ON THE NUMBER OF ITERATIONS IN DANTZIG-WOLFE DECOMPOSITION ALGORITHM

by

Ilan Adler and Aydin Ülkücü
Department of Industrial Engineering
and Operations Research
University of California, Berkeley

May 1972

INTRODUCTION

Given a linear program in standard form, Dantzig-Wolfe decomposition algorithm replaces the original polytope by another one with many more variables but fewer equations. In this note we investigate the relations between the combinatorial structures of the two polytopes. In particular we show that, contrary to what one intuitively expects, the diameter of the original polytope may be smaller than the diameter of the decomposed one. Since the diameter of a polytope gives the maximal number of iterations taken by the "ideal" vertex-following algorithm, this observation may provide a clue for the slow convergence of the decomposition algorithm reported by some authors.

PRELIMINARY RESULTS

Given a polytope $\bar{\underline{X}} = \{x \mid Ax = b, x \geq 0\}$ where A is an $m \times n$ matrix, b is an m-vector and x is an n-vector. Let us partition A and b such that $A = \begin{bmatrix} A_0 \\ A_1 \end{bmatrix}$, $b = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$ where A_0 is an $m_0 \times n$ matrix and b_0 is an m_0-vector. Define $\bar{\underline{X}}_1 = \{x \mid A_1 x = b_1, x \geq 0\}$; to simplify the discussion we assume that $\bar{\underline{X}}_1$ is bounded. Let $E = (x^1, \ldots, x^k)$

where $x^i (i=1, \ldots, k)$ are all the vertices of $\underline{\bar{X}}_1$ and define

$\underline{\bar{Y}} = \left\{ \lambda \mid A_0 E\lambda = b_0 , \sum_{j=1}^{k} \lambda_j = 1 , \lambda \geq 0 \right\}$ where λ is a k-vector. In the following proposition we present the well known correspondence between $\underline{\bar{X}}$ and $\underline{\bar{Y}}$ which is the basis for the Dantzig-Wolfe decomposition algorithm (See [1]).

Proposition 1

Let T be the linear mapping from $\underline{\bar{Y}}$ to $\underline{\bar{X}}$ defined by $T(\lambda) = E\lambda$. Then $T(\underline{\bar{Y}}) = \underline{\bar{X}}$ (i.e. T is *onto* $\underline{\bar{X}}$). The proof of Proposition 1 can be found in many linear programming textbooks and hence is omitted here.

The importance of Proposition 1 comes from its use in linear programming. Given a linear program P_0: min cx subject to $x \in \underline{\bar{X}}$, one instead can solve another linear program P_1: min $g\lambda$ subject to $\lambda \in \underline{\bar{Y}}$ where $g_i = cx^i (i=1, \ldots, k)$. It follows directly from Proposition 1 that if λ^* is an optimal solution of P_1 then $T(\lambda^*)$ is an optimal solution of P_0 . (See [1]).

Since the simplex method, which is used to solve either problem, P_0 or P_1 , is a vertex following algorithm (i.e. the algorithm moves from one vertex of the given polytope to another along the edges until the optimal solution is reached) it is interesting to find out the relations between the vertices and edges of $\underline{\bar{X}}$ and $\underline{\bar{Y}}$. Propositions 2, 3 and 4 shed some light on these relationships.

Although our main interest is the relation between vertices and edges, Proposition 2 is more general and contains some information about all faces of $\underline{\bar{X}}$ and $\underline{\bar{Y}}$. So let us first introduce the notion of faces of a polytope.

The *dimension* of a given polytope $\underline{\bar{X}}$ is the maximal number of affinely independent points contained in $\underline{\bar{X}}$. ($z^0, \ldots, z^\ell$ are affinely independent if and only if $z^1 - z^0, \ldots, z^\ell - z^0$ are linearly independent.) Given a polytope $\bar{X} = \{x \mid Ax = b , x \geq 0\}$ let A' be a submatrix of A obtained by omitting some columns from A. If $\bar{X}' = \{x' \mid A'x' = b , x' \geq 0\} \neq \phi$ then $\bar{X}'$ is called a *d'-dimensional face* of $\bar{X}$ where d' is the dimension of $\bar{X}'$. Zero and one dimensional faces are called *vertices* (or *extreme points*) and *edges* respectively.

