J. Austral. Math. Soc. 19 (Series B), (1975), 242–248.

ON THE SOLUTION OF A STRUCTURED NONLINEAR PROGRAMME

L. T. M. BERRY

(Received 14 July 1975)

(Revised 29 October 1975)

Abstract

Explicit formulae are derived for the projected gradient vector and trial dual variables required in the application of Rosen's method [4] to the solution of a Minimum Cost Network problem.

1. Introduction

This paper considers mathematical programmes of the form:

$$\text{Minimize} \quad f(\boldsymbol{x})$$

subject to:

$$\sum_{j=1}^{\phi(k)} x_j^k = b^k ; \qquad k = 1, \cdots, K.$$

$$x_j^k \geqq 0 ; \qquad j = 1, \cdots, \phi(k); \qquad (1)$$
$$k = 1, \cdots, K.$$

The components x_j^k of the point $\boldsymbol{x}$ may represent chain flows on a multiple origin-destination (OD) network. For example, the programme can be interpreted as the mathematical formulation of a problem of practical importance in the economic design of alternative routing telephone networks. For each distinct OD pair k there exist $\phi(k)$ routes (chains) available to carry traffic. The equality constraints ensure that the total traffic carried on these routes is a specified quantity b^k. A feasible set of chain flows on the network defines a chain flow pattern $\boldsymbol{x}$, which incurs a corresponding cost $f(\boldsymbol{x})$. The problem is to determine a chain flow pattern $\boldsymbol{x}^*$ with minimum cost $f(\boldsymbol{x}^*)$.

https://doi.org/10.1017/S0334270000001119 Published online by Cambridge University Press

Applications of the above type [1] give rise to large programmes, typically 5,000–10,000 variables. The non-linear function f, which represents the cost of circuits required to give specified OD performances, is a differentiable function [2] permitting the use of gradient solution methods.

When formulae developed in this paper are incorporated, the Gradient Projection algorithm becomes a very efficient method for solving the minimum cost circuit allocation problem for Metropolitan Telephone Networks ([1], [3]).

Rosen noted that to make the Gradient Projection method computationally practical for a *general* problem an efficient computational procedure was required to find the inverse matrix $(NN')^{-1}$, where the rows of N are the gradient vectors of the current active constraints. A new inverse obtained by adding or deleting rows of N is usually required at each step. Rosen suggested recursion relations to obtain the new inverse from the old inverse.

For the structured constraints given in (1) it is not necessary to perform either matrix multiplication or inversion, as both the projected gradient vector and trial dual variables (see (4)) can be expressed explicitly in terms of the components of $\nabla f(x)$.

2. Some properties of projection matrices

The projection matrix

$$P = I - N'[NN']^{-1}N \tag{2}$$

projects the vector $-\nabla f(\underset{\sim}{x})$ into the subspace satisfying the active constraints. The directions of search in Rosen's method are provided by the projected gradient vector

$$\boldsymbol{v} = -P\nabla f(\boldsymbol{x}). \tag{3}$$

Direct application of (2) and (3) in the projection algorithm is not practical. Instead a suitable transformation is chosen which gives an equivalent programme from which $\boldsymbol{v}$ can be deduced in explicit form. Suppose N is an $r \times n$ matrix. It is assumed that N has full row rank (i.e. $r < n$).

Lemma 1. *If $T(r \times r)$ is non-singular then TN determines the same projection matrix as N.*

Proof:

$$\begin{aligned} & I - (TN)'[TN(TN)']^{-1}TN \\ = \; & I - N'T'(T')^{-1}[NN']^{-1}T^{-1}TN \\ = \; & I - N'[NN']^{-1}N. \end{aligned}$$

https://doi.org/10.1017/S0334270000001119 Published online by Cambridge University Press

If $P_1, \cdots, P_r$ are projection matrices the transformation $P_1 \cdots P_r \nabla f(\boldsymbol{x})$ is called a "successive projection" of the gradient vector.

