

# INVESTIGAÇÃO OPERACIONAL

Dezembro 1988

Número 2

Volume 8

Publicação Científica da

**apolo**

Associação Portuguesa para o Desenvolvimento  
da Investigação Operacional.

# INVESTIGAÇÃO OPERACIONAL

Propriedade:

APDIO — Associação Portuguesa para o Desenvolvimento  
da Investigação Operacional

## ESTATUTO EDITORIAL

*«Investigação Operacional», órgão oficial da APDIO cobre uma larga gama de assuntos reflectindo assim a grande diversidade de profissões e interesses dos sócios da Associação, bem como as muitas áreas de aplicação da I. O. O seu objectivo primordial é promover a aplicação do método e técnicas da I.O. aos problemas da Sociedade Portuguesa.*

*A publicação acolhe contribuições nos campos da metodologia, técnicas, e áreas de aplicação e software de I. O. sendo no entanto dada prioridade a bons casos de estudo de carácter eminentemente prático.*

---

**Distribuição gratuita aos sócios da APDIO**

# INVESTIGAÇÃO OPERACIONAL

volume 8 - nº 2 - Dezembro 1988

Publicação semestral

Direcção : J.M. Pinto Paixão  
(Fac. Ciências - Universidade de Lisboa)  
Joaquim J. Júdice  
(Fac. Ciências e Tecnologia - Univ. de Coimbra)

## Comissão Editorial

Mordecai Avriel	(Israel)	A. Simões Monteiro	(NORMA)
João A. Branco	(IST - Univ. Técn. Lisboa)	Mohamed Najim	(ENSIAS - Argélia)
Josep Casanovas	(UPC - Espanha)	J. Manuel Oliveira	(EFACEC)
J. Dias Coelho	(FE - Univ. Nova Lisboa)	Fernando Pacheco	(Univ. Católica)
Nuno Crato	(NORMA - Açores)	A. Gouvêa Portela	(IST - Univ. Técn. Lisboa)
J.A. Romão Eusébio	(CIMPOR)	M. Baptista Rodrigues	(Partex - CPS)
A. Sousa Ferraria	(Petrogal)	A. Guimarães Rodrigues	(Univ. Minho)
D. V. Gokhale	(Estados Unidos)	Bernard Roy	(LAMSADE - França)
J. Borges Gouveia	(FE - Univ. Porto)	C. Moreira da Silva	(FE - Univ. Porto)
R. Campos Guimarães	(FE - Univ. Porto)	L. Valadares Tavares	(IST - Univ. Técn. Lisboa)
Masao Iri	(TU - Japão)	Isabel H. Themido	(IST - Univ. Técn. Lisboa)
Joaquim J. Júdice	(FC - Univ. Coimbra)	B. Calafate Vasconcelos	(FE - Univ. Porto)
A. Rinnoy Kan	(EU - Holanda)	José M. Viegas	(IST - Univ. Técn. Lisboa)
Nelson Maculan	(UFRJ - Brasil)	Andres Weintraub	(UC - Chile)

A Revista "INVESTIGAÇÃO OPERACIONAL" está registada na Secretaria de Estado da Comunicação Social sob o nº108335.

Esta Revista é distribuída gratuitamente aos sócios da APDIO. As informações sobre inscrições na Associação, assim como a correspondência para a Revista devem ser enviadas para a sede da APDIO - Associação Portuguesa para o Desenvolvimento da Investigação Operacional - CESUR, Instituto Superior Técnico, Av. Rovisco Pais, 1000 Lisboa.

Este Volume foi subsidiado por :

Instituto Nacional de Investigação Científica (INIC)  
Junta Nacional de Investigação Científica e Tecnológica (JNICT)  
Fundação Calouste Gulbenkian.

Para efeitos de dactilografia e composição, foram utilizados equipamentos gentilmente postos à disposição pelo CEAUL (DEIOC-Faculdade de Ciências de Lisboa).

Assinatura : 3000\$00



## PRIMAL AND DUAL GREEDY HEURISTICS FOR THE GENERALIZED SET COVERING PROBLEM

**José Paixão**

Faculdade de Ciências, Universidade de Lisboa  
Av. 24 de Julho, 134-5º, 1300 Lisboa

**Margarida Pato**

Instituto Superior de Economia, Universidade Técnica de Lisboa  
R. Miguel Lupi, 20, 1200 Lisboa

### Abstract

This paper reports on the development of primal and dual greedy heuristics for the generalized set covering problem (GSCP).

The primal heuristics provide a feasible solution and, consequently, an upper bound on the optimum for the GSCP. Dual based heuristics are used for obtaining and improving lower bounds at optimal value. Both, the primal and dual procedures are described in this paper and the corresponding computational complexity is studied.

Moreover, we present empirical results, obtained from computational experience with 34 instances of the GSCP. The test problems are related to the scheduling of bus drivers at Rodoviária Nacional a large transport operator in Portugal. In fact, this specific integer program, GSCP, has been widely used in crew scheduling applications.

These computational tests show that a combined primal-dual greedy heuristic procedure is a reasonably accurate and fast tool at least to tackle with bus driver scheduling GSCP instances. Taking this into account, another procedure, embedding the primal-dual greedy heuristics in a lagrangean relaxation based method, has been devised for the GSCP and is presented in a different paper.

**Keywords:** Heuristics, generalized set covering.

---

Research for this paper was supported by Instituto Nacional de Investigação Científica and Junta Nacional de Investigação Científica e Tecnológica.

## 1. INTRODUCTION

The generalized set covering problem can be stated as the following integer program:

$$\begin{aligned}
 \text{(GSCP)} \quad & \min \sum_{j \in N} c_j x_j \\
 \text{s.t.} \quad & \sum_{j \in N} a_{ij} x_j \geq b_i \quad (i \in M) \quad (1.1) \\
 & 0 \leq x_j \leq h_j \quad (j \in N) \quad (1.2) \\
 & x_j \text{ integer} \quad (j \in N), \quad (1.3)
 \end{aligned}$$

where  $a_{ij}$ , for  $i \in M$  and  $j \in N$ , are equal to 0 or 1. When the integer variable-bounds  $h_j$  are all equal to 1 the GSCP has been referred in the literature as the multiple set covering problem or the multicovering problem ([3]). When, on top of this, the integer values  $b_i$  are all equal to 1, the GSCP becomes the well known set covering model ([1], [2], [4]) where constraints (1.2) and (1.3) are replaced by  $x_j \in \{0, 1\}$  ( $j \in N$ ).

Like the classical set covering, the generalized and the multiple covering models are closely related to some real life situations. Multicovering problems arise in communication or distribution problems where reliability is important. The generalized set covering problem is mainly related to personnel scheduling ([5] and [9]).

According to the usual,  $\mathcal{V}(\text{GSCP})$  denotes the optimal value for GSCP and  $\mathcal{F}(\text{GSCP})$  stands for the set of feasible solutions for GSCP.

The GSCP is well known as a NP-complete problem ([10]) and so, heuristic methods come as a reasonable way of obtaining an approximation interval for  $\mathcal{V}(\text{GSCP})$ .

In this paper, we present several heuristic techniques developed in order to produce bounds, both from above and from below, on the optimal value for the GSCP. Hence, the next section is devoted to describing primal greedy heuristics followed by local search for the GSCP which provide feasible solutions and an upper bound on  $\mathcal{V}(\text{GSCP})$ . In section 3, dual greedy heuristics are studied for the purpose of obtaining lower bounds on  $\mathcal{V}(\text{GSCP})$  and, possibly, improving the upper bounds.

Results on the computational complexity for each one of the algorithms are also presented.

At last, in section 4, an empirical analysis obtained with a composed procedure embedding all the heuristics is reported.

The GSCP instances considered for the experience belong to a particular type arising in bus driver scheduling applications. For these problems,  $N$  is the index set for the feasible driver shifts or workdays, each one of them being assigned a cost,  $c_j$ . The set  $M$  corresponds to time periods requiring a minimal number of drivers defined through vector  $\mathbf{b} = (b_i)_{i \in M}$ . An element  $a_{ij}$  is equal to 1 if period  $i$  is a working period of shift  $j$  and 0 otherwise. The particularity of the bus driver scheduling problem comes from the fact that each column  $(a_{ij})_{i \in M}$  consists of one or two strings of consecutive ones.

The composed procedure proves reasonable efficiency for those real life applications. In fact it produced good upper bounds for the most of the test problems and even the optimum for some of them. However, the lower bounds were not so accurate, requiring further attempts to obtain improvements namely using lagrangean relaxation techniques as described in [8]. Besides the procedure is relatively fast when the algorithms, implemented in FORTRAN, run on a VAX 11/750 VMS (with FPA). Moreover, the primal-dual greedy procedure can easily be implemented on a PC as reported in [5].

## 2. PRIMAL GREEDY HEURISTICS WITH LOCAL SEARCH

Greedy heuristics have been studied in deep detail for the classic set covering model ([1]) and the multicovering version ([3]). An extension of those techniques for a generalized covering model related to scheduling problems has been presented in a previous paper ([9]).

As described in there, and following the general pattern, the greedy heuristics construct feasible solutions, step by step, selecting a row to be covered and, then, a variable to cover the unsatisfied demand for that row. Several row and column selection criteria can

be combined for building up feasible solutions ([7], [8] and [9]) and consequently, producing upper bounds on the optimal value  $v(\text{GSCP})$ . Relatively to the value given to the selected variable, different options can be taken too.

The primal greedy procedure is summarized below.

**Procedure PRIMAL GREEDY**

**input:**  $M, N, (c_j)_{j \in N}, (h_i)_{i \in M}, (b_i)_{i \in M}, (a_{ij})_{i \in M, j \in N}$

**output:**  $(x_j)_{j \in N}, z_u$  [feasible solution for GSCP and upper bound on  $v(\text{GSCP})$ ]

1. [Initializing]

$$x_j := 0 \quad (j \in N)$$

$$z_u := 0$$

$$N_i := \{j \in N : a_{ij} = 1\} \quad (i \in M) \quad [\text{indices for columns to cover row } i]$$

$$M_j := \{i \in M : a_{ij} = 1\} \quad (j \in N) \quad [\text{indices for rows to be covered by column } j]$$

$$p_j := \sum_{i \in M_j} b_i \quad (j \in N) \quad [\text{weight for variable } j]$$

2. [Selecting a row index  $i^*$  according to criterion  $g(i)$ ]

$$i^* := \arg \min_{i \in M} g(i)$$

3. [Selecting a column index  $j^*$  according to criterion  $f(c_j, p_j)$ ]

$$j^* := \arg \min_{j \in N_{i^*}} f(c_j, p_j)$$

4. [Updating sets and values]

$$\Delta_{j^*} := \min \{h_{i^*} - x_{j^*}, v(b_i, i \in M_{j^*})\} \quad [\text{the increase on the variable } j^* \text{ depends on upper bound } h_{j^*} \text{ and on the criterion } v(b_i)]$$

$$x_{j^*} := x_{j^*} + \Delta_{j^*} \quad ; \quad z_u := z_u + c_{j^*} \Delta_{j^*}$$

update  $M, N,$  [rows to cover]

$(b_i)_{i \in M},$  [columns not fixed at their upper bounds]

$(N_i)_{i \in M}, (M_j)_{j \in N}$  and  $(p_j)_{j \in N}$  [uncovered demand for row  $i$ ]

5. [Checking feasibility]

if  $M \neq \Phi$  then go to 2 endif

stop

The computational complexity for this algorithm, considering any one of the criteria explained in previous works, is  $O(\max\{|M|, |N|\} \sum_{i \in M, j \in N} a_{ij})$ .

After obtaining a greedy feasible solution, local improvements on the upper bound can be attempted, amongst others, through a search procedure based on three main steps:

- decreasing redundant columns and producing a prime solution;
- replacing one column in the current solution by a cheaper one;
- replacing a pair of columns in the current solution by a single column which covers the same at less cost.

Follows the description of the local search procedure.

**Procedure SEARCH**

**input:**  $M, N, (c_j, h_j, M_j)_{j \in N}, (b_i, N_i)_{i \in M}, (x_j)_{j \in N}, z_u$

**output:**  $(x_j)_{j \in N}, z_u$  [improved feasible solution and upper bound on  $v(\text{GSCP})$ ]

1. [Obtaining a prime solution]

calculate  $(s_i)_{i \in M}$  [overcovered demand for row  $i$ ]

$$z^o := z_u$$

```

S := {j : Δj := min {xj, si over i ∈ Mj} ≥ 0}      [ indices of redundant columns ]
while S̄ ≠ ∅ do
  choose j̄ ∈ S̄                                     [ j̄ is selected for decrease ]
  xj̄ := xj̄ - Δj̄ ; zu := zu - cj̄ Δj̄
  update (si)i ∈ Mj̄, (Δj)j ∈ S and S̄
enddo
2. [Search for an improved solution changing one variable by another]
for j := 1, ..., |N| such that xj > 0 do
  find a variable j̄ such that cj̄ < cj and covers binding constraints of Mj
  calculate Δ
  xj := xj - Δ ; xj̄ := xj̄ + Δ ; zu := zu - Δ (cj - cj̄)
  update (si)i ∈ M
enddo
3. [Search for an improved solution changing a pair of variables by another]
for j1 := 1, ..., (|N| - 1) such that xj1 > 0 do
  for j2 := (j1 + 1), ..., |N| do
    if xj2 > 0 and Mj1 ∩ Mj2 has no binding constraints then
      find variable j̄ such that, cj̄ < cj1 + cj2 and covers binding constraints of
      Mj1 ∪ Mj2
      calculate Δ
      xjλ := xjλ - Δ (λ = 1, 2) ; xj̄ := xj̄ + Δ ; zu := zu - Δ (cj1 + cj2 - cj̄)
      update (si)i ∈ M
    endif
  enddo
enddo
4. [Checking local optimization]
if zu ≠ zo then (zo := zu and goto 1) endif
stop

```

As shown in [6], the algorithm is polynomial on  $|M|$  and  $|N|$  provided that the number of iterations for the local search phase is previously fixed. In fact, there one proves that the complexity of that local search procedure is given by  $O(|N|^2 \sum_{i \in M, j \in N} a_{ij})$ .

### 3. DUAL GREEDY HEURISTICS

Now, let us consider the linear program resulting from removing the integrality conditions, that is, the linear relaxation for GSCP:

$$\begin{aligned}
 (\overline{\text{GSCP}}) \quad & \min \sum_{j \in N} c_j x_j \\
 \text{s.t.} \quad & \sum_{j \in N_i} x_j \geq b_i & (i \in M) \\
 & 0 \leq x_j \leq h_j & (j \in N).
 \end{aligned} \tag{3.1}$$

We denote by  $v(\overline{\text{GSCP}})$  the optimal value for  $\overline{\text{GSCP}}$  which has the following linear dual problem:

$$\begin{aligned}
 (\overline{DGSCP}) \quad & \max \sum_{i \in M} b_i u_i - \sum_{j \in N} h_j v_j \\
 \text{s.t.} \quad & \sum_{i \in M_j} u_i - v_j \leq c_j \quad (j \in N) \\
 & v_j \geq 0 \quad (j \in N) \\
 & u_i \geq 0 \quad (i \in M).
 \end{aligned} \tag{3.2}$$

The constraints (3.2) can be rewritten, as

$$c_j - \sum_{i \in M_j} u_i + v_j \geq 0 \quad (j \in N) \tag{3.3}$$

with the left-hand-side being the linear reduced cost for the variable  $x_j$ .

It is well known that, if GSCP is feasible, then  $\nu(\overline{GSCP}) = \nu(\overline{DGSCP})$  is a lower bound on  $\nu(\text{GSCP})$ .

However, optimally solving  $\overline{GSCP}$  (or  $\overline{DGSCP}$ ) can be expensive from a computational point of view and a reasonably good lower bound may be obtained with less effort using heuristics to compute feasible solutions for  $\overline{DGSCP}$ .

Such heuristics can be developed for the GSCP following a similar pattern to the dual greedy heuristics developed by Balas and Ho [1] for the set covering problem. Hence the dual greedy heuristic that we developed for the GSCP selects, in each iteration, a particular element, say index  $i^*$ , from a subset row  $R \subseteq M$  and assigns a feasible value to the corresponding dual variable,  $u_{i^*}$ . The row index  $i^*$  is then removed from  $R$  and the process repeats until  $R$  becomes empty. If the initial  $R$  is strictly contained in  $M$ , the rows of  $M-R$  can be considered in order to improve the lower bound.

Two possibilities were considered for initializing  $R$ :

$$R := M \tag{DUAL 0}$$

and

$$R := \{i \in M : \sum_{j \in N_i} \bar{x} = b_i, \quad \text{with } \bar{x} \in \mathcal{F}(\text{GSCP})\}. \tag{DUAL 1}$$

The version DUAL 1 explicitly requires a feasible solution for GSCP which can be obtained from the procedure presented in the previous section.

For selecting a row from  $R$ , several different criteria were considered in [8] and [9].

Concerning to the dual variables  $v_j$  ( $j \in N$ ), although they can be given positive values, our own experience has shown that they seldom differ from zero in the best heuristic solutions for the problems such that the variable-bounds  $h_j$  are not very tight. In fact, this is the case for the real life bus crew scheduling problems that we considered and, therefore, one forces the  $v_j$  to be always at zero level.

Both DUAL 0 and DUAL 1, are synthetized in the following:

**Procedure DUAL**

**input:**  $M, N, (c_j, h_j, M_j)_{j \in N}, (b_i, N_i)_{i \in M}, (\bar{x}_j)_{j \in N}$  (DUAL 1) [feasible solution for GSCP]

**output:**  $((u_i)_{i \in M}, 0), z$  [feasible solution for  $\overline{DGSCP}$  and lower bound on  $\nu(\text{GSCP})$ ]



1. [Initializing]

$u_i := 0$  (i ∈ M) [dual variable associated to row i]

$R := \{ M \text{ (DUAL 0)}$

$\{i \in M : \sum_{j \in N_i} \bar{x}_j = b_i, \bar{x} \in \mathcal{F}(\text{GSCP})\}$  (DUAL 1)

$z_\lambda := 0$

2. [Selecting a row]

if  $R = \emptyset$  then goto 4

else choose  $i^* \in R$

endif

3. [Assigning a value to the dual variable and updating]

$u_{i^*} := \min_{j \in N_{i^*}} c_j$

$z_\lambda := z_\lambda + u_{i^*} b_{i^*}$

$c_j - u_{i^*}$  (j ∈ N<sub>i<sup>\*</sup></sub>)

$c_j := \{ \text{[update j-th reduced cost]}$

$R := R - \{i^*\}$  (j ∈ N - N<sub>i<sup>\*</sup></sub>)

goto 2

4. [Checking the terminal conditions]

if (step 4 is entered for the first time) then ( $R := M - R$  and goto 2) endif

stop

In a rather straightforward implementation this procedure DUAL is  $O(|M| \max\{|M|, |N|\})$ .

An attempt to improve the bounds, both from below and from above, produced by the heuristics PRIMAL and DUAL, can be made by trying to impose the linear complementary conditions on the corresponding solutions. That is, if  $[\bar{x}, (\bar{u}, 0)]$  is a pair of primal and dual feasible solutions for  $\overline{\text{GSCP}}$  one aims to impose:

$$\sum_{i \in M} \bar{u}_i \left( \sum_{j \in N_i} \bar{x}_j - b_i \right) = 0 \quad (3.4)$$

$$\sum_{j \in N} \bar{x}_j \left( \sum_{i \in M_j} \bar{u}_i - c_j \right) = 0. \quad (3.5)$$

Obviously, if  $\bar{x}$  and  $(\bar{u}, 0)$  satisfy (3.4) and (3.5) then they are optimal solutions, respectively for  $\overline{\text{GSCP}}$  and  $\overline{\text{DGSCP}}$ . If  $\bar{x}$  is feasible for GSCP then  $\mathcal{V}(\text{GSCP})$

$$= \sum_{j \in N} c_j \bar{x}_j.$$

The procedure IMPROVE 1, which is briefly described next, performs linear complementary improvement tests for a pair of dual and primal solutions.

Firstly, the primal solution is modified in order to satisfy (3.5) by setting  $\bar{x}_j = 0$  if  $r_j > 0$ . The resulting solution is unlike to be feasible and a primal greedy heuristic technique is used to bring it into feasibility. Then, the row complementary relations (3.4) are imposed to the dual solution relatively to the new primal solution in the following way:

- set, all the dual variables associated with positive surplus equal to zero;
- keeping dual feasibility, increase first the dual variables associated with zero surplus and then increase the ones which are associated with positive surplus.

The process repeats itself until no changes in both solutions are produced or the number of iterations exceeds a maximum value which, in our case, was fixed at 3.

The procedure IMPROVE 1 has the same complexity that the primal heuristic procedure called as a subroutine.

Another bound improving technique, that we named IMPROVE 2, was devised as follows:

1. initialize  $\bar{N}$  as the index set of columns for a primal feasible solution;
2. compute both primal and dual feasible solutions for the problem restricted to  $\bar{N}$ ;
3. find, among  $N - \bar{N}$ , the variable with the most negative reduced cost for the restricted dual solution; if no negative reduced cost variable exists, stop; otherwise, include the selected variable in  $\bar{N}$  and resume to 2.

The algorithm IMPROVE 2 has a computational complexity equal to  $|N|$  times the complexity of the primal-dual procedure called in step 2.

#### 4. COMPUTATIONAL RESULTS

As we have seen in the previous sections of this paper the primal and dual heuristics are polynomial algorithms on the dimensions of the GSCP, although, for some of the heuristics, the number of computing steps is bounded above by a polynomial in  $|M| |N|$  with high degree.

Besides, this worst-case computational study, an analysis of the heuristics from an empirical point of view was carried out. The test problems that we considered are of the bus crew scheduling type and so, exhibit the particularities mentioned in section 1.

All the test problems have the same number of covering constraints,  $|M| = 36$ , corresponding to the working periods from 6 a.m. to midnight. The set of test problems, consists of 30 instances generated according to the driver's contract rules in Rodoviária Nacional (15 with 100 columns - R1 to R15 - and 15 with 865 - G1 to G15). The right-hand-side numbers were randomly generated in order to produce 3 different type of distributions - unimodal, bimodal and irregular - for the demand pattern. The last 4 test problems correspond to real bus crew scheduling situations at Rodoviária Nacional (RN1 to RN4 also with 865 columns). The cost of each working shift is defined according to the rules in use at Rodoviária Nacional and consists of a fixed cost added by extra-costs related either to the period of the day or overtime working periods.

For a more detailed description of these problems see [7].

The combined procedure includes all heuristics described in this paper and gives a feasible solution for the GSCP and simultaneously a lower bound that yields a measure by excess of the error associated to the solution.

After performing extensive tests, we opted by a final version consisting of a combination of procedure DUAL 0, DUAL 1 (both with two different selection criteria), GREEDY PRIMAL (several row and column selection criteria) with SEARCH, IMPROVE 1 and IMPROVE 2. Also dual linear penalties ([8]) were calculated in order to tighten the variable-bounds,  $h_j$ , or eventually eliminating the variables.

This experience is reported in Table I where the column (1) identifies the problem which dimensions ( $|M|$ ,  $|N|$  and density - number of ones over  $|M| |N|$ ) and optimal value are shown, respectively, in columns (2) and (3).

TABLE I  
Computing times and quality of the bounds obtained with the primal-dual heuristic procedure

Problem	Dimension	Optimal value	Lower %	Upper %	Time sec
(1)	(2)	(3)	(4)	(5)	(6)
R1	IMI = 36	51726	90.5	3.2	6.0
R2	INI = 100	52268	96.3	0.7	6.3
R3	55%	84184	99.7	2.2	7.1
R4		85080	93.8	4.4	9.7
R5		120930	99.5	0.1	6.8
R6		48990	100.0	0.0	2.3
R7		27062	98.4	1.3	4.8
R8		55236	96.6	2.9	6.4
R9		47470	98.9	1.0	5.8
R10		26612	87.5	1.1	4.2
R11		98930	95.8	1.3	5.5
R12		51248	89.3	0.0	4.3
R13		49096	93.2	0.0	6.1
R14		40862	97.1	4.2	4.8
R15		188598	94.4	3.6	6.6
G1	IMI = 36	49972	92.2	5.2	31.6
G2	INI = 865	49326	96.3	0.8	28.6
G3	56%	80248	99.0	1.0	32.2
G4		79670	92.1	3.0	39.1
G5		115540	100.0	0.4	36.2
G6		48434	100.0	0.0	16.8
G7		24084	91.2	2.6	21.1
G8		51900	96.3	1.2	30.1
G9		45658	99.8	0.5	21.9
G10		25248	91.3	3.0	26.4
G11		93078	99.9	0.1	22.1
G12		51134	89.2	1.0	28.9
G13		47720	89.6	0.0	17.2
G14		36520	100.0	0.0	9.2
G15		167728	99.99	0.1	23.2
RN1	IMI = 36	70418	96.0	0.1	40.4
RN2	INI = 865	45952	98.3	4.4	36.8
RN3	56%	33360	83.4	0.5	31.2
RN4		80316	100.0	0.0	29.1
Average			95.5	1.5	
Worst value			83.4	5.2	
Number of optimal values			5	7	

The remaining columns of Table I refer to the computing experience relative to the primal-dual heuristic procedure that was carried out on a VAX 11/750 VMS with FPA and FORTRAN compiler (similar to the one presented in [8]).

Columns (4) and (5) refer the quality for the final lower and upper bounds. Hence, in column (4), the value  $(z_l / \nu(\text{GSCP}))10^2$  is shown for each test problem and, for evaluating the upper bounds, we give the value  $10^2 (z_u - \nu(\text{GSCP})) / \nu(\text{GSCP})$  in column (5).

Finally column (6) shows the computing time in seconds.

From this experience we can see (column (5)) that the final upper bound is very good for most of the test problems. In fact, the worst value (problem G1) was 5.2% above the optimum, and the upper bound is equal to the optimal value for 7 out of 34 test problems. For the large real life problems (RN1 to RN4), the upper bound value proved even better. Concerning to the lower bound values, the procedure didn't prove as good as above. Although, the lower bound value is within 1% of the optimum for 11 test problems, the procedure failed to produce reasonably good results for some cases, in particular for problem RN3.

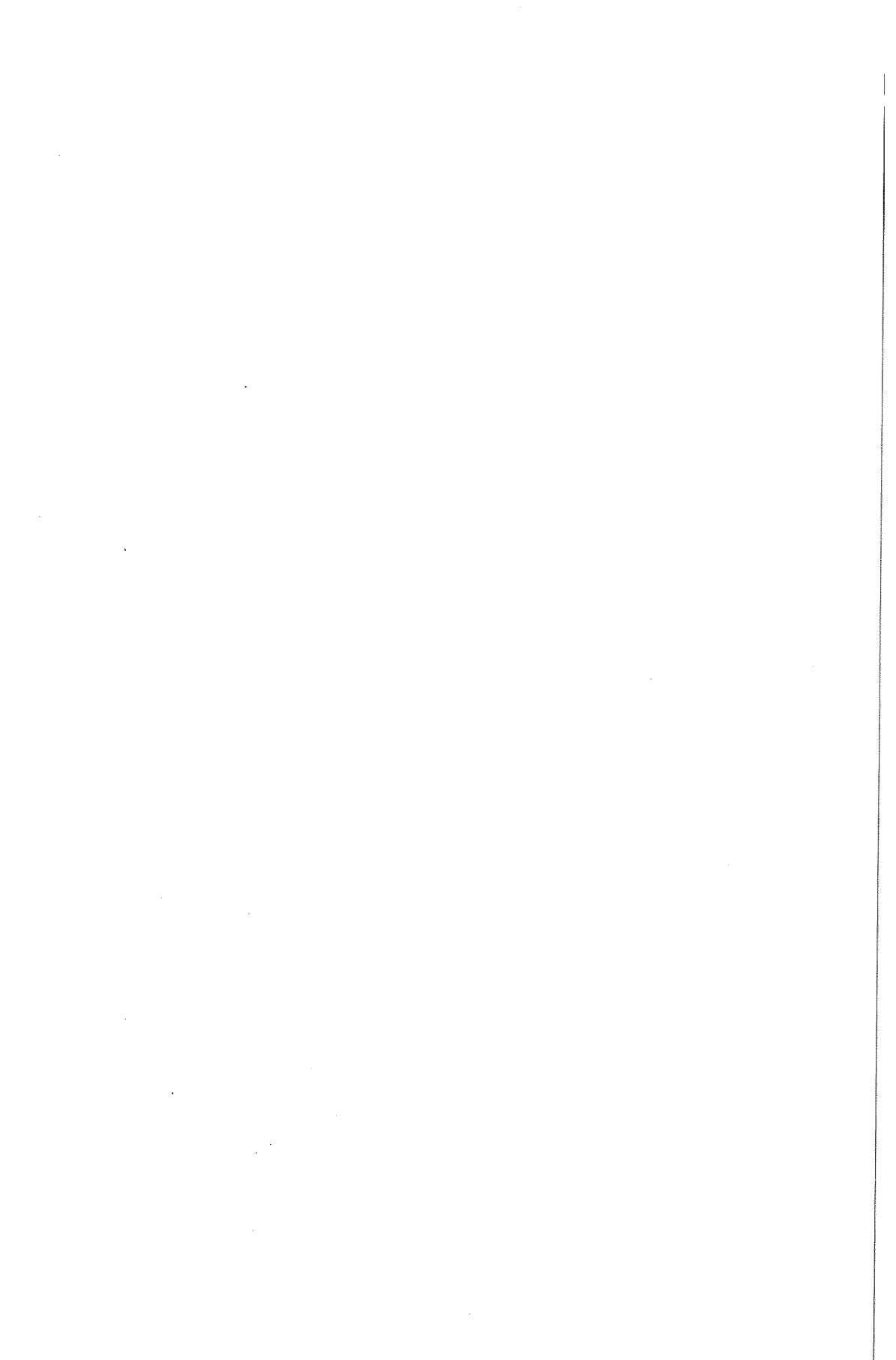
## 5. CONCLUSIONS

Finally, we can conclude that the primal-dual greedy procedure described in this paper is a relatively easy method to tackle with the generalized set covering problem. In particular, it proved to be reasonably fast at least for the two-duty period bus driver scheduling instances that we tried out to solve. The primal-dual heuristics implemented on a microcomputer have been in use at Rodoviária Nacional ([5]) with very good results for the real bus crew scheduling problems.

