  $\varepsilon$ -optimal solution to P. To describe the *dual program* D, we introduce the set of *functions*

$$\mathcal{F} = \{g: \mathbb{R}^m \rightarrow \mathbb{R} \cup \{\pm\infty\} \mid g(r) \leq g(r') \forall r, r' \in \text{RHS}: r \diamond r'\}, \tag{4}$$

and define

$$\text{D:} \quad \min_g \quad g(0) \tag{5}$$

$$\text{s.t.} \quad g(G(x)) \geq f(x) \quad \forall x \in X, \tag{6}$$

$$g \in \mathcal{F}. \tag{7}$$

Note that  $\mathcal{F}$  imposes a kind of monotonicity condition only with respect to the *inequality* constraints in (2). It is easily verified that  $g \in \mathcal{F}$  satisfies (6) if and only if  $g(r) \geq v(r) \forall r \in \mathbb{R}^m$ . Taking  $r = 0$  one obtains *weak duality*, which states that every feasible dual solution yields an upper bound to the optimal value of the primal program. By observing that  $v$  itself is feasible for D, one obtains *strong duality*, which says that the optimal primal and dual objective values meet. Finally, by defining  $P^0$  as the primal problem (1)-(3) with  $f(x)$  in (1) being replaced by 0 and applying weak and strong duality to  $P^0$  and its dual  $D^0$ , one obtains a proper generalization of the *Farkas Lemma*:

$$\exists x \in X: G(x) \diamond 0 \Leftrightarrow h(0) \geq 0 \quad \forall h \in \mathcal{F}: [h(G(x)) \geq 0 \quad \forall x \in X]. \tag{8}$$

Clearly, the major drawback of this particular duality theory is the *asymmetry* between P and D, in that the primal solution space  $X$  is *finite* dimensional, whereas the dual solution space  $\mathcal{F}$  is not. This asymmetry is resolved by restricting  $\mathcal{F}$  to some finite dimensional subset, generally at the expense of introducing a strictly positive *duality gap* however (i.e. giving up strong duality). A typical example is restricting  $\mathcal{F}$  to the (finite dimensional) space of *affine functions*; in that case, *Lagrangian duality* is recovered, which properly generalizes *Linear* and *Quadratic Programming duality*. In addition to that, a one-to-one correspondence between dual variables and primal constraints is obtained (cf. Tind and Wolsey, 1981). Fortunately, it turns out that the latter property holds for the general framework as well; the following theorem is easily proven (Flippo, 1991) and will be useful in the sequel.

**Theorem 2.1** (Additively separable duality). *Consider the dual program (5)-(7) and the set of functions  $\Gamma = \Gamma_1 + \dots + \Gamma_m \subseteq \mathcal{F}$  with*

$$\Gamma_i = \{g_i: \mathbb{R} \rightarrow \mathbb{R} \cup \{+\infty\} \mid g_i(r_i) \leq g_i(r'_i) \quad \forall r_i, r'_i \in \mathbb{R}: r_i \diamond_i r'_i\}$$

and with set and function addition being defined in the usual way. Assume that  $f(x) < +\infty \forall x \in X$  and that P is regular. Then there is an optimal dual solution  $g^* \in \Gamma$  with  $g^*(0) = g_1^*(0) + \dots + g_m^*(0) = v(0)$  for some  $g_i^* \in \Gamma_i$  ( $i = 1, \dots, m$ ).  $\square$

Thus, without eliminating the asymmetry between P and D, one can at least associate a dual variable to each primal constraint under hardly any conditions on the primal program at all.

### 3. Variable Decomposition

We are now ready to introduce *Variable Decomposition*, one of the two main decomposition approaches and a direct generalization of techniques that have been proposed under names such as (*Generalized*) *Benders Decomposition*, *Primal Decomposition* and *Resource Directive Decomposition*. Let us assume P can be written as:

$$P: \quad \max_{x,y} f(x, y) \tag{9}$$

$$\text{s.t.} \quad G(x, y) \diamond 0, \tag{10}$$

$$(x, y) \in U \cap (X \times Y), \tag{11}$$

with  $X \subseteq \mathbb{R}^{n_1}$ ,  $Y \subseteq \mathbb{R}^{n_2}$ ,  $U \subseteq \mathbb{R}^n$  ( $n_1 + n_2 = n$ ),  $\diamond \in \{\leq, =\}^m$ ,  $f: U \cap (X \times Y) \rightarrow \mathbb{R} \cup \{\pm\infty\}$ , and  $G: U \cap (X \times Y) \rightarrow \mathbb{R}^m$ . For reasons to be explained later (see the paragraph following (26)–(29) and the second remark after the algorithmic description of the Variable Decomposition Procedure), we also assume that:

$$(x, y) \text{ is feasible for P} \Rightarrow \exists y' \in Y: [(x, y') \text{ is feasible for P and } f(x, y') > -\infty]. \tag{12}$$

Note that (12) is redundant in case the objective function  $f$  in (9) is real-valued; the reason for allowing  $f$  to range to the extended reals is to account for the situation in which  $f$  itself is the value function of some (parametric) optimization problem. As a methodology, Variable Decomposition is captured by the notions of *projection*, *dualization*, *outer approximation* and *relaxation* (cf. Geoffrion, 1972a). *Projection* means that the joint optimization over the  $x$  and  $y$  variables in (9)–(11) is replaced with a nested optimization; by setting  $x = \bar{x} \in X$  we obtain a family of parametrized *primal subprograms*

$$P(\bar{x}): \quad \max_{x,y} f(x, y) \tag{13}$$

$$\text{s.t.} \quad G(x, y) \diamond 0, \tag{14}$$

$$x = \bar{x}, \tag{15}$$

$$(x, y) \in U \cap (\tilde{X} \times Y), \tag{16}$$

where  $\tilde{X}$  is any superset of  $X$  that is independent of  $\bar{x}$ . (For an example where a choice of  $\tilde{X}$  different from  $X$  is fruitful, we refer to Flippo (1991).) The original program  $P$  may then be written as

$$\max_{\bar{x}} \quad \varphi(P(\bar{x})) \tag{17}$$

$$\text{s.t.} \quad \varphi(P^0(\bar{x})) \geq 0, \tag{18}$$

$$\bar{x} \in X. \tag{19}$$

The notation  $\varphi(\cdot)$  is introduced to denote the optimal objective function value of a given optimization problem. As in Section 2,  $P^0(\bar{x})$  is an optimization problem that is obtained from  $P(\bar{x})$  by replacing the objective function  $f$  with the zero-function. Note that (18) ensures restriction to those  $\bar{x} \in X$  that can be extended to a feasible solution  $(\bar{x}, \bar{y})$  of (9)–(11) (cf. (8)). *Dualization* of  $P(\bar{x})$  now yields the *dual subprograms*

$$D(\bar{x}): \quad \min_g \quad g(0, \bar{x}) \tag{20}$$

$$\text{s.t.} \quad g(G(x, y), x) \geq f(x, y) \quad \forall (x, y) \in U \cap (\tilde{X} \times Y), \tag{21}$$

$$g \in \mathcal{F}. \tag{22}$$

The dual solution space  $\mathcal{F}$  is defined similarly as in Section 2:

$$\mathcal{F} = \{g : \mathbb{R}^{m+n_1} \rightarrow \mathbb{R} \cup \{\pm\infty\} \mid \\ g(r, x) \leq g(r', x') \quad \forall (r, x), (r', x') \in \text{RHS} : r \diamond r', x = x'\},$$

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

$$\text{RHS} = \{(r, x') \in \mathbb{R}^{m+n_1} \mid \exists (x, y) \in U \cap (\tilde{X} \times Y) : [G(x, y) \diamond r, x = x']\}.$$

Thanks to strong duality (cf. Section 2), (17)–(19) can now be replaced by

$$\max_x \quad \varphi(D(x)) \tag{23}$$

$$\text{s.t.} \quad \varphi(D^0(x)) \geq 0, \tag{24}$$

$$x \in X, \tag{25}$$

where  $D^0(x)$  is the dual to  $P^0(x)$ . (Note that we omitted the bar overlining the  $x$ -variables here.) A trivial reformulation of (23)–(25) renders

$$\max_{x, \theta} \quad \theta$$

$$\text{s.t.} \quad -\varphi(D(x)) + \theta \leq 0,$$

$$-\varphi(D^0(x)) \leq 0,$$

$$(x, \theta) \in X \times \mathbb{R}.$$

Now, if we let  $\Delta$  denote the feasible set of  $D(x)$  defined by (21)–(22) and similarly, if we let  $\Delta^0$  denote the feasible set of  $D^0(x)$ , then *outer approximation* yields the *master program*

$$VD(\Delta, \Delta^0): \quad \max_{x, \theta} \quad \theta \tag{26}$$

$$\text{s.t.} \quad -g(0, x) + \theta \leq 0 \quad \forall g \in \Delta, \tag{27}$$

$$-h(\theta, x) \leq 0 \quad \forall h \in \Delta^0, \tag{28}$$

$$(x, \theta) \in X \times \mathbb{R}, \tag{29}$$

which, under (12), is clearly still equivalent to P. Here, the constraints in  $\Delta$  can be thought of as *value cuts* (cf. (17)) and the constraints in  $\Delta^0$  as *feasibility cuts* (cf. (18)).