In the following lemma we give necessary and sufficient conditions for a subset of $\bar{X}$ to be a face. This lemma is widely used and easy to prove, therefore we state it without a proof.

Lemma 1

Let $\bar{X}$ be a polytope. A set F is a face of $\bar{X}$ if and only if there exists a vector c such that $F = \{z \in \bar{X} \mid cz = \min_{x \in \bar{X}} cx\}$.

MAIN RESULTS

Given a polytope $\bar{X}$ define $F_i(\bar{X})$ as the set of all i-dimensional faces of $\bar{X}$.

Proposition 2

(i) $F \in F_s(\bar{X})$ implies that $T^{-1}(F) \in F_t(\bar{Y})$ where $t \geq s$.

(ii) F is a face of $\bar{Y}$ does not necessarily imply that $T(F)$ is a face of $\bar{X}$.

Proof.

(i) By Lemma 1 there exists a vector c such that $F = \{z \in \bar{X} \mid cz = \min_{x \in \bar{X}} cx\}$. Let $cx^i = g_i \, (i=1, \ldots, k)$ and $g = (g_1, \ldots, g_k)$ then obviously, by Proposition 1, $T^{-1}(F) = \{\mu \in \bar{Y} \mid g\mu = \min_{\lambda \in \bar{Y}} g\lambda\}$. Hence, by Lemma 1, $T^{-1}(F)$

is a face of $\bar{Y}$. Since F is of dimension s there exist $s + 1$ points $z^0, \ldots, z^s$ in F which are affinely independent, i.e. $z^i - z^0 (i=1, \ldots, s)$ are linearly independent or, equivalently, $\sum_{i=1}^{s} \mu_i (z^i - z^0) = 0$ implies $\mu_i = 0 (i=1, \ldots, s)$ where μ_i are real numbers.

For $i=1, \ldots, s$ let $\lambda^i \varepsilon T^{-1}(z^i)$ then $0 = \sum_{i=1}^{s} \mu_i (E\lambda^i - E\lambda^0) = E \sum_{i=1}^{s} \mu_i (\lambda^i - \lambda^0)$ implies $\mu_i = 0 (i=1, \ldots, s)$. Hence, $\lambda^0, \ldots, \lambda^s$ are affinely independent and thus the dimension of $T^{-1}(F)$ is greater than or equal to s .

(ii) Consider the following example:

$$A = \begin{pmatrix} A_0 \\ A_1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ \hline 2 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{pmatrix} \text{ and } b = \begin{pmatrix} b_0 \\ b_1 \end{pmatrix} = \begin{pmatrix} 3/2 \\ \hline 5/2 \\ 1 \end{pmatrix}.$$

It is easy to verify that the extreme points of $\bar{X}_1$ are: (3/4,1,0,0,0); (5/4,0,0,0,1); (0,1,3/2,0,0); (0,1,0,3/2,0); (0,0,5/2,0,1); (0,0,0,5/2,1). So $\bar{Y}$ is the set of all solutions to the following system:

$$7/4\lambda_1 + 5/4\lambda_2 + 5/2\lambda_3 + \lambda_4 + 5/2\lambda_5 = 3/2$$

$$\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5 + \lambda_6 = 1$$

$$\lambda_j \geq 0 (j=1, \ldots, 5)$$

Moreover, $\bar{\lambda} = (6/7,0,0,0,0,1/7)$ is an extreme point of $\bar{Y}$ while $T(\bar{\lambda}) = (9/14,6/7,0,5/14,1/7)$ is not an extreme point of $\bar{X}$. ||

In Proposition 2 we proved that to every vertex of $\bar{\underline{X}}$ corresponds one or more vertices of $\bar{\underline{Y}}$. The effect of this fact on vertex following algorithms is twofold. The greater number of vertices in $\bar{\underline{Y}}$ suggests that such algorithms may take many more steps on $\bar{\underline{Y}}$ than on $\bar{\underline{X}}$. On the other hand a great number of vertices may present shorter routes on $\bar{\underline{Y}}$ (since every vertex has more neighbor vertices). In fact, it may be conjectured that if one uses an ideal vertex-following algorithm, then given any two vertices in $\bar{\underline{X}}$ and corresponding vertices in $\bar{\underline{Y}}$ the algorithms would take no more steps to connect those vertices in $\bar{\underline{Y}}$ than in $\bar{\underline{X}}$. In the following proposition we disprove this conjecture. This observation may partially explain the reported slow convergence of the Dantzig-Wolfe decomposition algorithm.