Moreover, this procedure can be used as an initial step for a subgradient optimization based technique to improve a lagrangean bound on the optimum. It also can be embedded, as the bounding tool, in a tree-search-method for optimally solving the GSCP.

## REFERENCES

- [1] Balas, E. and Ho, A., "Set covering Algorithms Using Cutting Planes, Heuristics and Subgradient Optimization: A Computational Study", *Mathematical Programming*, vol. 12, pp. 37-70 (1980).
- [2] Balas, E. and Padberg, M.W., "On the Set Covering Problem", *Operations Research*, vol. 20, pp. 1153-1161 (1972).
- [3] Hall, N. and Hochbaum, D., "The Multicovering Problem: The Use of Heuristics, Cutting Planes and Subgradient Optimization for a Class of Integer Programs", Working Paper WPS 85-73, College of Administrative Science, The Ohio State University (1985).
- [4] Paixão, J.P., "Algorithms for Large Scale Set Covering Problems", Ph.D.Thesis, Department of Management Science, Imperial College, London (1983)
- [5] Paixão J.P., Branco, I.M., Captivo, E., Pato, M.V., Eusébio, R. and Amado, L., "Bus and Crew Scheduling on a Microcomputer", in "Or Models on Microcomputers", J.D. Coelho and L. V. Tavares (eds), North Holland Amsterdam (1986).
- [6] Paixão, J.P. and Pato, M.V., "Análise de Pior-Caso no Comportamento de Heurísticas Greedy para o Problema de Cobertura Generalizado", nota 9/87 CEA, Faculdade de Ciências da Universidade de Lisboa (1987).
- [7] Paixão, J.P. and Pato, M.V., "Greedy Heuristics for Generalized Set Covering Problems: A Computational Study", nota 8/85 CEA, Faculdade de Ciências da Universidade de Lisboa (1985).
- [8] Paixão, J.P. and Pato, M.V., "Heuristics, Lagrangean Relaxation and Penalties for Two-Duty Period Scheduling Problems", nota 1/87 CEA, Faculdade de Ciências da Universidade de Lisboa (1987).
- [9] Paixão, J.P. and Pato, M.V., "Heuristics for an Operator Scheduling Problem", nota 10/85 CEA, Faculdade de Ciências da Universidade de Lisboa (1985).
- [10] Papadimitriou, C.H. and Steiglitz, K., "Combinatorial Optimization", Prentice Hall, Englewood Cliffs (1982).





## MODELO DE SIMULAÇÃO DE TRÁFEGO: CASO DO TRANSPORTE METROPOLITANO

J. Caliço Grosso  
Metropolitano de Lisboa

### Resumo

É apresentado um modelo de simulação de tráfego de um sistema de transportes. Embora o modelo se concretize no tipo de transporte metropolitano, a sua transposição para outros sistemas de transporte público é bastante fácil.

Neste artigo, após uma pequena introdução sobre o que é a simulação, expõe-se o esquema lógico representativo da circulação dos comboios numa rede de metropolitano. Seguidamente, indicam-se os dados de base definidores da estrutura da rede, da procura e da oferta de transporte.

Resumidamente, apresenta-se a lógica do processamento de algumas actividades, bem como o tratamento técnico de algumas variáveis aleatórias.

Finalmente, indicam-se os dados produzidos pela simulação, procurando evidenciar o apoio que este tipo de modelo pode prestar não só à gestão corrente (gestão operacional), como ao estudo de hipotéticas redes (gestão estratégica).

### 1. O que é a Simulação ?

A simulação é uma técnica que tem como objectivo a condução de experiências em modelos representativos de determinada realidade que se pretende estudar. Em sistemas de alguma complexidade, as variáveis são de tal modo numerosas e de difícil controle que a condução de experiências "reais" se torna impraticável. Por exemplo, qual a diferença entre dois tipos de horário e quais as informações estatísticas de exploração que resultarão de um novo horário num sistema de transportes públicos ? Por outro lado, situações há em que o analista pretende obter informações sobre funcionamento de um sistema a construir. Neste caso é obviamente impossível experimentar o real. Por exemplo, é absurdo pensar em obter dados de exploração em experiências conduzidas numa linha de metropolitano por construir ou ainda não aberta ao público. As técnicas de simulação pretendem precisamente resolver as situações em que a experimentação no real não faz sentido por ser impossível, cara ou desaconselhável por qualquer motivo.

No domínio da engenharia é já antiga a utilização da simulação de natureza analógica: estudos efectuados em modelos de barragens, pontes, simuladores de voo, etc...

O advento dos computadores levou a que um novo tipo de simulação ficasse disponível: a simulação digital. Neste tipo de simulação é concebido um modelo de determinada realidade através de um encadeado de decisões lógicas e cálculos matemáticos.

Esse modelo é programado para computador, instrumento que se torna necessário pelo grande número de cálculos a efectuar.

Uma vez introduzido o programa do modelo no computador são conduzidas experiências com vista a analisar o funcionamento do sistema debaixo de condições que se pretendem testar.

A estrutura do modelo de simulação traduzirá a realidade que pretende representar através de relações de natureza lógica, funções matemáticas e parâmetros caracterizadores da sua estrutura.

O modelo recebe um conjunto de valores (relativos a variáveis decisionais), que representam níveis dos elementos que o decisor pode arbitrar. Em consequência, é produzido um conjunto de resultados que uma vez analisados e correlacionados com os valores das variáveis decisionais que foram introduzidas, permitem ao decisor pôr em prática a política mais adequada aos seus objectivos. Assim, a simulação, como de resto qualquer técnica, não substitui o decisor constituindo, antes, uma ferramenta que pode ser de extrema utilidade.

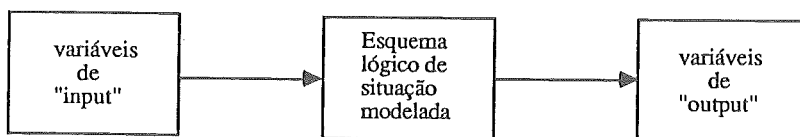
## 2. Simulação numa rede de transportes

Se há actividades em que a experimentação no "real" é difícil, senão impossível, os transportes estão entre elas. A complexidade dos sistemas de transportes, pelo grande número de variáveis, sua alteração temporal e implicações ao nível da colectividade fizeram com que um dos sectores privilegiados da simulação digital fosse o dos transportes. De facto, não é fácil ver as consequências, em termos de tráfego, da implementação de determinadas modificações na rede. E mesmo admitindo que era desejável proceder a essas mesmas alterações para observar as consequências, é obvio que a par de eventuais fracassos com todas as consequências negativas, a própria quantificação das consequências nas diferentes vertentes seria de grande dificuldade.

A simulação, uma vez definidos os parâmetros caracterizadores do sistema (paragens, troços, linhas, meios de locomoção e respectivos horários), permite obter as consequências em termos dos diversos indicadores do tráfego.

## 3. Modelo de simulação de tráfego para o Metropolitano de Lisboa (ML)

Em particular, uma rede de metropolitano muito pode beneficiar desta técnica. E é para preencher essa intenção que apresentamos um modelo de simulação do tráfego para o ML. Este modelo está programado em FORTRAN. Após uma breve apresentação do modelo, indicaremos o modo de utilização. Gráficamente podemos decompôr um modelo de simulação em três componentes



No nosso caso vamos apresentar muito sucintamente as três componentes referidas, começando pelo esquema lógico representado pelo fluxograma do modelo:

### a) Fluxograma do modelo

Nas páginas 4 a 6 apresenta-se o fluxograma do modelo. Para a leitura do fluxograma diremos que os rectângulos representam acções (chegada ou partida do comboio, entrada de passageiros, etc.). Os círculos representam entidades que deverão ter um estado determinado para que a acção se verifique (um comboio parte de acordo com o horário, um comboio só chega a uma estação se, no seu sentido, não houver nenhum comboio nessa estação, etc.).

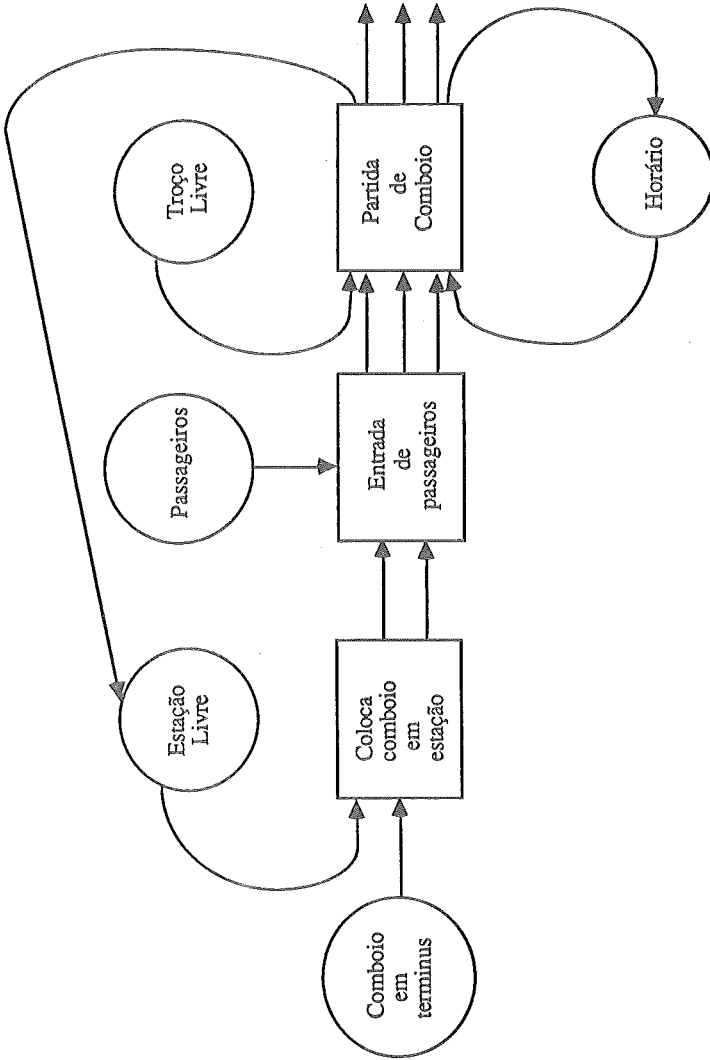
Vamos apresentar três situações:

- a. 1 Partida do início de linha
- a. 2 Chegada e partida de estação intermédia
- a. 3 Chegada a uma estação terminal e colocação em terminus

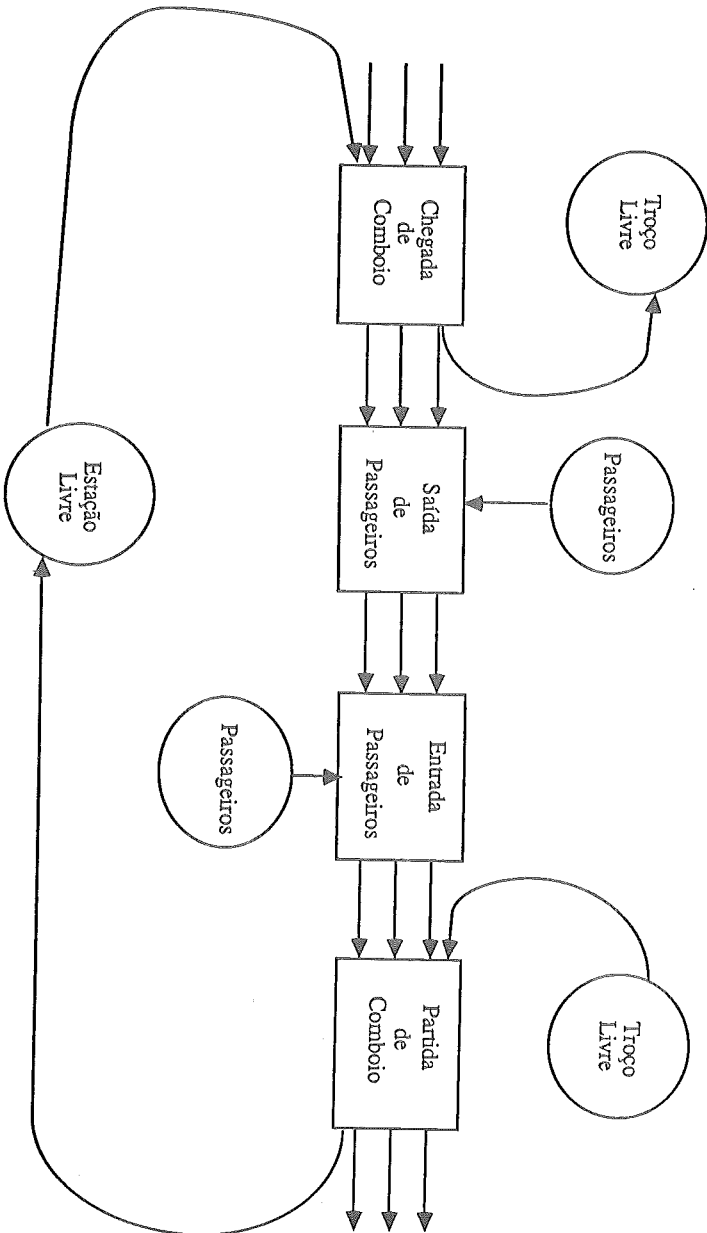
### b) Variáveis de "input"

Consideramos no conjunto destas entidades não só variáveis de decisão que podem ser definidas por critérios próprios do gestor responsável pela

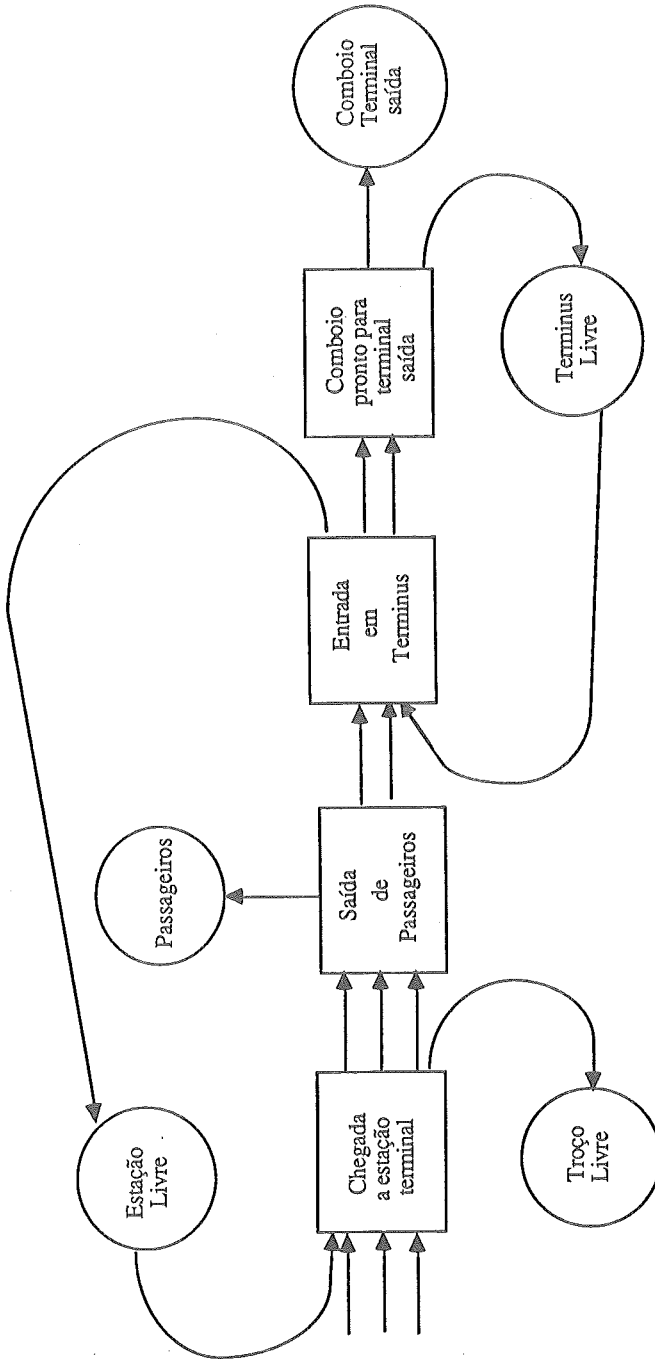
a) 1. Partida de comboio do início da linha



a) 2. Chegada e partida da estação intermediária



a) 3. Chegada a estação Terminal





administração do sistema de transporte, como entidades que têm a ver com a estrutura do sistema de transportes e que seriam melhor designadas de parâmetros. Todavia, vamos chamar a todas estas entidades variáveis, adjectivando-as para evidenciar o seu grau de "variabilidade". Assim, consideramos quatro conjuntos de elementos de input agrupados em outros tantos ficheiros: fixos, semi-fixos, exógenos e decisionais.

b. 1 Elementos fixos

Estes elementos tem a ver com a definição da estrutura da rede, sendo, na prática, e para dada rede, imutáveis. ELEFIX designa o ficheiro onde estão inscritos os dados relativos às estações, troços e seu comprimento, linhas e sua definição (através dos troços constituintes).

b. 2 Elementos semi-fixos

Estes elementos não são tão rígidos como os anteriores, podendo ser alterados por inovações tecnológicas e outras razões. O ficheiro chama-se ELESFIX e contém a indicação dos terminus, duração do percurso de cada troço, duração do embarque/desembarque dos passageiros, tempo de paragem mínimo em cada estação, tempo de deslocamento em vazio entre cada par de terminus e tempo de manobra nos terminus.

b. 3 Elementos exógenos

Apesar da importância destes elementos (são eles a razão de ser de qualquer sistema de transportes), chamamos-lhe exógenos pela sua fluidez e fraco poder de intervenção do ML relativamente à sua definição. São elementos a obter por contagens e sondagens.

Obviamente, são indispensáveis, constituindo um elemento fundamental à simulação. São eles: taxa de chegada dos passageiros a cada estação e em cada período horário, probabilidade de deslocamento entre estações, ao longo do dia.

b. 4 Elementos decisionais

Estes elementos constituem o conjunto de dados a fornecer ao modelo com maior variabilidade. São eles que definem os valores a experimentar, as estratégias do gestor.

Assim, enquanto os outros ficheiros, uma vez definidos, se manterão relativamente fixos por muito tempo, este pode ter tantas versões quanto as experiências que se quiserem conduzir. Assim, podem-se dispôr no computador vários ficheiros com designações diferentes que, a seu critério, o decisor faz introduzir. Por isso o próprio nome do ficheiro é variável: NOMFIC é uma designação que assume em cada "run" do programa o nome de diferentes ficheiros que o decisor constituiu.

Os dados nele existentes são instantes de partida de comboio em cada partida, lotação do comboio, instante de início e duração da simulação.

c) Lógica de funcionamento do modelo

Os elementos atrás apresentados representam os parâmetros caracterizadores da estrutura física da rede, das condições de circulação, da procura e da oferta.

Assim, considerada determinada hipótese de procura (taxa de chegada de passageiros a cada estação, ao longo do dia, e matriz de probabilidades de deslocamento) e da oferta (instantes de partida em cada linha e capacidade dos comboios afectos), descreveremos sucintamente a lógica de funcionamento do modelo, através da explicitação dos mecanismos lógicos de processamento das acções mais significativas.

Em termos de dinâmica de funcionamento diremos que a unidade tempo é o segundo e que iniciada a simulação no instante (segundo) zero o "relógio" avança de acordo com o tratamento sucessivo dos acontecimentos (acções) temporalmente mais próximos, terminando quando todas as acções se tiverem executado que, logicamente, coincidirá com o termo da exploração, ou seja, com a chegada do último comboio ao terminus, o que ocorrerá no fim do dia de exploração do Metropolitano.

Naturalmente, que a 1ª acção a ter lugar é a saída do primeiro comboio. Assim, o modelo detecta o instante da primeira partida "ventilando" os ficheiros das partidas em cada linha. Em seguida, calcula o número de passageiros que

chegaram (desde o instante zero) a esse terminus, determina o número de passageiros que vai para cada estação, dá entrada dos passageiros que podem utilizar aquele comboio, determina o instante em que o comboio chega à estação seguinte e considera o comboio fora da estação de partida e a ocupar o troço definido por essa estação e a estação de chegada.

Seguidamente, o modelo vai procurar a próxima acção que nesta altura só pode ser uma de duas: (a) chegada do comboio já em linha à estação fim do troço que está a percorrer ou (b) partida de outro comboio do início de uma linha.

Vamos, agora, indicar sinteticamente as acções mais significativas e a lógica com que elas são tratadas no modelo.

<u>Acção</u>	<u>Lógica de Processamento</u>
Partida de comboio início da linha	Quando o relógio do modelo regista um instante coincidente com a partida de um comboio do terminus, vê se há do comboio disponível nesse terminus. Se há, procede à acção anteriormente descrita, retirando esse comboio do terminus. Se não há comboio disponível no terminus, retira um comboio do "depósito geral" e lança-o na circulação de acordo com a lógica descrita anteriormente.
Chegada de comboio à intermédia.	Quando o tempo do "relógio" coincide com o instante de chegada de um comboio a uma estação intermédia, o modelo analisa se a estação está livre (se não está livre atrasa o instante de chegada do comboio de um segundo e vai procurar a próxima acção a ser executada) e, caso esteja livre, desembarca os passageiros que se destinam a essa estação e procede à entrada dos passageiros existentes nessa estação e servidos por aquele comboio. O instante de partida do comboio é determinado pela soma do instante actual ao resultado da multiplicação do número total de passageiros embarcados e desembarcados pelo tempo médio de embarque/desembarque de um passageiro.
Partida de comboio de estação intermédia.	Quando o "relógio" regista um tempo igual ao instante de partida, o comboio é lançado no troço seguinte e o tempo estimado para a duração do percurso desse troço é adicionado ao instante actual, produzindo, assim, o instante de chegada desse comboio à estação seguinte.
Chegada de comboio a estação terminus	Quando o "relógio" regista um tempo coincidente com a chegada de um comboio a um terminus, o modelo vê se a estação está livre (se não está livre atrasa a chegada de um segundo e vai ver qual a próxima acção a ser executada) e, caso esteja, procede ao desembarque dos passageiros que vão no comboio, calculando o tempo que essa operação dura. Adicionando essa duração ao instante actual, determina o instante em que o comboio pode entrar em manobra no terminus. Adicionando depois o tempo de manobra no terminus, determina o instante em que o comboio pode ser utilizado nas linhas que partem desse terminus.
Comboio é colocado em terminus.	Quando o "relógio" regista um tempo coincidente com o fim das manobras em terminus, coloca o comboio como disponível para partida na linha (ou linhas) de que essa estação é início.

Não sendo todas as acções tratadas pelo modelo as que estão referidas atrás, elas são as mais significativas.

No decorrer da simulação é acumulada toda a informação que permite produzir as estatísticas de tráfego e que apresentamos noutra secção. Algumas informações sobre as condições em que o tráfego simulado está a decorrer são produzidas como seja a situação em que o número de passageiros a entrar em determinada estação leva ao esgotamento da capacidade disponível do comboio. Nesta circunstância é enviada uma mensagem para um ficheiro registando-se toda a informação relacionada com esse facto (estação, instante, número em excesso, etc.), que será submetida a análise posterior.

A título de exemplo apresentamos de seguida os mecanismos utilizados no tratamento da afluência de passageiros às estações e sua distribuição pelas diferentes hipóteses de destino.

O número de passageiros numa estação é determinado adicionando o número de passageiros que ficou nessa estação, após a última partida do comboio, com um número que é obtido do seguinte modo:

- Multiplica-se o intervalo de tempo que decorreu após a última partida de comboio dessa estação pela taxa de chegada de passageiros a essa estação nesse período do dia, arredondando-se o número assim obtido para o inteiro mais próximo, N.

Se esse valor N é maior do que L (o número L que temos utilizado é 20 mas pode ser alterado se assim for entendido), obtêm-se o número definidor das chegadas simulando a distribuição NORMAL com média N e variância N. Se N é menor do que L, o número de chegadas resulta da simulação da distribuição POISSON, de média N.

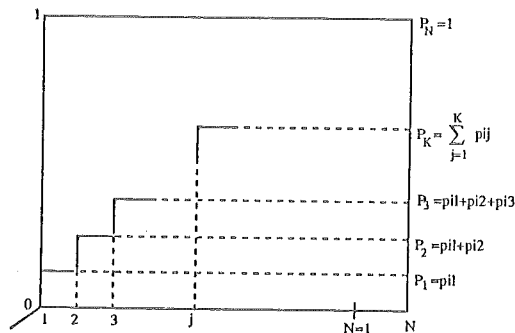
Para determinar o destino que têm os passageiros entretanto chegados a cada estação, funciona a seguinte lógica:

1. Para cada período do dia (30 minutos, 1 hora, ou outro período que se quiser usar), é dada uma matriz de probabilidade de deslocamento entre as estações do Metropolitano. Isto é, para cada período do dia, há um ficheiro que diz, em relação a cada estação, i, quais as probabilidades  $p_{ij}$  ( $j=1, N$ , número de estações) de um passageiro se deslocar da estação i para a estação j.
2. Assim, se desde a última partida de um comboio da estação i chegaram Q passageiros a essa estação, o modelo determina o número de passageiros que se deslocará para a estação j ( $j=1, N$ ) pela aplicação do critério seguidamente apresentado:

- a.  $K_j =$  maior inteiro contido no produto  $Q \times p_{ij}$  ( $j=1, N$ )
- b. Calcula a diferença

$$K' = Q - \sum_{j=1}^N K_j$$

- c. Define o histograma acumulado das probabilidades  $p_{ij}$



- d. Extraí sucessivamente  $K'$  número pseudo-aleatórios entre 0 e 1 e, de cada vez, um passageiro é acrescentado aos que vão para a estação j, desde que o número pseudo-aleatório, x, extraído verifique a condição  $P_{j-1} \leq x \leq P_j$

d) Variáveis de "output"

Os elementos produzidos pela simulação são praticamente todos os que interessam à análise do tráfego na rede.

Assim, e sem pretendermos ser exaustivos, indicamos alguns resultados do modelo de simulação.

- d.1 Passageiros entrados e saídos em cada estação em períodos de tempo, e, naturalmente, no total da simulação.
- d.2 Cargas em cada troço em períodos de tempo e no total.
- d.3 Total de lugares oferecidos.
- d.4 Total de lugares/Km oferecidos.
- d.5 Taxa de ocupação dos comboios.
- d.6 Matriz origem/destino, em períodos de tempo e no total.
- d.7 Transbordos na Rotunda.
- d.8 Gráfico de marcha dos comboios.
- d.9 Horários detalhados de cada comboio, com chegada e partida em cada estação. Estes horários, consideram a influência exercida pela intensidade da procura.
- d.10 Número de comboios em cada terminus ao longo do tempo.
- d.11 Transferências de comboios entre terminus.
- d.12 Mensagens sobre excesso de passageiros nas estações (oferta não satisfaz procura).
- d.13 Número mínimo de comboios a utilizar na implementação de qualquer esquema de horário.
- d.14 Diferentes rácios que se entendam por convenientes produzir, como o tempo médio de espera do passageiro na estação, tempo médio no metropolitano, etc...

Como procurámos transmitir, o modelo de simulação que apresentamos permite explorar as consequências no tráfego de qualquer rede existente, ou a implementar, dados os elementos exógenos e em função de uma estratégia particular (partidas de cada terminus e comboios a utilizar). Este modelo, portanto, pode ser utilizado não só para "experimental", antes de pôr em prática, novos esquemas de horários, produzindo a informação estatística de tráfego correspondente, como também para produzir informação e avaliar consequências em futuras redes de Metropolitano, a implementar.

Como curiosidade adiantamos que para a actual rede simulámos os horários actualmente em prática ao longo de todo o dia, tendo sido necessários 21 comboios: o mesmo que a Direcção de Exploração está a utilizar. Embora tivéssemos a esperança de propôr uma solução com menos de 21 comboios, tal foi demonstrado não ser teoricamente possível com os horários actualmente implementados pela DE.

Simplesmente se se puser aos técnicos da DE o problema de saber quantos comboios, no mínimo, são necessários para implementar um determinado horário para a rede, após ampliação da mesma, estamos em crer que vão ter dificuldades em responder (não são as pessoas em concreto que equacionamos, até porque a situação actual demonstra a sua perícia, mas a enormidade de cálculos que terão de desenvolver, sendo desumano exigir que estejam seguros da resposta).

No modelo de simulação a resposta será dada aproximadamente em 45 minutos. Esta é uma potencialidade do modelo de simulação, embora não seja a mais importante.

e) Dados a fornecer ao modelo

Como dissemos, o modelo está programado em FORTRAN.

O input existe em 4 ficheiros de input, sendo 3 não decisoriais, que designámos ELEFIX, (1), ELESFIX, (2), ELEVAR, (3) são relativamente estáveis; Os dados para 1 e 2 são fáceis de obter. Para o 3, haverá que proceder a contagens nas estações, não sendo necessário contar ao longo de todo o dia. Como nos interessa a taxa de chegada (definida como número, que pode ser fracção de 1), de passageiros por segundo, a contagem de períodos de 5 ou 10 minutos ao longo do dia, desde que criteriosamente definidos, pode dar os elementos necessários à simulação.

Analogamente a probabilidade de deslocamento entre estações não obriga a questionar todos os passageiros entrados na rede. O princípio a utilizar será o das sondagens e estas não observam todo o universo, como é sabido. Uma vez desenvolvido o esforço inicial de levantamento de dados iniciais, é de extrema

facilidade a manutenção das condições necessárias à produção das mais diversificadas estatísticas de tráfego.

f) Utilização do modelo. Mecanismos gerais

Em jeito de conclusão diremos que é apresentado um instrumento de análise muito útil, desculpe-se a imodéstia, quer em termos de apoio à gestão táctica (gestão corrente), quer em termos de gestão estratégica (longo prazo).

