La Revue de Modulad

Quelques Exemples d'Application Statistique du Logiciel d'Optimisation GINO

Antoine de Falguerolles
Laboratoire de Statistique et Probabilités,
Université Paul Sabatier,
118, Route de Narbonne,
31 062 Toulouse Cedex.

Résumé

Il arrive qu'une méthode d'estimation ne soit pas directement offerte par les logiciels statistiques usuels ou immédiatement disponibles ou que sa programmation dans les environnements définis par ces logiciels soit trop lourde pour être envisagée. Il se peut encore que l'on veuille contrôler les résultats d'un logiciel ou d'un programme ad hoc *sur des exemples tests de taille raisonnable. L'objet de cet article est de décrire, sous forme d'exemples, comment le logiciel d'optimisation non linéaire **GINO** peut être utilisé dans un contexte statistique pour obtenir des estimations variées.*

Mots clés : *Optimisation Non Linéaire, Tableaux Carrés, Modèle d'Homogénéité des Marges, Modèle de Quasi-Symétrie et d'Asymétrie, Analyse en Composantes Principales.*

1 Introduction

L'objet de cet article est de présenter des exemples d'utilisation d'un logiciel d'optimisation non linéaire dans le contexte statistique de recherche d'estimations (maximum de vraisemblance, moindres valeurs absolues ...). En effet il peut se produire que le traitement statistique correspondant ne soit pas directement offert par les logiciels usuels ou disponibles ou que sa programmation dans les environnements définis par ces logiciels soit trop lourde pour être envisagée. On peut encore vouloir contrôler les résultats d'un logiciel ou d'un programme *ad hoc* sur des exemples tests de taille raisonnable.

Une situation typique est celle de l'analyse des tables de contingence. Par exemple, l'analyse de la dépendance dans les tables de contingence à deux entrées peut être abordée en posant que les effectifs conjoints sont les valeurs observées de variables aléatoires indépendantes et distribuées selon des lois de Poisson dont les paramètres sont alors modélisés. Le modèle d'association $R \times C$ de Goodman et le modèle de corrélation canonique sont des exemples connus de cette démarche qui rejoint sous bien des aspects celle, plus exploratoire, de l'analyse factorielle des correspondances (Goodman, 1991). Dans toutes ces approches, un logiciel d'optimisation est susceptible de fournir des estimations pour les scores des modalités lignes ou colonnes, notamment dans des situations où ceux-ci sont assujettis à des contraintes spécifiques (Rao et Galigiuri, 1992). Il existe bien d'autres modèles dont l'estimation peut être envisagée au moyen d'un tel logiciel. Nous en présentons certains dans cet article.

Les traits principaux du logiciel **GINO** sont décrits dans la section 2. Puis nous rendons compte de l'application de ce logiciel à des problèmes statistiques concrets. Nous considérons dans la section 3 un exemple de tableau carré et étudions l'ajustement par maximum de vraisemblance du modèle d'homogénéité des marges, puis d'un modèle généralisant celui de quasi-symétrie pour y incorporer une forme d'asymétrie. Dans la section 4, nous rapportons une analyse en composantes principales de type L_1.

2 Le Logiciel GINO

Le logiciel présenté est un logiciel d'optimisation non linéaire appelé **GINO** (*The Scientific Press*, 540 University Avenue, Palo Alto, CA 94301, USA). Il permet d'optimiser une fonction objectif d'expression assez générale sous des contraintes assez générales aussi. **GINO** utilise une méthode de gradient réduit généralisé **GRG2** (Abadie (1978), Lasdon et Waren (1978)). Il existe différentes adaptations de ce logiciel suivant le type d'ordinateur (PC, Mac, Infocentre) ou le type d'application (étudiante, professionnelle).

La version que nous considérons ici est dite **Super GINO** pour PC (version du 3 Février 1988). La disquette contient les fichiers suivants:

```
Repertoire de A:\

GINO     EXE    217922 19/06/90    13:01
```

```
AUTOGN   DAT        523 23/09/87    10:42
NLQCO2   DAT        970 23/09/87    11:13
NLQUE1   DAT        743 23/09/87    11:07
NLPORT   DAT        645 23/09/87    10:59
NLHOUS   DAT        436 23/09/87    11:00
NLNBOY   DAT       1261 23/09/87    11:14
README   LNG       2375 12/04/88    12:20
README   VNO       3345 12/02/88    16:09
READ     ME        5895 01/06/89    13:13
       10 fichier(s)      234115 octets
```

L'installation de **GINO** ne pose pas de problème particulier. Il suffit pour cela :

- soit de créer un répertoire et modifier le PATH en conséquence (installation durable), soit de choisir un répertoire courant (installation provisoire),
- puis de recopier dans ce répertoire le contenu utile (GINO.EXE et AUTOGN.DAT) de la disquette (non protégée).