In (26)–(29), the  $y$ -variables have been eliminated at the expense of a significant increase in the number of constraints. Thus, one would typically solve  $VD(\Delta, \Delta^0)$  by *relaxation*, i.e. by considering a *relaxed master program*  $VD(\bar{\Delta}, \bar{\Delta}^0)$  for appropriately chosen subsets  $\bar{\Delta} \subseteq \Delta$  and  $\bar{\Delta}^0 \subseteq \Delta^0$ . It is easily verified that  $VD(\cdot, \cdot)$  may provide arbitrarily tight *upper bounds* and  $P(\cdot)$  arbitrarily tight *lower bounds* on  $\varphi(P)$ . Thus, one naturally arrives at the notion of an *iterative procedure* in which the relaxed master programs  $VD(\bar{\Delta}, \bar{\Delta}^0)$  generate candidate  $\bar{x}$ -values, which, through  $D(\bar{x})$  or  $D^0(\bar{x})$ , yield value or feasibility cuts that can be used to extend  $\bar{\Delta}$  or  $\bar{\Delta}^0$ . The following lemma summarizes the essential ingredients for such a procedure.

**Lemma 3.1.** Assume that  $\varphi(VD(\bar{\Delta}, \bar{\Delta}^0)) < +\infty$ .

(a) If  $VD(\bar{\Delta}, \bar{\Delta}^0)$  is infeasible, then so is P.

(b) If  $(\bar{x}, \bar{\theta})$  is feasible for  $VD(\bar{\Delta}, \bar{\Delta}^0)$ , then

- $\varphi(P(\bar{x})) < \bar{\theta} \Leftrightarrow \exists \bar{g} \in \Delta \setminus \bar{\Delta}: -\bar{g}(0, \bar{x}) + \bar{\theta} > 0$ ,

- $P(\bar{x})$  is infeasible  $\Leftrightarrow \exists \bar{h} \in \Delta^0 \setminus \bar{\Delta}^0: -\bar{h}(0, \bar{x}) > 0$ .

(c) If  $(\bar{x}, \bar{\theta})$  is  $\varepsilon_1$ -optimal for  $VD(\bar{\Delta}, \bar{\Delta}^0)$ ,  $\bar{y}$  is  $\varepsilon_2$ -optimal for  $P(\bar{x})$ , and  $\bar{g} \in \bar{\Delta}$  is  $\varepsilon_3$ -optimal for  $D(\bar{x})$ , then  $(\bar{x}, \bar{y})$  is an  $(\varepsilon_1 + \varepsilon_2 + \varepsilon_3)$ -optimal solution for P.

**Proof.** The proof of (a) is trivial, and (b) follows from strong duality, from the fact that  $-g(0, \bar{x}) + \bar{\theta} \leq 0 \quad \forall g \in \bar{\Delta}$  and  $-h(0, \bar{x}) \leq 0 \quad \forall h \in \bar{\Delta}^0$ , and from the generalized Farkas Lemma (cf. (8)). As to (c),

$$\varphi(P) \leq \bar{\theta} + \varepsilon_1 \leq \bar{g}(0, \bar{x}) + \varepsilon_1 \leq \varphi(D(\bar{x})) + \varepsilon_1 + \varepsilon_3$$

$$= \varphi(P(\bar{x})) + (\varepsilon_1 + \varepsilon_3) \leq f(\bar{x}, \bar{y}) + (\varepsilon_1 + \varepsilon_2 + \varepsilon_3). \quad \square$$

We are now in a position to describe the Variable Decomposition Procedure. Let the superscript  $k$  denote an iteration index, and let  $UB^k$ ,  $LB^k$  and  $(x^{inc,k}, y^{inc,k})$  be the values of the best upper bound on  $\varphi(P)$ , best lower bound on  $\varphi(P)$ , and best feasible solution to P found so far.

**Variable Decomposition Procedure.**

*Step 0. Initialization phase.*

$$k := 1; \text{UB}^0 := +\infty; \text{LB}^0 := -\infty;$$

Choose  $\Delta^k \subseteq \Delta$  and  $\Delta^{0,k} \subseteq \Delta^0$  such that  $\varphi(\text{VD}(\Delta^k, \Delta^{0,k})) < +\infty$ ;

*Step 1. Master program phase.*

if  $\varphi(\text{VD}(\Delta^k, \Delta^{0,k})) = -\infty$  then stop [P is infeasible]

else begin

Choose  $(x^k, \theta^k, \varepsilon_1^k)$  such that  $\varepsilon_1^k \geq 0$  and  $(x^k, \theta^k)$  is  $\varepsilon_1^k$ -optimal for  $\text{VD}(\Delta^k, \Delta^{0,k})$ ;

$$\text{UB}^k := \min\{\text{UB}^{k-1}, \theta^k + \varepsilon_1^k\}$$

end;

*Step 2. Subprogram phase.*

if  $\varphi(P(x^k)) = -\infty$  then begin

$$\varepsilon_2^k := 0; \varepsilon_3^k := 0; \text{LB}^k := \text{LB}^{k-1};$$

Choose  $h^k \in \Delta^0$  such that  $-h^k(0, x^k) > 0$ ;

Choose  $(\Delta^{k+1}, \Delta^{0,k+1})$  such that  $\Delta^k \subseteq \Delta^{k+1} \subseteq \Delta$  and  $\Delta^{0,k} \cup \{h^k\} \subseteq \Delta^{0,k+1} \subseteq \Delta^0$

end else begin

Choose  $(y^k, \varepsilon_2^k)$  such that  $\varepsilon_2^k \geq 0$  and  $y^k$  is  $\varepsilon_2^k$ -optimal for  $P(x^k)$ ;

Choose  $(g^k, \varepsilon_3^k)$  such that  $\varepsilon_3^k \geq 0$  and  $g^k$  is  $\varepsilon_3^k$ -optimal for  $D(x^k)$ ;

Choose  $(\Delta^{k+1}, \Delta^{0,k+1})$  such that  $\Delta^k \cup \{g^k\} \subseteq \Delta^{k+1} \subseteq \Delta$  and  $\Delta^{0,k} \subseteq \Delta^{0,k+1} \subseteq \Delta^0$ ;

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

if  $\text{LB}^k > \text{LB}^{k-1}$  then  $(x^{\text{inc},k}, y^{\text{inc},k}) := (x^k, y^k)$

else  $(x^{\text{inc},k}, y^{\text{inc},k}) := (x^{\text{inc},k-1}, y^{\text{inc},k-1})$

end;

*Step 3. Optimality verification.*

Choose  $\varepsilon_0^k \geq \varepsilon_1^k + \varepsilon_2^k + \varepsilon_3^k$ ;

if  $\text{UB}^k - \text{LB}^k \leq \varepsilon_0^k$  then stop [ $(x^{\text{inc},k}, y^{\text{inc},k})$  is  $\varepsilon_0^k$ -optimal for P]

else  $k := k + 1$ ; return to Step 1.

The following comments may be helpful.

- Finding initial sets  $\Delta^1$  and  $\Delta^{0,1}$  with  $\varphi(\text{VD}(\Delta^1, \Delta^{0,1})) < +\infty$  may be a non-trivial task. If such sets do not exist then clearly  $\varphi(P) = +\infty$ .

- In the subprogram phase  $h^k$  exists, since it is implied by (8) and (12) that  $\varphi(P(x^k)) = -\infty \Leftrightarrow \varphi(P^0(x^k)) = -\infty$ .

- The inaccuracy parameters  $\varepsilon_i^k$  ( $i = 1, 2, 3$ ) may vary between iterations and need not be specified prior to iteration  $k$ .

- In the optimality verification phase,  $(x^{\text{inc},k}, y^{\text{inc},k})$  is  $(\text{UB}^k - \text{LB}^k)$ -optimal for P because  $0 \leq \varphi(P) - f(x^{\text{inc},k}, y^{\text{inc},k}) \leq \text{UB}^k - \text{LB}^k$ . This solution is only considered accurate enough if it meets the bound  $\varepsilon_0^k$  on the overall inaccuracy.

- This framework not only allows for inaccuracies during the iterative process, but also for duality gaps between the primal and dual subprograms, since these gaps are in fact part of the inaccuracies  $\varepsilon_3^k$  of the dual subprogram solutions.

Before relating the above procedure to the existing literature, we first establish a few of its crucial properties.