Em termos de curto prazo, consideramos ser um instrumento de apoio não só na produção de dados estatísticos de tráfego, como em termos de suporte à elaboração de horários, respectiva gestão e produção de dados com vista à optimização de escalas de serviço.

Em termos de longo prazo, para o gestor estratégico, constitui um apoio potencialmente de grande utilidade já que permite obter, não só em termos globais, como de detalhe, as consequências de diferentes hipóteses de expansão e esquemas de horários, para diferentes variáveis exógenas. Analogamente podem-se equacionar as consequências na exploração da rede de alterações urbanísticas na área servida pela rede e que se admite virem a ter lugar.

O modelo está em condições de ser utilizado, todavia, a bondade dos seus resultados depende bastante da vontade dos potenciais utilizadores em o quererem utilizar. Para já, vemos dois utilizadores indiscutíveis: A Direcção de Exploração, numa óptica de curto prazo, e a Direcção de Desenvolvimento, numa óptica de longo prazo.

Os procedimentos a utilizar para fazer "correr" o programa são de extrema simplicidade. Consideramos actuações a três níveis:

f.1 Alteração dos ficheiros de input

Para isso chama-se o ficheiro que se pretende alterar, com o comando ex (espaço) (nome de ficheiro)

chama-se a linha que se pretende alterar digitando o respectivo número. Em consequência aparece no écran a linha a modificar. Faz-se agora s/ (informação que está) / (informação que se quer passe a estar)

Faz-se agora, x, e o ficheiro é guardado com a alteração feita.

f.2 Instruções para correr o programa

Para isso basta digitar o nome de programa. Por exemplo

Teste 100

de seguida aparece no écran

"DIGA NOME DO FICHEIRO"

o utilizador digita então o nome do ficheiro que tem os dados a usar na simulação e que designamos de "decisionais": instantes de partida de cada comboio em cada linha e tipo de viajante.

Uma vez digitado o nome desse ficheiro o programa corre até ao fim sem mais intervenção do utilizador.

f.3 Ficheiros de resultados produzidos pela simulação

Uma série de ficheiros são gerados aquando da simulação. Entre os mais relevantes temos:

O ficheiro "GERAL" contem eventuais anomalias encontradas no decorrer da simulação, como o excesso de passageiros na estação, e, também, informação relativa à entrada dos comboios em circulação, cargas nos troços, passageiros entrados e saídos nas estações, taxa de ocupação dos comboios, lugares/Km oferecidos.

O ficheiro "PEMDES" contem o número de passageiros embarcados em cada estação, bem como os desembarcados.

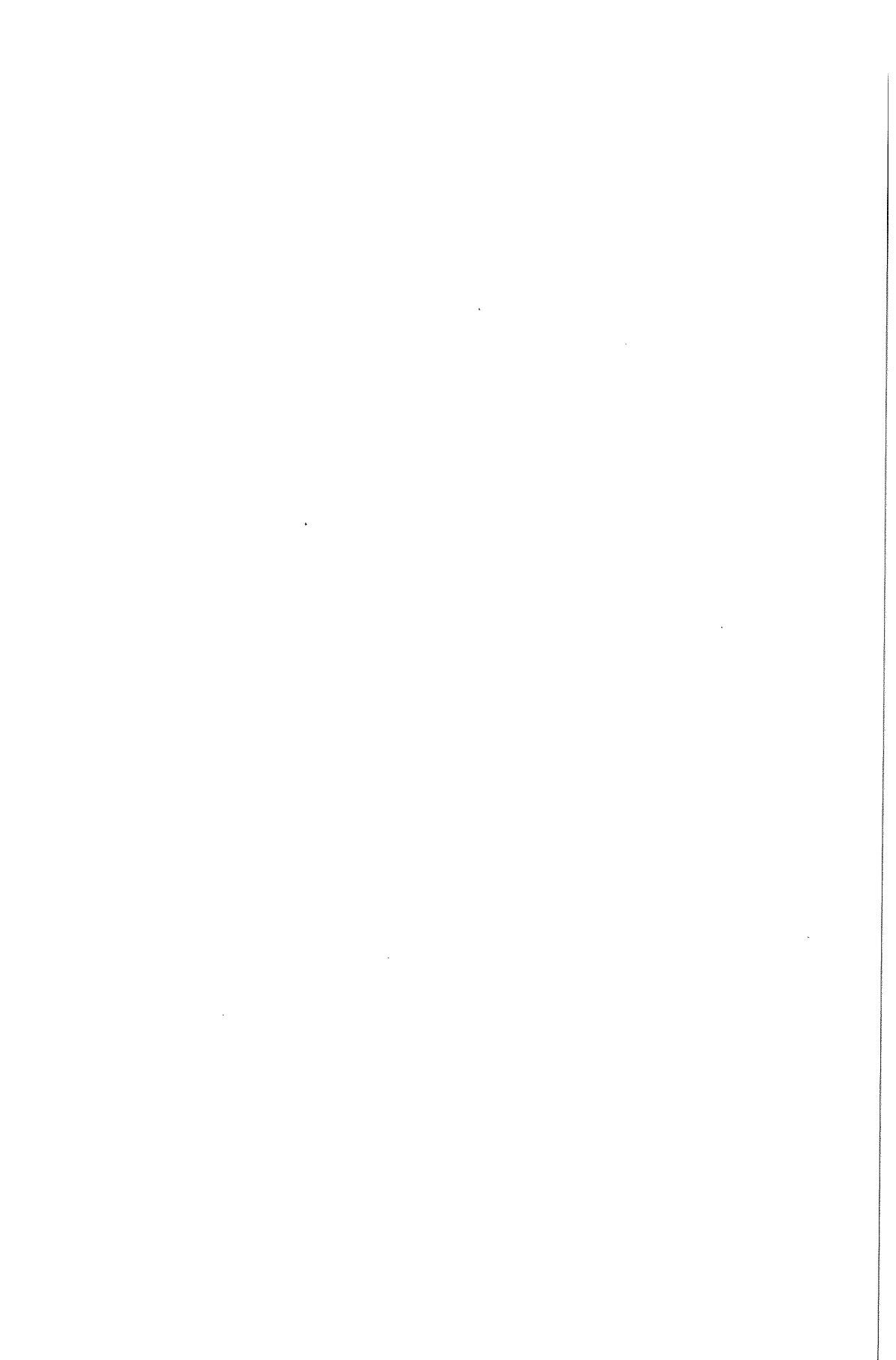
O ficheiro "MATROD" contem a matriz origem/destino dos deslocamentos.

O ficheiro "PRCHT" contem os horários genéricos de cada comboio (número de ordem da circulação, terminus de partida e instante, terminus de chegada e instante).

O ficheiro "VIATDEPi" indica o número de comboios que estão no terminus i ao longo da simulação, registando os instantes de saída dos comboios e de colocação dos comboios em terminus em condições de iniciar circulação.



Os horários detalhados são produzidos com um programa chamado "horar", que utiliza os elementos produzidos na simulação. Para os obter basta digitar no terminal "horar" e aguardar cerca de 5 minutos. Os ficheiros HORAR*i* (i número de linha), contêm, para cada linha e cada comboio, os instantes de chegada e de partida em cada estação.



# UMA APLICAÇÃO DE COMBINAÇÃO DE PREVISÕES SUBJETIVAS PARA O MERCADO INTERNACIONAL DE PETRÓLEO - ABORDAGEM BAYESIANA

R. C. Souza

DEE, PUC/RJ  
Rua Marquês de São Vicente, 225

A. L. B. Squadri

PESOP, PETROBRÁS  
Av. Chile, 65 - Rio de Janeiro, Brasil

## Abstract

Since 1984 Petrobrás, by its Operational Research Group, has been looking for the utilization of the oil products traders personal information in the generation of price forecasts to the international market. A design was developed to the business Department with tools from the classical statistics.

Then in 1986, with the experience of the above design and the reading of Bayesian papers about combining probability distributions, Petrobrás has developed a new design to the international petroleum market which is the subject of this paper.

The choosen model was that developed by Winkler (1981), which we programmed in SAS with a original expert selection criterium. On a second step it will be developed the calibration module based on Morris papers.

## Resumo

Desde 1984 a Petrobrás, através de uma equipe de pesquisa operacional, vem pesquisando o aproveitamento das informações pessoais de seus negociadores de derivados de petróleo na geração de previsões quantitativas de preços no mercado internacional de derivados de petróleo.

Um projeto foi desenvolvido para o Departamento Comercial, utilizando as ferramentas da estatística clássica.

Em 1986, acumulando a experiência do acompanhamento do projeto citado e a leitura da literatura Bayesiana voltada para a solução do problema de combinação de julgamentos, a Petrobrás desenvolveu um novo projeto para a área de petróleo, o qual é o objetivo desse trabalho.

O modelo escolhido foi o apresentado por Winkler (1981), implementado em SAS e com critério próprio de seleção de "experts". Numa segunda fase será desenvolvido o módulo de calibração dos "experts" baseado nos trabalhos de Morris.

## 1 - INTRODUÇÃO

A partir de 1983 a divisão de pesquisa operacional (PESOP) estruturou um grupo de trabalho na área de previsão de séries economicas (consumo, preços internacionais, etc). A primeira experiência desse grupo com séries de preço do mercado internacional de derivados de petróleo reforçou mais idéias iniciais de que as previsões dessas séries deveriam, de alguma forma, incorporar a opinião de "experts" na área, "traders", analistas, etc.

Algumas dificuldades tiveram de ser contornadas na modelagem dessas séries a partir da abordagem "clássica", de combinação de previsão Granger [6]. Uma análise mais aprofundada mostrou que tais dificuldades eram consequência da inadequacidade dessa abordagem.

Em 1986 à PESOP foi encomendado um projeto de previsão de preços de petróleo no mercado internacional. Na busca de um modelo adequado para a utilização da opinião dos "experts" escolhemos aquele desenvolvido por Winkler [15].

Foi desenvolvido um sistema, "on line", em SAS que além do modelo matemático contém algumas idéias do autor. Os resultados estão expostos no item 03.

## 2 - MODELO DE CONSENSO COM DEPENDÊNCIA

Suponha que a variável de interesse, a qual poderia ser um parâmetro de um modelo estatístico ou uma observação futura, é denominada  $\theta$ . Distribuições para  $\theta$  são acessadas por  $k$  "experts", e por conveniência é assumido que  $\theta$  é real e ilimitado e que todas as  $k$  funções de distribuição são contínuas. Então  $g_1, \dots, g_k$  são densidades sobre  $R$  as quais representam os julgamentos de  $k$  "experts" sobre  $\theta$ . Considerando a média de  $g_i = \mu_i$  uma boa estimativa pontual de  $\theta$ , definimos o erro de estimação como  $u_i = \mu_i - \theta$ , e então, a densidade de  $u_i$  para o "expert"  $i$  é  $g_i(\mu_i - u_i)$ ;  $i=1, 2, \dots, k$ .

No modelo desenvolvido aqui, qualquer dependência entre os "experts" é representada pela dependência entre os erros de estimação.

Seja  $f$  densidade de  $u=(u_1, u_2, \dots, u_k)^t$  onde  $t$  significa transposição. Antes dos "experts" serem consultados, ambos  $\theta$  e  $\mu=(\mu_1 \dots \mu_k)^t$  são variáveis aleatórias, e é suposto que  $u$  é "location-invariant" no sentido de que o conhecimento de  $\theta$  não muda  $f$ . Essa hipótese poderia ser relaxada ao custo de aumentar o encargo de acessar densidade (para cada  $\theta$ ) ou atrasar o acesso de  $f$  pelo tomador de decisão até que  $\mu$  seja observado.

Alternativamente o tomador de decisão pode trabalhar com uma variável diferente. Por exemplo, se  $v_i = \mu_i / \theta$  é invariante, então uma transformação  $w_i = \log v_i = \log \mu_i - \log \theta$  faz possível trocar-se  $u_i$  por  $w_i$  e usa-se o modelo aqui desenvolvido. Nesse sentido, o modelo é mais geral do que se possa parecer à primeira vista.

Dado  $f$ , as informações do  $k$  "experts" podem ser usadas para revisar a distribuição de  $\theta$  do tomador de decisão. Das hipóteses do modelo,  $f(\mu_1 - \theta, \dots, \mu_k - \theta)$  pode ser considerado como a função de verossimilhança, e a distribuição de consenso é a distribuição a posteriori de  $\theta$  com base em uma densidade a priori difusa, ou seja :

$$h(\theta | g_1, \dots, g_k, f) \propto f(\mu_1 - \theta, \dots, \mu_k - \theta)$$

isto porque

$$h(\theta | \mu) \propto h_0(\theta). \quad g(\mu | \theta), \quad \text{no caso} \quad h_0(\theta) \text{ é difusa e } g(\mu | \theta)$$

pode ser deduzida como segue :

$$G(\mu | \theta) = \Pr | \mu \leq \mu | \theta | = \Pr | u + e \leq \mu | \theta | = \Pr | u \leq \mu - e | \theta | = F(\mu - e | \theta) \quad (*)$$

$$g(\mu | \theta) = \frac{\partial G(\mu | \theta)}{\partial \mu} = f(\mu - e | \theta), \quad \text{onde } e = (1, 1, \dots, 1)$$

(\*) Como a função  $F$  não depende de  $\theta$  a relação é a mesma para todo  $\theta$ .

Se o tomador de decisão tem uma priori não difusa  $h_0(\theta)$  então a posteriori de interesse é

$$h(\theta | g_1, \dots, g_k, f) \propto f(\mu_1 - \theta, \dots, \mu_k - \theta). \quad h_0(\theta)$$

Desde que  $f$  é fundamental para o modelo de consenso aqui desenvolvido, sua determinação é um passo importante e todos os dados relevantes disponíveis devem ser utilizados com esse objetivo. Se os valores passados de  $\theta$  e as médias correspondentes  $\mu_1, \dots, \mu_k$  dos "experts" foram observadas, então os vetores  $\underline{u}$  devem trazer bastante informação. Se os vetores de erro ( $\underline{u}$ ) e o vetor de erro para a presente situação em mãos são julgados "exchangeable", então a distribuição de frequência dos vetores de erros

observados deve ser usada como uma estimativa de  $f$ . Quando nenhum dado é disponível o tomador de decisão deve acessar  $f$  de maneira subjetiva.

Incertezas sobre  $f$  podem ser incluídas no modelo de consenso em termos de uma distribuição a priori de  $f$ .

O tomador de decisão deve sempre ter em mente o fato de que evidências empíricas indicam que indivíduos algumas vezes exibem vícios em suas probabilidades acessadas. Com a experiência, ele deve descobrir se seu "expert" é otimista ou pessimista em alguns aspectos e temperar (calibrar) seus julgamentos.

Afortunadamente, a calibração de um "expert" pode ser conseguida ajustando-se  $g_i$  apropriadamente e, então, tratando a função  $g_i$  como se essa tivesse sido acessada pelo "expert". Morris mostra que a calibração da densidade  $g_i$  pode ser realizada pela multiplicação desta por uma função de calibração  $c_i(\theta)$  e discute o "acessamento" de  $c_i$  a partir de dados passados ou julgamentos subjetivos. De Groot desenvolveu um modelo para auto-calibração e melhorar "acessamento" no tempo.

Para o modelo aqui considerado, se  $g_1, \dots, g_k$  representam densidades calibradas,  $f$  então representa a densidade conjunta dos erros de estimação determinados das distribuições calibradas.

**2.1 - Um Modelo Normal:**

A aplicação do modelo da seção 1 sem qualquer restrição na forma de  $f$  é bastante difícil. Aqui restringiremos  $f$  à família das densidades normais  $k$  - variadas com o vetor de média consistindo de  $k$  zeros e uma matriz de covariância positiva definida  $\Sigma$  consistindo de variâncias  $\sigma_i^2$  para  $i=1, k$  e covariâncias  $\sigma_{ij}$ ;  $i, j=1, k$ ;  $i \neq j$ . Isso nem sempre será uma distribuição apropriada na prática, mas a "teoria normal dos erros" sugere que a normalidade pode resultar em boa aproximação em muitos casos, e que calibração pode ser usada para termos vetor de média igual a zero.

A distribuição de  $\theta$  do "expert"  $i$  é aproximada por uma normal média  $\mu_i$  e variância de  $\sigma_i^2$  onde uma calibração pode ter ajustado  $\mu_i$  e/ou  $\sigma_i^2$  acessado pelo "expert"  $i$ .

Se  $\sigma_i^2 < \sigma_j^2$  então deve ser assumido que o "expert"  $i$  tem mais informação que  $j$ , a qual se reflete por uma estimativa mais precisa.

As implicações do modelo normal para consenso será investigado para dois casos:  $\Sigma$  conhecido e  $\Sigma$  desconhecido.

**2.1.1 - Consenso com  $\Sigma$  Conhecido**

Se a densidade priori para  $\theta$  é uma difusa imprópria e  $u$  é multinormal com média  $(0, \dots, 0)^t$  e matriz de covariância  $\Sigma$  então a densidade posteriori para  $\theta$  é

$$h(\theta | \mu) \propto \phi [(\theta - \mu^*) / \sigma^{*2}]$$

onde  $\phi$  é uma normal standard e

$$\mu^* = e^t \Sigma^{-1} \mu / e^t \Sigma^{-1} e$$

$$\sigma^{*2} = 1 / e^t \Sigma^{-1} e$$

$$e = (1 \dots 1)^t$$

para mostrar isso, note que tem termos de  $\theta$

$$f(\mu - \theta e) \propto \exp \{ -(\mu - \theta e)^t \Sigma^{-1} (\mu - \theta e) / 2 \}$$

completando o quadrado em  $\theta$ :

$$i) -(\mu - \theta e)^t \Sigma^{-1} (\mu - \theta e) =$$

$$= -[\mu^t \Sigma^{-1} \mu - \theta e^t \Sigma^{-1} \mu - \mu^t \Sigma^{-1} e \theta + e^t \Sigma^{-1} e \theta^2]$$

$$ii) -(\theta - \mu^*)^2 / \sigma^{*2} =$$

$$= -[\theta^2 e^t \sum_A^{-1} e - 2\theta e_B^t \sum^{-1} \mu + \frac{(e^t \sum^{-1} \mu)^2}{e^t \sum^{-1} e}]$$

de i) e ii) temos :

$$f(\mu - \theta e) \propto \exp[-(\theta - \mu^*)^2 / 2\sigma^{*2}]$$

A média da distribuição de consenso,  $\mu$ , é uma combinação linear, não necessariamente convexa, de  $\mu_1, \mu_2, \dots, \mu_k$ , com pesos :

$$(\omega_1 = \sum_{j=1}^k \alpha_{ij} \sum_{m=1}^k \sum_{j=1}^k \alpha_{mj}), \sum^{-1} = (\alpha_{ij}).$$

Por exemplo, se  $k=2$  temos:

$$\mu^* = [(\sigma_2^2 - \rho \sigma_1 \sigma_2) \mu_1 + (\sigma_1^2 - \rho \sigma_1 \sigma_2) \mu_2] / (\sigma_1^2 + \sigma_2^2 - 2\rho \sigma_1 \sigma_2)$$

Assumindo, sem perda de generalidade, que  $\sigma_1 < \sigma_2$ , temos que  $\mu_1$  sempre receberá peso positivo mas o peso para  $\mu_2$  é negativo para  $\rho > \sigma_1/\sigma_2$ . Dessa forma uma alta correlação faz com que provavelmente  $\mu_1$  e  $\mu_2$  estejam do mesmo lado de  $\theta$ , e como o "expert" 1 é mais preciso que 2 dá-se um peso negativo para aproximar o resultado para  $\theta$ .

Embora  $\rho > \sigma_1/\sigma_2$  seja possível para o caso geral, isso não poderá ocorrer para observadores independentes com "overlapping" na informação.

Para um exemplo numérico com  $k=3$ , suponha que o investidor consulte três analistas de mercado de capital, que acessam distribuição para  $\theta$ , o preço de uma particular ação daqui a um ano. A distribuição acessada pode ser aproximada por distribuições normais com  $\mu_1 = 60$ ,  $\sigma_1^2 = 36$ ,  $\mu_2 = 62$ ,  $\sigma_2^2 = 25$ ,  $\mu_3 = 70$  e  $\sigma_3^2 = 49$ . O investidor acha que os três analistas são bem calibrados e, baseado na experiência passada com os três analistas, acessa correlações entre os erros de previsão  $\rho_{12}=0.6$ ,  $\rho_{13}=0.5$  e  $\rho_{23}=0.6$ . Aqui  $\mu = (60, 62, 70)$  e

$$\Sigma = \begin{bmatrix} 36 & 18 & 21 \\ 18 & 25 & 21 \\ 21 & 21 & 49 \end{bmatrix}$$

pela fórmula de  $\mu^*$  e  $\sigma^{*2}$  temos:

$\mu^*=60, 2$  e  $\sigma^{*2}=22, 83$  e os pesos assinalados para  $\mu_1, \mu_2$  e  $\mu_3$  são respectivamente 0.263, 0.669 e 0.068. Se a dependência entre os erros de previsão é ignorada na combinação das distribuições dos 3 analistas, a distribuição de consenso é normal com média = 63.22 e variância  $\sigma^{*2}=11.34$ , e os pesos dados para  $\mu_1, \mu_2$  e  $\mu_3$  são 0.315, 0.454 e 0.231 respectivamente.

Então, ignorando-se a dependência temos um peso maior para o terceiro analista e uma muito menor variância (porque independência implica em não coincidência de informação).

Quando  $k$  cresce, o número de parâmetros de  $\Sigma$  cresce rapidamente e torna-se difícil expressar  $\Sigma^{-1}$  em termos desses parâmetros.

Objetivando estudar o impacto da dependência, é conveniente supor alguma estrutura especial para  $\Sigma$ . Por exemplo, suponha que  $\rho_{ij} = \rho$  para  $i, j = (1, \dots, k)$ . Agora  $\Sigma^{-1}$  pode ser encontrada via teorema da Binomial Inversa e

$$\mu^* = \frac{\sum_{i=1}^k \{ [1 + (k-1) \rho] \sigma_i^{-2} - \rho \sigma_i^{-1} \sum_{j=1}^k \sigma_j^{-1} \} \mu_i}{\{ [1 + (k-1) \rho] \sum_{i=1}^k \sigma_i^{-2} - \rho (\sum_{j=1}^k \sigma_j^{-1})^2 \}}$$

e

$$\sigma^{*2} = (1-\rho) \quad [1 + (k-1) \rho] / \{ [1 + (k-1) \rho] \sum_{i=1}^k \sigma_i^{-2} - \rho (\sum_{j=1}^k \sigma_j^{-1})^2 \}$$

Aqui os pesos não são necessariamente iguais, e  $\omega_1 < 0$  se  $\sigma_i^{-1} < \sum_{j \neq i} \sigma_j^{-1} (k-2+\rho^{-1})$  o

que implica que se  $\rho \rightarrow 1$ , a precisão do "expert"  $i$  é menor que a precisão média dos outros  $k-1$  "experts".

Muitos trabalhos citados na bibliografia envolvem distribuições de consenso que são misturas de distribuições de "experts".

Tal procedimento levaria a uma distribuição de consenso que seria uma mistura das três distribuições normais ao invés de uma única distribuição normal.

OBS: Se duas variáveis aleatórias são marginalmente normais mas não conjuntamente normais então a soma delas não é necessariamente normal.

Uma mistura pode levar a inferirmos resultados consideravelmente diferentes sobre  $\theta$  daqueles atingidos via o procedimento de consenso aqui desenvolvido. Por exemplo: o investidor pode estar interessado em probabilidades tais como  $P(\theta > 70)$ . Essa probabilidade é 0.05 da distribuição de consenso levando-se em conta a dependência: 0.02 da distribuição de consenso supondo independência ; 0.08 de uma mistura com pesos 0.263, 0.669 e 0.068 (pesos usados na determinação de  $\mu^*$  no caso com dependência) ; e 0.20 de uma mistura com pesos iguais. Além disso, se o tomador de decisão tem uma distribuição priori normal para  $\theta$ , com média  $\mu_0$  e variância  $\sigma_0^2$ , então a distribuição a posteriori para  $\theta$  é normal com média  $(\sigma_0^{-2} \mu_0 + \sigma^{*-2} \mu^*) / (\sigma_0^{-2} + \sigma^{*-2})$  e variância  $(\sigma_0^{-2} + \sigma^{*-2})^{-1}$ .

### 2.2 - Consenso com $\Sigma$ Desconhecido:

A hipótese que  $\Sigma$  é conhecido é bastante forte. Entretanto, julgamentos subjetivos e dados passados podem ser usados para determinar uma priori para  $\Sigma$ . A distribuição de Wishart invertida, a qual é uma priori conjugada natural para a matriz de covariância de um processo normal, pode ser considerada uma aproximação satisfatória para a distribuição de  $\Sigma$ . Uma independência a priori de  $\Sigma$  e  $\theta$  parece razoável, desde que  $\Sigma$  descreve o processo de geração de erro e não o processo gerador de  $\theta$ .

Supondo que  $\theta$  e  $\Sigma$  tenham uma densidade difusa a priori e uma Wishart invertida respectivamente.

$$p(\Sigma) \propto |\Sigma^{-1}|^{(\delta_0 + 2k)/2} \exp \{ -\delta_0 \text{tr} [\Sigma^{-1} \Sigma_0] / 2 \text{ com } \delta_0 > 0 \text{ e } \Sigma_0$$

positiva definida e simétrica. Se  $u$  é normal com média  $(0, \dots, 0)^t$  e matriz covariância desconhecida  $\Sigma$ , então a densidade posterior para  $\theta$  é:

$$h(\theta | \mu) \propto [1 + (\theta - m^*)^2 / (\delta_0 + k - 1) \delta^{*2}]^{-(\delta_0 + k)/2}$$

onde

$$m^* = e^t \Sigma_0^{-1} \mu / e \Sigma_0^{-1} e ;$$

$$\delta^{*2} = [\delta_0 + (m^* e - \mu)^t \Sigma_0^{-1} \mu] / (\delta_0 + k - 1) e^t \Sigma_0^{-1} e$$

Aqui a distribuição de consenso é uma  $t$  com  $\delta_0 + k - 1$  graus de liberdade, média  $m^*$  e variância  $(\delta + k - 1) s^{*2} / (\delta + k - 3)$ . Se a distribuição priori do tomador de decisão para  $\theta$  é normal ou  $t$ , a distribuição posteriori é poly- $t$  e pode ser estudada numericamente. Demonstra-se o resultado, para priori difusa do tomador de decisão, utilizando a seguinte relação:

$$h(\theta, \Sigma/\mu) \propto h_0(\theta) \cdot h_1(\Sigma/\theta) \cdot f(\mu/\theta, \Sigma) \text{ ou}$$

$$h(\theta, \Sigma/\mu) \propto h_0(\theta) \cdot p(\Sigma) \cdot f(\mu/\theta, \Sigma)$$

Naturalmente, dados disponíveis das situações passadas podem ser utilizados para a fixação de  $\Sigma_0$ . A distribuição de  $\Sigma$  pode ser revisada assim que novos dados estiverem disponíveis.

### 2.3 - Comentário:

O modelo de consenso investigado aqui permite formalmente a dependência entre "experts", e os resultados obtidos demonstram que a distribuição posteriori seguida das informações dos "experts" é bastante sensível ao grau de dependência entre eles.

### 3 - APLICAÇÃO:

O Departamento Comercial da Petrobrás solicitou à Divisão de Pesquisa Operacional um projeto para a previsão das séries de preços médios semanais dos seguintes petróleos: Bonny (Nigéria), Brent (Mar do Norte), Dubai (Arábia Saudita), Ural (URSS) e WTI (USA).

Escolhida a abordagem apresentado no item 2.2, a qual é tema da tese de mestrado de Squadri, este desenvolveu um sistema, "on line", em SAS que a partir das opiniões dos "experts" fornece previsões de preços médios semanais para horizontes de 4,8 e 12 semanas.

Para cada previsora (6 ao todo), petróleo e horizonte, são gerados quadros e gráficos conforme anexo I. Isso permite que os "experts" possam se "calibrar", evitando erros sistemáticos.

O quadro I apresenta os erros médios absolutos percentuais dos "experts" e do modelo para o horizonte igual a 8 semanas (para os outros horizontes ainda não há resultado disponível).

Mesmo o número de erros observados pequeno ( $N=8$ ), o resultado é aparentemente decepcionante. Contudo uma análise do Quadro II aponta para o esclarecimento da situação. No período anterior às previsões do modelo, os analistas D e E foram os melhores previsores e durante o período em análise no Quadro I foram os escolhidos para compor a previsão combinada. Para o "azar" do modelo os outros analistas melhoraram seu desempenho, enquanto D e E pioraram.

### 4 - CONCLUSÃO:

Do que foi exposto nos itens anteriores podemos concluir:

- a) até 22/04/87 o sistema de previsão de preços médios de petróleo apresenta um número muito pequeno de erros observados, impossibilitando uma análise acurada.
- b) o baixo desempenho do sistema de previsão no período em análise deve-se basicamente à curva de aprendizado dos "experts".
- c) a abordagem bayesiana apresentada é adequada pois contempla a incorporação de opiniões geradas em ambientes com muitas informações comuns.

