# A DUAL DECOMPOSITION PRINCIPLE

by

C.E. Lemke and T.J. Powers

Part of the research underlying this paper was undertaken for ONR Contract Nonr-591(13).

Department of Mathematics
Rensselaer Polytechnic Institute
Troy, New York

RPI Math Report No. 48
December 4, 1961

## SUMMARY

A decomposition principle for linear programming is presented. The technique may be viewed as a dual of the Dantzig - Wolfe decomposition principle for linear programs [7]. The program matrix in what we may call the basic problem is considered as having many (an infinite number of) columns. As in the Dantzig - Wolfe treatment, one visualizes a basic problem, in primal form, where, however the set of permissible columns is not finite, as in the usual primal form, but is a given convex polyhedron. The basic problem is solved by the modified simplex method [11], but at each iteration the column to enter the current basis emerges as the solution to an auxiliary linear program and is, in fact, an extreme point of the given convex polyhedron.

# A DUAL DECOMPOSITION PRINCIPLE

## I. Introduction and Notation

One of the main obstacles to the full application of linear programming is the inability of most current computational methods to cope with problems yielding program matrices of large order. However, many of these problems have program matrices which exhibit special structures, e.g. diagonal or block triangular and which are very sparse in non-zero elements. A large order transportation problem is an example. The size of the problems we have in mind make it infeasible to attempt to solve them by a general technique even on the most modern digital computers.

Much current research is being directed toward developing special techniques taking advantage of these matrix structures in order to facilitate solution of such problems as: (i) traffic flow problems and network problems [1, 4, 12], (ii) gasoline blending problems [3], (iii) communications problems [9], (iv) ware-housing problems [2], (v) production planning problems [5, 13], (vi) the caterer problem [8], and (vii) von Neumann's model of a constantly expanding ecomomy [14].

One recent technique for handling certain intrinsically large order problems is embodied in the Dantzig - Wolfe decomposition principle. We shall review the basic notions of their principle,

taking the liberty of using our own notation. We then develop our dual decomposition principle, so as to compare the two. Dantzig and Wolfe apply their technique to some problems of block triangular form.

Consider the general linear programming problem in the form:

(1) $\max X_0$

where $X_0 P_0 + \bar{A} X = \hat{b}$

$A X = b, X \geq 0 .$

We think of $A$ as being of large order, but fairly easy to handle. For example, we may have:

$$A = \begin{bmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & A_T \end{bmatrix}$$

where, without any coupling constraints, $A X = b$ would break down into:

$A_i X_i = b_i, \; i = 1,2, \ldots, T .$

In general, our aim is to treat efficiently the case where $A$ has blocks of zeros, variously placed.

We consider one "iteration" of the Dantzig - Wolfe algorithm. We assume, which Dantzig and Wolfe do not, that the convex set

$\{X \mid A X = b, X \geq 0\}$ is bounded.* So, in our discussion of their algorithm we will not consider the possibility of homogeneous solutions to (1) which may arise when the above set is not bounded.

We may pose the problem in the form:

$$(2) \qquad \max X_0$$

$$\text{where} \quad X_0 P_0 + P = \hat{b} ,$$

where $P$ belongs to $C = \{P \mid P = \overline{A} X;\ A X = b,\ X \geq 0\}$ .

Dantzig and Wolfe assume that one has a feasible solution $(\hat{X}_0, \hat{X})$, and in fact points $X_1, X_2, \ldots, X_m$ such that:

$$(3) \qquad X = \sum_{i=1}^{m} \hat{\lambda}_i X_i \ ; \ \sum_{i=1}^{m} \hat{\lambda}_i = 1, \quad \hat{\lambda}_i \geq 0 .$$

where the $X_i$ are extreme points of $\{X \mid A X = b, X \geq 0\}$ .

Thus, if $P_i = \overline{A} X_i$, then:

$$(4) \qquad \hat{P} = \sum_{i=1}^{m} \hat{\lambda}_i P_i .$$

One may then pose the problem in the form:

$$(5) \qquad \max X_0$$

$$\text{where} \quad X_0 P_0 + \lambda P + \sum_{i=1}^{m} \lambda_i P_i = \hat{b},$$

$$\lambda + \sum_{i=1}^{m} \lambda_i = 1 ; \qquad \lambda, \lambda_i \geq 0,$$

---

* This is readily realized by the technical device of adding the constraint $X_1 + \ldots + X_n \leq K$ where $K$ is sufficiently large.

where $P$ is any point in $C$. We thus think of (5) as the "basic problem."