**Theorem 3.1** (Non-repetition of complete solutions).

(a) *In the master program phase no solution  $(x^k, \theta^k)$  will be generated more than once.*

(b) *In the subprogram phase no solution  $h^k$  will be generated more than once, and as soon as a solution  $g^k$  is generated for a second time, the algorithm will terminate.*

**Proof.** (a) If  $\varphi(P(x^k)) > -\infty$  then  $g^k(0, x^k) - \varepsilon_2^k - \varepsilon_3^k \leq f(x^k, y^k) \leq LB^k$ . Since  $UB^k \leq \theta^k + \varepsilon_1^k$ , we have that  $(\theta^k + \varepsilon_1^k) - (g^k(0, x^k) - \varepsilon_2^k - \varepsilon_3^k) \geq UB^k - LB^k$ . Now, if the algorithm does not terminate at iteration  $k$ , then  $UB^k - LB^k > \varepsilon_0^k$ , which implies that  $-g^k(0, x^k) + \theta^k > 0$ . Adding  $g^k$  to  $\Delta^{k+1}$  ensures that  $(x^k, \theta^k)$  will not be considered in any subsequent iteration (cf. (27)). Similarly, if  $\varphi(P(x^k)) = -\infty$  then adding  $h^k$  to  $\Delta^{0,k+1}$  yields that for any value of  $\theta$ ,  $(x^k, \theta)$  will not be generated in any subsequent iteration (cf. (28)).

(b) If  $\varphi(P(x^k)) > -\infty$  then  $g^k \in \Delta^k$  implies that  $-g^k(0, x^k) + \theta^k \leq 0$ . If, in addition,  $UB^k - LB^k > \varepsilon_0^k$ , then, as in (a),  $-g^k(0, x^k) + \theta^k > 0$ , which is a contradiction. Therefore,  $UB^k - LB^k \leq \varepsilon_0^k$ , so the procedure will stop. In a similar way, if  $\varphi(P(x^k)) = -\infty$  then  $h^k \in \Delta^{0,k}$  would imply that  $-h^k(0, x^k) \leq 0$ , contradicting the fact that  $-h^k(0, x^k) > 0$ .  $\square$

It is worth noting the essential role played in the above proof by our (intuitively reasonable) assumption that the overall inaccuracy  $\varepsilon_0^k$  is at least equal to the sum of the inaccuracies in the composing parts of the procedure.

**Theorem 3.2** (Non-repetition of subprograms). *If for all  $k$  satisfying  $\varphi(P(x^k)) > -\infty$ ,  $\varepsilon_0^k \geq \varepsilon_1^k + \max_{1 \leq j \leq k} \{\varepsilon_2^j + \varepsilon_3^j\}$ , then the procedure terminates as soon as a solution  $x^k$  is generated for the second time.*

**Proof.** If  $x^l = x^k$  for some  $l > k$ , then we must have that  $\varphi(P(x^l)) = \varphi(P(x^k)) > -\infty$  (see the proof of Theorem 3.1(a)), implying that  $-g^k(0, x^l) + \theta^l \leq 0$  (cf. (27)). Hence,

$$\theta^l \leq g^k(0, x^l) = g^k(0, x^k) \leq g^k(0, x^k) + \varepsilon_0^l - \varepsilon_1^l - (\varepsilon_2^k + \varepsilon_3^k). \tag{30}$$

In addition

$$f(x^k, y^k) + \varepsilon_2^k \geq \varphi(P(x^k)) = \varphi(D(x^k)) \geq g^k(0, x^k) - \varepsilon_3^k. \tag{31}$$

Together, (30) and (31) imply that

$$\begin{aligned} UB^l - LB^l &\leq UB^l - LB^k \leq (\theta^l + \varepsilon_1^l) - f(x^k, y^k) \\ &\leq (\theta^l + \varepsilon_1^l) - (g^k(0, x^k) - \varepsilon_2^k - \varepsilon_3^k) \leq \varepsilon_0^l \end{aligned}$$

and the procedure will stop.  $\square$

**Theorem 3.3** (Finite convergence). *If for all  $k$  with  $\varphi(P(x^k)) > -\infty$ ,*

$$\varepsilon_0^k \geq \varepsilon_1^k + \max_{1 \leq j \leq k} \{\varepsilon_2^j + \varepsilon_3^j\},$$

*then the procedure terminates in a finite number of steps if at least one of the following two conditions is met:*

- (a) *all  $x^k$  generated belong to a finite subset of  $X$ ;*
- (b) *all  $g^k$  and  $h^k$  generated belong to finite subsets of  $\Delta$  and  $\Delta^0$ .  $\square$*

The proof of Theorem 3.3 is trivially obtained from its two predecessors. Thus, the general procedure has the properties typically associated with its special cases. As we shall see below, its asymptotic convergence can also be guaranteed, once certain conditions are verified.

However, let us first draw attention to a crucial property of the dual subprograms  $D(\bar{x})$  and  $D^0(\bar{x})$ , which is that *their feasible sets do not depend on  $\bar{x}$* . A moment's reflection shows that if that property did not hold, then  $\Delta$  and  $\Delta^0$  in (27)–(28) would depend on  $x$ , and as a consequence, serious computational difficulties would arise in solving the relaxed master programs. Now, if the dual subprograms are only allowed to depend on  $\bar{x}$  through their objective functions, then the primal subprograms  $P(\bar{x})$  and  $P^0(\bar{x})$  are only allowed to depend on  $\bar{x}$  through their right-hand-sides. This observation explains the appearance of (15) in  $P(\bar{x})$ . It is instructive to note that generalizations of Benders Decomposition which do not comply with this observation frequently have to include additional regularity conditions as a result. For instance, Lazimy (1982, 1985) imposes unnecessarily restrictive conditions on the Convex Quadratic Programs considered (see Flippo and Rinnooy Kan, 1990), Geoffrion (1972b) relies on a qualitative “*Property P*” to escape from the same computational difficulties, and Burkard et al. (1985) restrict themselves to separable programs for which primal subprograms of the required form are easily obtained without the addition of constraints like  $x = \bar{x}$ .

It is not difficult to see that our approach properly generalizes the original Benders Decomposition (cf. Benders, 1962) and the aforementioned generalizations by Geoffrion (1972b) and Lazimy (1982, 1985). A comparable generalization by Wolsey (1981) can be seen to lead to a master program identical to ours, but with many redundant constraints removed; since these constraints are not necessarily redundant in *relaxations* of the master program, it is not clear beforehand whether removing these is fruitful. In the algebraic optimization approach of Burkard et al. (1985), inaccuracies and duality gaps have not (and could not have) been accommodated, due to the absence of a metric space. Indeed, incorporation of these is a particularly attractive feature, especially since computational considerations sometimes dictate such severe restrictions on the dual solution space  $\mathcal{F}$  that strong duality is no longer assured (e.g. Rouhani et al., 1985). For a more detailed account on the relation with existing literature we refer to Flippo (1991).

We now return to the issue of asymptotic convergence. Let us first focus on the sequence of feasible solutions  $x^k \in FS_X = \{x \in X \mid \varphi(P(x)) > -\infty\}$  as generated by

the algorithm, and let us view the generation of this feasible sequence  $\{x^k \mid k \in I \subseteq \mathbb{N}\}$  as the result of repeated application of the composed *point-to-set map*  $\alpha \circ \beta$ , where

$$\alpha : \Delta \rightarrow X \times \mathbb{R}$$

$$\text{with } \alpha(g) = \{(x, \theta) \in X \times \mathbb{R} \mid -g(0, x) + \theta \leq 0\}$$

and

$$\beta : \text{FS}_X \times \mathbb{R}_+ \rightarrow \Delta$$

$$\text{with } \beta(x, \varepsilon_3) = \{g \in \Delta \mid g \text{ is } \varepsilon_3\text{-optimal for } D(x)\}.$$

Of course, if feasibility rather than value cuts have to be added to a relaxed master program (a process that can be described similarly in terms of point-to-set maps), then the generation of the sequence of feasible solutions is (temporarily) interrupted.

Let us assume that  $|I| = +\infty$ , i.e. that the procedure continues to generate feasible solutions  $(x^k, y^k)$ , and that P is regular. To ensure asymptotic convergence, we shall require  $\alpha \circ \beta$  to be a *closed* point-to-set map (cf. Zangwill, 1969). In addition to this, we shall have to impose certain continuity and compactness conditions as well.

**Theorem 3.4** (Asymptotic convergence). *Assume that P is regular and that  $|I| = +\infty$ . Furthermore, assume that*

- $U \cap (X \times Y)$  is compact;
- $f$  is upper semi-continuous on  $U \cap (X \times Y)$ ;
- $G_i$  is lower semi-continuous on  $U \cap (X \times Y)$  in the case that  $\diamond_i \in \{\leq\}$ , and continuous on  $U \cap (X \times Y)$  in the case that  $\diamond_i \in \{=\}$ ;
- $\varphi(P(\cdot))$  is lower semi-continuous on  $\text{FS}_X$ ;
- $\forall k \in I: \varepsilon_i^k \in E_i \subseteq \mathbb{R}_+$  with  $E_i$  compact ( $i = 1, 2, 3$ );
- $\alpha \circ \beta$  is a closed point-to-set map on  $\text{FS}_X \times E_3$ .