### BIBLIOGRAFIA :

- 1 - Bates, J.M. e Granger, C.W.J. (1969) - "The Combination of Forecasts", Oper. Res. Q.20.
- 2 - De Groot, M.H., "Reaching a Consensus" J. Amer. Statist. Assoc. VI. 69 (1974). pp.118-121.
- 3 - De Groot, M.H., "Improving Predictive Distributions", Technical Report nº 159, Department of Statistics, Carnegie-Mellon University, Pittsburg, 1979.
- 4 - Drése, J.H., "Bayesian Regression Analysis Using Poly-t Densities", CORE



- Discussion Paper 7613, University Catholique de Louvin, Heverlee, Belgium, 1976.
- 5 - Geisser S., "A Bayes Approach for Combining Correlated Estimates", J. Amer. Statist. Assoc. Vol. 60 (1965), pp.602-607.
  - 6 - Granger, C.W.J. e Newbold, Paul (1977) - "Forecasting Economic Time Series", Academic Press.
  - 7 - Hogartii, R.M., "Cognitive Processes and the Assignment of Subjective Probability Distributions", J. Amer. Statist. Assoc., Vol. 70 (1975), pp.271-294.
  - 8 - Hogartii, R.M., "Methods for Aggregaing Opinions", in Decision Making and Change in Human Affairs, eds H. Jungermann and G. de Zeeuw, Reidie, Dordrecht, 1977, pp.231-235.
  - 9 - Hotelling, H. (1933) - "Analysis of a Complex of Statistical Variables in to Principal Components", Journal of Educational Psychology.
  - 10 - Johnson, R.A. e Wichern, D.W. (1982) - "Apllied Multivariate Statistical Analysis", Prentice-Hall.
  - 11 - Morris, P.A., "Decision Analysis Expert Use", Management Sci, Vol. 20 (1974), pp.1233-1241.
  - 12 - Morris, P.A., "Combinig Expert Judgements: A Bayesian Approach", Management Sci, Vol. 23 (1977), pp.679-693.
  - 13 - Pree, S.J., "Applied Multivariate Analysis", Holt, Rinihart and Witson, New York 1972.
  - 14 - Morris, P.A., "Combining Probability Distributions from Dependent Information Sources", Management Sci, Vol. 27 (1981), pp.479-488.

ANEXO I  
RELATÓRIO DO SISTEMA DE PREVISÃO DE PREÇO DE PETRÓLEO  
ACOMPANHAMENTO DE PREVISÕES DE PREÇOS  
PETRÓLEO BRENT  
HORIZONTE 8 SEMANAS  
PREVISOR MODELO

de	para	minfmo	máximo	esperado	observado	intervalo?	er%	er.abs.
15/10/86	10/12/86	-	-	-	-	-	-	-
22/10/86	17/12/86	-	-	-	14.95	-	-	-
29/10/86	24/12/86	-	-	-	15.82	-	-	-
05/11/86	31/01/87	-	-	-	16.99	-	-	-
12/11/86	07/01/87	-	-	-	17.89	-	-	-
19/11/86	14/01/87	-	-	-	18.21	-	-	-
26/11/86	21/01/87	-	-	-	18.71	-	-	-
03/12/86	28/02/87	-	-	-	18.50	-	-	-
10/12/86	04/2/87	-	-	-	18.33	-	-	-
17/12/86	11/02/87	-	-	-	18.06	-	-	-
24/12/86	18/02/87	-	-	-	17.65	-	-	-
31/12/86	25/03/87	-	-	-	17.16	-	-	-
07/01/87	04/03/87	-	-	-	16.34	-	-	-
14/01/87	11/03/87	-	-	-	16.93	-	-	-
21/01/87	18/03/87	-	-	-	17.79	-	-	-
28/01/87	25/03/87	-	-	-	18.23	-	-	-
04/02/87	01/04/87	15.62	20.50	18.06	18.38	-	-	-
11/02/87	08/04/87	16.15	20.79	18.47	18.48	SIM	0.42	2.27
18/02/87	15/04/87	-	-	-	19.08	SIM	0.61	3.27
25/02/87	22/04/87	-	-	-	17.71	-	-	-
04/03/87	29/04/87	-	-	-	18.10	-	-	-
11/03/87	06/05/87	14.05	19.98	17.02	-	-	-	-
18/03/87	13/05/87	14.47	20.23	17.35	-	-	-	-
25/03/87	20/05/87	15.22	20.81	18.01	-	-	-	-
01/04/87	27/06/87	15.60	21.19	18.39	-	-	-	-
08/04/87	03/06/87	14.63	20.20	17.41	-	-	-	-
15/04/87	10/06/87	15.27	20.85	18.06	-	-	-	-
22/04/87	17/06/87	15.94	21.64	18.79	-	-	-	-

ANEXO I (continuação)

QUADRO I  
DESEMPENHO DOS EXPERTS E DO MODELO  
(ERRO MÉDIO ABSOLUTO PERCENTUAL) (\*)

	EXPERT A	EXPERT B	EXPERT C	EXPERT D	EXPERT E	MODELO
BONNY	3.73	5.30	4.19	4.69	7.64	9.14
BRENT	4.67	6.17	4.44	5.80	6.64	6.76
DUBAI	1.16	3.57	1.55	1.89	3.45	3.49
URAL	4.27	5.21	2.74	4.54	5.06	4.93
WTI	3.72	4.62	4.36	4.44	7.48	5.98

QUADRO II  
DESEMPENHO DOS "EXPERTS" PARA O PETRÓLEO BRENT  
(ERRO MÉDIO ABSOLUTO PERCENTUAL)

N	EXPERT	ANTES DE 14/01/87 (%)	DEPOIS DE 14/01/87 (%)
13	A	9.69	4.67
13	B	8.09	6.17
13	C	8.41	4.44
13	D	5.63	5.80
13	E	5.10	6.64

(\*) Previsões nas semanas entre 14/01/87 a 11/02/87 e 11/03/87 a 25/03/87.

## AMOSTRAGEM DESCRITIVA: UMA NOVA ABORDAGEM EM SIMULAÇÃO

Eduardo Saliby  
(COPPEAD/UFRJ)

### Resumo

O uso da amostragem descritiva, em lugar da amostragem aleatória simples, representa uma grande mudança conceitual na simulação por Monte Carlo. A amostragem descritiva caracteriza-se por uma seleção totalmente determinística dos valores amostrais e na permutação aleatória; sendo assim, ela implica na modificação na ideia de que uma simulação deva reproduzir integralmente o comportamento probabilístico de um sistema.

Este trabalho tem por objectivo apresentar: uma síntese dos desenvolvimentos teóricos e empíricos que nos levaram a questionar o uso da amostragem aleatória simples em simulação, um procedimento geral para a geração de amostras descritivas e um resumo dos principais testes já realizados.

Os resultados obtidos confirmaram a maior eficiência estatística da amostragem descritiva. Além disso, foi constatado que o tempo de processamento normalmente se reduz. Desta forma, a amostragem descritiva mostra-se bastante útil em aplicações práticas da simulação.

### Abstract

The use of descriptive sampling, instead of simple random sampling, represents an important conceptual change in Monte Carlo simulation. Descriptive sampling is based on a fully deterministic selection of the sample values and their random permutation; therefore, it departs from the basic idea that a simulation has to entirely reproduce the probabilistic random behaviour of a real system.

This work presents: an outline of the theoretical development and empirical results that conduced us to question the use of simple random sampling in simulation; a general procedure for descriptive sampling generation; and a summary of the main empirical results already gathered.

All results do confirm the higher statistical efficiency of descriptive sampling. Even more, the processing time is usually lower. Therefore, descriptive sampling seems to be very useful in practice.

## 1. INTRODUÇÃO

Como uma técnica de pesquisa operacional (P.O.), a simulação caracteriza-se pela experimentação numérica com modelos lógico-matemáticos. Embora ela se aplique a outras situações, vamos subentendê-la como uma técnica para o estudo de problemas probabilísticos; neste caso, que corresponde ao seu enfoque usual em P.O., ela é também denominada simulação por Monte Carlo.

Esta denominação decorre do uso da amostragem aleatória simples como sua abordagem padrão, fazendo da simulação um método impreciso. Por este motivo, muitos autores apresentam-na como o último recurso a ser tentado na solução de um problema.

Embora concordemos com esta postura em relação à simulação, a maior complexidade dos problemas com que nos deparamos faz deste último recurso o único disponível. Não é pois, sem razão, que a simulação é hoje uma das técnicas de P.O. mais utilizadas.

Existem dois caminhos para se contornar suas deficiências:

- a) Aumentando a eficiência computacional da simulação.
- b) Aumentando a eficiência estatística da simulação.

Sem subestimarmos a importância do primeiro caminho, voltamos nossa atenção ao problema da eficiência estatística. Após estudarmos as fontes de variação das estimativas de simulação (Saliby, 1980 e 1986), constatamos que aumentaríamos a sua precisão pela simples substituição da amostragem aleatória simples pela amostragem descritiva (Saliby, 1980 e 1982b). A amostragem descritiva representa uma importante mudança conceitual em simulação, uma vez que se baseia numa seleção totalmente determinística dos valores amostrais.

Apresentamos aqui uma síntese dos estudos que resultaram na amostragem descritiva e um resumo dos principais testes comparativos já realizados.

Nosso ponto de partida será a formulação de um problema de simulação.

## 2. O PROBLEMA DE SIMULAÇÃO

No contexto do presente trabalho, uma simulação tem por objetivo estimar parâmetros que descrevam o comportamento probabilístico de um sistema. Estas estimativas são calculadas a partir de uma amostra de valores para cada uma das variáveis de resposta do problema; por simplicidade de apresentação, vamos supor que ela é única.

A distribuição de probabilidades da variável de resposta, que é desconhecida, é caracterizada pelos parâmetros

$$\Theta_j, \quad j = 1, \dots, L$$

Definiremos uma corrida de simulação como sendo o esforço computacional que origina uma estimativa para cada parâmetro em estudo.

Também por simplicidade de apresentação, vamos supor que temos uma única variável de entrada,  $X$ . Neste caso, uma corrida de simulação equivale à avaliação numérica de um conjunto de funções

$$y_j = F_j(X_1, \dots, X_n), \quad j = 1, \dots, L,$$

onde,

$y_j$ ,  $j = 1, \dots, L$ , é a estimativa de  $\Theta_j$ ,

$x_i$ ,  $i = 1, \dots, n$ , é a amostra de entrada, e

$F_j$ ,  $j = 1, \dots, L$ , é a função que define as estimativas a partir da amostra de entrada.

Esta função é normalmente descrita por um programa de computador.

Com base nesta formulação, uma corrida de simulação pode ser vista como composta de duas etapas:

- a) a seleção de um ponto  $R^n$ , correspondente à amostra de entrada, e
- b) a avaliação numérica do conjunto de funções que definem cada uma das estimativas.

Qualquer que seja o método de amostragem utilizado numa simulação, faremos uso de corridas replicadas para avaliar a variância das estimativas. Trata-se de um método com aplicação geral que corresponde a gerar, de forma independente, um diferente ponto de avaliação a cada corrida. Um conjunto de  $M$  ( $M \geq 1$ ) corridas independentes definirá um experimento.

### 3. AS FONTES DE VARIAÇÃO DAS ESTIMATIVAS

Uma amostra de entrada pode ser decomposta em:

- a) um conjunto de valores, independente de sua ordem de ocorrência, e
- b) a sequência em que estes valores são utilizados.

No caso da amostragem aleatória simples, ambos os fatores são gerados aleatoriamente. Assim sendo, ela equivale a:

- a geração aleatória de um conjunto de valores, e
- a geração aleatória da permutação destes valores.

Um estudo da contribuição de cada um destes fatores na variabilidade das estimativas (Saliby, 1980 e 1986) permitiu constatar que:

- a) Uma característica comum à maioria dos problemas de simulação é que, no caso da amostragem aleatória simples, as propriedades de conjunto da amostra de entrada respondem por uma substancial parcela da variação das estimativas. A esta parcela denominou-se efeito de conjunto.
- b) O efeito de um conjunto pode ser explicado, em sua quase totalidade, através de uma equação de regressão denominada Modelo Linear de Resposta. Esta regressão representa uma extensão da ideia das variáveis de controle, uma das técnicas de redução de variância mais eficientes.
- c) Quanto à sequência, sua contribuição foi isolada apenas em alguns casos particulares. Ao contrário do que ocorreu com efeito de conjunto, não foi possível identificar um modelo que explicasse sua influência.

### 4. DA AMOSTRAGEM ALEATÓRIA SIMPLES À AMOSTRAGEM DESCRITIVA

O estudo das fontes de variação mostrou que, ao utilizarmos a amostragem aleatória simples, as estimativas de uma simulação terão boa parte de sua variabilidade explicada em função dos desvios entre os momentos da amostra e seus respectivos valores teóricos. A parcela de variância explicada pelo efeito de conjunto é muitas vezes da ordem de 50% ou superior; além disso, a experiência mostra que ela independe do tamanho da corrida.

Enquanto a técnica das variáveis de controle busca ajustar as estimativas em função dos desvios entre a amostra e a respectiva distribuição teórica, seguimos um caminho oposto: procuramos eliminar estes desvios na sua própria origem.

Uma vez que temos, em teoria, um total conhecimento a respeito das distribuições das variáveis de entrada, não há qualquer inferência a ser feita acerca dos seus parâmetros. Desta forma, não é para este fim que utilizamos a amostragem aleatória numa simulação, mas sim o interesse em representarmos um comportamento probabilístico.

Isso, no entanto, poderemos fazer de forma mais eficiente sem recorrer à amostragem aleatória simples. De facto, sendo conhecido o tamanho da amostra, poderemos selecionar deterministicamente os seus valores de modo a descreverem, com a máxima precisão possível, a distribuição de probabilidades representada. Por outro lado, somente a sequência destes valores continuaria sendo gerada aleatoriamente. A este método denominamos amostragem descritiva.

A amostragem descritiva objetiva eliminar, ou pelo menos minimizar, a variação de conjunto associada às amostras de entrada. Desta forma, as estimativas variarão unicamente em função das diferentes sequências geradas, sendo mais precisas.

### 5. GERAÇÃO DE AMOSTRAS DESCRITIVAS

A principal e única diferença entre a amostragem aleatória simples e a amostragem descritiva reside na seleção dos valores amostrais: nesta última, totalmente determinística e intencional. Em ambos os casos, no entanto, a sequência dos valores é aleatória.

A única exigência adicional para a geração de amostras descritivas é que  $n$ , o tamanho da amostra, seja previamente conhecido. Na prática, há muitas situações em que isto não ocorre; nestes casos, desde que se tenha um conhecimento aproximado de  $n$ , a amostragem descritiva poderá ser ainda utilizada, porém não mais resultando num controle total da variação de conjunto.

A geração de uma amostra descritiva consiste de duas etapas, cuja ordem de execução é teoricamente irrelevante:

- a) Geração do conjunto de valores descritivos, e
- b) Geração de uma permutação aleatória:

Geração dos valores descritivos

Sendo

$$xd_i, \quad i = 1, \dots, n,$$

um conjunto ordenado de valores, pode-se mostrar que para

$$xd_i = F^{-1} \left[ (i - 0.5) / n \right], \quad i = 1, \dots, n,$$

onde  $F^{-1}(R)$  é a inversa da função de distribuição, as discrepâncias entre as características de conjunto da amostra e da população são minimizadas.

Desta forma, tal como ocorre com a amostragem aleatória simples, dispomos de um método geral para a geração de amostras descritivas. Basta que, em lugar de um conjunto aleatório de valores uniformes, utilizemos como argumento da transformada inversa um conjunto descritivo de valores uniformes:

$$ud_i = (i - 0.5) / n, \quad i = 1, \dots, n.$$

Os seguintes pontos merecem atenção:

- O tempo de processamento para a geração de um conjunto descritivo é aproximadamente o mesmo da amostragem aleatória simples, desde que também utilizando a transformada inversa.
- A transformada inversa é válida tanto para a amostragem aleatória simples como para a amostragem descritiva. No entanto, esta dupla utilidade não se estende aos demais métodos para a geração de amostras aleatórias.
- Frequentemente, não se dispõe de uma expressão analítica para a inversa da função de distribuição. Nestes casos poderemos utilizar fórmulas aproximadas ou utilizar métodos numéricos de integração (Almeida, 1983).

Geração de permutações aleatórias

A geração de permutações aleatórias, embora também mereça atenção, consiste num problema relativamente mais simples.

O procedimento normalmente utilizado equivale a um sorteio de bingo onde  $n$  bolas numeradas de 1 a  $n$ , são embaralhadas e retiradas de uma urna.

Os passos envolvidos na sua geração são os seguintes:

- Inicializar um vetor de índices, fazendo
 
$$P(i) = i, \quad i = 1, \dots, n.$$
- Embaralhar os índices:
 

Para  $k = 1, \dots, n-1$

  - Gerar ao acaso um valor inteiro  $iaux \in [k, n]$ ;
  - Trocar os valores de  $P(k)$  com  $P(iaux)$ ;
  - Incrementar  $k$  e repetir o processo a partir de "b1";

Fim da repetição.

- Ao término do processo, o vetor  $P(i)$ ,  $i = 1, \dots, n$ , conterà os índices já permutados.

Os seguintes pontos merecem atenção:

- O processo de permutação encerra-se com a geração do penúltimo índice; com isso, o último ficará também determinado.
- O tempo de processamento é praticamente proporcional ao número de elementos permutados, ou seja, de ordem  $O(n)$ .
- Ao gerarmos uma permutação aleatória a cada corrida de um experimento, não precisamos reinicializá-la. Basta tomarmos a permutação da corrida anterior como ponto de partida para uma nova permutação.

A utilização de amostras descritivas

Replicações independentes com a amostragem descritiva correspondem ao uso de diferentes permutações a cada corrida, mantendo-se porém o mesmo conjunto de valores de entrada. Como o tempo de geração de uma permutação aleatória é geralmente menor que na geração de uma amostra aleatória simples, a amostragem descritiva resulta muitas vezes numa sensível redução do tempo total de processamento. Este ganho torna-se mais acentuado quanto maior for o número de corridas num experimento ( $M$ ).

**6. VERIFICAÇÃO EMPÍRICA**

Nosso objectivo é o de mostrar que a amostragem descritiva é estatisticamente mais eficiente do que a amostragem aleatória simples. Para isso, deveremos ter:

- Estimativas não-tendenciosas

Sendo o erro padrão das estimativas inversamente proporcional à raiz quadrada do tamanho da amostra de entrada, elas poderão ser tendenciosas desde que o termo de viés seja de magnitude inferior. Desta forma, é suficiente que as médias das estimativas - para ambos os métodos amostrais - não difiram significativamente, para que esta propriedade seja satisfeita.

b) Estimativas de menor variância.

Uma vez confirmada sua não-tendenciosidade, a redução de variância proporcionada pela amostragem descritiva, em relação à amostragem aleatória simples, define sua eficiência estatística.

Utilizando ambos os métodos amostrais, foi feito um grande número de corridas - da ordem de 100 para cada problema da tabela 1 - e, a partir delas, calculados os seguintes valores:

- a) Média global das estimativas, para todas as corridas de um experimento.
- b) Redução de variância com a amostragem descritiva.
- c) Coeficiente de determinação do Modelo Linear de Resposta.

Este coeficiente, calculado apenas para a amostragem aleatória simples, fornece uma idéia da parcela de variação associada ao efeito de conjunto.

Os resultados, sumariados na tabela 1, nos mostram que:

- a) Para ambos os métodos amostrais, nenhuma diferença significativa entre as médias globais das estimativas foi constatada.
- b) Para todas as estimativas observaram-se reduções de variância com a amostragem descritiva.
- c) Confirmando as previsões teóricas, as reduções de variância com a amostragem descritiva foram da mesma ordem de grandeza que a variação explicada pelo Modelo Linear de Resposta.
- d) O problema do jornaleiro é de particular interesse para nós. Ele se refere a uma situação em que o valor da estimativa independe da sequência da amostra. Assim sendo, ao fixarmos um conjunto de entrada, o resultado de uma corrida será totalmente determinado, resultado este que corresponde à solução exata do problema.

## 7. PONTOS ADICIONAIS E CONSIDERAÇÕES FINAIS

Embora se soubesse que um maior controle do processo de amostragem poderia produzir estimativas mais precisas, a idéia de que uma simulação deva ser uma imitação da realidade e de todas as suas incertezas, sempre prevaleceu. Assim, com exceção da amostragem seletiva (Brenner, 1963), da qual falaremos adiante, as demais técnicas de redução de variância não atuam sobre a variabilidade amostral dentro de uma corrida.

Até o momento, todo controle amostral restringiu-se apenas à variabilidade entre diferentes corridas, como no caso das variáveis antitéticas e das mesmas sequências de números aleatórias ("Common random numbers"). Desta forma, o princípio de que uma simulação é uma imitação da realidade foi sempre obedecido. Com a amostragem descritiva, no entanto, dele nos afastamos.

Em consequência, observamos normalmente uma reação inicial contrária à amostragem descritiva. De fato, o controle na seleção da amostra parece ser excessivo, o que poderia conduzir a estimativas distorcidas. Esta foi também nossa própria atitude inicial. No entanto, ao encontrarmos sua justificativa conceitual e com base nos novos resultados empíricos e teóricos obtidos, mudamos totalmente de opinião.

### Amostragem seletiva

Embora não houvesse sido claramente explicitada, a idéia de se eliminar a variação de conjunto numa simulação foi previamente sugerida por Brenner (1963), ao propor a amostragem seletiva.

No entanto, apesar de apontar na direção correta, o trabalho de Brenner careceu da necessária fundamentação teórica. De acordo com o autor, tratava-se de uma sugestão para se chegar a estimativas mais precisas, porém de aplicação restrita. Na falta de uma justificativa apropriada, esta idéia foi abandonada.

### Considerações finais

As técnicas de redução de variância que atuam sobre o processo de amostragem, e o princípio da imitação sempre estiveram em contradição. Mesmo com o controle da amostragem restrito a diferentes corridas, ainda assim o princípio da imitação estaria sendo violado.

TABELA 1  
 SUMÁRIO DOS RESULTADOS DOS TESTES COMPARATIVOS  
 ENTRE A AMOSTRAGEM ALEATÓRIA SIMPLES (A. A. S.)  
 E A AMOSTRAGEM DESCRITIVA (A. D.)

Problema	Estimativa	Média Global		Redução de Variância	RZ do MLK	Referência
		A.A.S.	A.D.			
Rede PERI	$\bar{D}_i$	11.214	11.201	63%	57%	(Saliby, 1987)
	S	1.863	1.899	10%	9%	
	$D_i$					
Fila (= 0.75)	$\bar{SIZE}$	2.991	2.875	71%	62%	(Saliby, 1980)
	S	3.235	3.178	45%	29%	
	SIZE					
Estoques	$\bar{COST}$	78.530	78.526	52%	48%	(Saliby, 1980) (ROF=50,EOQ=150)
	S	69.497	69.499	33%	36%	
	COST					
Jornaleiro	$\bar{LUCRO}$	344.760	344.600	100%	97%	(Saliby, 1981)
	S	82.245	82.128	100%	81%	
	LUCRO					
Inteoral múltipla	$\bar{7}$	1.079	1.070	64%	62%	(Ferreira Sobrinho, 1982; n = 128)
Frota de distribuição	$\bar{D}$	1.187	1.201	65%	82%	(Ferreira Sobrinho, 1982; cap. = 26)
	S	0.392	0.393	90%	81%	
	D					
Amplitude amostral	$\bar{R}$	0.6664	0.6675	49%	47%	(Saliby, 1982)
	S	0.1774	0.1780	12%	5%	
	R					
Análise de risco	$\bar{VP}$	392.82	390.44	97%	94%	(Saliby, 1982)
	S	576.57	506.67	45%	23%	
	VP					
Fila M/G/2	$\bar{WT}$	1.400	1.420	30%	40%	(Saliby e Funchal, 1987; T = 10 anos)
Fila M/S/3	$\bar{WT}$	0.150	0.152	44%	40%	(Saliby e Funchal, 1987; T = 10 anos)

Apesar de pouco eficientes, não há nada de errado com as técnicas de redução de variância, mas sim com o próprio princípio da imitação. A nosso ver, esta imitação é que é exagerada e não o controle dos valores numa amostra descritiva. Este excesso decorre de uma confusão entre um modelo e a realidade por ele representada. Assim, embora sempre tenhamos incertezas acerca dos parâmetros de um sistema em estudo, os parâmetros do respectivo modelo serão sempre supostos conhecidos, caso contrário não poderíamos obter uma solução.

Ora, uma simulação é feita com um modelo e não com a realidade. Utilizando a amostragem aleatória simples, nos afastamos das premissas do modelo; em consequência, as estimativas também se afastam dos parâmetros em estudo, tornando-se menos precisas.

Com a amostragem descritiva, evitamos que isto ocorra, pois nos atemos exatamente às hipóteses formuladas.

A teoria da amostragem descritiva, que se estende a todas as aplicações do método de Monte Carlo, pode ser facilmente testada. Até o momento, após diversos estudos, não observamos um único resultado que lhe fosse contraditório. Assim, embora não consideremos os testes definitivos, trata-se de uma teoria com plena confirmação empírica.

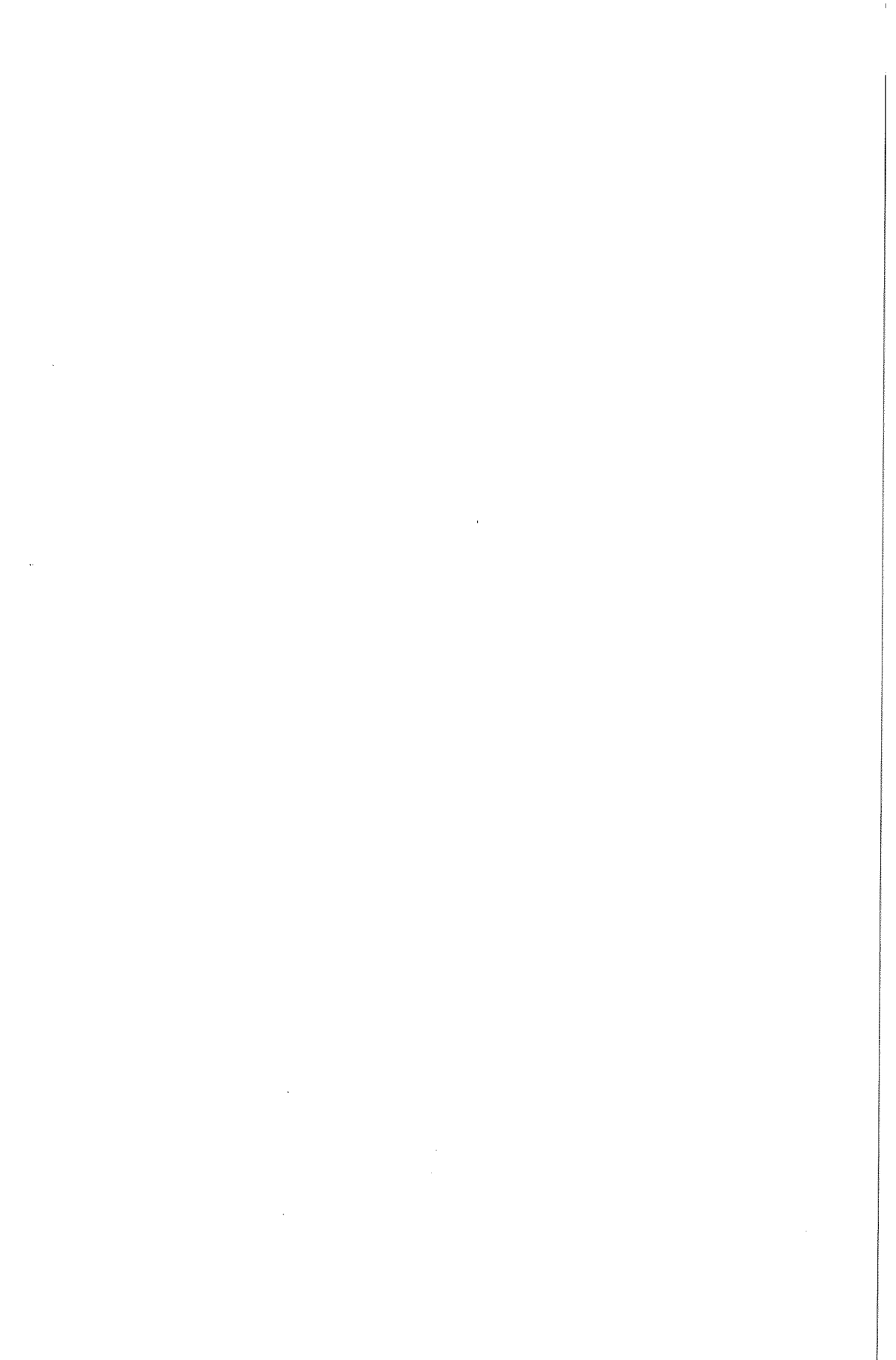
Sendo a amostragem descritiva mais eficiente em termos estatísticos e também, geralmente mais eficiente em termos computacionais, ela demonstra ser de grande



utilidade prática nas mais diversas aplicações da simulação por Monte Carlo.

#### REFERÊNCIAS

- Almeida, G. C. P. Geração de amostras descritivas em simulação. Rio de Janeiro, COPPE/UFRJ, 1983. Tese de Mestrado.
- Brenner, M. E. Selective sampling - a technique for reducing sample size in simulation of decision - making problems. J. Ind. Eng. 14 (6) : 291 - 296, Nov./Dec. 1963.
- Ferreira Sobrinho, W. O. Utilização de amostragem descritiva em simulação : em estudo empírico. Rio de Janeiro, COPPEAD/UFRJ, 1982. Tese de Mestrado.
- Saliby, E. A reappraisal of some simulation fundamentals. Lancaster, Universidade de Lancaster, 1980. Tese de Doutorado.
- \_\_\_\_\_. Uma nova comprovação de que o uso de amostragem aleatória simples em simulação é incorreto : estudo do caso particular do problema do jornaleiro. Rio de Janeiro, COPPEAD/UFRJ, 1981. (Relatório de Pesquisa, 29).
- \_\_\_\_\_. Sobre o uso incorreto de amostragem aleatória simples em simulação : um estudo da distribuição da amplitude de uma amostra uniforme. Revista de Administração, 17 (3) : 78-86, jul./set. 1982a.
- \_\_\_\_\_. Uma revisão de fundamentos da simulação : o uso incorreto de amostragem aleatória simples. Pesquisa Operacional, 2 (2) : 1-16, dez. 1982b.
- \_\_\_\_\_. Sources of variation of simulation estimates : an empirical study. Rio de Janeiro, COPPEAD/UFRJ, 1986. (Relatório de Pesquisa, 64).
- \_\_\_\_\_. Aplicação da amostragem descritiva da análise de risco. Rio de Janeiro, COPPEAD/UFRJ, 1987. (Relatório Técnico, 108).
- \_\_\_\_\_. Aplicação da amostragem descritiva na simulação de uma rede PERT. 1987. Aguardando publicação.
- \_\_\_\_\_. & Funchal, G. Estudo da eficiência da amostragem descritiva na simulação de um problema de filas. 1987. Aguardando publicação.