One further assumes that the matrix:

$$(6) \qquad B = \begin{bmatrix} P_0, & P_1, & \ldots & , & P_m \\ 0, & 1, & \ldots & , & 1 \end{bmatrix}$$

is non-singular.

We shall now review one "iteration" of the Dantzig - Wolfe technique. If we write:

$$(7) \qquad (B^{-1})^T = \begin{bmatrix} \pi_1, & P^1, & \ldots & , & P^m \\ \beta, & \beta^1, & \ldots & , & \beta^m \end{bmatrix}$$

then $\pi = \begin{pmatrix} \pi_1 \\ \beta \end{pmatrix}$ is the set of "simplex multipliers" associated with the "basic feasible solution" obtained by setting $X_0 = \hat{X}_0$, $\lambda_i = \hat{\lambda}_i$ and $\lambda = 0$ in (5). Hence $\pi$ is the unique point satisfying the conditions:

$$P_0^T \pi_1 = 1$$

$$P_i^T \pi_1 + \beta = 0 \quad , \quad i = 1, 2, \ldots, m .$$

We assume that the inverse (7) has been computed.

The solution $(\hat{X}_0, \hat{X})$ is optimal if:

$$(8) \qquad \pi_1^T P + \beta \geq 0$$

for all $P \in C$.

This suggests that we consider the auxiliary problems:

$$(9) \qquad \min \gamma^T X$$

$$\text{where} \quad A X = b,\ X \geq 0,$$

where $\gamma = \bar{A}^T \pi_1$ .

If $\bar{X}$ is an (extreme point) solution to this auxiliary problem, then $(\hat{X}_0, \hat{X})$ is the optimal solution if $\gamma^T \bar{X} + \beta \geq 0$ . If not, then

$$(10) \qquad \bar{P} = \bar{A} \bar{X}$$

is introduced into the basis $B$ in the usual simplex method manner and the iteration is then completed.

In summary, with reference to the "basic problem" (5), one examines, in the usual simplex method way, the "relative costs" of each "column" $P$, using the multiplier $(\pi_1, \beta)$ to ascertain whether or not optimality is reached, as indicated by $\pi_1^T P + \beta \geq 0$ for all such columns. However, here this means examining all the points $P$ of a convex polyhedron $C$, suggesting the auxiliary problem.

Now with reference to (9), notice that at each iteration only the functional in the auxiliary problem changes. When one has a new functional, and wishes to solve (9) one starts from the optimal solution to the previous auxiliary problem and continues on (possibly in a few iterations) to the minimum of the new functional.

The Dual Algorithm

Both our algorithm and the Dantzig - Wolfe algorithm agree in (i) having the "basic problem" in the form (5), (ii) having the

non-singular, current basis matrix in the form (6); (iii) assuming that the current inverse, given by (7), has been computed; and (iv) having the same convex set C.

Our algorithm, which is in one sense, a dual of the Dantzig-Wolfe technique, is applied to the dual of the "basic problem" defined by (5), namely:

$$
\begin{aligned}
(11) \qquad & \min \ (\hat{b}^T \pi_1 \quad + \beta) \\
\text{where} \qquad & P_0^T \pi_1 \quad = 1 \\
& P_j^T \pi_1 \quad + \beta \geq 0 \ , \ j = 1, 2, \ldots, m \\
& P^T \pi_1 \quad + \beta \geq 0, \text{ for all } P \in C \ .
\end{aligned}
$$

We of course assume that both (11) and (1) have solutions.

At the start of one iteration we also assume, as so Dantzig and Wolfe, that we have a starting feasible solution $(\pi_1, \beta)$ to (11) and a $\pi_2$ such that $b^T \pi_2 = \beta$.

As was mentioned previously, we assume that the convex set $\{X \mid A X = b, X \geq 0\}$ is bounded. Dantzig and Wolfe do not make this assumption which is an essential part of the derivation of our auxiliary problem.

Our algorithm first of all differs from the Dantzig - Wolfe technique in assuming that their optimality criterion, given by (8), is satisfied. The optimality criterion for our dual technique is given by the following:

Theorem: Let $r$ be such that $\hat{\lambda}_r = \min_i \{\hat{\lambda}_i\}$

If $\hat{\lambda}_r \geq 0$, then $(\hat{X}_0, \hat{X})$ is optimal.

So at the start of each iteration, we compute $\hat{\lambda}_r$. If $\hat{\lambda}_r \geq 0$, then we are finished. If $\hat{\lambda}_r < 0$, then the column $\begin{pmatrix} P_r \\ 1 \end{pmatrix}$ is the VGO (vector going out of the current basis).