*Then the Variable Decomposition Procedure converges asymptotically, in the sense that:*

(a) every accumulation point of  $(x^k)_I$  is a  $\limsup(\varepsilon_1^k + \varepsilon_3^k)$ -optimal solution to (17)–(19);

(b) every accumulation point of  $(x^k, y^k)_I$  is a  $\limsup(\varepsilon_1^k + \varepsilon_2^k + \varepsilon_3^k)$ -optimal solution to P;

(c) every accumulation point of  $(x^{\text{inc},k})_I$  is a  $\limsup(\varepsilon_1^k + \varepsilon_2^k + \varepsilon_3^k)$ -optimal solution to (17)–(19), and every accumulation point of  $(x^{\text{inc},k}, y^{\text{inc},k})_I$  is a  $\limsup(\varepsilon_1^k + \varepsilon_2^k + \varepsilon_3^k)$ -optimal solution to P;

$$(d) 0 \leq \lim(\text{UB}^k - \text{LB}^k) \leq \limsup(\varepsilon_1^k + \varepsilon_2^k + \varepsilon_3^k).$$

**Proof.** (a) Let  $x^\infty \in \text{FS}_X$  be any accumulation point. It is easy to see that all  $\theta^k$  lie in some compact set, and hence, there is a subsequence  $(p(k))_I$  of  $I$  such that

$$\lim(x^{p(k)}, \theta^{p(k)}, \varepsilon_1^{p(k)}, \varepsilon_3^{p(k)}) = (x^\infty, \theta^\infty, \varepsilon_1^\infty, \varepsilon_3^\infty)$$

for some  $(\theta^\infty, \varepsilon_1^\infty, \varepsilon_3^\infty) \in \mathbb{R} \times E_1 \times E_3$ . Since  $(x^{p(k+1)}, \theta^{p(k+1)}) \in (\alpha \circ \beta)(x^{p(k)}, \varepsilon_3^{p(k)})$ , the closedness of  $\alpha \circ \beta$  implies that  $(x^\infty, \theta^\infty) \in (\alpha \circ \beta)(x^\infty, \varepsilon_3^\infty)$ . Hence, there is an  $\varepsilon_3^\infty$ -optimal solution  $g^\infty$  to  $D(x^\infty)$  such that

$$\theta^\infty - \varepsilon_3^\infty \leq g^\infty(0, x^\infty) - \varepsilon_3^\infty \leq \varphi(D(x^\infty)) = \varphi(P(x^\infty)) \leq \varphi(P). \quad (32)$$

Since clearly  $\theta^\infty + \varepsilon_1^\infty \geq \varphi(P)$  (that inequality being satisfied for every  $(\theta^{p(k)}, \varepsilon_1^{p(k)})$ ), we obtain the required inequality

$$\varphi(P) - \varepsilon_1^\infty - \varepsilon_3^\infty \leq \theta^\infty - \varepsilon_3^\infty \leq \varphi(P(x^\infty)) \leq \varphi(P). \quad (33)$$

(b) As in (a) we consider an accumulation point and a subsequence converging to it. Then upper semi-continuity of  $f$ , lower semi-continuity of  $\varphi(P(\cdot))$  and (33) imply that

$$\begin{aligned} f(x^\infty, y^\infty) &\geq \limsup f(x^{p(k)}, y^{p(k)}) \geq \limsup (\varphi(P(x^{p(k)})) - \varepsilon_2^{p(k)}) \\ &\geq \liminf \varphi(P(x^{p(k)})) - \varepsilon_2^\infty \geq \varphi(P(x^\infty)) - \varepsilon_2^\infty \\ &\geq \varphi(P) - \varepsilon_1^\infty - \varepsilon_2^\infty - \varepsilon_3^\infty \\ &\geq \varphi(P) - \limsup (\varepsilon_1^k + \varepsilon_2^k + \varepsilon_3^k). \end{aligned} \quad (34)$$

(c) Once again we consider an accumulation point, and a subsequence converging to it. From (34), the upper semi-continuity of  $f$  and, by definition of  $I$ , the feasibility of  $(x^{\text{inc}, \infty}, y^{\text{inc}, \infty})$  we deduce that

$$\begin{aligned} \varphi(P(x^{\text{inc}, \infty})) &\geq f(x^{\text{inc}, \infty}, y^{\text{inc}, \infty}) \geq \limsup f(x^{\text{inc}, p(k)}, y^{\text{inc}, p(k)}) \\ &\geq \limsup f(x^{p(k)}, y^{p(k)}) \geq \varphi(P) - \limsup (\varepsilon_1^k + \varepsilon_2^k + \varepsilon_3^k). \end{aligned}$$

(d) Due to monotonicity both  $\text{LB}^k$  and  $\text{UB}^k$  converge to, say  $\text{LB}^\infty$  and  $\text{UB}^\infty$  respectively. As in (34) we deduce that

$$\text{LB}^\infty \geq \limsup f(x^{p(k)}, y^{p(k)}) \geq \varphi(P(x^\infty)) - \varepsilon_2^\infty.$$

Since  $\text{UB}^{p(k)} \leq \theta^{p(k)} + \varepsilon_1^{p(k)}$  we also deduce that  $\text{UB}^\infty \leq \theta^\infty + \varepsilon_1^\infty \leq \varphi(P(x^\infty)) + \varepsilon_1^\infty + \varepsilon_3^\infty$  (cf. (32)). Combining these two inequalities yields the result desired.  $\square$

We note that Theorem 3.4(d) implies that, if  $\varepsilon_0^k$  is chosen to exceed  $\varepsilon_1^k + \varepsilon_2^k + \varepsilon_3^k$  by a fixed, strictly positive amount independent of  $k$ , convergence to an  $\varepsilon_0^k$ -optimal solution can be guaranteed in a *finite* number of steps as soon as the procedure will no longer be interrupted by the generation of feasibility cuts.

Although the assumptions of Theorem 3.4 may appear to be fairly restrictive, it can be argued that they are essentially the minimal ones under which asymptotic convergence can reasonably be expected. That is certainly the case for the compactness assumptions and for the upper and (lower semi-) continuity of  $f$  and  $G_i$  respectively ( $i = 1, \dots, m$ ). The lower semi-continuity of  $\varphi(P(\cdot))$ , which is a much more restrictive assumption (see e.g. Flippo and Minoux, 1989), is necessitated by the fact that, on the one hand, Variable Decomposition in fact tries to solve (17)–(19), and on the other hand, the commonly used notion of asymptotic convergence is

meaningless without lower semi-continuity of the objective function because in the absence of this lower semi-continuity, asymptotic convergence to an optimal *solution* does not necessarily imply asymptotic convergence to the optimal *objective function value*. Finally, the closedness condition on  $\alpha \circ \beta$  is more or less standard ever since it was identified by Zangwill (1969) as a crucial ingredient for general asymptotic convergence to hold.

It is natural to explore next what can be said about convergence if a *finitely parametrizable* representation of the dual space is available. Roughly speaking, this means that the relevant dual solutions  $g \in \Delta$  are fully characterized by only a finite number of parameters (e.g. belong to some *finite* dimensional functional space). Finite parametrizability is important since without it there is little hope of computing (optimal) dual solutions at all.

**Theorem 3.5** (Closedness of  $\alpha \circ \beta$ ). *Suppose that  $X$  is closed and  $\varphi(P(\cdot))$  is upper semi-continuous on  $FS_X$ . In addition, let there be a non-empty and compact set  $T \subseteq \mathbb{R}^\tau$  ( $\tau \in \mathbb{N}$ ) and function  $w: T \times \mathbb{R}^{m+n_1} \rightarrow \mathbb{R} \cup \{\pm\infty\}$  which is continuous on  $T \times \{(r, x)\} \forall (r, x) \in RHS$  and on  $T \times \{0\} \times FS_X$ , which is upper semi-continuous on  $T \times \{0\} \times X$ , and which satisfies*

$$\forall (x, \varepsilon_3) \in FS_X \times E_3: [(z, \theta) \in (\alpha \circ \beta)(x, \varepsilon_3) \\ \Rightarrow \exists t \in T: w(t, \cdot, \cdot) \in \beta(x, \varepsilon_3) \wedge (z, \theta) \in \alpha(w(t, \cdot, \cdot))].$$

Then  $\alpha \circ \beta$  is closed on  $FS_X \times E_3$ .

**Proof.** Let  $(x^k, \varepsilon_3^k)_{\mathbb{N}} \in FS_X \times E_3$  converge to  $(x^\infty, \varepsilon_3^\infty) \in FS_X \times E_3$  and let  $(z^k, \theta^k)_{\mathbb{N}}$  converge to  $(z^\infty, \theta^\infty)$ , where  $(z^k, \theta^k) \in (\alpha \circ \beta)(x^k, \varepsilon_3^k) \forall k \in \mathbb{N}$ . We have to prove that  $(z^\infty, \theta^\infty) \in (\alpha \circ \beta)(x^\infty, \varepsilon_3^\infty)$ .