## PROGRAMAÇÃO DE ALOCAÇÃO DE ESTÁGIOS DE LINHA REMOTOS EM REDES TELEFÓNICAS ANÁLOGO-DIGITAIS

Cassilda M. Ribeiro  
Jurandir F.R. Fernandes  
Hermano de Medeiros F. Tavares

Universidade Estadual de Campinas  
Faculdade de Engenharia Elétrica

### Resumo

Nos próximos anos ocorrerão importantes mudanças nos sistemas telefônicos, devido à transição da tecnologia analógica para a digital. Neste sentido, a consideração de Estágios de Linha Remotos (ELR) será de fundamental importância. Estes equipamentos têm por função concentrar determinado número de assinantes, em seguida ligados a uma central de comutação através de cabos pulso-modulados. O programa ERAH (Estágios Remotos no Ano Horizonte) desenvolvido, determina quantos e onde devem ser implantados os ELR's. A decisão é tomada através de análises custo/benefício, baseadas na solução de problemas de fluxo de custo mínimo. O sistema telefónico é representado por uma rede de fluxos, com capacidade definida. Um exemplo de aplicação a uma cidade de médio porte, é descrito.

### Abstract

In the next years important changes in telephone systems will occur due the transition from analog to digital technology. In this sense it is important to consider the use of (DRU) \*. These equipments can concentrate subscribers to a switching-center by pulse code modulation cables. A computational tool, the RUHY program, was developed to determine whether and where installation of a (DRU) is required. The decision is taken after a cost/benefit analysis is performed, based on a solution of minimal cost network flow problems. The telephone system mathematical model is represented by a capacitated network flow problem. Example of application to a Brazilian urban network is reported.

### 1. INTRODUÇÃO

Com a crescente digitalização do sistema telefónico urbano, torna-se imperativo o uso de estágios de linha remotos (ELR). Estes equipamentos são capazes de concentrar assinantes, ligando-os a uma central-digital (central-mãe) através de cabos PCM.

Um dos problemas de se instalar estágios de linha remotos numa rede telefónica, advém do facto de seu custo não ser baixo. Surge então a necessidade de se desenvolver uma ferramenta capaz de indicar em que pontos da rede há viabilidade económica para esta instalação.

O trabalho a ser descrito neste artigo, foi desenvolvido no sentido de fornecer ao analista uma ferramenta que o auxilie no estudo de alternativas de expansão de uma rede telefónica, num determinado ano horizonte. O programa ERAH-Estágios Remotos no Ano Horizonte - resultado deste trabalho, busca um compromisso ótimo entre economia de rede e custos associados à instalação do ELR.

\* DRU = Digital Remote Unit

## 2. ASPECTOS GERAIS DO ESTÁGIO DE LINHA REMOTO

O ELR é um equipamento sem autonomia de encaminhamento, ou seja, seus assinantes devem ser encaminhados à central-mãe a que está ligado. Por outro lado, um ELR oferece a seus assinantes as mesmas facilidades existentes na central-mãe.

O equipamento do ELR é composto de vários módulos de concentração. Neste trabalho considerou-se um ELR capaz de concentrar um total de 8192 assinantes, através de módulos com capacidade de atender 1024 assinantes cada.

O emprego do ELR acarreta um acréscimo de equipamentos na central-mãe, com custo adicional. Este acréscimo está representado de forma simplificada nas figuras 1 e 2.

## 3. CARACTERIZAÇÃO DO PROBLEMA

A partir da rede existente no ano base de planejamento, o problema consiste em fornecer uma fotografia para o ano horizonte, dos pontos da rede onde a relação benefício-custo seja positiva devido à instalação de ELR's.

No desenvolvimento do programa ERAH foram feitas as seguintes considerações:

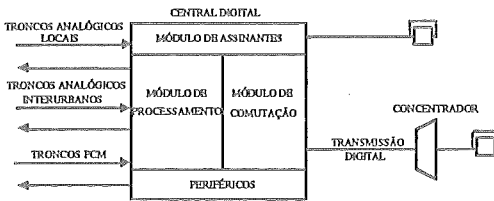


Fig. 1: Atendimento através de uma central digital sem ELR

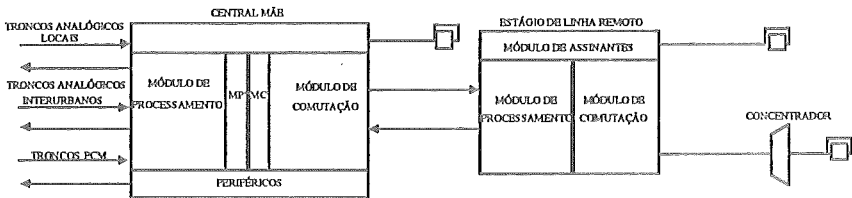


Fig. 2: Atendimento através de uma central digital com ELR

cade: MP - acréscimo do módulo de processamento devido ao atendimento remoto.

MC - acréscimo do módulo de comutação.

- é possível instalar um ELR no prédio de uma estação analógica. Como esta não comuta assinantes provenientes de um ELR, isto será feito pela central-mãe mais próxima, através da rede de troncos existente e que interliga todas as centrais. Assim a estação analógica simplesmente estará servindo de abrigo ao ELR.
- com base na consideração acima, o custo de entroncamentos dos ELR só leva em conta o custo dos cabos PCM e dos regeneradores necessários para ligá-los à estação mais próxima, mesmo que seja analógica.
- a instalação do ELR é feita módulo a módulo.
- o custo do primeiro módulo instalado é maior que os demais, pois neste caso considera-se o custo da infraestrutura necessária (terreno, obras civis).
- cada nó candidato pode receber um tipo diferente de ELR, de acordo com o desejo do planejador, ou seja, o programa ERAH permite a análise de um cenário com equipamentos de fabricantes distintos.

#### 4. MODELO DA REDE TELEFÔNICA

Adota-se um modelo de grafo, cujos nós representam os armários (seções de serviço) e as centrais de comutação. Cada armário apresenta uma demanda por terminais (assinantes) e cada central possui uma dada capacidade de comutação. Os arcos representam as ligações entre estes elementos.

O problema consiste em circular um fluxo na rede que satisfaça a filiação de todos os assinantes, respeitada a capacidade global de comutação, a um mínimo custo.

O modelo considera apenas a rede primária; o custo dos cabos é definido sobre os arcos do grafo e o objetivo é minimizar simultaneamente os custos de rede e de comutação. A representação dos custos de comutação é apresentada no item 5.3.

A técnica básica de otimização empregada pelo programa ERAH é um algoritmo de fluxo de custo mínimo [Kunnington, J.L. e outros, 1980], implementado na rotina PFCM (Problema de Fluxo de Custo Mínimo).

#### 5. A ESTRUTURA DO PROGRAMA ERAH

O programa está estruturado em quatro fases:

##### 5.1 FASE I : Seleção de candidatos

Nesta fase o programa processa a concentração de assinantes em determinados nós, que apontarão regiões caras da rede propícias a conter centros de fios. Estará assim facilitada ao analista, a escolha de nós candidatos a receberem estágios de linha remotos. Esta fase é executada pela rotina NUVEM [Ribeiro, C.M. 1985.]

##### 5.2 FASE II : Cálculo das distâncias entre os candidatos e as estações mais próximas

Aplica-se aqui uma rotina de cálculo de fluxo de custo mínimo, sem levar em conta a disponibilidade de fios existentes na rede. A rotina PFCM - Problema de Fluxo de Custo Mínimo - é responsável por esta Fase.

##### 5.3 FASE III : Preparação da rede

São criados arcos artificiais, chamados arcos de fechamento, para cada uma das centrais, levando-se em conta as capacidades e custos de comutação das mesmas. Estes arcos de fechamento são ligados a um nó fictício chamado nó de fechamento.

Os nós candidatos à localização de ELR também são ligados ao nó de fechamento. Seus arcos de fechamento contém os custos, por assinante, referentes ao ELR mais os custos de comutação da central-mãe.

Para ilustrar representa-se na figura 3 um trecho de rede antes da instalação de um ELR.

Suponha-se que os assinantes dos nós I e II dirijam-se ao nó de fechamento (NF) através da estação C. Neste caso o custo de comutação referente à central está sendo computado no arco de fechamento (AF).

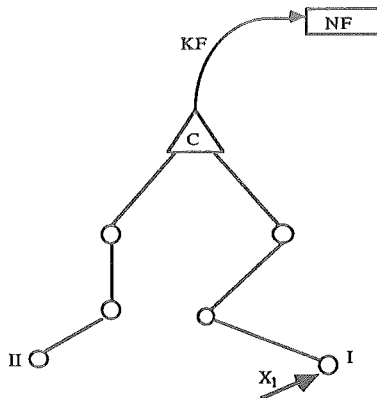


Fig. 3 : Arco e nó de fechamento de uma estação

Ao se instalar um ELR no nó I (Fig. 4), cria-se um novo arco de fechamento (AF'), ligando o estágio remoto diretamente ao nó de fechamento.

Supondo-se que seja mais económica a ligação dos assinantes, de I até (NF), através do estágio remoto, e que  $X_1$  seja menor que a capacidade do ELR, então os  $X_1$  assinantes não mais passarão pelo arco (AF) e sim por (AF'). Justifica-se pois a necessidade de se computar o custo de comutação de  $X_1$ , referente à central C, no arco AF.

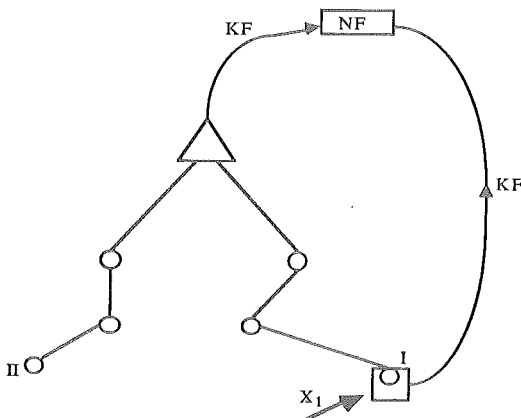


Fig. 4 : Arco e nó de fechamento de estação e de ELR

Criados os arcos de fechamento, novamente aplica-se a rotina PFCM, levando-se agora em conta as disponibilidades de pares de fios já existentes na rede no ano base.

Para obter-se um custo de rede sem ELR, a aplicação acima se dá atribuindo-se capacidades nulas aos arcos de fechamentos dos estágios remotos.

#### 5.4 FASE IV : Implementação dos ELR

Nesta fase são realizados testes para se confirmar ou não a instalação de um ELR num determinado nó candidato.

Os nós candidatos são testados a partir dos de maior potencial na rede, ou seja, daqueles mais distantes de centros de fios, se suposta a ausência de fios ociosos. Iterativamente executa-se um programa de fluxo de custo mínimo, para se calcular o lucro obtido pela instalação de um módulo de ELR, num determinado nó candidato. O arco de fechamento deste nó terá capacidade igual à do módulo.

A implantação de um módulo será confirmada se na relação abaixo o lucro  $L_n$  for positivo:

$$L_n = (\text{custo da rede sem ELR}) - (\text{custo da rede com ELR})$$

onde  $n$  é o número de módulos do ELR em teste.

Se a capacidade do primeiro módulo foi atingida, um novo módulo deve ser testado mesmo se  $L_1 \leq 0$ , uma vez que o custo do primeiro é maior que dos demais.

Adicionar um novo módulo num só significa aumentar a capacidade do seu arco de fechamento e executar novamente o PFCM, partindo-se da solução ótima anterior, obtendo-se  $L_2$ .

Generalizando, suponha-se que o  $n$ -ésimo módulo de um ELR esteja sendo testado num determinado nó candidato. Pode ocorrer:

$L_n > L_{n-1}$ $L_n > \emptyset$	: O $n$ -ésimo módulo é confirmado. Se sua capacidade foi atingida testa-se mais um módulo.
$L_n < L_{n-1}$ $\emptyset < L_{n-1}$	: O $n$ -ésimo módulo não é confirmado. Testa-se outro nó candidato.

O programa ERAH mediante a ocorrência de  $L_1 < 0$  e  $L_2 < 0$ , descarta a implantação de ELR no nó candidato em teste.

Após a aplicação da rotina NUVEM na FASE I e feita a seleção de nós candidatos a receberem ELR, o programa ERAH apresenta a estrutura mostrada na figura 5.

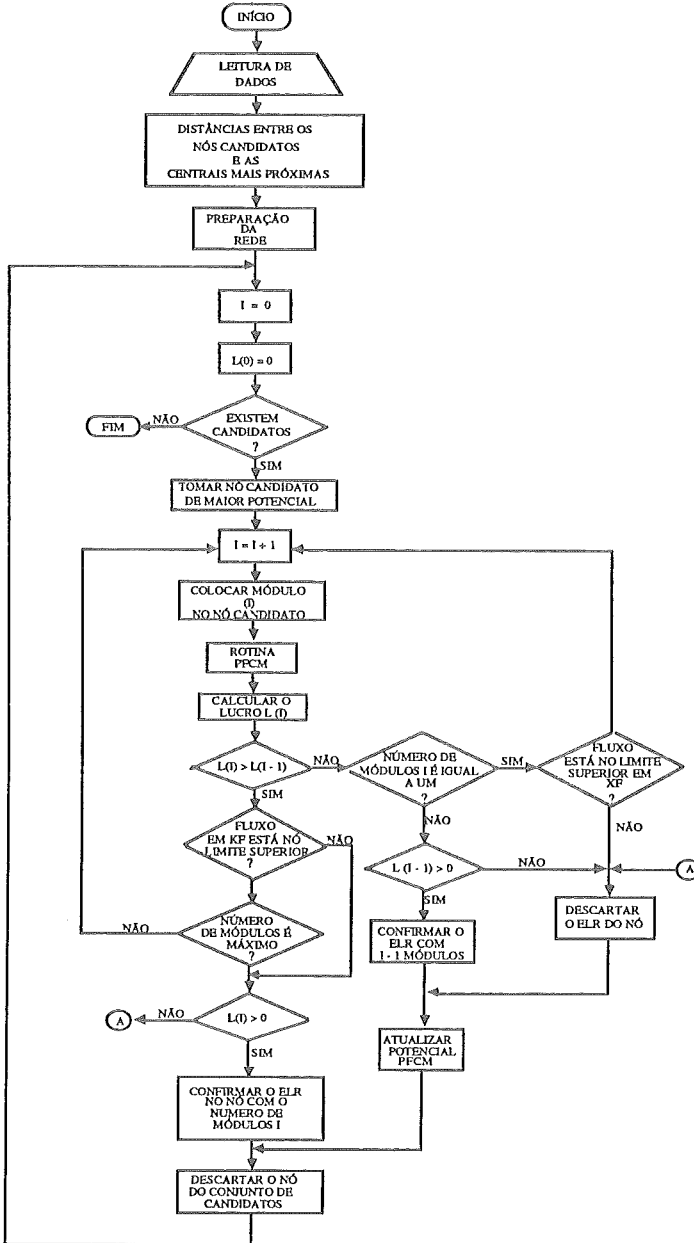


Fig. 5 : Diagrama simplificado do programa ERAH

## 6. ASPECTOS COMPUTACIONAIS

### 6.1. Dados Necessários

- . número de nós, arcos e estações no horizonte;
- . número estimado de assinantes, por seção de serviço, no ano horizonte;
- . rede primária prevista para o ano horizonte;
- . rede primária atual;
- . localização e capacidade das centrais existentes, bem como a capacidade de expansão de cada uma delas;
- . capacidade de cada módulo do ELR, número máximo de módulos e custo fixo de implantação de cada módulo;
- . custo de comutação do ELR e das centrais;
- . área requerida para o ELR e custo do m<sup>2</sup> de terreno;
- . custo e distância entre regeneradores;
- . custo do parâmetro do cabo PCM e de fio de cobre;

### 6.2. Relatório de Saída

Fornecer os nós que recebem ELR; a estação mais próxima de cada um deles; o número de módulos e de regeneradores instalados; o número de assinantes ligados via ELR; o custo da rede com e sem ELR e o lucro obtido com a implantação dos ELR.

## 7. APLICAÇÃO A UMA REDE REAL

A rede telefônica de Curitiba serviu de base para os testes do programa ERAH. O ano base considerado foi o de 1985, ocasião em que a rede da cidade possuía 218.207 terminais e 10 estações de comutação analógicas. O ano 200 foi considerado o horizonte de planejamento, com uma demanda prevista de 410.713 terminais e 10 novas estações, todas de tecnologia digital. Foi considerada ainda a possibilidade de ampliação de todas as estações antigas, exceto uma localizada na área mais central de Curitiba. Os custos envolvidos estão de acordo com os dados registrados em (FERNANDES, J.F.R. e TAVARES, H.M.F, 1983).

O modelo representativo desta rede resultou num grafo com 586 nós e 4464 arcos.

Como resultado, o programa ERAH localizou 22 estágios de linha remotos, mostrados na figura 6. O número ao lado do nó com ELR fornece a quantidade de módulos implantados. Indicou ainda que 13,6% dos terminais estariam sendo filiados através do ELR, resultando em cerca de 6% de economia em investimentos, se comparado com um cenário de planejamento sem estágios remotos.

## 8. COMENTÁRIOS E CONCLUSÕES

Durante os testes notou-se grande sensibilidade do problema de localização de ELR em função dos custos envolvidos. Ou seja, um nó candidato pode ou não ser rejeitado em função dos custos relativos à comutação na estação e no ELR.

Via de regra os testes indicaram economias globais em torno de 6% a favor de cenários contemplando a presença de ELR. Particularmente interessante foram as economias resultantes nos investimentos em redes (pares de fios de cobre), em média, da ordem de

É importante ainda assinalar que expansões de sistemas telefônicos urbanos que levem em conta a implantação de estágios remotos, propiciam uma utilização mais adequada e intensa dos cabos (pares de fios) lançados, evitando perdas financeiras.

## REFERÊNCIAS BIBLIOGRÁFICAS

- (1) Fernandes, J.F.R.; Tavares, H.M.F., (1985). "Relatório sobre Custos de Centrais Analógicas e Digitais e de Estágios de Linha Remotos", RT - 031/83, FEC, UNICAMP.
- (2) Kennington, J.L.; Helgason, R.V., (1980). "Algorithms for Network Programming", John Wiley & Sons, Inc.
- (3) Ribeiro, C.M., (1985). "Planejamento de Redes Telefônicas : Alocação de Concentradores, Indicação de Centros de Fios e Localização de Estágios de Linha Remotos", Dissertação de Mestrado, Publicação FEC 092/85, UNICAMP.



# OTIMIZAÇÃO DE FUNÇÕES NÃO-LINEARES DE FLUXOS EM REDES UM MÉTODO EFICIENTE PARA O PLANEJAMENTO DA OPERAÇÃO A MÉDIO PRAZO DE SISTEMAS HIDROELÉTRICOS

C. Lyra  
H. Tavares

Depto. de Engenharia de Sistemas  
Faculdade de Engenharia Elétrica  
UNICAMP

## Resumo

Apresenta-se um algoritmo para a otimização de funções não lineares em redes. O método foi desenvolvido com o propósito de viabilizar o planejamento da produção de energia a médio prazo de grandes sistemas hidroelétricos. A estrutura de fluxos das restrições permite dispensar manipulações de matrizes no processo de cálculo, proporcionando eficiência e estabilidade numérica. Resultados computacionais asseguram a boa performance do algoritmo.

## Abstract

An algorithm for optimization of nonlinear functions of flows in a network is presented. The method was developed with the purpose of turning possible energy production planning of large scale hydroelectric systems over midterm horizons. The network structure of constraints allow to avoid matrices manipulations in the computation process, what leads to both efficiency and numerical stability. Computational results assures the good performance of the algorithm.

## 1. INTRODUÇÃO

Este trabalho discute um algoritmo para otimização de funções não lineares de fluxos conservativos em redes. Seu projeto foi realizado no quadro de pesquisas para elaboração de uma metodologia que viabilizasse a obtenção de políticas ótimas no planejamento da operação energética a médio prazo de grandes sistemas hidroelétricos. A identificação de uma estrutura de fluxos intrínseca ao problema, aliada ao projeto do algoritmo aqui apresentado, permitiu que este objetivo fosse atingido.

O algoritmo é aplicável a qualquer problema de otimização de fluxos em redes conservativas. Não se exige propriedades especiais da função objetivo além de uma certa suavidade. Mas, naturalmente, a otimalidade global só é garantida sob condições de convexidade.

O item 2 expõe a formulação do problema. No item seguinte apresenta-se o método de solução. Resultados computacionais são comentados no item 4. Uma breve discussão acerca de generalizações (item 5) precede as conclusões (item 6).

## 2. O Problema

Políticas ótimas para a operação a médio prazo de um sistema de potência predominantemente hidroelétrico podem ser encontradas através da solução do problema abaixo [1]:

$$\begin{aligned} \text{Min } & f(x) & (1) \\ \text{s. a. } & Ax = b & (2) \\ & \underline{x} \leq x \leq \bar{x} & (3) \end{aligned}$$

onde  $f$  é um funcional em  $\mathbb{R}^n$ ,  $A$  é uma matriz de incidência nó-arco (de dimensão  $m \times n$ ) que representa expansão no tempo da rede hidroelétrica [1 - 2], e  $b$  é o vetor de recursos

associados a cada nó. Supõe-se a rede balanceada, i.e.,  $\sum b_i = 0$ .

Na expansão no tempo da rede hidroelétrica representada por  $A$ , cada nó corresponde a um reservatório, durante certo intervalo. Em cada nó há um arco convergente cujo fluxo é o volume d'água armazenado no final de cada intervalo anterior e um arco divergente onde flui o volume armazenado no final do intervalo atual; existem também arcos convergentes onde fluem as defluências de usinas imediatamente a montante e um arco divergente onde flui a defluência total do reservatório durante o intervalo. Um nó artificial, a raiz, é acrescentado à rede para o balanceamento; a raiz coleta todos os fluxos de volumes no final do horizonte de planejamento e as defluências das usinas mais a jusante, em todos os intervalos. Assim, para um sistema hidroelétrico de  $K$  reservatórios e um período de planejamento de  $T$  intervalos discretos,  $A$  representa uma rede de  $m=(K.T+1)$  nós e  $n=2(K.T)$  arcos.

## 3. UM ALGORITMO PARA OTIMIZAÇÃO DE FUNÇÕES NÃO-LINEARES DE FLUXOS EM REDES

Nas situações de maior interesse o problema (1) - (3) é de muito grande porte. Esta característica exige uma exploração cuidadosa de sua estrutura matemática específica.

O algoritmo aqui apresentado pode ser visualizado como uma síntese da versão revisada de gradiente reduzido [1, 3] com o método simplex para o problema de fluxos de custo mínimo [4].

Seja  $I$  um conjunto de índices. No desenvolvimento a seguir adota-se a notação abaixo:

$A_i^j$  = elemento da coluna  $j$  e linha  $i$  da matriz  $A$

$A^j$  = coluna  $j$  de  $A$

$A^I$  = matriz formada pela união de colunas  $i$ ,  $i \in I$

$x_i$  =  $i$ -ésimo componente de vetor  $x$

$x_I$  = vetor de fluxos cujas componentes são  $x_i$ ,  $i \in I$

$\nabla f$  = vetor de derivadas parciais  $\frac{\partial f}{\partial x_i}$ ,  $i \in I$ .

Produtos de vetores e matrizes estão sempre ajustados; não se utiliza a notação de transposição.

Dado um valor de fluxos  $x$ , deve-se determinar uma partição  $(I, J)$  tal que  $A^I$  represente uma árvore contendo todos os nós.

$$\underline{x}_I \leq x_I \leq \bar{x}_I \quad (4)$$

Os fluxos em  $x_I$  são dependentes (ou básicos) e os fluxos em  $x_J$  são independentes (ou não básicos).

O gradiente reduzido é calculado convenientemente em duas etapas:

- resolve-se o sistema (5),

$$\lambda A^I = \nabla f(x) \quad (5)$$

- usa-se o vetor  $\lambda$ , de variáveis duais, em (6),

$$e_j = \nabla_j f(x) - (\lambda_p - \lambda_q), \quad \forall j \in J \quad (6)$$

onde  $\lambda_p$  e  $\lambda_q$  são, respectivamente, variáveis duais associadas ao nó-origem e nó-destino do arco representado por  $A^j$ .

As componentes independentes,  $\Delta x_j$ , da direção de busca são calculadas por projeção do gradiente reduzido sobre as restrições de canalização. As componentes dependentes,

$\Delta x_j$ , são obtidas através do sistema (2)

$$\Delta x_I = \sum_{j=1}^n [-\hat{A}^j] \Delta x_j \quad (7)$$

onde  $\hat{A}^j$  é solução de (8).

$$A^j = A^I \hat{A}^j \quad (8)$$

A característica de fluxos em redes das restrições permite que o sistema (5) e o sistema (8) sejam resolvidos sem a necessidade de inversão de matrizes [4].

A partição (I, J) deve ser atualizada sempre que uma variável básica deixar de satisfazer a condição (4), isto é, atingir um dos seus limites.

#### Sumário de Algoritmo

Dispondo-se de um vetor de fluxos factíveis  $x$  e de uma partição (I, J), uma iteração do algoritmo pode ser sintetizada na sequência de passos descritos a seguir:

- 1) Calcula-se o vetor gradiente  $\Delta f(x)$ .
- 2) Encontra-se as variáveis duais através da solução de (5).
- 3) Calcula-se todas as componentes do gradiente reduzido, através de (6).
- 4) Projeta-se o gradiente reduzido sobre as restrições de canalização, isto é, faz-se:

$$\Delta x_I = \begin{cases} e_j, & \text{se } x_j \leq \bar{x}_j \leq \bar{x}_j \\ & \text{ou } x_j = \bar{x}_j \text{ e } e_j \leq 0 \\ 0, & \text{se } x_j = \bar{x}_j \text{ e } e_j \geq 0 \\ & \text{ou } x_j = \bar{x}_j \text{ e } e_j \leq 0 \end{cases}$$

- 5) se  $\Delta x_j = 0$ , para-se; a solução atual é ótima. Caso contrário, encontra-se  $\Delta x_I$  através de (7).

- 6) Encontra-se  $\alpha_1, \alpha_2, \alpha_3$  satisfazendo respectivamente

$$\begin{aligned} & \max \{ \alpha \text{ t.q. } \underline{x}_I \leq x_I - \alpha \Delta x_I \leq \bar{x}_I \} \\ & \max \{ \alpha \text{ t.q. } \underline{x}_J \leq x_J - \alpha \Delta x_J \leq \bar{x}_J \} \\ & \min \{ f(x - \alpha \Delta x) \text{ t.q. } 0 \leq \alpha \leq \alpha_1, 0 \leq \alpha \leq \alpha_2 \} \end{aligned}$$

Substitui-se o vetor  $x$  pelo novo vetor  $x - \alpha_3 \Delta x$ .

- 7) Se  $\alpha_3 < \alpha_1$  volta-se para o passo 1.

Caso contrário, troca-se de base e em seguida volta-se ao passo 1.

#### Comentários

- a situação de otimalidade declarado no passo 5 decorre da satisfação das condições de Kuhn-Tucker [3].
- a eficiência do algoritmo é baseada na estrutura de redes, explorada através de um sistema de indexação denominado ppai/fio [1, 4].
- não havendo a necessidade de inverter matrizes, inexistem erros de arredondamento e, conseqüentemente, problemas de estabilidade numérica.
- quando não é possível a obtenção a priori de um fluxo factível  $x$  e uma partição (I, J), é necessário adotar-se um procedimento tipo fase I do método simplex [3].

#### 4. RESULTADOS COMPUTACIONAIS

O algoritmo foi codificado em FORTRAN e executado em equipamento computacional DEC 10. Sua utilização permitiu que se encontrasse políticas ótimas para operação a médio prazo do sistema hidroelétrico da bacia do rio Paraná, composto de 10 usinas com reservatório, 13 usinas a fio d'água, e uma potência total de quase 37800 MW.

A aplicação mais promissora consistiu em encontrar políticas ótimas para operação do sistema durante seu "período crítico", caracterizado pelos 60 meses mais secos do histórico de vazões conhecidas - de Janeiro de 1952 a Dezembro de 1956. O comportamento do sistema sob "período crítico" é um cenário fundamental em estudos de operação e expansão.

Manipulações do modelo [1, 5] permitiram reduzir a matriz de incidência do problema de operação ótima do sistema da bacia do Paraná sob período crítico a apenas 1081 nós e 2160 arcos. Considerou-se soluções iniciais onde os reservatórios desconsideraram suas capacidades de regularização, defluindo o total das vazões afluentes. Nessas circunstâncias, num caso típico, a solução ótima foi encontrada após 448 iterações, em cerca de 14 minutos de C.P.U.

O resultado acima ilustra a potencialidade de aplicação do algoritmo aos problemas que motivaram seu projeto. Melhores performances computacionais podem ser obtidas através de um pré-condicionamento do problema por considerações de engenharia [5].

## 5. GENERALIZAÇÕES

O método discutido neste trabalho pode ser generalizado a problemas de fluxos em redes com perdas (ou ganhos). Estes problemas surgem, por exemplo, em planejamento da operação de sistemas hidroelétrico onde são consideráveis as perdas por evaporação nos reservatórios.