In order to determine a vector $\begin{pmatrix} \bar{P} \\ 1 \end{pmatrix}$ coming into the basis, VCI, we solve the auxiliary problem given by:

$$(12) \qquad \min \{ y^T \gamma + \eta \beta \}$$

$$\text{where } A y - \eta \beta = 0$$

$$\text{and } y^T (\bar{A} P^r) + \eta \beta' = -1 \quad ; \quad y \geq 0, \eta \geq 0 .$$

The solution of (12) yields $\bar{P} = \bar{A} \bar{y}$ which is the VCI.

Expressing $\bar{P}$ in terms of the basis, we get:

$$(13) \qquad \begin{pmatrix} \bar{P} \\ 1 \end{pmatrix} = \bar{p}_0 \begin{pmatrix} P_0 \\ 0 \end{pmatrix} + \sum_{i=1}^{m} \bar{p}_i \begin{pmatrix} P_i \\ 1 \end{pmatrix} .$$

The new $\lambda_i$'s, i.e. the $\bar{\lambda}_i$'s are computed from the formulas:

$$\bar{\lambda}_i = \hat{\lambda}_i - \frac{\hat{\lambda}_r}{\bar{p}_r} \bar{p}_i \quad , \quad i \neq r$$

and $$\bar{\lambda}_r = \hat{\lambda}_r / \bar{p}_r$$

where $P_i = \begin{pmatrix} \bar{P} \\ 1 \end{pmatrix}^T \begin{pmatrix} P^i \\ \beta^i \end{pmatrix}$ .

$(\bar{B}^{-1})^T$ is computed by algorithm from $B^{-T}$ by the usual simplex formula:

$$\bar{B}^{-T} = B^{-T} - \frac{1}{\begin{pmatrix} P^r \\ \beta^r \end{pmatrix}^T \begin{pmatrix} \bar{P} \\ 1 \end{pmatrix}} \begin{pmatrix} P^r \\ \beta^r \end{pmatrix} \left\{ \begin{pmatrix} \bar{P} \\ 1 \end{pmatrix} - \begin{pmatrix} P_r \\ 1 \end{pmatrix} \right\} B^{-T} .$$

The new functional value $\bar{X}_0$ is computed by means of:

$$\bar{X}_0 = \begin{pmatrix} \hat{b} \\ 1 \end{pmatrix}^T \begin{pmatrix} \bar{\pi}_1 \\ \bar{\beta} \end{pmatrix} = X_0 + \theta \hat{\lambda}_r \geq X_0$$

where

$$\begin{pmatrix} \bar{\pi}_1 \\ \bar{\beta} \end{pmatrix} = \begin{pmatrix} \pi_1 \\ \beta \end{pmatrix} + \theta \begin{pmatrix} P^r \\ \beta^r \end{pmatrix}.$$

This completes one iteration.

Now, unlike the Dantzig - Wolfe Technique our auxiliary problems differ not only in the functional but also in the additional constraint. So, if we solve (12), we will not be able to take maximum advantage of the previous computations as do Dantzig and Wolfe when they solve their auxiliary problems. This suggests that we solve the dual of (12) as a primal problem. The dual of (12) is given by

$$\max \theta$$

$$(14) \quad \text{where} \quad \begin{aligned} -A^T \pi_2 \quad - \theta(\bar{A}^T P^r) &\leq A^T \pi_1 \\ b^T \pi_2 \quad - \theta \, \beta^r \quad &\leq \beta \end{aligned}$$

We solve (14) by the modified simplex technique [11]. We will now show that by means of one artificial iteration we can avoid

the problem caused by the additional constraint and hence start from the optimum solution to the previous problem and continue on to the maximum of the new functional thereby taking maximum advantage of the previous iteration.

At the start of one iteration in the solution of (14) we assume that we have a basis $B_1$. The columns of $B_1$ are selected from the matrix:

$$\begin{bmatrix} -A^T & I & 0 & -A^{-T} P^r \\ b^T & 0 & 1 & -\beta^r \end{bmatrix},$$

where $\begin{bmatrix} I & 0 \\ 0 & 1 \end{bmatrix}$ is associated with slack variables introduced into (14). Let $B_1$ be given by:

$$(15) \qquad B_1 = \begin{bmatrix} -B^* & -\bar{A} P^r \\ b^* & -\beta^r \end{bmatrix}$$

where the columns of $\begin{bmatrix} -B^* \\ b^* \end{bmatrix}$ are selected from the columns of $\begin{bmatrix} -A^T & I \\ b^T & 0 \end{bmatrix}$. We also assume that we have a $\begin{pmatrix} \pi_1 \\ \beta \end{pmatrix}$ and $\theta$ determined from the previous solution of (14).