We know that  $\forall k \in \mathbb{N}, \exists t^k \in T: -w(t^k, 0, z^k) + \theta^k \leq 0$ .  $T$  is compact, so without loss of generality we may assume that  $\lim t^k = t^\infty \in T$ . The (upper semi-) continuity assumptions on  $w$  imply that  $-w(t^\infty, 0, z^\infty) + \theta^\infty \leq 0$  and that  $w(t^\infty, \cdot, \cdot) \in \Delta$ . Now,  $w(t^k, \cdot, \cdot)$  is  $\varepsilon_3^k$ -optimal for  $D(x^k)$ , which, combined with the continuity of  $w$  and the upper semi-continuity of  $\varphi(P(\cdot))$ , implies that  $w(t^\infty, \cdot, \cdot)$  is  $\varepsilon_3^\infty$ -optimal for  $D(x^\infty)$ . This concludes the proof.  $\square$

We can use a similar approach to investigate the case in which the algorithm is only able to generate a finite number of feasible solutions from  $FS_X$  (i.e.  $|I| < +\infty$ ). If the feasible set of  $D^0$  can be restricted to finitely parametrizable solutions, then under appropriate assumptions one can show that every accumulation point of  $(x^k)_{\mathbb{N}, I}$  belongs to  $FS_X$ , so that at least asymptotic feasibility is assured (for details, see Flippo, 1991). Thus, the possibility of finite parametrization, a notion that does *not* necessarily require the usual *convexity properties*, enables strong properties of convergence to be established in a natural way.

#### 4. Constraint Decomposition

The Constraint Decomposition approach that we are about to introduce next, generalizes Dantzig-Wolfe Decomposition in much the same way that Variable Decomposition generalizes Benders's original work. The idea underlying the approach is also known as *Column Generation*, *Generalized Linear Programming*, (*Generalized*) *Dantzig-Wolfe Decomposition*, *Dual Decomposition* and *Price Directive Decomposition*. As a methodology, Constraint Decomposition is captured by the notions of *Inner Approximation* and *Relaxation* (cf. Geoffrion, 1972a). Let us consider the original *primal program* (1)-(3) under the additional assumption that

$$\forall x \in X: f(x) < +\infty. \tag{35}$$

For all  $\bar{X} \subseteq X$  the following dual pair of programs are defined

$$P(\bar{X}): \max_x f(x) \tag{36}$$

$$\text{s.t. } G(x) \diamond 0, \tag{37}$$

$$x \in \bar{X}, \tag{38}$$

$$D(\bar{X}): \min_g g(0) \tag{39}$$

$$\text{s.t. } g(G(x)) \geq f(x) \quad \forall x \in \bar{X}, \tag{40}$$

$$g \in \Gamma, \tag{41}$$

where  $\Gamma$  consists of all functions in  $\mathcal{F}$  with the exception of those that can take on the value  $-\infty$  (cf. (4)); this does not affect strong duality between  $P(\bar{X})$  and  $D(\bar{X})$ , provided the value of the former is strictly greater than  $-\infty$  (see Flippo, 1991). The reason for introducing (35) and (41) is to ensure that  $f(x) - g(G(x))$ , an expression that will occur frequently in this section, is well-defined for every  $x \in X$  and  $g \in \Gamma$ .

$P(X)$  and  $D(X)$  are clearly equivalent to  $P$  and  $D$ , and are called the *primal* and *dual master program* respectively;  $P(\bar{X})$  and  $D(\bar{X})$  are the *restricted primal* and *relaxed dual master programs* respectively. The latter programs provide (arbitrarily tight) *lower bounds* on  $\varphi(P)$ . For given  $\bar{g} \in \Gamma$ , an *upper bound* will be provided through a *generalized version of Lagrangean Relaxation*:

$$CD(\bar{g}): \max_x f(x) + \bar{g}(0) - \bar{g}(G(x))$$

$$\text{s.t. } x \in X.$$

Indeed, if  $\bar{g} \in \Gamma$  and  $\bar{g}(0) \in \mathbb{R}$  then

$$\begin{aligned} \varphi(CD(\bar{g})) &= \sup_x \{f(x) + \bar{g}(0) - \bar{g}(G(x)) \mid x \in X\} \\ &\geq \sup_x \{f(x) + \bar{g}(0) - \bar{g}(G(x)) \mid G(x) \diamond 0, x \in X\} \\ &\geq \sup_x \{f(x) \mid G(x) \diamond 0, x \in X\} = \varphi(P). \end{aligned}$$

These upper bounds can also be as tight as desired because if  $\bar{g}$  is  $\varepsilon$ -optimal for D, it follows that (cf. (40))

$$\begin{aligned}\varphi(\text{CD}(\bar{g})) &= \bar{g}(0) + \sup_x \{f(x) - \bar{g}(G(x)) \mid x \in X\} \\ &\leq \bar{g}(0) \leq \varphi(D) + \varepsilon = \varphi(P) + \varepsilon.\end{aligned}$$

It is now natural to consider an iterative procedure in which the relaxed dual master programs  $D(\bar{X})$  generate candidate  $\bar{g}$ 's which, through  $\text{CD}(\bar{g})$ , yield improvements to  $\bar{X}$ . We describe this Constraint Decomposition Procedure now, using the same notation as in Section 3.

### Constraint Decomposition Procedure.

*Step 0. Initialization phase.*

$$k := 1, \text{UB}^0 := +\infty; \text{LB}^0 := -\infty;$$

Choose  $X^k \subseteq X$  such that  $\varphi(P(X^k)) > -\infty$ ;

*Step 1. Master program phase.*

**if**  $\varphi(P(X^k)) = +\infty$  **then stop** [ $\varphi(P) = +\infty$ ]

**else begin**

Choose  $(x^k, \varepsilon_1^k)$  such that  $\varepsilon_1^k \geq 0$  and  $x^k$  is  $\varepsilon_1^k$ -optimal for  $P(X^k)$ ;

Choose  $(g^k, \varepsilon_2^k)$  such that  $\varepsilon_2^k \geq 0$  and  $g^k$  is  $\varepsilon_2^k$ -optimal for  $D(X^k)$ ;

$$\text{LB}^k := \max\{\text{LB}^{k-1}, f(x^k)\};$$

**if**  $\text{LB}^k > \text{LB}^{k-1}$  **then**  $x^{\text{inc},k} := x^k$  **else**  $x^{\text{inc},k} := x^{\text{inc},k-1}$

**end;**

*Step 2. Subprogram phase.*

**if**  $\varphi(\text{CD}(g^k)) = +\infty$  **then begin**

$$\varepsilon_3^k := 0; \text{UB}^k := \text{UB}^{k-1};$$

Choose  $z^k \in X$  such that  $f(z^k) + g^k(0) - g^k(G(z^k)) > g^k(0)$ ;

Choose  $X^{k+1}$  such that  $X^k \cup \{z^k\} \subseteq X^{k+1} \subseteq X$

**end else begin**

Choose  $(z^k, \varepsilon_3^k)$  such that  $\varepsilon_3^k \geq 0$  and  $z^k$  is  $\varepsilon_3^k$ -optimal for  $\text{CD}(g^k)$ ;

$$\text{UB}^k := \min\{\text{UB}^{k-1}, f(z^k) + g^k(0) - g^k(G(z^k)) + \varepsilon_3^k\};$$

Choose  $X^{k+1}$  such that  $X^k \cup \{z^k\} \subseteq X^{k+1} \subseteq X$

**end;**

*Step 3. Optimality verification.*

Choose  $\varepsilon_0^k \geq \varepsilon_1^k + \varepsilon_2^k + \varepsilon_3^k$ ;

**if**  $\text{UB}^k - \text{LB}^k \leq \varepsilon_0^k$  **then stop** [ $x^{\text{inc},k}$  is  $\varepsilon_0^k$ -optimal for P]

**else**  $k := k + 1$ ; **return to 1.**

A few comments are in order.

- Finding an initial  $X^1 \subseteq X$  for which  $\varphi(P(X^1)) > -\infty$  may be a non-trivial task. If no such set exists then clearly  $\varphi(P) = -\infty$ .
- Since  $f(z^k) + g^k(0) - g^k(G(z^k)) + \varepsilon_3^k$  is an upper bound on  $\varphi(\text{CD}(g^k))$ , it is also one on  $\varphi(P)$ .

- The inaccuracy parameters  $\varepsilon_i^k$  ( $i = 1, 2, 3$ ) may vary between iterations and need not be specified prior to iteration  $k$ .

- In the optimality verification phase,  $x^{\text{inc},k}$  is  $(UB^k - LB^k)$ -optimal for P because  $0 \leq \varphi(P) - f(x^{\text{inc},k}) \leq UB^k - LB^k$ . Termination occurs if the bound  $\varepsilon_0^k$  on the overall inaccuracy is met.