A diferença fundamental do método seria na solução dos sistemas (5) e (8), que necessitariam de uma abordagem um pouco mais elaborada para o aproveitamento da estrutura de redes. Um bom estudo acerca das peculiaridades e do método de solução desses sistemas pode ser encontrado em [4].

O algoritmo é também generalizável a problemas onde restrições de fluxos são acompanhadas de outras restrições lineares, sem estrutura particular. O procedimento para solução dos sistemas (5) e (8) nessas circunstâncias é discutido no capítulo 7 de referência [4].

## 6. CONCLUSÕES

Apresentou-se algoritmo para otimização de funções não lineares de fluxos em redes. Seu desenvolvimento teve o propósito de permitir a obtenção de políticas ótimas para o planejamento da operação a médio prazo de grandes sistemas hidroelétricos. Não obstante, evitou-se usar especificidades das redes características destes problemas hidroelétricos, o que alarga a gama de aplicações do método. Além disso, extensões do algoritmo a problemas com estruturas "próximas" a do gradiente reduzido são realizáveis.

Resultados computacionais demonstram o bom funcionamento do algoritmo; ele permitiu, por exemplo, uma abordagem por otimização para o problema de operação sob período crítico do sistema da bacia do Paraná, até então tratado por simulação. Vale lembrar que, sendo a energia de origem hidroelétrica responsável por mais de 90% da energia elétrica produzida no País, o uso dos recursos hidráulicos no limite de suas capacidades é importante não apenas para a reprodução de custos imediatos, mas sobretudo para evitar situações de racionamento de energia.

## REFERÊNCIAS

- [1] C.Lyra "Contribuição ao Planejamento da Produção de Energia em Sistemas de Potência", Tese de Doutorado, UNICAMP, Publicação FEC 040/84, Campinas, 1984.
- [2] R.E. Rosenthal, "A Nonlinear Network Flow Algorithm for Maximization of Benefits in a Hydroelectric Power System", *Operations Research*, Vol. 29, pp.763-786, 1981.
- [3] D.G. Luenberger, "Linear and Nonlinear Programming", Addison Wesley, Reading, Massachusetts, 1984.
- [4] J.L. Kennington e R.V. Helgason, "Algorithms for Network Programming", John Wiley and Sons, New York, 1980.
- [5] C. Lyra e H. Tavares, "A Contribution to the Midterm Scheduling of Large Scale Hydrothermal Power Systems", paper 87 SM 484-9, IEEE/PES 1987 Summer Meeting, São Francisco, 1987.

## LA INVESTIGACION OPERATIVA EN LA AGRICULTURA ESPANOLA

Francisco Amador Hidalgo  
Manuel Cabanes Fuentes  
José J. Romero Rodríguez

Escuela Superior de Técnica Empresarial  
Agrícola (E T E A)  
Córdoba

**Resumo:** Las aplicaciones agrícolas de Investigación Operativa (IO) pueden generalmente incluirse en dos grupos: (1) aplicaciones a nivel de empresa agrícola y (2) aplicaciones a nivel de sector agrario, bien en un contexto nacional o en regiones o zonas concretas.

Se argumenta en el artículo que en países como España, la empresa agrícola (al igual que ocurre con la pequeña y mediana empresa industrial o comercial) está condenada, debido a sus especiales características, a no poder beneficiarse en absoluto, o a hacerlo solamente de forma limitada, de la utilidad que el enfoque de "IO" supone en materia de "Gestión y Dirección de Empresas" (beneficio del que sí disfrutaban en mayor medida las grandes empresas).

La experiencia con que se cuenta, aunque corta, sugiere que la introducción de la "IO" en la empresa agrícola española requiere el desarrollo de acciones y medios especiales.

A nivel sectorial, se señala que el número de aplicaciones en España es muy reducido y presentan algunos casos problemas metodológicos importantes.

### 1. Introducción

Con la incorporación de la Investigación Operativa (IO) al ámbito académico, aparece una cierta tendencia natural hacia la formulación de problemas teóricos y su manipulación, en detrimento del espíritu práctico y multidisciplinar que fueron esenciales a los primeros tiempos de esta disciplina.

Este doble aspecto de teoría y práctica sirve para diferenciar los enfoques de los distintos departamentos de IO en varios países, según el énfasis que los mismos ponen en uno u otro aspecto de la formación en esta área.

En general se pueden decir que en la mayoría de los países se intenta corregir esta tendencia natural, (a) concibiendo la formación en IO como independiente de los departamentos de Ciencias Exactas (Matemáticas Puras), (b) desarrollando unos programas de formación en IO, en los que el contenido práctico (casos, aplicaciones, proyecto) supone un elevado porcentaje del tiempo total de formación, (c) favoreciendo la publicación de casos prácticos en las revistas de las respectivas Sociedades de I.O. (incluso con premios especiales). España es quizás uno de los pocos países donde esta tendencia teórica no se ha corregido, y donde ninguna de las medidas correctoras, mencionadas anteriormente, se ha aplicado.

La Agricultura, supone un campo de aplicación y no es de extrañar en consecuencia que las aplicaciones agrícolas de Investigación Operativa en España sean prácticamente nulas. A excepción de algunos trabajos aislados (1) sin continuidad temporal, la única aportación significativa en este campo, conocida a los autores de este artículo, es la de la Escuela Superior de Técnica Empresarial Agrícola (E.T.E.A.) de Córdoba. Esta aportación no ha sido siempre publicada debido a que en la mayoría de los casos es el resultado de investigaciones contratadas con Entidades Públicas o Privadas.

En este artículo se hace una breve historia de las aplicaciones agrícolas de I.O. en E.T.E.A., presentando con mayor detalle algunas de ellas.

## 2. La Empresa Agrícola y el Sector Agrícola

La problemática decisional en la agricultura afecta fundamentalmente a dos grupos de decisores distintos: los empresarios agrícolas (a nivel de empresas individuales), y las autoridades administrativas (a nivel de política agraria). Ambos niveles presentan potencialmente situaciones decisionales para la aplicación de I.O., sin embargo, las dificultades de realizaciones prácticas son mucho mayores en el primer caso que en el segundo.

### La Empresa Agrícola

Aún hoy, es frecuente oír hablar de la necesidad de reforma en el empresa agraria. En el caso de España no se trataya, y cada vez menos, de una reforma técnica, sino, de una forma empresarial y de planificación.

En este sentido, las aplicaciones de I.O. y otras técnicas de decisión y gestión se han visto siempre dificultadas por el bajo nivel socio-cultural-empresarial del medio agrícola, lo cual constituye en general una característica especial de este tipo de empresa (al igual que lo es de la pequeña y mediana empresa industrial o comercial).

Estas dificultades peculiares se resumen diciendo que el cuadro humano actual de la empresa agrícola no está en condiciones de poder ser formado en IO o áreas similares. Al mismo tiempo este tipo de empresas carece de tradición (y posiblemente recursos) para acceder a un asesoramiento externo contínuo.

Consiente de esas limitaciones, la labor de E.T.E.A., en los años 70 estuvo encaminada a colaborar con otras entidades de ámbito nacional, Fundación para el Desarrollo de la Función Social de las Comunicaciones (FUNDESCO) y Asociación para el Desarrollo Empresarial de la Agricultura (ADEA) en una instrumentación de medios (Teleproceso Agrario) que sí estuviera al alcance de la empresa agraria.

El proyecto de colaboración pretendía textualmente:

"Llegar a montar a nivel nacional un servicio de información y cálculo para la resolución de problemas de gestión de la empresa agropecuaria.

Dicho servicio exigiría en su total concepción:

a) *HARWARE*; una unidad central de cálculo y sus correspondientes unidades periféricas situadas en puntos del territorio nacional.

- una red de terminales distribuidas por el territorio y que podrían revertir un doble modelo.

- un telepio de salida escrita, que podría estar localizado en centrales telefónicas, agencias del Servicio de Extensión Agraria, Centros Docentes y de Investigación especializados, Centros de Gestión, y donde quiera que exista un personal especializado en su manejo que atienda las consultas de los interesados.

- un auricular con salida de datos oral, que podría ser conectado a la red telefónica en cualquier punto.

b) *SOFTWARE*; integrado por una colección de programas de fácil manejo por el usuario, que permita la información adecuada o el cálculo matemático de problemas de gestión cuya resolución resulta inaccesible hoy día por métodos normales".

Como resultado de este proyecto de investigación se elaboraron un número de modelos (no publicados hasta la fecha) y sus correspondientes códigos de ordenador, entre los que se pueden citar:

- Formulación de raciones para ganado

- Fórmulas de abonado a coste mínimo

- Alternativas de financiación: Optimización del empleo de medios financieros en una explotación ganadera.

- Gestión de stocks de materias auxiliares agrícolas

- Optimización de resultados en el cebo ganado

- Sustitución de la maquinaria agrícola

- Distribución óptima de entregas de trigo a través de los canales del S.E.M.P.A.

- Renovación de ponedoras

- Previsión de la producción física e ingresos de un lote de ponedoras de huevos comerciales.

- Matriz de carácter máximo

Al igual que ocurre en otras áreas un porcentaje importante de aplicaciones agrícolas lo son de programación matemática.

A título ilustrativo se hacen seguidamente algunos comentarios sobre uno de estos modelos.

#### Matriz de Carácter Máximo: Modelo Pletea

Pletea es un modelo clásico de PL. (La primera formulación de PLETEA se realizó en 1975 y no ha sido publicado con anterioridad).

$$\begin{aligned} \text{Óptimo de } Z &= [C] \quad (X) \\ \text{Sometido a: } [A] (X) &\leq (b) \\ &(X) \leq (0) \end{aligned}$$

En onde:

- $(X)$  = Vector de actividades (Vector columna)
- $[C]$  = Vector de coeficientes de la función objetivo (vector fila).  
En el modelo PLETEA estos coeficientes son precios de productos y de factores.
- $[A]$  = Matriz de coeficientes técnicos
- $(b)$  = Vector de limitaciones (Vector columna)

#### Características más Acusadas

PLETEA es un modelo de programación lineal continua, estático y estacionario. El mismo puede integrarse dentro de "la familia de los programas de adaptación" dado que su objetivo se circunscribe a la determinación de unos planes de producción óptimos en el marco de un ejercicio económico y sin que intervengan modificaciones en la estructura productiva de la empresa.

Consta de 340 actividades y 315 relaciones de condición o restricciones. La dimensión del mismo es, por tanto, de 315 filas por 340 columnas, exclusión hecha de las variables de holgura y artificiales. (Véase figura 1, en la página 5).

De las 340 actividades definidas en el modelo, 134 tienen el carácter de actividades de producción; 166 miden el nivel de consumo (demanda) de factores y servicios externos; 24 se han reservado para la medición de los distintos niveles de producción alcanzados en el desarrollo del programa óptimo; 7 están destinadas a medir el crédito bancario a corto plazo mínimo necesario en cada uno de los siete períodos en que se subdivide el de planificación, para poder hacer frente a los compromisos de pago; siete miden el nivel del saldo de caja al final de cada uno de los siete períodos considerados; las dos actividades restantes se han destinado a la medición de los pagos y cobros a efectuar en el período de planificación con independencia de cual sea el sistema de producción resultante (pagos y cobros fijos).

EL conjunto de relaciones de condición puede partirse en los siguientes subconjuntos e submatrices:

#### Filas

1. Ocupación del suelo.....	15
2. Rotación de cultivos.....	57
3. Limitaciones a la dimensión de las actividades de producción, basadas en consideraciones de tipo psicológico, legal o de mercado.....	38
4. Demanda y disponibilidades de mano de obra, tracción y maquinaria.....	48
5. Demanda fertilizantes.....	21
6. Demanda semillas.....	26
7. Demanda y disponibilidades de agua de riego.....	16
8. Demanda de tratamientos fitosanitarios y otros conceptos que impliquen gasto variable.....	34
9. Transformación de las horas de tracción propia y de las horas de maquinaria propia demandadas, en demanda de gas-oil y gastos en reapaciones y recambios.....	26
10. Niveles de producción y venta.....	24
11. Pagos y cobros fijos.....	2
12. Circuito financiero de la empresa a corto plazo.....	8

Se consideran en el modelo dos funciones objetivo alternativas:

1º Maximización del Margen Bruto Total (Ingreso Total Gasto Variable Total).

2º Optimización de desviaciones sobre metas preestablecidas.

*Utilización del Modelo "Pletea": Datos necesarios e información aportada por el mismo.*

### Datos necesarios

La utilización del modelo "PLETEA" en la planificación de la producción a corto plazo permite al usuario hacer uso de valores standard o definir sus propios valores respecto de:

- 1º Precios previstos de productos y de factores necesarios para su obtención (elementos del vector [C]).
- 2º Coeficientes factor-producto, definidores de las cantidades de factores requeridas por unidad de actividad (elementos de la matriz [A]).
- 3º Disponibilidades en factores fijos o estructurales (elementos del vector {b}).

### Información Aportada por el Modelo

La información suministrada por "PLETEA" comprende los siguientes aspectos:

- 1º Alternativa óptima de cultivos sobre cada una de las clases de suelo consideradas en el modelo.
- 2º Margen Bruto (Ingresos totales menos Gastos Variables totales) asociado a esas alternativas óptimas de cultivo.
- 3º Presupuesto periodificado de ingresos por naturaleza.
- 4º Presupuesto periodificado de gastos, distinguiendo entre gastos fijos y gastos variables y dentro de cada una de estas categorías según naturaleza.
- 5º Presupuesto periodificado de tesorería (cobros - pagos).

### Conclusión

Interés empresarial del modelo.

La utilización de "PLETEA" en planificación de la producción a corto plazo presenta las siguientes ventajas:

- 1º El modelo considera el conjunto de la explotación como un todo único.
- 2º Constituye, así mismo, un notable instrumento de síntesis del funcionamiento de la empresa.
- 3º Planifica las relaciones de input-output con el exterior. Para los factores estructurales el modelo permite conocer los que son más limitantes y la prioridad en que dejen de serlo.
- 4º La posibilidad de utilización del programa por multitud de empresarios sin necesidad de tener que recorrer las distintas fases de la modelización en cada caso, supone un gran ahorro de tiempo y una disminución sustanciosa del costo de utilización dejando de ser, por tanto, el coste de utilización, barrera que impida el acceso de la empresa agraria a estas modernas técnicas de gestión.

Recientemente se ha incorporado como nueva faceta la Programación por Objetivos.

### Perspectivas de Futuro

El Teleproceso Agrario, que sí ha tenido éxito en otros países, no acaba de ver la luz en España. Por el contrario, el campo de los miniordenadores está alcanzando un desarrollo vertiginoso incluso en la empresa agrícola. Los miniordenadores representan actualmente la vía instrumental más prometedora para hacer llegar (con bibliotecas de software adecuadas) la Metodología y las técnicas de I.O. a la empresa Agraria.

### El Sector Agrícola

El sector agrícola aún habiendo perdido peso dentro del conjunto económico de los países desarrollados, sigue jugando un papel fundamental en países como España.

Al mismo tiempo, el intervencionismo estatal en la planificación del sector ha sido creciente, así como creciente es la complejidad de esa labor. Esto ha dado lugar a que las autoridades administrativas se interesen progresivamente por metodologías que les permitan conocer no ya tanto decisiones óptimas como las consecuencias de decisiones "políticas".

La experiencia de E.T.E.A. en este campo se centra fundamentalmente en dos áreas: (a) modelos de planificación de zonas, (b) modelos de desarrollo basados en el sector agrícola. Estos últimos son modelos de programación lineal multi objetivo sobre tablas Input-Output (5).

Con respecto a los modelos de zona son varios los realizados:

- a) Modelo para el análisis de las potencialidades de una zona. Realizado en 1975. Es un modelo de PL multiperiodica (3 años), con un total de 425 columnas, 299 filas y una densidad de 1'76%. El objetivo fundamental del



modelo era evaluar los efectos, desde el punto de vista de producción y empleo, de la puesta en riego de una zona rural.

- b) Modelo de colonización. Modelo a 2 años de programación lineal por metas realizado en 1976. El modelo permite analizar socio-económicamente la planificación y utilización de recursos en una zona de colonización.

La matriz incluye en su estructura el análisis de los siguientes aspectos:

- Dimensión de las explotaciones y su influencia
  - Actividades productivas a realizar (agrícolas y ganaderas).
  - Estudio de la mecanización y su influencia
  - Decisiones de inversión en maquinaria
  - Sistemas alimenticios a seguir
  - Decisiones de inversión en instalaciones y consiguiente nivel de actividades ganaderas.
  - Nivel de empleo y estacionalidad del mismo
  - Necesidades financieras de corto y largo plazo
  - Análisis de la posible realización colectiva de ciertas actividades (actividades ganaderas, fábrica de piensos etc.).
- c) Modelo multiobjetivo para la planificación de una comarca. Este modelo realizado en 1980 describe a continuación en mayor detalle.

### Modelo de Planificación de una Comarca Agrícola a Corto Plazo

El modelo que se presenta seguidamente formaba parte (planificación a corto plazo) de un proyecto mucho más ambicioso que abarcaba también el medio plazo (programación sobre tablas input-output) y el largo plazo (Dinámica de Sistemas) (el proyecto completo lleva como título "Elaboración de un Sistema de Modelos de Desarrollo Regional basado en el Sector Agrario") (2).

El modelo multiobjetivo es a su vez el resultado de una labor de trabajo previo, que puede resumirse como sigue:

#### 1º Recogida de información empírica.

La zona objeto del estudio, es la "Campiña" de Córdoba, aproximadamente unas 300,000 Has. de magnífica tierra de cultivo. Las explotaciones agrícolas de esta comarca han sufrido una gran transformación técnica en las últimas décadas. No obstante, hoy día representan una actitud empresarial con poca capacidad de riesgo. De hecho su estructura productiva, a base de cultivos tradicionales como cereales, viña y olivo, sólo ha sufrido algunas innovaciones significativas de nuevos cultivos (como son la remolacha, el algodón y el girasol) en la 2ª mitad del siglo XX.

Esta evolución, inmersa en la política económica nacional, ha tenido graves consecuencias sociales de migración, empleo estacionario y desempleo.

La recogida de datos necesarios sobre las explotaciones de la zona se hizo mediante una encuesta a una muestra. La información obtenida es de dos tipos: estructural - económica para permitir hacer una clasificación de las explotaciones, y técnica para obtener los coeficientes técnicos de la matriz.

#### 2º Clasificación (3)

Este programa es el encargo de clasificar las empresas encuestadas en grupos homogéneos. Estos grupos representan las características de la agricultura de la zona objeto de estudio.

#### 3º El Modelo

El modelo en sí, es un Programa Lineal Multiobjetivo. Como puede verse en la figura 2 presenta dos bloques diferenciados de actividades y restricciones. Aquellas que hacen referencia a los distintos tipos de explotaciones, y las que se refieren a la totalidad de la comarca (actividades y restricciones de agregación, tales como disponibilidades de capital, mercados, etc.).

- (3) El modelo de clasificación empleado es el del cálculo de funciones discriminantes biclásticas, cuyo autor es Antonio Vasquez Muñoz, del Centro de Investigación de IBM en la Universidad Autónoma de Madrid.

El cuadro 1 presenta los distintos tipos de explotaciones considerados y la extensión total que cada uno de ellos ocupa en la zona. Así mismo, el cuadro 2 señala las distintas posibilidades de cultivo en cada tipo de explotación.

Cuadro 1. Tipos Básicos de Explotaciones

TIPO	CARACTERISTICAS	EXTENSION TOTAL (HA)
I	Explot. pequeñas de regadío con cultivos herbáceos	5.240
II	Explot. grandes de regadío con cultivos herbáceos	6.460
III	Explot. pequeñas de secano con cultivos herbáceos	45.060
IV	Explot. medianas de secano con cultivos herbáceos	30.450
V	Explot. grandes de secano con cultivos herbáceos	57.690
VI	Explot. pequeñas con olivar y herbáceos	30.450
VII	Explot. medianas con olivar y herbáceos	20.570
VIII	Explot. grandes con olivar y herbáceos	38.980
IX	Explotaciones pequeñas de olivar	23.150
X	Explotaciones grandes de olivar	21.850
XI	Explotaciones de viña	12.000
XII	Explotaciones de viña y herbáceos	3.000

Cuadro 2. Relación de Cultivos

TIPOS DE FINCAS	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
CULTIVOS												
TRIGO	X	X	X	X	X	X	X	X				X
MAIZ	X	X										
SORGO	X	X										
HABAS	X	X	X	X	X	X	X	X				X
PATATAS	X	X										
REMOLACHA	X	X	X	X	X	X	X	X				X
ALGODON	X	X										
SOJA	X	X										
GARBANZOS	X	X	X	X	X	X	X	X				X
GIRASOL	X	X	X	X	X	X	X	X				X
CARTAMO	X	X	X	X	X	X	X	X				X
MELON	X	X	X	X	X	X	X	X				X
SANDIA	X	X	X	X	X	X	X	X				X
AJO	X	X	X	X	X	X	X	X				X
OLIVO						X	X	X	X	X		
VID											X	X

Las principales submatrices incluidas en el modelo son:

1) Submatriz técnica:

Recoge las limitaciones de tierra y de sucesión de cultivos.

2) Submatriz de consumo de factores:

Incluye el consumo de mano de obra, tracción, abonado, semillas, agua, productos fitosanitarios etc.

3) Submatriz de producción y venta.

4) Submatriz de riesgo

Recoge la incertidumbre en rendimientos que acompaña a la actividad productiva agraria. La metodología seguida es la de Boussard (2).

5) Submatriz financiera

Tiene la doble vertiente de poner de manifiesto las necesidades de crédito en la comarca, y poner límite a las mismas según los casos. El cuadro 3 recoge el esquema de esta submatriz.

6) Submatriz de empleo

Es un aspecto fundamental en el modelo. Las autoridades administrativas intentan encontrar una alternativa productiva para la comarca que genere más empleo que la actual, pero al mismo tiempo, intentando reducir la estacionalidad del mismo. El cuadro 4 recoge la estructura de esta submatriz.

7) La función objetivo

El algoritmo usado (4) permite optimizar un conjunto de objetivos y metas dentro de una estructura de prioridades. El usuario puede definir niveles de prioridad para los distintos objetivos y puede asignar distintos pesos a aquellos que están en un mismo nivel.

Así mismo el algoritmo distingue entre objetivos y metas, es decir entre funciones a optimizar y funciones a satisfacer. En el segundo caso el método

Quadro 3. Submatriz Financiera

	GASTOS	INGRE.	CDTO 1	SCAJA 1	CDTO 2	SCAJA 2	CDTO 3	SCAJA 3	CDTO 4	SCAJA 4	CDTO 5	SCAJA 5	CDTO 6	SCAJA 6	CDTO 7	SCAJA 7	RHS
PER1 <sub>n</sub>	(+)	(-)	-1	1													+DISP
PER2 <sub>n</sub>	(+)	(-)	1.02	-1	-1	1											-0
PER3 <sub>n</sub>	(+)	(-)			1.02	-1	-1	1									+0
PER4 <sub>n</sub>	(+)	(-)					1.02	-1	-1	1							+0
PER5 <sub>n</sub>	(+)	(-)							1.02	-1	-1	1					+0
PER6 <sub>n</sub>	(+)	(-)									1.02	-1	-1	1			+0
PER7 <sub>n</sub>	(+)	(-)											1.02	-1	-1	1	+0
PER1 <sub>tot</sub>			1														+DISP
PER2 <sub>tot</sub>				1													+DISP
PER3 <sub>tot</sub>					1												+DISP
PER4 <sub>tot</sub>								1									+DISP
PER5 <sub>tot</sub>									1								+DISP
PER6 <sub>tot</sub>													1				+DISP
PER7 <sub>tot</sub>															1		+DISP



permite elegir entre alcanzar la meta exactamente, alcanzarla o sobrepasarla, alcanzarla pero no sobrepasarla, alcanzarla pero que no sea menor, etc.

Los objetivos utilizados han sido el "Margen Bruto" (MB), el "Valor Añadido" (VA), la "Estacionalidad en el Empleo", y el "Riesgo".

El juego de estos objetivos y los niveles de prioridad permite simular distintas políticas agrarias para la comarca. La columna uno del cuadro 5 es el resultado de considerar una estructura de objetivos que supuestamente refleja la situación actual. Esta estructura incluye en el primer nivel de prioridad las 12 funciones económicas de los distintos tipos de explotaciones (margen bruto para las explotaciones grandes y valor añadido para las familiares). El segundo, tercer y cuarto

Cuadro 5. Resultados

	1	2	3	4
EXP-TYPE 1	$524 \times 10^6$ (AV)	$262 \times 10^6$ (GM)	$524 \times 10^6$ (AV)	$524 \times 10^6$ (AV)
EXP-TYPE 2	$388 \times 10^6$ (GM)	$388 \times 10^6$ (GM)	$581 \times 10^6$ (AV)	$388 \times 10^6$ (GM)
EXP-TYPE 3	$2.703 \times 10^6$ (AV)	$1.577 \times 10^6$ (GM)	$2.703 \times 10^6$ (AV)	$2.703 \times 10^6$ (AV)
EXP-TYPE 4	$1.065 \times 10^6$ (GM)	$1.065 \times 10^6$ (GM)	$1.827 \times 10^6$ (AV)	$1.065 \times 10^6$ (GM)
EXP-TYPE 5	$2.192 \times 10^6$ (GM)	$2.192 \times 10^6$ (GM)	$3.400 \times 10^6$ (AV)	$2.192 \times 10^6$ (GM)
EXP-TYPE 6	$1.827 \times 10^6$ (AV)	$790 \times 10^6$ (GM)	$1.827 \times 10^6$ (AV)	$1.827 \times 10^6$ (AV)
EXP-TYPE 7	$1.234 \times 10^6$ (AV)	$530 \times 10^6$ (GM)	$1.234 \times 10^6$ (AV)	$1.234 \times 10^6$ (AV)
EXP-TYPE 8	$1.013 \times 10^6$ (GM)	$1.013 \times 10^6$ (GM)	$2.300 \times 10^6$ (AV)	$1.013 \times 10^6$ (GM)
EXP-TYPE 9	$1.400 \times 10^6$ (AV)	$324 \times 10^6$ (GM)	$1.400 \times 10^6$ (AV)	$1.400 \times 10^6$ (AV)
EXP-TYPE 10	$306 \times 10^6$ (GM)	$305 \times 10^6$ (GM)	$1.311 \times 10^6$ (AV)	$306 \times 10^6$ (GM)
EXP-TYPE 11	$840 \times 10^6$ (AV)	$78 \times 10^6$ (GM)	$840 \times 10^6$ (AV)	$840 \times 10^6$ (AV)
EXP-TYPE 12	$210 \times 10^6$ (AV)	$60 \times 10^6$ (GM)	$210 \times 10^6$ (AV)	$210 \times 10^6$ (AV)
T.G. Margin	$7.800 \times 10^6$	$8.100 \times 10^6$	$7.600 \times 10^6$	$8.100 \times 10^6$
Empl. (UTH)	27.700	25.900	28.300	25.500
Risk	$6.900 \times 10^6$	$6.700 \times 10^6$	$7.200 \times 10^6$	$6.400 \times 10^6$

nivel recogen objetivos de empleo (respectivamente: (a) minimizar las desviaciones por debajo en el nivel de empleo de cada período con respecto a la media del año. (b) Minimizar en el nivel de empleo de cada período las desviaciones por encima con respecto a la media del año. (c) Minimizar la suma en valor absoluto de las desviaciones anteriores.

En el quinto nivel de prioridad se considera el riesgo.

Los resultados obtenidos con esta hipótesis son congruentes con la realidad.

Las columnas 2-4 reflejan otras hipótesis en la estructura de objetivos (distintas políticas agrarias en definitivas).

En la columna 2 se ha considerado el margen bruto como función económica de todos los tipos de explotación. En la columna 3 se considera el valor añadido como función económica general, lo que evidentemente incrementa el empleo. Finalmente la columna 4 recoge una hipótesis de un crecimiento relativamente lento en los precios de los factores agrícolas.

### Perspectivas de Futuro

A nivel de sector agrario la planificación de comarcas, y posibilidad de simulación de distintas "políticas" constituyen un aspecto fundamental. La programación matemática parece ser un instrumento adecuado a este respecto, en la problemática decisional a corto plazo. El medio plazo, y la ligazón de la planificación del sector agrícola con los demás sectores parece estar también resuelta con la programación sobre tablas Input-Output. Es en el largo plazo, en poner a disposición de la construcción del futuro, donde todavía queda un largo camino por recorrer.

### 3. Conclusion

La metodología de IO no se ha incorporado todavía a la mecánica decisional de la agricultura en España.

A nivel de sector agrario se ha avanzado algo más, porque el entorno es más favorable. Pero a nivel de empresa agrícola las dificultades ambientales son enormes. El miniordenador y programas adecuados representan quizás la vía de incorporación más prometedora en la situación actual.

### Referencias

- (1) ANDRADA y otros. "Estudio del impacto socioeconómico y análisis de rentabilidad del Plan de Puesta en Riego de la zona de Fuente Palmera (Córdoba). Edición Mimeografiada. INIA, 1983.
- (2) BOUSSARD J.M. and PETIT M. - "Representation of Farmer Behavior Under Uncertainty with Focus-Loss Constraint". *Journal of Farm Economics*, vol. 49 (1967).
- (3) CALATRAVA, Javier. - Metodología para construir un modelo lineal de decisión que optimice el empleo de recursos en una comarca agrícola. Aplicación a la zona de Alcaraz (Albacete). Tesis doctoral en la ESTIA de Madrid, dirigida por Jean Claud Tirel (INRA-FRANCIA), 1972-1973.
- (4) CORTES A. and VAZQUEZ A. - "Solution of Large Scale Linear Programming Problems with several Objective Functions". *Cahier d'Etudes et de Recherche*, vol. 80, 1980.
- (5) RODRIGUEZ ALCAIDE, J.J. y ROMERO RODRIGUEZ, J.J. - "La Economía Andaluza". *Papeles de Economía Española*, Nº 6, pp. 280, 1981.
- (6) RODRIGUEZ ALCAIDE, J.J.- "Ajuste óptimo a largo plazo de una empresa de secano de la provincia de Córdoba, mediante el uso de programación lineal dinámica". *Avigan*, números 158, 159, 160, 161 y 162, 1.958.
- (7) SANTIAGO BUJALANCE, Julián. - "Metodología para la planificación de cultivos en una zona de nuevos regadíos. Aplicación a la zona regable Genil-Cabra". Tesis doctoral. ETSIA, Córdoba, 1979.
- (8) SPITZER, H. y BLAS, J. - "Agricultural development. Plan Calatayud, Spain" (diciembre 1971).