LEMMA 2. *If the row vectors of N are mutually orthogonal then successive projection of a vector onto each individual active constraint is equivalent to a single projection onto the intersection of the active constraints.*

PROOF: We show that the two matrices transforming the vector are identical. First consider projection onto the intersection. Let

$$N = \begin{bmatrix} \boldsymbol{n}'_1 \\ \vdots \\ \boldsymbol{n}'_r \end{bmatrix}.$$

As

$$\boldsymbol{n}'_i \boldsymbol{n}_j = 0 \qquad (i \neq j),$$

$$NN' = \text{diag.}\,(\boldsymbol{n}'_i \boldsymbol{n}_i).$$

Thus

$$I - N'[NN']^{-1} N = I - \sum_{j=1}^{r} \boldsymbol{n}_j \, (\boldsymbol{n}'_j \boldsymbol{n}_j)^{-1} \, \boldsymbol{n}'_j.$$

The transformation matrix which projects a vector successively onto the r active constraints is given by

$$\prod_{j=1}^{r} [I - \boldsymbol{n}_j \, (\boldsymbol{n}'_j \boldsymbol{n}_j)^{-1} \, \boldsymbol{n}'_j] = I - \sum_{j=1}^{r} \boldsymbol{n}_j \, (\boldsymbol{n}'_j \boldsymbol{n}_j)^{-1} \, \boldsymbol{n}'_j,$$

the terms $\boldsymbol{n}_{j1} (\boldsymbol{n}'_{j1} \boldsymbol{n}_j)^{-1} \boldsymbol{n}'_{j1} \boldsymbol{n}_{j2} (\boldsymbol{n}'_{j2} \boldsymbol{n}_{j2})^{-1} \boldsymbol{n}'_{j2}$ vanishing due to the orthogonality of distinct vectors $\boldsymbol{n}_{j1}$, $\boldsymbol{n}_{j2}$.

During the application of the projection algorithm whenever $\boldsymbol{v}$ becomes a null vector it is necessary to examine the trial dual variables, that is the components of

$$[NN']^{-1} N \nabla f(\boldsymbol{x}). \tag{4}$$

If each component is non-negative the Kuhn–Tucker optimality conditions are satisfied and the algorithm terminates. Otherwise, a row of N corresponding to a negative component of (4) is deleted and a new projection matrix determined. The following lemma is used to obtain explicit expressions for the trial dual variables in terms of the components of $\nabla f(\boldsymbol{x})$.

LEMMA 3. *If the matrix $T(r \times r)$ is non-singular,*

$$[NN']^{-1} N = T'[(TN)(TN)']^{-1} TN.$$

https://doi.org/10.1017/S0334270000001119 Published online by Cambridge University Press

PROOF:

$$T'[(TN)(TN)']^{-1}\, TN = T'(T')^{-1}\,[NN']^{-1}\,(T^{-1}\, T)N = [NN']^{-1}\, N.$$

3. The projection vector

Let the component of $\boldsymbol{v}$ corresponding to variable x_j^k be denoted by v_j^k.

THEOREM: *Let $p(k)$ be the number of routes with positive flow between* OD *pair k at the current point $\boldsymbol{x}$, that is*

$$p(k) = \sum_{j=1}^{\phi(k)} \delta_j^k$$

where

$$\delta_j^k = \begin{cases} 1 & \text{if} \quad x_j^k > 0 \\ 0 & \text{otherwise} \end{cases}$$

then

$$v_j^k = \begin{cases} 0 \quad \text{if} \quad x_j^k = 0, \\ \dfrac{1}{p(k)} \displaystyle\sum_{l=1}^{\phi(k)} \delta_l^k \frac{\partial f(\boldsymbol{x})}{\partial x_l^k} - \frac{\partial f(\boldsymbol{x})}{\partial x_j^k} & \text{otherwise.} \end{cases} \tag{5}$$

PROOF: Suppose the active constraints are

$$\begin{aligned} \sum_{j=1}^{\phi(k)} x_j^k = b^k; \qquad k &= 1, \cdots, K. \\ x_{j_1}^{k_1} &= 0 \\ &\vdots \\ x_{j_s}^{k_s} &= 0. \end{aligned} \tag{6}$$

Clearly an equivalent system of active constraints having the property that the gradient vectors are mutually orthogonal is given by

$$\begin{aligned} \sum_{j=1}^{\phi(k)} \delta_j^k x_j^k = b^k; \qquad k &= 1, \cdots, K. \\ x_{j_1}^{k_1} &= 0 \\ &\vdots \\ x_{j_s}^{k_s} &= 0. \end{aligned} \tag{7}$$

The result (5) was obtained by projecting $-\nabla f(\boldsymbol{x})$ 'successively' onto first the

https://doi.org/10.1017/S0334270000001119 Published online by Cambridge University Press

active non-negativity constraints and then the new equality constraints in turn. To verify equation (5) we first show that the transformation of $-\nabla f(\boldsymbol{x})$ to $\boldsymbol{v}$ is a projection.