Now, at the start of the modified simplex technique, we replace, in $B_1$, the column $\begin{pmatrix} -\bar{A} P^r \\ -\beta^r \end{pmatrix}$ by the column $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$. This

initial replacement eliminates the problem of the additional constraint in (12) and yields a new basis for (14) as we will now show.

(14) may be written in the form:

(16) $\max \theta$

where $$\begin{bmatrix} -B^* & -\bar{A}^T P^r \\ b^* & -\beta^r \end{bmatrix} \begin{pmatrix} \gamma \\ \theta \end{pmatrix} = \begin{pmatrix} \bar{A}^T \pi_1 \\ \beta \end{pmatrix} .$$

$\gamma$ is composed of elements of $\pi_2$ and the slack variables corresponding to the basis columns. But we may write:

$$\begin{bmatrix} -B^* \\ b^* \end{bmatrix} \gamma = \begin{pmatrix} \bar{A}^T \pi_1 \\ \beta \end{pmatrix} + \theta \begin{pmatrix} \bar{A}^T P^r \\ \beta^r \end{pmatrix}$$

$$= \begin{pmatrix} \bar{A}^T \bar{\pi}_1 \\ \bar{\beta} \end{pmatrix}$$

or finally:

(17) $$\begin{bmatrix} -B^* & 0 \\ b^* & 1 \end{bmatrix} \begin{pmatrix} \gamma \\ 0 \end{pmatrix} = \begin{pmatrix} \bar{A}^T \bar{\pi}_1 \\ \bar{\beta} \end{pmatrix}$$

and the replacement is complete.

Now we let

(18) $$\bar{B}_1 = \begin{bmatrix} -B^* & 0 \\ b^* & 1 \end{bmatrix} .$$

In order to show that $\bar{B}_1$ is a basis we express $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ in terms of the previous basis $B_1$ to get:

$$(19)\qquad \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{bmatrix} -B^* \\ b^* \end{bmatrix} \epsilon + \begin{pmatrix} A^T P^r \\ \beta^r \end{pmatrix} \epsilon_0 = B_1 \begin{pmatrix} \epsilon \\ \epsilon_0 \end{pmatrix}.$$

$\overline{B}_1$ will be a basis if and only if $\epsilon_0 \neq 0$. Now, utilizing the fact that the constraints in (12) may be written in the form:

$$(y^T, \eta) \begin{bmatrix} -A^T & -A^T & P^r \\ b & & -\beta^r \end{bmatrix} = (0^T, 1),$$

we may write, with (19):

$$(20)\qquad (y^T, \eta) \begin{pmatrix} 0 \\ 1 \end{pmatrix} = (y^T, \eta) \; B_1 \begin{pmatrix} \epsilon \\ \epsilon_c \end{pmatrix}$$

or

$$\eta = (0^T, 1) \begin{pmatrix} \epsilon \\ \epsilon_0 \end{pmatrix} = \epsilon_0$$

If we assume that $A X = b$, $X \geq 0$ has a solution $\hat{X}$ and $Ay = 0$, $y \geq 0$ has a solution $\hat{y} \neq 0$, then, for all $\alpha \geq 0$, $A(\hat{X} + \alpha \hat{y}) = b$, $\hat{X} + \alpha \hat{y} \geq 0$. But this contradicts our assumption that the convex set $\left\{ X \middle| A X = b, X \geq 0 \right\}$ is bounded. Hence $Ay = 0$, $y \geq 0$ if and only if $y = 0$. So we will not get $\eta = 0$ in (12). Therefore $\epsilon_0 > 0$ and $B_1$ is a basis for (14). We get (automatically) a starting feasible solution $\begin{pmatrix} V \\ 0 \end{pmatrix}$ for (14) from which we proceed with the modified simplex technique.

The solution of the Dantzig-Wolfe auxiliary problem yields an X, say X*, which in turn yields a P* = A X* as the VCI. The VGO in the Dantzig - Wolfe technique is determined after the auxiliary problem has been solved. In our dual technique the VGO is determined at the beginning of the iteration. The solution of our

auxiliary problem, given by (15), yields a solution $(\bar{y}, \bar{\eta})$ to (12) and, hence, a $\bar{P} = \bar{A}\,\bar{y}/\bar{\eta}$ as the VCI. As we noted before, $\bar{\eta} > 0$.