- Optimality verification in this procedure is essentially verification of *dual optimality*. To see why, consider the most recent improvement in the upper bound up until iteration  $k$  (say, in iteration  $j \leq k$ ) and define  $\tilde{g}$  by

$$\tilde{g}(\cdot) = g^j(\cdot) + f(z^j) - g^j(G(z^j)) + \varepsilon_3^j.$$

Note that  $UB^{j-1} > UB^j = UB^k = \tilde{g}(0)$ . Now,  $\varepsilon_3^j$ -optimality of  $z^j$  can easily be seen to imply that  $f(x) - \tilde{g}(G(x)) \leq 0 \quad \forall x \in X$ , and it follows readily that  $\tilde{g}$  is feasible for D. Furthermore,  $\tilde{g}(0) - \varphi(D) \leq UB^k - LB^k$ , so  $UB^k - LB^k \leq \varepsilon_0^k$  implies that  $\tilde{g}$  is  $\varepsilon_0^k$ -optimal for D.

- As in the previous section, this framework allows for inaccuracies during the iterative process, as well as for duality gaps between the primal and dual master programs that are encountered during the solution procedure, since these gaps are in fact part of the inaccuracies  $\varepsilon_2^k$  of the relaxed dual master program solutions.

As in the case of Variable Decomposition, we easily obtain non-repetition and finite convergence under appropriate assumptions.

**Theorem 4.1** (Non-repetition of complete solutions).

- (a) *In the master program phase no solution  $g^k$  will be generated more than once.*
- (b) *As soon as a solution  $z^k$  is generated for a second time, the algorithm will terminate.*

**Proof.** (a) To prove the result, it suffices to establish that (cf. (40))

$$f(z^k) - g^k(G(z^k)) > 0. \tag{42}$$

If  $\varphi(\text{CD}(g^k)) = +\infty$ , (42) follows trivially. Else,

$$\begin{aligned} & (f(z^k) + g^k(0) - g^k(G(z^k)) + \varepsilon_3^k) - (g^k(0) - \varepsilon_1^k - \varepsilon_2^k) \\ & \geq UB^k - f(x^k) \geq UB^k - LB^k. \end{aligned} \tag{43}$$

If the right-hand side of (43) is strictly larger than  $\varepsilon_0^k$ , then (42) follows immediately; if not, the algorithm terminates anyway.

(b) If  $z^l = z^k$  for some  $l > k$ , then (cf. (40))

$$f(z^l) = f(z^k) \leq g^l(G(z^k)) = g^l(G(z^l))$$

and so, in view of the argument under (a),  $UB^l - LB^l$  cannot be strictly larger than  $\varepsilon_0^l$ .  $\square$

To establish the equivalent of Theorem 3.2, let us call two functions  $g$  and  $g'$  *essentially identical* if their difference is a constant. Clearly,  $CD(g)$  is then equivalent to  $CD(g')$ .

**Theorem 4.2** (Non-repetition of subprograms). *Assume that a real-valued upper bound has been obtained at iteration  $k_0$ , and that from that iteration onwards,  $z^k$  is chosen so as to ensure that  $f(z^k) + g^k(0) - g^k(G(z^k)) \geq UB^k$  if  $\varphi(CD(g^k)) = +\infty$ , and  $\varepsilon_0^k \geq \varepsilon_1^k + \varepsilon_2^k + \max_{k_0 \leq j \leq k} \{\varepsilon_3^j\}$  if  $\varphi(CD(g^k)) < +\infty$ . Then the procedure terminates if a solution  $g^k$  is generated which is essentially identical to a previously generated solution  $g^j$  ( $k > j \geq k_0$ ).*

**Proof.** We know that  $z^j \in X^k$ , implying that  $f(z^j) + g^k(0) - g^k(G(z^j)) \leq g^k(0)$  (cf. (40)). Moreover, if  $g^j$  is essentially identical to  $g^k$ , then  $g^j(0) - g^j(G(z^j)) = g^k(0) - g^k(G(z^j))$ . If  $\varphi(CD(g^j)) = \varphi(CD(g^k)) = +\infty$ , the above two statements imply that

$$\begin{aligned}
 UB^k - LB^k &\leq UB^j - f(x^k) \\
 &\leq (f(z^j) + g^j(0) - g^j(G(z^j))) - (g^k(0) - \varepsilon_1^k - \varepsilon_2^k) \leq \varepsilon_0^k \quad (44)
 \end{aligned}$$

and the procedure terminates after iteration  $k$ . On the other hand, if  $\varphi(CD(g^j)) = \varphi(CD(g^k)) < +\infty$ , then the proof is similar, provided  $\varepsilon_3^j$  is incorporated in (44) from the second inequality onwards.  $\square$

**Theorem 4.3** (Finite convergence). *The procedure terminates in a finite number of steps if at least one of the following two conditions is met:*

- (a) *all  $z^k$  generated belong to a finite subset of  $X$ ;*
- (b) *the assumptions from Theorem 4.2 apply and, in addition, all  $g^k$  generated are essentially equivalent to a member of a finite subset of  $\Gamma$ .*  $\square$

The proof of Theorem 4.3 trivially follows from its two predecessors. It is not difficult to see that Dantzig-Wolfe Decomposition (1960) for linear programs, as well as its extension in Dantzig (1963) to convex ones, are both subsumed by the above approach. In the former LP approach,  $X$  is a polyhedral set and  $\bar{X}$  corresponds to the set of all convex combinations of a subset of all the extreme points of  $X$  plus the set of all non-negative combinations of a subset of all the extreme rays of  $X$  ("plus" in the sense of set addition). In the later extension where  $X$  is a general convex set,  $\bar{X}$  is chosen to be some *finite* subset of  $X$  and the dual solution space  $\Gamma$  is restricted to the set of *affine* functions only. The relaxed dual master programs then turn out to be *linear* programs, so that it is only natural for their LP duals to replace the initial restricted primal master programs (a process which is also known as *convexification*; cf. Magnanti et al., 1976). In Burkard et al. (1985) an algebraic extension of Constraint Decomposition is discussed. The additional generality of their framework is due to the fact that the primal spaces are not required to be metric spaces. A consequence of this is that in their framework inaccuracies and

duality gaps are not accounted for, and an asymptotic convergence result like Theorem 4.4 is absent. For more details we refer to Flippo (1991).

The case covered by Dantzig (1963) once again illustrates the crucial role played by *finite parametrization*. If, indeed, we assume that  $D(\bar{X})$  allows such finite parametrization, asymptotic convergence is guaranteed without any finiteness conditions at all.

**Theorem 4.4** (Asymptotic convergence). *Assume that*

- $P$  is regular;
- $X$  is compact;
- $f$  and  $G$  are continuous on  $X$ ;
- $\forall k: \varepsilon_i^k \in E_i \subseteq \mathbb{R}_+$  with  $E_i$  compact ( $i=1, 2, 3$ );
- there is a non-empty and compact set  $T \subseteq \mathbb{R}^\tau$  ( $\tau \in \mathbb{N}$ ) and a function  $w: T \times \mathbb{R}^m \rightarrow \mathbb{R} \cup \{\pm\infty\}$  which is continuous and real-valued on  $T \times \text{RHS}$ , such that  $w(t, \cdot) \in I \forall t \in T$  and such that if  $g^k$  is generated in the master program phase, then  $\exists t^k \in T: w(t^k, \cdot) = g^k$ .

*Then the Constraint Decomposition Procedure converges asymptotically, in the sense that*

- (a) for every accumulation point  $t^\infty$  of  $(t^k)_\mathbb{N}$ ,  $w(t^\infty, \cdot) + \varepsilon_3^\infty$  is an  $(\varepsilon_2^\infty + \varepsilon_3^\infty)$ -optimal solution for  $D(X)$  for some accumulation point  $(\varepsilon_2^\infty, \varepsilon_3^\infty)$  of  $(\varepsilon_2^k, \varepsilon_3^k)_\mathbb{N}$ ;
- (b) every accumulation point of  $(x^k)_\mathbb{N}$  is a  $\limsup(\varepsilon_1^k + \varepsilon_2^k + \varepsilon_3^k)$ -optimal solution to  $P$ ;
- (c) every accumulation point of  $(x^{\text{inc},k})_\mathbb{N}$  is a  $\limsup(\varepsilon_1^k + \varepsilon_2^k + \varepsilon_3^k)$ -optimal solution to  $P$ ;
- (d)  $0 \leq \lim(\text{UB}^k - \text{LB}^k) \leq \limsup(\varepsilon_1^k + \varepsilon_2^k + \varepsilon_3^k)$ .

**Proof.** By way of preliminary observation, let us note that under (35), (41) and the continuity and compactness conditions of the theorem, all  $\text{CD}(g^k)$ 's are regular programs.