$$\begin{aligned}(P \cdot P\nabla f)_j^k &= (P \cdot (-\boldsymbol{v}))_j^k \\ &= \frac{\partial f(\boldsymbol{x})}{\partial x_j^k} - \frac{1}{p(k)} \sum_{s=1}^{\phi(k)} \delta_s^k \frac{\partial f(\boldsymbol{x})}{\partial x_s^k} \\ &\quad - \frac{1}{p(k)} \sum_{l=1}^{\phi(k)} \delta_l^k \left\{ \frac{\partial f(\boldsymbol{x})}{\partial x_l^k} - \frac{1}{p(k)} \sum_{s=1}^{\phi(k)} \delta_s^k \frac{\partial f(\boldsymbol{x})}{\partial x_s^k} \right\} \\ &= \frac{\partial f(\boldsymbol{x})}{\partial x_j^k} - \frac{2}{p(k)} \sum_{s=1}^{\phi(k)} \delta_s^k \frac{\partial f(\boldsymbol{x})}{\partial x_s^k} \\ &\quad + \left\{ \frac{1}{p(k)} \sum_{l=1}^{\phi(k)} \delta_l^k \right\} \frac{1}{p(k)} \sum_{s=1}^{\phi(k)} \delta_s^k \frac{\partial f(\boldsymbol{x})}{\partial x_s^k} = -v_j^k.\end{aligned}$$

Next, we show that $\boldsymbol{v}$ is orthogonal to the gradient vectors for the new constraint set. The case $x_j^k = 0$ is immediate. If $x_j^k > 0$ we have

$$\begin{aligned}\sum_{j=1}^{\phi(k)} \delta_j^k v_j^k &= \sum_{j=1}^{\phi(k)} \delta_j^k \left\{ \frac{1}{p(k)} \sum_{l=1}^{\phi(k)} \delta_l^k \frac{\partial f(\boldsymbol{x})}{\partial x_l^k} - \frac{\partial f(\boldsymbol{x})}{\partial x_j^k} \right\} \\ &= \left\{ \sum_{j=1}^{\phi(k)} \delta_j^k \right\} \frac{1}{p(k)} \sum_{l=1}^{\phi(k)} \delta_l^k \frac{\partial f(\boldsymbol{x})}{\partial x_l^k} \\ &\quad - \sum_{j=1}^{\phi(k)} \delta_j^k \frac{\partial f(\boldsymbol{x})}{\partial x_j^k} = 0.\end{aligned}$$

Equations (5) show that the projection algorithm attempts to transfer flow onto the j'th route for OD pair k if and only if its marginal cost is less than the average marginal cost for routes with positive flow. On the other hand, if the marginal cost on some route with positive flow is greater than the average marginal cost for routes with positive flow then flow is removed.

4. The transformation matrix producing the orthogonalization

Let $q(k) = \phi(k) - p(k)$, i.e. the number of routes with zero flow for OD pair k. Suppose that there are $q = \sum_{k=1}^{K} q(k)$ active non-negativity constraints.

We define the matrix

$$Q(K \times q) = \begin{bmatrix} -1 & -1 \cdots -1 & 0 & \cdots & & & & \\ 0 & 0 \cdots & 0 & -1 & -1 \cdots -1 & 0 & \cdots & \\ & & & & & & & . \\ 0 & \cdots & & & & & 0 & -1 \cdots -1 \end{bmatrix}$$

with row i having $q(i)$ non-zero elements, each -1. Then the matrix which

https://doi.org/10.1017/S0334270000001119 Published online by Cambridge University Press

transforms the constraint set (6) into the orthogonal system (7) is given by

$$T = \left[\begin{array}{c|c} I(K \times K) & Q(K \times q) \\ \hline 0(q \times K) & I(q \times q) \end{array}\right]$$

5. Trial dual variables

Consider the active constraint set given by (7). Denote the i'th row of TN by $\hat{\boldsymbol{n}}'_i$. For the equality constraints it is clear that $(\hat{\boldsymbol{n}}'_i\hat{\boldsymbol{n}}_i)^{-1} = 1/p(i)$, whereas for the active non-negativity constraints $(\hat{\boldsymbol{n}}'_i\hat{\boldsymbol{n}}_i)^{-1} = 1$. From Lemma 2, the components of $[(TN)(TN)']^{-1}TN$ are of the form $(\hat{\boldsymbol{n}}_i\hat{\boldsymbol{n}}_i)^{-1}\hat{\boldsymbol{n}}'_i$, thus

$$[(TN)(TN)']^{-1}\,TN\nabla f(\boldsymbol{x}) = \begin{bmatrix} \dfrac{1}{p(1)}\displaystyle\sum_{l=1}^{\phi(1)} \delta_l^1\, \partial f(\boldsymbol{x})/\partial x_l^1 \\ \vdots \\ \dfrac{1}{p(K)}\displaystyle\sum_{l=1}^{\phi(K)} \delta_l^K\, \partial f(\boldsymbol{x})/\partial x_l^K \\ \partial f(\boldsymbol{x})/\partial x_{j_1}^{k_1} \\ \vdots \\ \partial f(\boldsymbol{x})/\partial x_{j_s}^{k_s} \end{bmatrix}$$