We make no direct use of the X's in our auxiliary problem as do Dantzig and Wolfe in the direct technique. However $\bar{y} = \bar{\eta}\,\bar{X}$. $\bar{X}$ is an extreme point of $\{X \mid A\,X = b,\ X \geq 0\}$. For, if it were not, then there would be no lower bound for $y^T\gamma + \eta\beta$ and the above convex set would not be bounded, contrary to our assumption.

In the development of the dual decomposition technique no mention was made of the structure of the matrix A. However, in the direct decomposition principle the composition of the current basis matrix B changes as the structure of A changes. For example, for a matrix A of the form:

$$A = \begin{bmatrix} A_1 & 0 & 0 \\ 0 & A_2 & 0 \\ 0 & 0 & A_3 \end{bmatrix},$$

the current basis matrix is no longer given by an expression comparable to (6) but assumes the form:

$$B = \begin{bmatrix} P_{11} & P_{12} \cdots P_{1k_1} & P_{21} & P_{22} \cdots P_{2k_2} & P_{31} & P_{32} \cdots P_{3k_3} \\ 1 & 1 \cdots 1 & 0 & 0 \cdots 0 & 0 & 0 \cdots 0 \\ 0 & 0 \cdots 0 & 1 & 1 \cdots 1 & 0 & 0 \cdots 0 \\ 0 & 0 \cdots 0 & 0 & 0 \cdots 0 & 1 & 1 \cdots 1 \end{bmatrix}$$

In the dual technique the structure of B will essentially remain unchanged as the structure of A changes.

In summary, our dual technique is essentially the modified simplex technique, as was stated before, with the exception that the column to enter the current basis is obtained by solving an auxiliary problem given by (14). By means of one artificial iteration we are able to eliminate the presence of the additional constraint in the dual to (14) and obtain an initial feasible solution for the auxiliary problem while, at the same time, taking maximum benefit of the previous computations.

Assuming non-degeneracy the iterative process is finite since the number of extreme points of the convex set $\{ X \mid AX = b,\ X \geq 0 \}$ is finite.

## Reference


1. Charnes, A. and Cooper, W. W., "Extremal Principles for Simaulating Traffic Flows in a Network", Proc. Nat'l. Acad. Sci., Vol.44, No. 2, February, 1958.

2. Charnes, A. and Cooper, W. W., "Generalizations of the Warehousing Model," Operational Research Quarterly, Vol.6, No. 4, December, 1955.

3. Charnes, A. and Cooper, W. W., "Management Models and Industrial Applications of Linear Programming," Management Science, Vol.14, No. 1, October, 1957.

4. Charnes, A. and Lemke, C. E., "Multi-Copy Generalized Networks and Multi-Page Programs," R.P.I. Math Rep. No. 41, December 23, 1960.

5. Dantzig, G. B., "Recent Advances in Linear Programming", Management Science, Vol.2, No. 2, January, 1956.

6. Dantzig, G. B., "Upper Bounds, Secondary Constraints, and Block Triangularity in Linear Programming," The Rand Corporation, Research Memorandum, R.M. -1367, October, 1954.

7. Dantzig, G. B. and Wolfe, P., "Decomposition Principle for Linear Programs," Operations Research, Vol. 8, No. 1 January, February, 1960.

8. Jacobs, W., "The Caterer Problem," Naval Research Logistics Quarterly, Vol. 1, No. 2, June, 1954.

9. Juncosa, M. and Kalaba, R. E., "Optimal Design and Utilization of Communications Networks," Management Science, Vol. 3, No. 1, October, 1956.

10. Lemke, C. E., "The Dual Method of Solving the Linear Programming Problem", Naval Research Logistics Quarterly, Vol. 1, No. 1, March, 1954.

11. Lemke, C. E. and Charnes, A., "Extremal Problems In Linear Inequalities," Carnegie Institute of Technology, Technical Report No. 36, May, 1953.

12. Lemke, C. E., and Powers, T. J., "A Computer Program for a Class of Network-type Optimization Problems," R.P.I. Math Report No. 36, September, 1960.

13. Manne, Alan s., "A Note of the Modigliani - Hohn Production Smoothing Model," Management Science, Vol. 3, No. 4, July, 1957.

14. von Newmann, John, "A Model of General Economic Equilibrium," The Review of Exonomic Studies, Vol. XIII, No. 1, 1945-1946.