### Notas:

- (1) - Véase referencias (1), (3), (6), (7), (8)
- (2) - En esta investigación participaron como Entidades colaboradoras El Centro de Investigación UAM - IBM (Madrid) IDR (Seilla), ETEA (Córdoba).

# AN APPLICATION OF THE METHODS OF SUCCESSIVE ORTHOGONAL PROJECTIONS TO A FEASIBILITY CONTROL PROBLEM

José Mário Martinez  
Antônio Carlos Moretti

Universidade Estadual de Campinas  
Departamento de Matemática Aplicada  
Caixa Postal 6065  
13081 Campinas - S.P.  
BRASIL

**Abstract:** In this paper the Method of Successive Orthogonal Projections of Gubin, Polyak and Raik is used to solve a feasibility control problem. At each subiteration of the method a Least Distance Quadratic Programming Problem is solved. Numerical experiments are presented, which confirm that this algorithm is useful tool to solve large scale convex feasibility problems.

## 1 - Introduction

In this article we analyze an application of the method of Successive Orthogonal Projections (S.O.P) to solve a problem related to educational planning.

The Simplified Mathematical Educational Model (S.M.E.M) (see [5]) simulates the student's flow through the educational system using the following **Dynamic Equations**:

$$Sc(t+1) = a1Sc(t) + a2Es(t) - V(1, t+1) - V(2, t+1) \quad (1)$$

$$Eu(1, t+1) = (a3-a4) Eu(1, t) + V(1, t+1) \quad (2)$$

$$Eu(2, t+1) = (a3-a5) Eu(2, t) + V(2, t+1) \quad (3)$$

$$Gu(t+1) = a6Gu(t) + a4Eu(1, t) + a5Eu(2, t) \quad (4)$$

for  $t = 0, 1, 2, \dots, T-1$

where:

$Sc(t)$  represents the number of persons which finished junior high school during year  $t$ , excluding those who became university students.

$Eu(1, t)$  and  $Eu(2, t)$  represent the number of undergraduate university students that work and do not work, respectively.

$Gu(t)$  represents the number of students that completed undergraduate studies.

$Es(t)$  represents the number of junior high school students.

$V(1, t)$  and  $V(2, t)$  are the decision variables of the model in year  $t$ . We assume that the Educational Planning Board has control of the rate of the access of the working and not working students to the universities. Of course, the rates of egress of students that work and do not work are different ( $a4 < a5$ ).

The rates  $a1, a3$  and  $a6$  are survival rates and  $a2, a4$  and  $a5$  are the egress rates.

The decisions have the purpose of satisfying the **Supply Restrictions**:

$$Sc(t) + wEu(1, t) \geq b1pop(t) + b2Ep(t) \quad (5)$$

$$Gu(t) \geq b3pop(t) + b4Es(t) + b5(Eu(1, t) + Eu(2, t)) \quad (6)$$

for  $t = 1, 2, \dots, T$  and  $0 \leq w \leq 1$

Inequalities (5) and (6) represent the population's necessity of the two types of professionals at period  $t$ . Coefficient  $w$ , which is less than 1, is used to express the fact that university students in the class  $EU(1, \cdot)$ , dedicate less time to professional activities than professionals which do not study.

The variables  $pop(t)$  and  $Ep(t)$  are exogenous variables of the model and represent the total population and the number of elementary school students in the year  $t$ . The coefficients  $b_1, \dots, b_5$  have the following meaning:

$b_2Ep(t)$  = number of primary teachers which is considered satisfactory.

$b_1Pop(t)$  = idem, professionals of secondary level, other than teachers.

$b_4Es(t)$  = idem, professors of high school.

$b_5(EU(1, t) + EU(2, t))$  = idem, university professors.

$b_3Pop(t)$  = idem, professionals of university level.

In addition to (5) and (6), all the variables of the model, except  $V(1, \cdot)$  and  $V(2, \cdot)$  are submitted to restrictions that avoid rough variations between two consecutive years:

$$c_1X(t) \leq X(t+1) \leq c_2X(t), \quad t = 0, 1, \dots, T-1 \quad (7)$$

where  $X(t)$  represents any of the variables  $Sc(t)$ ,  $EU(1, t)$ ,  $EU(2, t)$  and  $Gu(t)$ .

Moreover, the model requires that:

$$V(1, t+1) \geq V(1, t) \quad (8)$$

$$V(1, t+1) + V(2, t+1) \geq V(1, t) + V(2, t) \quad (9)$$

The restrictions (1-9) define a polytope in a finite dimension space whose elements are sequence of values for the variables of the model (except the exogenous variables). It is easy to see that the dimension of this space is  $6T$  and the total number of restrictions is  $16T-9$ .

The S.M.E.M. is a simplified version of a large scale model with additional refinements, which is considered in [5].

The system generated by the restrictions (1-9) has a very sparse pattern. For a big planning horizon it may be impossible to solve the problem using a program that handles all information at once. In this case the S.O.P. method may be the unique practical way to solve the system.

## 2 - The Method of Successive Orthogonal Projections

The method of Successive Orthogonal Projections (S.O.P) was introduced by Gubin et al [4] to solve the problem of Recently, Garcia - Palomares [3] applied the method to the problem of solving a set of linear equalities and inequalities, considering the associated convex set as defined by subsets of constraints. Essentially, the method consists in the following: we start with an arbitrary point, and we perform orthogonal projections in a cyclic way from a subsystem to another. The convergence of the sequence generated in this way is guaranteed if the intersection of the family is not empty (see [3, 4]).

To analyze the application of the S.O.P. method to our problem, we consider the system (1-9) within a more operational structure:

Let us write, for  $t = 0, 1, 2, \dots, T$ :

$$Z(t) = \begin{bmatrix} Sc(t) \\ Eu(1, t) \\ Eu(2, t) \\ Gu(t) \\ V(1, t) \\ V(2, t) \end{bmatrix}$$

We can consider the system as a set of three types of restrictions: Dynamic, Continuity and Spully Restrictions. Each variables of the model appears in restrictions of the three types. Let  $K(i, j)$  be the convex defined by restrictions where the variables corresponding to the years  $i, i+1, \dots, j$  appear. Thus,  $K(1, T)$  represents the convex defined by the equations (1-9). Note that  $Z(0)$  is a known initial condition. Avoiding unnecessary complications and cumbersome notation, we may describe the S.O.P. method as follows:

### Initialization Procedure:

- (1) Choose an arbitrary point ( $Z^*(t), t=1, 2, \dots, T$ ).
- (2) Let  $\epsilon$  be a small positive number.



- (3) Divide the system in  $q$  subsystems
- (4) Set  $h = (\text{Total number of years}) / q$
- (5) Set  $k = 0$

**Main Procedure:**

- (1) Set  $m = 1$
- (2) Set  $Z^* = Z^k$
- (3) If  $m = 1$  then set  $i = 1$  ; otherwise  $i = (m-1)h$
- (4)  $j = mh$
- (5) Replace  $Z^*$  by its orthogonal projection on  $k(i, j)$
- (6) If  $m = q$  then go to step 9
- (7) Set  $m = m + 1$
- (8) Go to step 3
- (9) Set  $k = k + 1$
- (10) Set  $Z^k = Z^*$
- (11) If  $\|Z^k(t) - Z^{k-1}(t)\|_{\infty} < \text{Eps}$ ,  $t = 1, 2, \dots, T$

then convergence was achieved ; Stop. Otherwise go to step 12. (12) Go to step 1

We note that step 5 in the main procedure is a Least Distance Quadratic Programming Problem, where only the variables  $Z(t)$ ,  $t = i, i+1, \dots, j$  are involved.

### 3 - The Numerical Solution of the Simplified Mathematical Educational Model

The Simplified Mathematical Educational Model can be solved using the method described in the previous section. In our experiments the coefficients  $a_i, b_j, c_k$  simulated the real conditions of the state of São Paulo, Brasil (a population of about 25 million people). Furthermore, we considered a period of  $T=15$  years and we tested the model with  $q=3$  and  $q=5$ . The Quadratic Programming Problems described in the previous section were solved using the MINOS package [8].

We obtained the starting point of the method setting values of the control variables  $V(1, t)$  and  $V(2, t)$  and computing the rest of the variables recursively using (1-4). Of course, this first trajectory does not necessarily satisfy inequalities (5-9).

Let us survey the results of the tests:

#### Case 1 : $q=3$ and $T=15$

We display the results obtained in this case in figures 1 to 5. They show the behaviour of the variables  $Sc(t)$  in iterations 1, 15, 30, 45 and 60, respectively.

In this case, the empirical convergence rate  $(\|Z^{k+1} - Z^k\|_{\infty} / \|Z^k - Z^{k+1}\|_{\infty})$ , for large  $k$  of the method was 0.943.

#### Case 2 : $q=5$ and $T=15$

Figures 6 to 10 show the behaviour of the variables  $Sc(t)$  in the same iterations as in the previous case.

The empirical convergence rate was 0.9629.

**Observation:** In both cases the other variables of the model convergence to their final shape in a similar way as  $Sc(t)$ . The corresponding graphs are available from the authors.

### 4 - Conclusions

The method of S.O.P. was used and generalized by several authors (see [1]). The slowness is its obvious limitation and convergence rates near of 1.0 are usual in this method. This is a fact which can be justified theoretically (see [2, 6, 7]).

The first iteration of the method provide a much better progress than the subsequent ones (see figure 7). This features limits the range of application of the method.

On the other hand, there are cases in which this method is the only feasible one and in which the high accuracy of the solution is not a major requirement.

The problem which we presented in this paper belongs to that class of problems and, in this sense, the results may be considered very promising. In fact, a very accurate solution is meaningless in economical or social models where uncertainty in data is high and the very structure of the model is dubious.

Furthermore, in models where decisions are taken each year only the solution in the first years is important, since, of course, the model will be run every year to update the decisions and the data. In our problem we observed that a trajectory close to the solution in the first years is obtained in iteration 20 for the case 1 and in the iteration 35 for case 2.

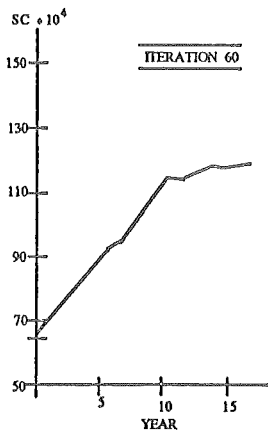
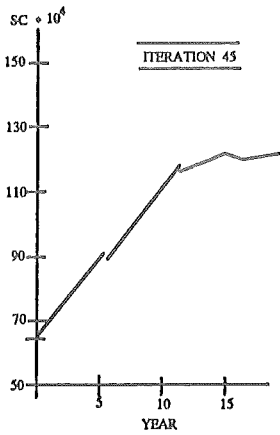
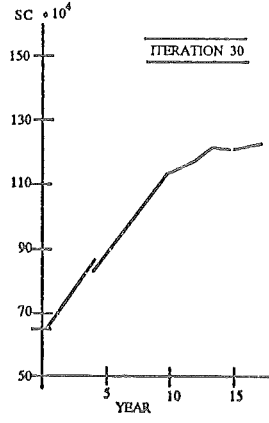
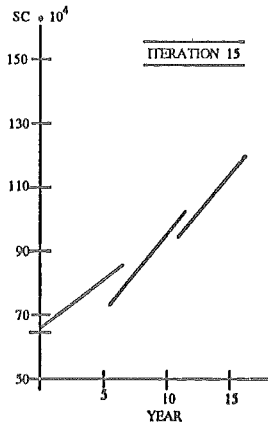
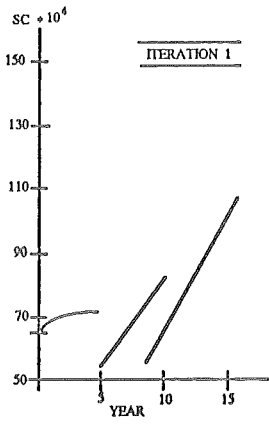
Finally, the comparison of the results between cases 1 and 2, confirms the intuitive feeling that the system must be divide in a small quantity of subsystems, in order to obtain better convergence rate (see figure 7).

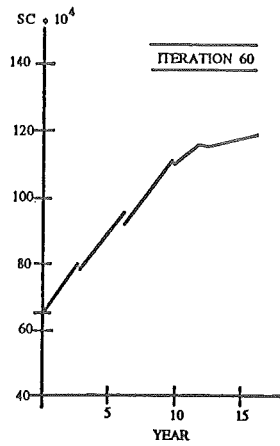
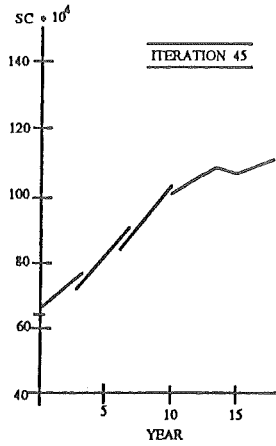
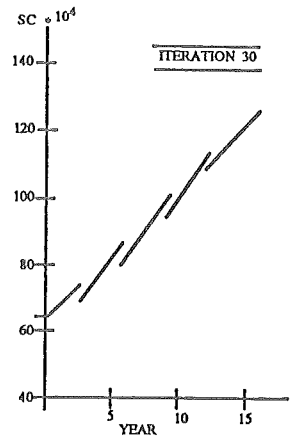
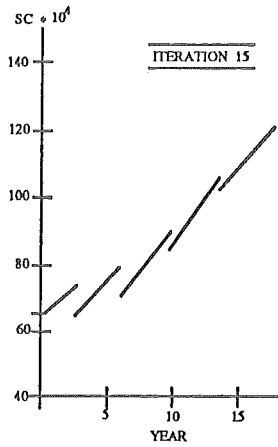
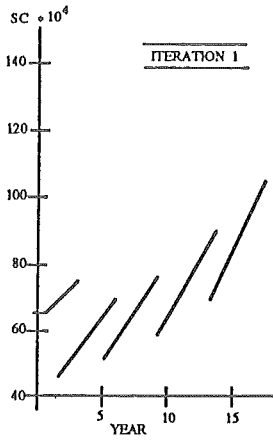
## 5 - Acknowledgement

The author are indebted to Flávia Brito Garboggini Moretti for the Drawings.

## References

- [1] Censor, Y.  
"Row-Action Methods for Huge and Sparse Systems and their Applications",  
SIAM Review 23, 444-446, 1981.
- [2] Gastinel, N.  
"Analyse Numérique Linéaire", Herman, Paris, 1966.
- [3] García-Palomares, U.  
"Q-Relaxation Method for Solving a System of Linear Inequalities",  
Tech.Mem nº19, Math and Computer Science Division, Argonne National  
Laboratory, November 1983.
- [4] Gubin, L.G., Polyak, R.T. & Raik, E.V.  
"Methods of Projections for Finding the Common Point of Convex Sets",  
U.S.S.R. Computacional Mathematics and Mathematical Physics, 7, 1-24,  
1967.
- [5] Martinez, J.M.  
"Modelo de Educação: Um Problema para o G.T.O.", Relatório Interno nº240  
IMECC, UNICAMP, 1983.
- [6] Martinez, J.M.  
"Solving Systems of Nonlinear Equations by Means of an Accelerated  
Successive Orthogonal Projections Method", to appear in Journal of  
Computacional and Applied Mathematics.
- [7] Martinez, J.M. and Sampaio, R.J.B.  
"Parallel and Sequential Kaczmarz Methods for Solving Under-Termined  
Nonlinear Equations", to appear in Journal of Computacional and Applied  
Mathematics.
- [8] Murtagh, B. and Saunders  
"Large Sclae Linearly Constrained Optimization", Mathematical Programming  
14, 41-72, 1978





## A SEQUENTIAL LCP ALGORITHM FOR BILINEAR AND CONCAVE QUADRATIC PROGRAMMING

J.J. Júdice\*

A.M. Faustino\*\*

\* Depart<sup>o</sup> de Matemática, Universidade de Coimbra

\*\* Depart<sup>o</sup> de Engenharia Civil, Universidade do Porto

### Resumo:

Neste artigo mostra-se que um algoritmo Sequencial LCP desenvolvido pelos autores para a resolução de Programas Lineares de Dois Níveis pode ser aplicado com pequenas modificações para a determinação de um mínimo global do Problema de Programação Bilinear (BLP). O algoritmo consiste em resolver uma sucessão de Problemas Lineares Complementares (LCP). O método pode também ser usado para a determinação de um mínimo global de um Programa Quadrático Côncavo (CQP) se este problema for primeiramente reduzido a um BLP. Experiência Computacional com CQPs e BLPs de pequena e média dimensões indica que o método é muito eficaz para obter um mínimo global. Contudo para problemas de dimensões mais elevadas torna-se difícil estabelecer que o mínimo global foi realmente encontrado.

### Abstract:

In this paper it is shown that a sequential LCP algorithm developed by the authors for the Bilevel Linear Programming Problem can be successfully applied with minor modifications to find a global minimum of a Bilinear Programming Problem (BLP). The algorithm relies on the solution of a finite number of Linear Complementarity Problems (LCP). The method can also be applied to find a global minimum of a Concave Quadratic Program (CQP) if this latter problem is firstly reduced to a BLP. Computational experience with medium-scale BLPs and CQPs indicates that this approach is quite efficient to find a global minimum. However, for problems of larger dimensions it is difficult to establish that the global minimum has been achieved.

### Keywords:

Bilinear Programming, Concave Quadratic Programming, Linear Complementarity Problem, SLCP algorithms, sparse matrices.

## 1 - INTRODUCTION

The Bilinear Programming Problem (BLP) can be defined as

$$\left. \begin{array}{l} \text{Minimize } c^T x + d^T y + x^T Q y \\ \text{subject to } \quad A x \geq a, E y \geq b \\ \quad \quad \quad x \geq 0 \quad y \geq 0 \end{array} \right\} \quad (1)$$

where  $c, x \in \mathbb{R}^{n_1}$ ,  $d, y \in \mathbb{R}^{n_2}$ ,  $A \in \mathbb{R}^{t_1 \times n_1}$ ,  $E \in \mathbb{R}^{t_2 \times n_2}$ ,  $a \in \mathbb{R}^{t_1}$ ,  $b \in \mathbb{R}^{t_2}$  and  $Q \in \mathbb{R}^{n_1 \times n_2}$ . The BLP is nowadays a well-known established optimization problem. A large number of applications of the BLP have been proposed [11] and a number of algorithms have been developed for its solution [6, 9, 10, 17, 18, 19].

Another important optimization problem is the so called Concave Quadratic Program (CQP) which is defined as

$$\left. \begin{array}{l} \text{Minimize } 2c^T x + x^T Q x \\ \text{subject to } \quad A x \geq b, x \geq 0 \end{array} \right\} \quad (2)$$

where  $c, x \in \mathbb{R}^n$ ,  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$ , and  $Q \in \mathbb{R}^{n \times n}$  is a symmetric negative semi-definite matrix, that is, satisfies  $x^T Q x \leq 0$  for all  $x \in \mathbb{R}^n$ . The CQP also has a large number of applications and several algorithms for the CQP have been proposed in recent years. We suggest [15] for a survey of the main applications and algorithms of the CQP. In [12], Konno has shown that the CQP (2) can always be reduced to a BLP of the following form

$$\left. \begin{array}{l} \text{Minimize } c^T x + d^T y + x^T Q y \\ \text{subject to } \quad A x \geq a, Ay \geq b \\ \quad \quad \quad x \geq 0, y \geq 0 \end{array} \right\} \quad (3)$$

As stated before, there are a large number of algorithms for finding global minima of BLP and CQP. Branch-and-bound, cutting-plane and ranking extreme points are usually the techniques that have been recommended for such purpose. These kind of algorithms may perform well for small problems but their performance tends to deteriorate enormously with an increase of the dimension. Recently, Pardalos and Rosen [15] have proposed a hybrid method for finding a global minimum of a CQP which relies on linear and 0-1 integer programming techniques. They report computational experience with medium-scale CQPs. It is however quite difficult to draw any final conclusions of the behaviour of this method, since the authors only provide the time as measure for efficiency.

The BLP can be shown to be equivalent to a Minimum Linear Complementarity Problem (MLCP) [9]. The MLCP is a nonconvex optimization problem that consists of minimizing a linear function subject to linear constraints and a complementarity condition  $z^T w = 0$  among nonnegative variables  $z$  and  $w$ . In [8] we have designed a sequential LCP algorithm for the solution of a MLCP that arises in the solution of the Bilevel Programming Problem [3]. The SLCP algorithm is a modification of a method proposed by Bialas and Karwan [3] and consists of solving a sequence of Linear Complementarity Problems (LCP). These LCPs are solved by a hybrid enumerative method [7,8] which incorporates Al-Khayyal modified reduced gradient procedure [1] among a number of other procedures and heuristic rules.

In this paper we show that this SLCP algorithm can be used with minor modifications to the solution of the MLCP that is equivalent to the BLP. This enables the solution of the BLP, and the CQP can also be processed by exploring its reduction to a BLP. We also describe a number of experiments with this approach for the solution of small and medium scale BLPs and CQPs with sparse structure. The computational experience indicates that the SLCP approach is quite efficient if the optimal value of the objective function of the

BLP and CQP is known. If this does not occur, then the method should be incorporated in a more general global solver. For instance the algorithm can be used as a procedure for finding an incumbent for the branch-and-bound method for the BLP and CQP described in [9].

The organization of the paper is as follows. In Section 2 the SLP algorithm for the BLP is described. A procedure for the solution of the first LCP and a hybrid enumerative method for the remaining LCPs are discussed in the next two sections. Finally, computational experience with small and medium scale BLPs and CQPs is presented in the remaining sections of the paper.

## 2 - A SLP ALGORITHM FOR THE BLP AND CQP

Consider the BLP as stated in (1) and assume that the constraint sets for the  $x$  and  $y$  variables

$$H_1 = \{x \in \mathbb{R}^{n_1} : Ax \geq a, x \geq 0\}$$

$$H_2 = \{y \in \mathbb{R}^{n_2} : Ey \geq b, y \geq 0\}$$

are bounded. Hence the BLP has a finite global minimum provided  $H_1$  and  $H_2$  are nonempty [10]. A global minimum  $(\bar{x}, \bar{y})$  of the BLP is also an optimal solution of the following Minimum Linear Complementarity Problem (MLCP) [9]

$$\left. \begin{array}{ll} \text{Minimize} & d^T y + a^T u \\ \text{subject to} & \alpha = c - A^T u + Qy \\ & \beta = -a + Ax \\ & \gamma = -b + Ey \end{array} \right\} \quad (4)$$

$$\alpha, \beta, \gamma, x, y, u \geq 0, \quad \alpha^T x = \beta^T u = 0$$



As in [8], we propose a modification of the Bialas and Karwan algorithm [3] for the solution of the MLCP (4). In this method, a parameter  $\lambda$  is introduced and the objective function is replaced by the constraint  $d^T y + a^T u \leq \lambda$  to obtain the following parametric LCP

LCP( $\lambda$ ):

$$\begin{bmatrix} \alpha \\ \beta \\ \gamma \\ v_0 \end{bmatrix} = \begin{bmatrix} c \\ -a \\ -b \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \lambda + \begin{bmatrix} 0 & -A^T & 0 \\ A & 0 & 0 \\ 0 & 0 & E \\ 0 & -a^T & -d^T \end{bmatrix} \begin{bmatrix} x \\ u \\ y \end{bmatrix} \tag{5}$$

$$x, y, u, \alpha, \beta, \gamma, v_0 \geq 0, \quad \alpha^T x = \beta^T u = 0$$

The global minimum  $(\bar{x}, \bar{y})$  of the BLP is the solution of the LCP  $(\bar{\lambda})$ , where  $\bar{\lambda}$  is the smallest value of  $\lambda$  such that LCP( $\lambda$ ) has a solution. To find  $(\bar{x}, \bar{y})$ , the method starts by solving the LCP( $\lambda_0$ ) obtained from LCP( $\lambda$ ) by omitting the constraint  $d^T y + a^T u \leq \lambda$ . This is done by a procedure to be explained in the next section. Let  $(x^0, y^0, u^0)$  be the solution of this LCP and let  $\lambda_0 = d^T y^0 + a^T u^0$ . Then the algorithm solves a sequence of LCPs( $\lambda_k$ ), where  $(\lambda_k)$  is a decreasing sequence defined by

$$\lambda_k = a^T u^{k-1} + d^T y^{k-1} - \gamma_k (a^T u^{k-1} + d^T y^{k-1}) \tag{6}$$

with  $(x^{k-1}, y^{k-1}, u^{k-1})$  the solution of LCP( $\lambda_{k-1}$ ) and  $\gamma_k$  a small positive number. The method terminates in a iteration  $k$  such that LCP( $\lambda_k$ ) has no solution. When this occurs, the solution  $(x^{k-1}, y^{k-1}, u^{k-1})$  of the LCP( $\lambda_{k-1}$ ) satisfies

$$0 \leq a^T u^{k-1} + d^T y^{k-1} - \text{VAL} \leq \gamma_k (a^T u^{k-1} + d^T y^{k-1}) \tag{7}$$

where VAL is the value of the objective function of the BLP at the optimal solution. Hence, if the constraint sets  $H_1$  and  $H_2$  of the BLP are nonempty and bounded, then the SLCP algorithm finds an  $\epsilon$ -optimal solution of the BLP, where

$$\epsilon = | \gamma_k (a^T u^{k-1} + d^T y^{k-1}) | \quad (8)$$

In practice, if  $\gamma_k$  is quite small, then the solution  $(x^{k-1}, y^{k-1})$  of the last LCP( $\lambda_{k-1}$ ) is usually the optimal solution of the BLP. As in [8] we recommend the choice of  $\gamma_k = 0.01$  in each iteration  $k$ .

The steps of the SLCP algorithm are as follows:

**1 Step 0** - Let  $k = 0$ .

**General Step** - Solve the LCP( $\lambda_k$ ). If LCP( $\lambda_k$ ) has no solution go to Exit. Otherwise let  $(x^k, y^k, u^k)$  be the solution of this LCP. Set

$$\lambda_{k+1} = a^T u^k + d^T y^k - | \gamma_{k+1} (a^T u^k + d^T y^k) | \quad (9)$$

where  $\gamma_{k+1}$  is a fixed value.

Set  $k = k+1$  and repeat.

**Exit** - If  $k = 0$ , then the BLP is infeasible, that is,  $H_1$  or  $H_2$  is empty. Otherwise  $(x^{k-1}, y^{k-1})$  is an  $\epsilon$ -global minimum of the BLP, where  $\epsilon$  is given by (8).

The efficiency of the SLCP algorithm depends essentially on the procedures to solve the LCPs( $\lambda_k$ ). In the next two sections we propose two procedures to solve the initial LCP( $\lambda_0$ ) and the LCPs( $\lambda_k$ ) that occur during the SLCP algorithm and we discuss their applicability to the solution of small and medium scale BLPs (and CQPs) with sparse structure.

3 - AN ALGORITHM FOR THE LCP( $\lambda_0$ ) [ 10 ]

Since the constraint  $a^T u + d^T y \leq \lambda$  does not exist in the LCP( $\lambda_0$ ), then we can write this problem in the following form

$$\begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} c \\ -a \\ -b \end{bmatrix} + \left[ \begin{array}{cc|cc} 0 & -A^T & 0 & 0 \\ A & 0 & 0 & 0 \\ \hline 0 & 0 & I & E \end{array} \right] \begin{bmatrix} x \\ u \\ y \end{bmatrix} \tag{10}$$

$$\begin{bmatrix} x \\ u \end{bmatrix} \geq 0, \quad \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \geq 0, \quad y \geq 0, \quad \gamma \geq 0, \quad \begin{bmatrix} x \\ u \end{bmatrix}^T \begin{bmatrix} \beta \\ \alpha \end{bmatrix} = 0$$

Due to the special structure of the LCP(10), then a solution to this LCP can be obtained by the following procedure:

- (i) Find a feasible solution  $\bar{y} \in H_2$ .
- (ii) Solve the LCP

$$\left. \begin{array}{l} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} c + Q\bar{y} \\ -a \end{bmatrix} + \begin{bmatrix} 0 & -A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix} \\ \begin{bmatrix} x \\ u \end{bmatrix} \geq 0, \quad \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \geq 0, \quad \begin{bmatrix} x \\ u \end{bmatrix}^T \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = 0 \end{array} \right\} \tag{11}$$

By the duality theory of linear programming [13, chapter 4], solving the LCP (11) is equivalent to finding an optimal solution of the following linear program

$$\left. \begin{array}{l} \text{Minimize } (c + Q\bar{y})^T x \\ \text{subject to } Ax \geq a, x \geq 0 \end{array} \right\} \quad (12)$$

Furthermore, since the choice of the feasible solution  $\bar{y} \in H_2$  is arbitrary, then we can compute  $\bar{y}$  by solving the linear program

$$\left. \begin{array}{l} \text{Minimize } (c + Q^T \bar{x})^T y \\ \text{subject to } Ey \geq b, y \geq 0 \end{array} \right\} \quad (13)$$

with  $\bar{x} = 0$ . Since  $\lambda_0 = (c + Qy^0)^T x^0 + d^T y^0 = (d + Q^T x^0)^T y^0 + c^T x^0$ , then a procedure to find a value of  $\lambda_0$  as small as possible consists of solving the linear programs (13) and (12) by turns until vectors  $\bar{x}$  and  $\bar{y}$  are found such that  $\bar{x}$  and  $\bar{y}$  are the optimal solutions of the linear programs (12) and (13). Since the value of  $(a^T u + d^T y)$  reduces when each one of the linear programs (12) and (13) is solved, then the procedure always terminates with a solution  $(x^0, y^0)$  of the LCP(10), provided the constraints sets  $H_1$  and  $H_2$  of the linear programs (12) and (13) are nonempty. Then we set  $\lambda_0 = (c + Qy^0)^T x^0 + d^T y^0$ , and we proceed with the SLCP algorithm.