(a) Let  $t^\infty$  be any accumulation point of  $(t^k)_\mathbb{N}$ . Define the function  $g^\infty = w(t^\infty, \cdot)$ , and consider the subsequence  $(p(k))_\mathbb{N}$  for which

$$\lim(t^{p(k)}, \varepsilon_2^{p(k)}, \varepsilon_3^{p(k)}, z^{p(k)}) = (t^\infty, \varepsilon_2^\infty, \varepsilon_3^\infty, z^\infty)$$

for some  $(\varepsilon_2^\infty, \varepsilon_3^\infty, z^\infty) \in E_2 \times E_3 \times X$ . Now,  $\varepsilon_3^{p(k)}$ -optimality of  $z^{p(k)}$  implies that

$$\forall x \in X: f(x) - w(t^{p(k)}, G(x)) \leq f(z^{p(k)}) - w(t^{p(k)}, G(z^{p(k)})) + \varepsilon_3^{p(k)}$$

and, since this inequality carries over to the limit,  $z^\infty$  is  $\varepsilon_3^\infty$ -optimal for  $\text{CD}(w(t^\infty, \cdot))$ . Also,  $z^{p(k)} \in X^{p(k+1)}$ , so (cf. (40))

$$w(t^{p(k+1)}, G(z^{p(k)})) \geq f(z^{p(k)}).$$

Taking limits again we find that

$$w(t^\infty, G(z^\infty)) \geq f(z^\infty), \tag{45}$$

This, in combination with  $\varepsilon_3^\infty$ -optimality of  $z^\infty$ , yields that

$$\forall x \in X: w(t^\infty, G(x)) + \varepsilon_3^\infty \geq f(x)$$

so that  $g^\infty(\cdot) + \varepsilon_3^\infty$  is feasible for D. Moreover,

$$w(t^{p(k)}, 0) - \varepsilon_2^{p(k)} \leq \varphi(D(X^{p(k)})) \leq \varphi(D). \tag{46}$$

In the limit, (46) implies that, indeed,  $g^\infty(\cdot) + \varepsilon_3^\infty$  is  $(\varepsilon_2^\infty + \varepsilon_3^\infty)$ -optimal for D.

(b) As before, we consider an accumulation point and a subsequence converging to it. Now,

$$\begin{aligned} f(x^\infty) &= \lim f(x^{p(k)}) \geq \limsup(\varphi(P(X^{p(x)})) - \varepsilon_1^{p(k)}) \\ &\geq \lim(w(t^{p(k)}, 0) - \varepsilon_1^{p(k)} - \varepsilon_2^{p(k)}) \\ &= (w(t^\infty, 0) + \varepsilon_3^\infty) - (\varepsilon_1^\infty + \varepsilon_2^\infty + \varepsilon_3^\infty) \\ &\geq \varphi(P) - (\varepsilon_1^\infty + \varepsilon_2^\infty + \varepsilon_3^\infty), \end{aligned} \tag{47}$$

which, combined with the primal feasibility of  $x^\infty$ , establishes the result; the last inequality follows from the dual feasibility of  $g^\infty(\cdot) + \varepsilon_3^\infty$ .

(c) The proof is as above, making use of the fact that  $f(x^{\text{inc},k}) \geq f(x^k)$ .

(d) With  $UB^k$  and  $LB^k$  converging to, say  $UB^\infty$  and  $LB^\infty$  respectively, and with converging subsequences constructed as before, (45) implies that on the one hand,

$$UB^\infty \leq f(z^\infty) + w(t^\infty, 0) - w(t^\infty, G(z^\infty)) + \varepsilon_3^\infty \leq w(t^\infty, 0) + \varepsilon_3^\infty. \tag{48}$$

On the other hand (cf. (47)),

$$LB^\infty \geq f(x^\infty) \geq w(t^\infty, 0) - \varepsilon_1^\infty - \varepsilon_2^\infty. \tag{49}$$

Together, (48) and (49) establish the result.  $\square$

As in Section 3, *finite convergence* to an  $\varepsilon_0^k$ -optimal solution can be assured if under the above conditions,  $\varepsilon_0^k$  exceeds  $\varepsilon_1^k + \varepsilon_2^k + \varepsilon_3^k$  by a strictly positive constant independent of  $k$ . Under these conditions one can also establish the equivalent of a well known property of Lagrangean Relaxation, in that every accumulation point  $z^\infty$  of  $(z^k)_\mathbb{N}$  is a  $\limsup \varepsilon_3^k$ -optimal solution to the program P in which the right-hand-side has been changed from 0 to  $G(z^\infty)$  (for details, see Flippo, 1991).

The reader may have noticed a great deal of similarity between the two types of decomposition procedures introduced above; indeed, in the next section they will be seen to be dual to each other in a well defined sense. From that perspective, the absence of an analogue to the general asymptotic convergence result in Theorem 3.4 is disturbing. On second thought however, it is perhaps not too surprising. The analogous construction of point-to-set maps would involve an appropriate definition of closedness of point-to-set maps in infinite dimensional functional spaces, a definition that is not at all obvious. Indeed, the standard norms that one could impose on such spaces (e.g.  $\sup\{|g(r)|\}$  or  $\int |g(r)| dr$ ) to introduce a notion of convergence, all have the unpleasant side effect that the dual solutions, and hence the value function, should have “bounded” or even “small tails” (i.e.  $\sup\{|g(r)|\} \leq M$  for some  $M \geq 0$  or  $g(r) \rightarrow 0$  for  $\|r\| \rightarrow \infty$ ). This would require strong conditions on the original problem P; if, for instance, the latter norm is used, then inequality

constraints can no longer be accounted for. Thus, one could argue that the asymmetry noted between the two types of decomposition methods simply reflects the fundamental asymmetry — finite versus infinite dimensionality — between the primal and the dual problem in the Tind–Wolsey duality framework, and that the best one can hope for is perfect symmetry in the case that both primal and dual are essentially finite dimensional. Theorems 3.5 and 4.4 accomplish precisely that. The lack of a point-to-set mapping framework also implies that for an asymptotic convergence analysis we have to rely on what actually happens during the iterative procedure. This might explain why Theorem 3.4 only requires the *existence* of finitely parametrizable dual solutions, whereas the proof of Theorem 4.4 requires their actual *generation*. From a computational point of view this is, of course, not at all restrictive.

### 5. Duality between decomposition methods

The above presentation of Variable and Constraint Decomposition brings out the striking symmetry between these procedures, in what appears to be an attractive didactical setting. In Variable Decomposition the master program produces relaxations and the subprograms produce restrictions of the given problem; in Constraint Decomposition it is the other way round. In Variable Decomposition the master program sends primal solutions to the subprograms and receives dual solutions in return; in Constraint Decomposition, once again, the reverse is the case. Indeed, as we shall see below, these two approaches can be seen as dual to each other, thus generalizing the relationship between Benders and Dantzig–Wolfe Decomposition for LP as outlined in Lasdon (1970). Our starting point will be the (regular) primal program

$$P: \quad \max_x f(x) \tag{50}$$

$$\text{s.t.} \quad G(x) \diamond 0, \tag{51}$$

$$H(x) \circ 0, \tag{52}$$

$$x \in X, \tag{53}$$

where  $f$ ,  $G$  and  $H$  are functions from  $X$  to  $\mathbb{R} \cup \{-\infty\}$ ,  $\mathbb{R}^{m_1}$  and  $\mathbb{R}^{m_2}$  respectively,  $\diamond \in \{\leq, =\}^{m_1}$  and  $\circ \in \{\leq, =\}^{m_2}$ . The additively separable dual reads (cf. Theorem 2.1)

$$D: \quad \min_{g,h} g(0) + h(0) \tag{54}$$

$$\text{s.t.} \quad g(G(x)) + h(H(x)) \geq f(x) \quad \forall x \in X, \tag{55}$$

$$g \in \Gamma_G, \quad h \in \Gamma_H, \tag{56}$$

with  $\Gamma_G$  and  $\Gamma_H$  being defined appropriately (cf. Section 4). Let us first apply *Constraint Decomposition* to (50)–(53) by restricting  $x$  to a subset  $\bar{\Pi} \subseteq \Pi = \{x \in X \mid H(x) \circ 0\}$ . In doing so, we obtain restricted primal and relaxed dual master

programs  $P(\bar{\Pi})$  and  $D(\bar{\Pi})$ , and subprograms  $CD(\bar{g})$ , which are all similarly defined as their counterparts in Section 4 with  $X, \bar{X}$  and  $\Gamma$  replaced by  $\Pi, \bar{\Pi}$  and  $\Gamma_G$ . Now let us apply *Variable Decomposition* to (54)–(56). By fixing  $g = \bar{g}$  we obtain

$$D(\bar{g}): \quad \min_h \quad \bar{g}(0) + h(0)$$

$$\text{s.t.} \quad h(H(x)) \geq f(x) - \bar{g}(G(x)) \quad \forall x \in X,$$

$$h \in \Gamma_H.$$

Note that, apart from  $\bar{g}(0)$ ,  $D(\bar{g})$  is in fact the dual program of  $CD(\bar{g})$  (cf. Section 2), so provided that  $\bar{g}(0) \in \mathbb{R}$ , *projection* and *dualization* yields

$$\min_{\bar{g}} \quad \max_x \quad f(x) + \bar{g}(0) - \bar{g}(G(x))$$

$$\text{s.t.} \quad H(x) \circ 0,$$

$$x \in X,$$

$$\text{s.t.} \quad \bar{g}(0) \in \mathbb{R}, \quad \bar{g} \in \Gamma_G.$$

Introducing a dummy variable  $\theta$  yields the following equivalent formulation

$$\min_{\bar{g}, \theta} \quad \theta$$

$$\text{s.t.} \quad \theta \geq f(x) + \bar{g}(0) - \bar{g}(G(x)) \quad \forall x \in \Pi,$$

$$\bar{g}(0) \in \mathbb{R}, \quad \bar{g} \in \Gamma_G, \quad \theta \in \mathbb{R}.$$