By Lemma 3 the trial dual variables for the original constraint set (6) are now given by

$$\begin{bmatrix} I(K \times K) & 0(K \times q) \\ Q'(q \times K) & I(q \times q) \end{bmatrix} \begin{bmatrix} \dfrac{1}{p(1)}\displaystyle\sum_{l=1}^{\phi(1)} \delta_l^1\, \partial f(\boldsymbol{x})/\partial x_l^1 \\ \vdots \\ \partial f(\boldsymbol{x})/\partial x_{j_s}^{k_s} \end{bmatrix}$$

$$= \begin{bmatrix} \dfrac{1}{p(1)}\displaystyle\sum_{l=1}^{\phi(1)} \delta_l^1\, \partial f(\boldsymbol{x})/\partial x_l^1 \\ \vdots \\ \dfrac{1}{p(K)}\displaystyle\sum_{l=1}^{\phi(K)} \delta_l^K\, \partial f(\boldsymbol{x})/\partial x_l^K \\ \partial f(\boldsymbol{x})/\partial x_{j_1}^{k_1} - \dfrac{1}{p(k_1)}\displaystyle\sum_{l=1}^{\phi(k_1)} \delta_l^{k_1}\, \partial f(\boldsymbol{x})/\partial x_l^{k_1} \\ \vdots \\ \partial f(\boldsymbol{x})/\partial x_{j_s}^{k_s} - \dfrac{1}{p(k_s)}\displaystyle\sum_{l=1}^{\phi(k_s)} \delta_l^{k_s}\, \partial f(\boldsymbol{x})/x_l^{k_s} \end{bmatrix}$$

6. Equilibrium conditions

The necessary optimality conditions follow directly from the above results. Firstly, the vector $\boldsymbol{v}$ is a null vector, hence if x_j^k is positive,

$$\frac{\partial f(\boldsymbol{x})}{\partial x_j^k} = \frac{1}{p(k)} \sum_{l=1}^{\phi(k)} \delta_l^k \frac{\partial f(\boldsymbol{x})}{\partial x_l^k}.$$

From the non-negativity of the dual variables, if $x_j^k = 0$,

$$\frac{\partial f(\boldsymbol{x})}{\partial x_j^k} \geqq \frac{1}{p(k)} \sum_{l=1}^{\phi(k)} \delta_l^k \frac{\partial f(\boldsymbol{x})}{\partial x_l^k}.$$

It follows that at the optimal point for each OD pair k, there is an ordering of the component of the gradient vector

$$\frac{\partial f(\boldsymbol{x})}{\partial x_{j_1}^k} = \frac{\partial f(\boldsymbol{x})}{\partial x_{j_2}^k} = \cdots = \frac{\partial f(\boldsymbol{x})}{\partial x_{j_\nu}^k} \leqq \frac{\partial f(\boldsymbol{x})}{\partial x_{j_{\nu+1}}^k} \leqq \cdots \leqq \frac{\partial f(\boldsymbol{x})}{\partial x_{j_{\phi(k)}}^k},$$

where the routes $j_1, \cdots, j_\nu$ are those with positive flow.

7. Remark

The above results generalize in a straightforward way to include upper bounds on the variables and arbitrary coefficients of the variables in the equality constraints. The inclusion of link capacity constraints introduces some interesting problems which are currently under consideration.

References

[1] L. T. M. Berry, 'A mathematical model for optimizing telephone networks', (Ph.D. Thesis, University of Adelaide, December 1971).

[2] L. T. M. Berry, 'An explicit formula for dimensioning links offered overflow traffic', *Australian Telecommunication Research* 8(1) (1974), 13–17.

[3] R. B. Potts and L. T. M. Berry, *Adelaide Telephone Network Study*, (University of Adelaide — Australian Post Office Reports, Sept. 1970–Dec. 1971).

[4] J. B. Rosen, 'The gradient projection method for nonlinear programming, Part 1. Linear Constraints', *J. Soc. Indust. Appl. Math.* 8 (1960), 181–217.

Department of Applied Mathematics,
University of Adelaide,
Adelaide, S.A., 5000,
Australia.