In this START procedure only linear programs are solved, whence the simplex method can be used to solve such programs. Since the two problems (12) and (13) are solved a number of times, then the optimal Basis for a linear program (12) ((13)) can be the initial Basis for the next program of the same form. Furthermore, in order to exploit the sparsity of the BLP or CQP, the techniques for reinversion and updating LU decompositions of simplex Bases [16] can be used throughout the procedure START.

As is shown in the last section of this paper, computational experience with the algorithm START indicates that in general a small number of linear programs have to be solved. Furthermore, the value  $\lambda_0$  attained by this procedure is in general not very far from the optimal value of the objective function of the BLP or CQP.

### 4 - A HYBRID ENUMERATIVE METHOD FOR THE LCP( $\lambda_k$ )

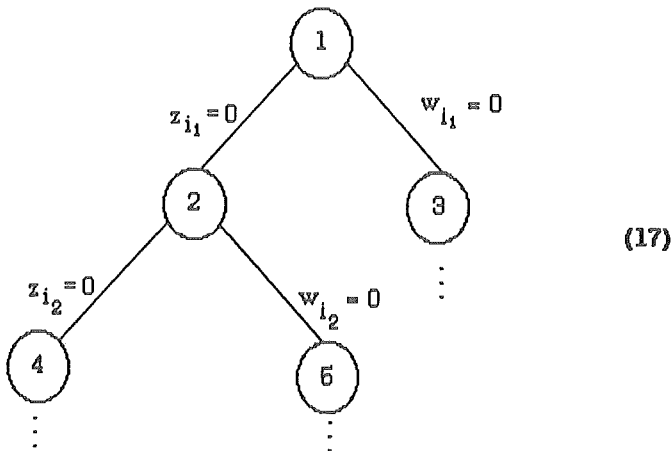
In this section we describe a hybrid enumerative method capable of solving the LCP( $\lambda_k$ ) with  $k \geq 1$ , that is required by the SLCP algorithm. Consider the LCP( $\lambda_k$ ) in the form

$$w = q + Mz + Ny \tag{14}$$

$$w \geq 0, z \geq 0, y \geq 0, \tag{15}$$

$$z_i w_i = 0, i = 1, 2, \dots, n. \tag{16}$$

where  $w \in \mathbb{R}^m$ ,  $z \in \mathbb{R}^n$  and  $y \in \mathbb{R}^{n_2}$ , with  $m = t_1 + t_2 + n_1 + 1$ ,  $n = t_1 + n_1$  and  $t_1, t_2, n_1, n_2$  defined as in (1). As in linear programming, a solution  $(z, w, y)$  satisfying the linear constraints (14) and (15) is called Feasible. A solution is Complementary if the variables  $z_i$  and  $w_i$  satisfy (16). An enumerative method attempts to find a complementary solution by only using basic feasible solutions of the system (14). To achieve this, the following tree is explored



where  $i_1, i_2, \dots$  are integer numbers of  $\{1, \dots, n\}$ . An initial feasible solution is obtained in node 1 by solving the linear program

$$\left. \begin{array}{ll} \text{Minimize} & z_0 \\ \text{subject to} & w = q + p z_0 + M z + N y \\ & z_0, z, y, w, \geq 0 \end{array} \right\} \quad (18)$$

where  $z_0$  is an artificial variable and  $p$  is a nonnegative vector satisfying  $p_i > 0$  for all  $i$  such that  $q_i < 0$ . This linear program is solved by a modification of the Phase 1 with a single artificial variable [13, pages 136-138]. This modified procedure is described in [8] and consists of minimizing the variable  $z_0$  in such a way that whenever possible the entering variable is chosen among the nonbasic variables whose complementary is also nonbasic. By doing this, we control the number of pairs of basic complementary variables and even a complementary solution can be obtained at the end of this procedure.

Each one of the nodes  $k$  with  $k \geq 2$  is generated by solving a subproblem that consists of minimizing a variable  $z_i$  or  $w_i$  subject to the linear constraints of the LCP and some constraints  $z_j = 0$  or  $w_j = 0$ . For instance, to generate node 4 of the tree (17) it is necessary to solve the linear program

$$\left. \begin{array}{ll} \text{Minimize} & z_{i_2} \\ \text{subject to} & w = q + M z + N y, z \geq 0, y \geq 0, w \geq 0 \\ & z_{i_1} = 0 \end{array} \right\} \quad (19)$$

Such linear program is solved by a modification of Phase 2 of the simplex method that exploits the same idea of controlling the number of pairs of basic complementary variables. Two cases may occur:

- (i) If the variable minimized has value equal to zero, then is fixed at zero in all descendent paths of the tree.

- (ii) If the minimum value of the variable is positive, then the branch is pruned and the node is fathomed.

The enumerative method attempts to solve the LCP by generating successive nodes of the tree, according to the process explained above. The algorithm either finds a solution for the LCP (it is the first complementary feasible solution) or establishes that the LCP has no solution (all the generated nodes of the tree are fathomed).

The enumerative method is efficient only if few nodes are generated before a complementary solution is found or it is verified that none exists. There exist some heuristic rules and some procedures that usually improve the efficiency of the algorithm. The heuristic rules are related with the choice of the node of the tree and of the pair  $(z_{i_k}, w_{i_k})$  and are discussed in [7]. The algorithm for generating nodes described before is an example of a procedure that reduces the search for a complementary solution. Next, we briefly discuss another technique that has been introduced by Al-Khayyal [1] and is applied in each node for the same purpose of reducing the overall search.

Al-Khayyal's algorithm is a modified reduced gradient method that finds a local star minimum of the function

$$f(z,w,y) = \sum_{i=1}^n z_i w_i$$

that is, a basic feasible solution  $(\bar{z}, \bar{w}, \bar{y})$  such that

$$f(\bar{z}, \bar{w}, \bar{y}) \leq f(z,w,y)$$

for all its adjacent basic feasible solutions  $(z,w,y)$  of the feasible set of the LCP

$$\{(z, w, y) : w = q + Mz + Ny, z \geq 0, w \geq 0, y \geq 0\}$$

If  $(\bar{z}, \bar{w}, \bar{y})$  is a basic feasible solution and  $(\tilde{z}, \tilde{w}, \tilde{y})$  is an adjacent basic solution, then we can write

$$(\tilde{z}, \tilde{w}, \tilde{y}) = (\bar{z}, \bar{w}, \bar{y}) + \mu_0 (d^z, d^w, d^y)$$

where  $\mu_0$  is the minimum ratio of the simplex method and  $d = (d^z, d^w, d^y)$  is a feasible direction such that  $d^z$ ,  $d^w$  and  $d^y$  are the vectors of the components of  $d$  corresponding to the  $z$ ,  $w$  and  $y$  variables respectively. Due to the special structure of the LCP( $\lambda_k$ ) presented in (5), we can use a proof similar to [8] to show that

$$f(\tilde{z}, \tilde{w}, \tilde{y}) - f(\bar{z}, \bar{w}, \bar{y}) = \mu_0 \sum_{i=1}^n (\bar{z}_i d_i^w + \bar{w}_i d_i^z) = -\mu_0 \bar{e}_s$$

where  $e_s$  is the reduced-cost coefficient of the nonbasic variable of index  $s$  associated with the linear function

$$\sum_{i=1}^n (\bar{z}_i w_i + \bar{w}_i z_i) \quad (20)$$

Therefore Al-Khayyal's modified reduced-gradient algorithm is in this case a simplex-type method in which the reduced cost coefficients  $\bar{e}_j$  of the nonbasic variables are associated with the linear function (20). We also incorporate a Bland's rule [4] to avoid the occurrence of cycling in the degenerate pivot operations [7].

We suggest [8] for a complete description of the hybrid enumerative method that incorporates this latter algorithm, the procedures for generation of nodes and the heuristic rules for choice of nodes and the



complementary pair of variables. The hybrid enumerative method can be implemented for solving large and sparse LCPs. The implementation uses reinversion and updating techniques for the LU decomposition of the Basis matrices [16] used by the simplex-type procedures and special data structures. We recommend [7] for a description of this implementation.

As discussed in section 2, the hybrid enumerative method is used to solve the LCPs( $\lambda_k$ ) required by the SLCP algorithm. For any two values  $\lambda_k < \lambda_{k-1}$ , the LCP( $\lambda_{k-1}$ ) and LCP( $\lambda_k$ ) only differ in the last component of the vector  $q$ . Furthermore, if  $B$  is the Basis associated with the solution  $(\bar{z}, \bar{w}, \bar{y})$  of the LCP( $\lambda_{k-1}$ ) obtained by the hybrid enumerative method and  $q$  is the right-hand side of the LCP( $\lambda_k$ ), then the vector  $\bar{q} = B^{-1} q$  satisfies

$$\bar{q}_j \geq 0 \text{ for all } j = 1, \dots, m-1 \quad \text{and} \quad \bar{q}_m < 0 \quad (21)$$

Hence the solution  $(\bar{z}, \bar{w}, \bar{y})$  of the LCP( $\lambda_{k-1}$ ) can be used as the initial basic solution for the LCP( $\lambda_k$ ). Since the vector  $B^{-1} q$  satisfies (21), then the vector  $p$  introduced in (17) can be defined by

$$p = B \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

where  $0 \in \mathbb{R}^{m-1}$  is the null vector. The choice of this initial basic solution usually provides great reductions in the computational effort of the hybrid enumerative method. In fact the algorithm requires quite often a unique node and a quite small number of pivot operations to find the solution of the LCP( $\lambda_k$ ). This is shown by the computational results presented in the next section.

**5 - COMPUTATIONAL EXPERIENCE**

In this section we present some computational experience with the SLCP algorithm described in this paper in the solution of small and medium scale BLPs and CQPs. The experiences have been performed in a CDC CYBER 180-830 of University of Porto and their results are presented in tables 1,2,3 and 4. In these tables, there are some parameters with the following meanings:

<b>NVX</b>	= number of variables $x$
<b>NRX</b>	= number of constraints of $H_1$ = number of rows of $A$
<b>NVY</b>	= number of variables $y$
<b>NRV</b>	= number of constraints of $H_2$ = number of rows of $E$
<b>NC</b>	= number of complementary pairs of variables of $LCP(\lambda_k)$
<b>NRW</b>	= number of rows of the matrix $[M \mid N]$ of $LCP(\lambda_k)$
<b>NCL</b>	= number of columns of the matrix $[M \mid N]$ of $LCP(\lambda_k)$
<b>NLP</b>	= number of linear programs to be solved by the algorithm START
<b>NLCP</b>	= number of $LCPs(\lambda_k)$ to be solved by the SLCP algorithm
<b>NI</b>	= total number of simplex pivot operations required by the START or SLCP algorithms
<b>NIMIN</b>	= minimum number of simplex pivot operations that a $LCP(\lambda_k)$ has required
<b>NIMAX</b>	= maximum number of simplex pivot operations that a $LCP(\lambda_k)$ has required
$\lambda_0$	= value of $\lambda_0$ after the algorithm START.
$\lambda_l$	= value of $\lambda_l$ for the last $LCP(\lambda_l)$ with a complementary solution that has been solved by the SLCP algorithm
<b>T</b>	= CPU time in seconds for the START or SLCP algorithms

$\overline{TL}$  = CPU time in seconds for the  $LCP(\lambda_k)$  without solution  
 (N means that no conclusion about the existence of  
 solution to the  $LCP(\lambda_k)$  has emerged after 2000 pivot  
 operations)

Table 1 contains the results of the first experience of solving BLPs by the SLCP algorithm. The matrices  $A$ ,  $E$  and  $Q$  of the test problems have been randomly generated by a technique similar to that described in [8]. The results confirm the experiences of solving Bilevel programs by this kind of approach [8]. As before, the efficiency of the algorithm depends on the dimension of the BLP. However, it is the number  $NC$  of pairs of complementary variables that mainly influentiates the behaviour of the SLCP algorithm. Since in the reformulation of a BLP into a MLCP [9] does not matter to choose the variables  $x$  or  $y$  to become the complementary variables, then we recommend the choice of the variables  $x$  if  $(NRX + NVX) < (NRY + NVY)$  and the  $y$  variables otherwise.

Another important conclusion of the results presented in Table 1 is the good behaviour of the START algorithm. The method usually provides a value  $\lambda_0$  that is not very far from the optimal value of the objective function of the BLP. The SLCP algorithm also performs well and usually requires the solution of a small number of  $LCPs(\lambda_k)$ . The value NIMIN is always equal to 3 and we should add that this number has been achieved more than once for some BLPs. The value NIMAX is usually large and is in general close to the total number of pivot operations required by the SLCP algorithm. This means that for each BLP the SLCP algorithm has to solve a difficult  $LCP(\lambda_k)$ . The experience also shows that this difficult LCP usually occurs in the last stages of the SLCP algorithm. The computational effort to establish that the last  $LCP(\lambda_k)$  has no solution is usually much superior to the remaining part of the SLCP algorithm and becomes prohibitive even for small BLPs. This is because of the great amount of tree search that is necessary to stop the hybrid enumerative method in this case.

Based on the results of the first experience, we have decided to generate some test problems for which the optimal value of the objective function is known. To do this, we have used an equivalence between BLPs and LCPs established in [2]. By this result, any LCP of the form

$$w = q + Mz, \quad z \geq 0, \quad w \geq 0, \quad z^T w = 0 \quad (22)$$

with  $z \in \mathbb{R}^n$  and  $w \in \mathbb{R}^n$  is equivalent to the BLP

$$\left. \begin{array}{l} \text{Minimize} \quad g(z,x) = e^T z + q^T x + x^T (M - I_n) z \\ \text{subject to} \quad Mz + q \geq 0, \quad x \leq e \\ \quad \quad \quad z \geq 0, \quad x \geq 0 \end{array} \right\} \quad (23)$$

where  $e = (1, \dots, 1)^T \in \mathbb{R}^n$  and  $I_n$  is the identity matrix of order  $n$ . Furthermore,  $\bar{z}$  is solution of the LCP (22) if and only if  $(\bar{z}, \bar{x})$  is global minimum of the BLP (23) and  $g(\bar{z}, \bar{x}) = 0$ . In our second experience we considered four LCPs taken from [7], transformed them into their equivalent BLPs and solved the resulting problems by the SLCP algorithm. In this experience we have decided not to allow more than 1000 pivot operations for each LCP( $\lambda_k$ ). The results of the experience are presented in Table 2. The SLCP and START algorithms have a similar behaviour than before. The SLCP algorithm has achieved a global minimum of the BLP in two cases and a value close to zero in the two remaining test problems. So these results confirm the ability of the SLCP algorithm for achieving a feasible solution for the BLP quite close to the global minimum in a reasonable amount of work.

In Tables 3 and 4 we present the results of the experiences of solving QCPs with a known optimal value by the SLCP algorithm. To generate such problems, we use a property [5] that states that any LCP is equivalent to a Quadratic Programming Problem of the form

$$\left. \begin{array}{l} \text{Minimize} \quad h(z) = q^T z + \frac{1}{2} z^T (M + M^T) z \\ \text{subject to} \quad q + M z \geq 0, z \geq 0 \end{array} \right\} \quad (24)$$

in the sense that  $\bar{z}$  is a solution of the LCP (22) if and only if  $\bar{z}$  is a global minimum of the quadratic program (24) and  $h(\bar{z}) = 0$ . Based on this equivalence, we can consider LCPs whose matrix  $M$  is negative semi-definite to generate Concave Quadratic Programs with a known optimal value. The test problems presented in Table 3 have been generated in this way by using matrices taken from [14] and multiplying all the entries of these matrices by  $-1$ . We have obviously to reduce these CQPs to BLPs of the form (3) in order to apply the SLCP algorithm.

The results presented in Table 3 show that the START and SLCP algorithms are quite efficient for these test problems. The SLCP algorithm always finds a global minimum for the CQP in a small amount of work.

The last experience has been performed with CQPs taken from [15, pages 15-16] of the form

$$\begin{array}{l} \text{Minimize} \quad (c - e)^T x - \frac{1}{2} x^T I_n x \\ \text{subject to} \quad 0 \leq x \leq 2e \end{array}$$

where  $e = (1, \dots, 1)^T \in \mathbb{R}^n$ ,  $I_n$  is the identity matrix of order  $n$  and  $c \in \mathbb{R}^n$  is a randomly generated positive vector. As stated in [15, page 16], there are  $3^n$  Kuhn-Tucker points for this quadratic program. Based on this, Pardalos and Rosen claim that an approach relying on complementarity is not efficient. The results have shown that even for this kind of problems the SLCP algorithm can do the job efficiently. The method always finds an  $\epsilon$ -optimal solution with  $\epsilon$  defined by (8). The results presented in Table 3 also show a similar behaviour of the SLCP algorithm for these test problems. However, there is a bigger gap between NI and NIMAX and a slightly increase in the number of LCPs( $\lambda_k$ ) required by the SLCP algorithm. These two

differences are not surprising because of the large number of Kuhn-Tucker points (that is, complementary solutions of the respective MILCP).

As a final conclusion from this computational study, we can claim that the SLCP algorithm incorporating the START algorithm is an efficient procedure to find global minima for BLPs and CQPs. In almost all the problems with a known optimal value the algorithm was capable of finding a global minimum. The SLCP algorithm behaves much better for more realistic problems than for totally randomly generated ones. The major drawback of the method relies on the establishment that the last  $LCP(\lambda_k)$  has no solution. This is not obviously necessary if the optimal value of the nonconvex problem is known a priori. If we do not have such information, then the SLCP algorithm should be used as an efficient way of finding an excellent incumbent for a branch-and-bound method like the one described in [9]. We are currently doing research on this topic.

P r o b l e m	DIMENSIONS							START				SLCP						TL
	NRX	NVX	NRV	NVY	NC	NRW	NCL	NLP	NI	T	$\lambda_0$	NLCP	NI	NIMIN	NIMAX	T	$\lambda_k$	
1	1	20	30	50	21	52	71	3	38	.3	-27	9	76	3	19	2.3	-38.5	12.3
2	1	20	30	100	21	52	121	5	135	21	-119.2							31.9
3	1	30	30	100	31	62	131	3	110	1.9	-148.3	2	288	3	285	13.9	-184.8	56.3
4	1	40	30	100	41	72	141	3	134	2.5	-215.9							92.4
5	1	40	50	200	41	92	241	3	498	24	-218.8	52	3108	3	1485	266.2	-434.4	N
6	1	40	30	200	41	72	241	3	111	2.2	-346.1	12	674	3	569	44	-4299	N
7	1	50	30	200	51	82	251	3	144	3.4	-332.1	14	768	6	469	55	-532.6	N
8	10	30	30	100	40	71	140	3	128	2.0	-91.16	9	680	4	628	36	-121.4	N
9	10	30	30	150	40	71	190	5	193	3.4	-153.5							N

TABLE 1

P r o b l e m	DIMENSIONS								START				S L C P					
	NRX	NVX	NRV	NVY	NC	NRW	NCL	NLP	NI	T	$\lambda_0$	NLCP	NI	NIMIN	NIMAX	T	$\lambda_\lambda$	
10	40	40	40	40	80	121	120	3	89.93	15.4	15	861	3	726	50.	0.8		
11	40	40	40	40	80	121	120	5	85.10	17.8	49	2314	3	967	243.	0.		
12	50	50	50	50	100	151	150	5	1604.6	15.6	46	380	3	88	50.8	0.83		
13	30	30	30	30	60	91	90	3	53.38	16.6	25	2129	3	861	110.	0.		

TABLE 2

P r o b l e m	DIMENSIONS								START				S L C P					
	NRX	NVX	NRV	NVY	NC	NRW	NCL	NLP	NI	T	$\lambda_0$	NLCP	NI	NIMIN	NIMAX	T	$\lambda_\lambda$	
14	30	30	30	30	60	91	90	3	30.4	14.8	1	18	18	18	1.1	0.		
15	30	30	30	30	60	91	90	3	28.3	10.1	7	47	3	19	2.9	0.		
16	100	100	100	100	200	301	300	3	1615.5	10.8	8	60	3	32	10.1	0.		
17	100	100	100	100	200	301	300	5	30033	11.4	20	265	3	102	73.4	0.		
18	200	200	200	200	400	601	600	3	673.3	0.								
19	200	200	200	200	400	601	600	9	249.21	3.9	14	193	3	84	74.9	0.		
20	300	300	300	300	600	901	900	3	409.77	2.3	2	16	3	13	5.2	0.		
21	300	300	300	300	600	901	900	7	336.29	25.9	23	259	3	95	117.	0.		

TABLE 3

P r o b l e m	DIMENSIONS								S L C P				
	NRX	NVX	NRV	NVY	NC	NRW	NCL	NLCP	NI	NIMIN	NIMAX	T	
22	50	50	50	50	100	151	150	46	1783	6	149	166.	
23	50	50	50	50	100	151	150	46	418	6	38	41.	
24	100	100	100	100	200	301	300	60	1727	6	183	278.	
25	100	100	100	100	200	301	300	28	1949	6	310	329.	
26	150	150	150	150	300	451	450	101	3953	6	333	759.	

TABLE 4

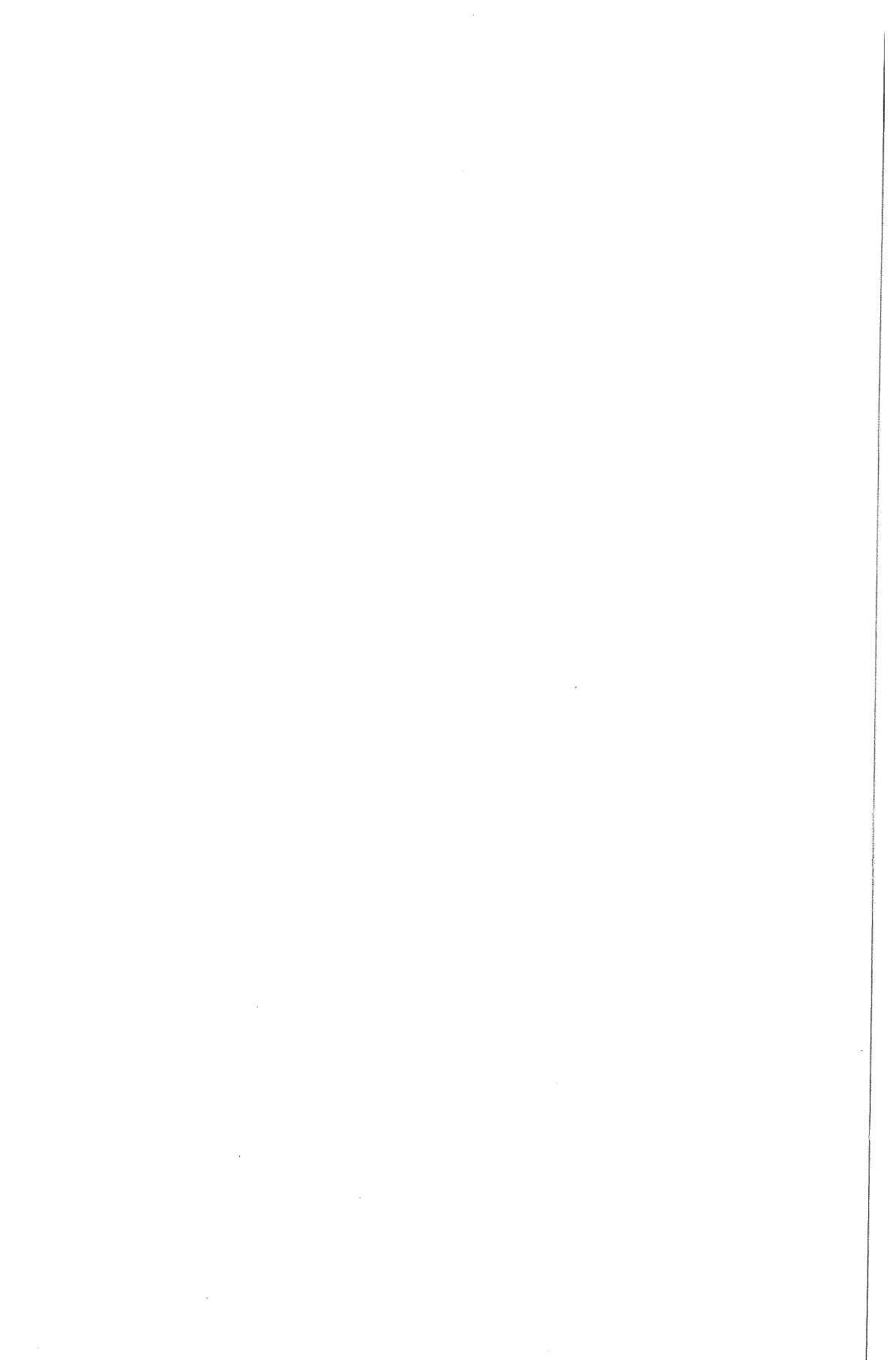
## REFERENCES

- [1] - F. A. Al-Khayyal, An implicit enumeration procedure for the general linear complementarity problem, *Mathematical Programming Study* 31 (1987) 1-20.
- [2] - F. A. Al-Khayyal, Further simplified characterizations of Linear Complementarity Problems solvable by linear programs, Technical Report PDRC 86-08, School of Industrial and Systems Engineering, Georgia Institute of Technology, 1986.
- [3] - W. F. Bialas and M. H. Karwan, Two-level linear programming, *Management Science* 30 (1984) 1004-1020.
- [4] - R. C. Bland, New finite pivoting rules for the simplex method, *Mathematics of Operations Research* 2 (1977) 103-107.
- [5] - R. W. Cottle, Note on a fundamental theorem in quadratic programming, *Journal SIAM* 12 (1964) 663-665.
- [6] - G. Gallo and A. Ulkucu, Bilinear Programming: an exact algorithm, *Mathematical Programming* 12 (1977) 173-194.
- [7] - J. J. Júdice and A. M. Faustino, An experimental investigation of enumerative methods for the linear complementarity problem, *Computers and Operations Research* 15 (1988) 417-426.
- [8] - J. J. Júdice and A. M. Faustino, The solution of the Linear Bilevel Programming Problem by using the linear complementarity problem, to appear in *Investigação Operacional*.
- [9] - J. J. Júdice and G. Mitra, Reformulation of mathematical programming problems as linear complementarity problems and investigation of their solution methods, *Journal of Optimization Theory and Applications* 57 (1988) 123-149.
- [10] - H. Konno, A cutting-plane algorithm for solving bilinear programs, *Mathematical Programming* 11 (1976) 14-27.
- [11] - H. Konno, Bilinear Programming: Part 2 - Applications of Bilinear Programming, Technical Report 71-10, Department of Operations Research, Stanford University, 1971.
- [12] - H. Konno, Maximization of a convex quadratic function under linear constraints, *Mathematical Programming* 11 (1976) 177-127.
- [13] - K.G. Murty, *Linear and Combinatorial Programming*, Wiley, 1976.
- [14] - J. S. Pang, A new and efficient algorithm for a class of portfolio selection problems, *Operations Research* 28 (1980) 769-767.
- [15] - P. M. Pardalos and J. B. Rosen, *Constrained Global Optimization: Algorithms and Applications*, Lecture Notes in Computer Science 268,



Springer-Verlag, Berlin Heidelberg, 1987.

- [16] -J. K. Reid, A sparsity-exploitation variant of the Bartels-Golub decomposition for linear programming bases, *Mathematical Programming* 24 (1982) 55-69.
- [17] -H. D. Sherali and C. M. Shetty, A finitely convergent algorithm for bilinear programming problems using polar cuts and disjunctive face cuts, *Mathematical Programming* 19 (1980) 14-31.
- [18] -T. V. Thieu, A note on the solution of bilinear programming problems by reduction to concave minimization, *Mathematical Programming* 41 (1988) 249-260.
- [19] -H. Vaish and C. M. Shetty, A cutting-plane algorithm for the bilinear programming problem, *Naval Research Logistics Quarterly* 24 (1977) 83-94.





## ÍNDICE

<i>José Paixão, Margarida Pato</i> Primal and Dual Greedy Heuristics for the Generalized Set Covering Problem . . . . .	3
<i>J. Caliço Grosso</i> Modelo de Simulação de Tráfego: Caso do Transporte Metropolitano . . . . .	13
<i>R. C. Souza, A. L. B. Squadri</i> Uma Aplicação de Combinação de Previsões Subjectivas para o Mercado Internacional de Petróleo — Abordagem Bayesiana . . . . .	25
<i>Eduardo Saliby</i> Amostragem Descritiva: Uma Nova Abordagem em Simulação . . . . .	33
<i>Cassilda M. Ribeiro, Jurandir F. R. Fernandes</i> <i>Hermano de Medeiros F. Tavares</i> Programação de Alocação de Estágios de Linha Remotos em Redes Telefónicas Análogo-Digitais . . . . .	41
<i>C. Lyra, H. Tavares</i> Optimização de Funções Não-Lineares de Fluxos em Redes: Um Método Eficiente para o Planeamento da Operação a Médio Prazo de Sistemas Hidroelétricos . . . . .	47
<i>Francisco Amador Hidalgo, Manuel Cabanes Fuentes</i> <i>José J. Romero Rodríguez</i> La Investigación Operativa en la Agricultura Espanola . . . . .	51
<i>José Mário Martinez, António Carlos Moretti</i> An Application of the Methods of Successive Orthogonal Projections to a Feasibility Control Problem . . . . .	61
<i>J. J. Júdice, A. M. Faustino</i> A Sequential LCP Algorithm for Bilinear and Concave Quadratic Programming . . . . .	67



Associação Portuguesa para o Desenvolvimento  
da Investigação Operacional.

CÉSUR – Instituto Superior Técnico – Avenida Rovisco Pais  
1000 Lisboa – Telef. 80 74 55