After a change of variables  $g(\cdot) = \bar{g}(\cdot) - \bar{g}(0) + \theta$  and restriction to the subset  $\bar{\Pi} \subseteq \Pi$ , we obtain a relaxed master program

$$VD(\bar{\Pi}): \quad \min_g \quad g(0)$$

$$\text{s.t.} \quad g(G(x)) \geq f(x) \quad \forall x \in \bar{\Pi},$$

$$g(0) \in \mathbb{R}, \quad g \in \Gamma_G,$$

which, provided that  $\varphi(P(\bar{\Pi})) > -\infty$ , is exactly identical to  $D(\bar{\Pi})$ ! Thus, the dual relationship between the two decomposition approaches has been made precise. It is now illuminating to think of the role of the subprogram in *Variable Decomposition* as providing information on how  $P$  is affected by fixing some of the *primal* decision variables, whereas in *Constraint Decomposition* the effect on  $P$  of fixing some *dual* variables is revealed; apparently, both the primal feasible set and the primal objective function are affected in that case.

The dual relationship between the two approaches suggests that mixtures of the two ought to be feasible, and indeed they are. A particularly interesting example is a generalized version of *Cross Decomposition* (Van Roy, 1980, 1983), which applies whenever the primal program would allow both *Variable* and *Constraint Decomposition* (cf. (9)–(11)). In their regular implementations, the relaxed (dual) master

programs would supply the input for their respective subprograms. But since the former are usually much harder to solve, it is tempting to iterate between the two types of subprograms instead. This is exactly what Cross Decomposition does. As before, upper and lower bounds are generated each iteration, and thus the usual termination criterion applies. This attractive simplification, however, comes at a price; non-repetition, let alone convergence, cannot be guaranteed. The intuitive reason for this is that the subprograms rely only on recently obtained information, whereas the master programs accumulate all information over time. Thus, at least an occasional call on a full master program is required to guarantee convergence for Cross Decomposition. In general, this also applies to other hybrid techniques such as *Kornai-Lipták Decomposition*, although for specific *linear* programs the latter method is proven to converge asymptotically (cf. Kornai and Lipták, 1965; and Holmberg, 1990; see also Dirickx and Jennergren, 1979; Aardal and Ari, 1990; Flippo and Rinnooy Kan, 1991; and Holmberg, 1992).

Other decomposition techniques can also be recovered as simple variations on one of the two main types. For example, if one duplicates all or some variables, adding constraints to force equality between a variable and its duplicate, and one applies Constraint Decomposition with  $X$  equal to all but the latter constraints, then *Variable Splitting* (Jörnsten et al., 1985) or *Lagrangean Decomposition* (Guignard and Kim, 1987) is obtained. If in Constraint Decomposition the master program is replaced by a (heuristic) dual update mechanism, then one readily recognizes the popular technique of *Lagrangean Relaxation*.

Our conclusion would be that all this confirms that the above framework is sufficiently general to capture a large variety of well known optimization techniques, and yet does not lack the analytical content necessary to derive essential formal properties.

## Acknowledgements

We would like to thank Michel Minoux, Gerard Van Der Hoek and Richard Wong for many fruitful discussions and helpful comments during the course of this research.

## References

- K. Aardal and A. Ari, "On the resemblance between the Kornai-Lipták and cross decomposition techniques for block-angular linear programs," *European Journal of Operational Research* 46 (1990) 393-398.
- J.F. Benders, "Partitioning procedures for solving mixed variables programming problems," *Numerische Mathematik* 4 (1962) 238-252.
- R.E. Burkard, H.W. Hamacher and J. Tind, "On general decomposition schemes in Mathematical Programming," *Mathematical Programming Study* 24 (1985) 238-252.
- G.B. Dantzig, *Linear Programming and Extensions* (Princeton University Press, Princeton, NJ, 1963).

- G.B. Dantzig and P. Wolfe, "Decomposition principle for Linear Programming," *Operations Research* 8 (1960) 101-111.
- Y.M.I. Dirickx and L.P. Jennergren, *System Analysis by Multi-Level Methods with Applications to Economics and Management* (Wiley, New York, 1979).
- O.E. Flippo, *Stability, Duality and Decomposition in General Mathematical Programming*, CWI Tract No. 76 (Centre for Mathematics and Computer Science, Amsterdam, 1991).
- O.E. Flippo and M. Minoux, "Continuity of the value-function in Mathematical Programming: a unifying framework with new results," Report 8903/A, Econometric Institute, Erasmus University Rotterdam (Rotterdam, 1989); Document No. 52, LAMSADE, Université de Paris Dauphine (Paris, 1989).
- O.E. Flippo and A.H.G. Rinnooy Kan, "A note on Benders decomposition in mixed-integer quadratic programming," *Operations Research Letters* 9 (1990) 81-83.
- O.E. Flippo and A.H.G. Rinnooy Kan, "Variable decomposition, constraint decomposition and cross decomposition in general mathematical programming," in: Akgül et al., eds., *Combinatorial Optimization*, NATO ASI Series Vol. F 82 (Springer, Heidelberg, 1992) pp. 1-18.
- A.M. Geoffrion, "Elements of large-scale mathematical programming," in: A.M. Geoffrion, ed., *Perspectives on Optimization: A Collection of Expository Articles* (Addison-Wesley, Reading MA, 1972) pp. 25-64.
- A.M. Geoffrion, "Generalized Benders decomposition," *Journal of Optimization Theory and Application* 10 (1972) 237-260.
- M. Guignard and S. Kim, "Lagrangean decomposition: a model yielding stronger Lagrangean bounds," *Mathematical Programming* 39 (1987) 215-228.
- K. Holmberg, "A convergence proof for linear mean value cross decomposition," Research Report No. 90/12, Matematisk Institut, Aarhus University (Aarhus, 1990).
- K. Holmberg, "Linear mean value cross decomposition: A generalization of the Kornai-Lipták method," *European Journal of Operational Research* 62 (1992) 55-73.
- K.O. Jörnsten, M. Näsberg and P.A. Smeds, "Variable Splitting — a new Lagrangean relaxation approach to some mathematical programming models," Working Paper LiTh-MAT-R-85-04, Linköping University (Linköping, 1985).
- J. Kornai and Th. Lipták, "Two-level planning," *Econometrica* 33 (1965) 141-169.
- L.S. Lasdon, *Optimization Theory for Large Systems* (Macmillan, London, 1970).
- R. Lazimy, "Mixed-integer quadratic programming," *Mathematical Programming* 22 (1982) 332-349.
- R. Lazimy, "Improved algorithm for mixed-integer quadratic programs and a computational study," *Mathematical Programming* 32 (1985) 100-113.
- T.L. Magnanti, J.F. Shapiro and M.H. Wagner, "Generalized linear programming solves the dual," *Management Science* 22 (1976) 1195-1203.
- R. Rouhani, L. Lasdon, W. Lebow and A.D. Waren, "A generalized Benders decomposition approach to reactive source planning in power systems," *Mathematical Programming Study* 25 (1985) 62-75.
- J. Tind, "Decomposition procedures for nonlinear and discrete models in hierarchical planning," Publication No. 88/2, Department of Operations Analysis, Aarhus University (Aarhus, 1988).
- J. Tind, "Decomposition in global optimization," Publication No. 90/7, Department of Operations Analysis, Aarhus University (Aarhus, 1990).
- J. Tind and L.A. Wolsey, "An elementary survey of general duality theory in Mathematical Programming," *Mathematical Programming* 21 (1981) 241-261.
- T.J. Van Roy, "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, 1980).
- T.J. Van Roy, "Cross decomposition for mixed integer programming," *Mathematical Programming* 25 (1983) 46-63.
- L.A. Wolsey, "A resource decomposition algorithm for general mathematical programs," *Mathematical Programming Study* 14 (1981) 244-257.
- W.I. Zangwill, *Nonlinear Programming: A Unified Approach* (Prentice-Hall, Englewood Cliffs, NJ, 1969).