<u>Proposition 3</u>

Let $x^1, x^2 \in F_0(\bar{\underline{X}})$ such that x^1, x^2 are joined by an edge of $\bar{\underline{X}}$. This does not necessarily imply the existence of $\lambda^i \in T^{-1}(x^i)$, i=1, 2 such that $\lambda^1, \lambda^2 \in F_0(\bar{\underline{Y}})$ and λ^1 and λ^2 are joined by an edge of $\bar{\underline{Y}}$.

<u>Proof</u>

Consider the following example:

$$A = \begin{pmatrix} A_0 \\ A_1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 4 & 2 & 3 & 4 & 5 & 6 \\ \hline 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} \quad b = \begin{pmatrix} b_0 \\ b_1 \end{pmatrix} = \begin{pmatrix} 4/3 \\ 11 \\ \hline 1 \\ 1 \\ 1 \end{pmatrix}$$

Then $\bar{\underline{X}}_1$ is a 3-dimensional cube with 8 vertices and $\bar{\underline{Y}}$ is given by the set of all solutions to the following system:

$$2\lambda_1 + 2\lambda_2 + 2\lambda_3 + 2\lambda_4 + \lambda_5 + \lambda_6 + \lambda_7 + \lambda_8 = 4/3$$

$$15\lambda_1 + 15\lambda_2 + 12\lambda_3 + 12\lambda_4 + 12\lambda_5 + 12\lambda_6 + 9\lambda_7 + 9\lambda_8 = 11$$

$$\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5 + \lambda_6 + \lambda_7 + \lambda_8 = 1$$

$$\lambda_i \geq 0 \ , \ i=1, \ldots, 8.$$

Take $x^1 = (0,2/3,2/3,1,1/3,1/3)$, $x^2 = (1,2/3,2/3,0,1/3,1/3)$ obviously x^1 and x^2 are vertices of $\bar{\underline{X}}$ which share a common edge. However the two vertices of $\bar{\underline{Y}}$ in $T^{-1}(x^1)$ are $(0,0,1/3,0,1/3,0,1/3,0)$ and $(1/3,0,0,0,0,0,2/3,0)$ while the two vertices of $\bar{\underline{Y}}$ in $T^{-1}(x^2)$ are $(0,0,0,1/3,0,1/3,0,1/3)$ and $(0,1/3,0,0,0,0,0,2/3)$. Obviously no vertex of $T^{-1}(x^1)$ shares a common edge with some vertex of $T^{-1}(x^2)$ which proves the proposition.||

The significance of the last example becomes clearer when stated in terms of diameters of polytopes. The *diameter* of a given polytope $\bar{\underline{X}}$ is defined as the smallest integer k such that any two vertices of $\bar{\underline{X}}$ can be joined by a vertex-following path of length of less than or equal to k . In a sense, the diameter of a polytope gives the maximal number of iterations taken by the "ideal" vertex-following algorithm. The following proposition is a direct result of the example given in the proof of Proposition 3.

Proposition 4

Given polytopes $\bar{\underline{X}}$ and $\bar{\underline{Y}}$ as defined in this paper, the diameter of $\bar{\underline{Y}}$ might be greater than or equal to the diameter of $\bar{\underline{X}}$.

Remark:

The examples used in Propositions 2, 3 and 4 are by no means unique. In fact we can construct classes of such examples. In particular, it is interesting to note that these classes are constituted primarily of linear programs with bounded variables.

Reference

[1] Dantzig, G. B. "Linear Programming and Extensions," Princeton University Press, 1963.