La disquette contient en outre des exemples (NLQC02.DAT, NLQUE1.DAT, NLPORT.DAT, NLHOUS.DAT et NLNBOY.DAT), des informations utiles (READ.ME) et de la publicité sur d'autres produits (README.LNG et README.VNO).

GINO est très simple Ilà exploiter (commande COM listant les commandes, commande HELP suivi du nom d'une commande donnant la signification et la syntaxe de cette commande). La commande QUIT permet de quitter le logiciel. L'importation (TAKE) d'un problème saisi dans un fichier de type ASCII ou l'exportation (DIVE) des résultats dans un fichier de même type sont sans malice. La commande SETP, d'emploi délicat, contrôle différentes options (choix de la méthode de calcul des dérivées partielles, de la méthode de gradient conjugué, critère d'arrêt ...) mais les valeurs fournies par défaut sont souvent suffisantes, voire même trop strictes. La commande GUESS permet de proposer des valeurs initiales pour tout ou partie des paramètres à estimer. Les commandes SLB et SUB permettent de fixer une borne inférieure et supérieure aux différentes variables.

GINO est bien documenté. Le manuel d'accompagnement (Liebman, Lasdon, Schrage et Waren, 1989) décrit clairement l'installation et la syntaxe du logiciel. Il donne des exemples intéressants et variés d'application (réseaux, systèmes de queues, gestion de stocks ...).

Enfin ce manuel vulgarise de façon très pédagogique l'algorithme général d'opti-misation sous-jacent ainsi que les notions de coûts réduits des variables et de prix des contraintes. La convention de signe est qu'un coût réduit r positif (resp. négatif) signifie qu'une variation ϵ, $\epsilon > 0$, de la variable détériore (resp. améliore) la valeur de la fonction objectif d'une quantité $\epsilon \times r$ approximativement. Un prix p positif (resp. négatif) signifie qu'une variation δ, $\delta > 0$, du second membre de la contrainte entraîne approximativement une amélioration (resp. détérioration) $\delta \times p$ de la valeur de la fonction objectif.

Lors de l'exécution d'un problème d'optimisation **GINO** rapporte l'une des situations suivantes :

- pas de solution réalisable,
- pas d'optimum fini,
- plafond du nombre d'itérations atteint,
- solution optimale (locale),

Dans ce dernier cas, il est précisé si les conditions duales sont satisfaites ou non. Il est à noter que cette constatation dépend assez largement des valeurs choisies avec la commande SETP.

3 Tableaux Carrés

Pour illustrer certaines des possibilités offertes par **GINO**, on considère le cas particulier d'un tableau carré c'est-à-dire d'une table de contingence dans laquelle les modalités des lignes et des colonnes sont identiques et où, le cas échéant, les éléments diagonaux sont absents. On choisit alors d'étudier l'ajustement du modèle d'homogénéité des marges puis d'un modèle généralisant celui de quasi-symétrie en y incorporant une forme simple d'asymétrie.

Dans ce qui suit, les lignes et les colonnes du tableau carré sont respectivement indicées par i, $i = 1, \ldots, n$ et j, $j = 1, \ldots, n$. On note classiquement λ_{ij} le paramètre de la loi de Poisson associée à la cellule (i, j) et n_{ij} l'effectif conjoint observé.

3.1 Exemple

Soit donc un exemple de tableau carré constitué par la table de mobilité étudiée par Hout(1983) croisant professions du père et du fils.

```
! donnees de Featherman and Hauser (1978) reproduites dans Hout (1983)     I
!-------------------------------------------------------------------------
!  profession du fils                                                      I
!-------------------------------------------------------------------------I
!  UNM  LNM   UM   LM    F        profession du pere                       I
!-------------------------------------------------------------------------I
  1414  521  302  643   40! intellectuelle superieure (upper nonmanual)   I
   724  524  254  703   48! intellectuelle subalterne (lower nonmanual)   I
   798  648  856 1676  108! manuelle superieure (upper manual)            I
   756  914  771 3325  237! manuelle subalterne (lower manual)            I
   409  357  441 1611 1832! agricole (farm)                               I
!-------------------------------------------------------------------------I
```

3.2 Modèle d'Homogénéité des Marges

Dans ce modèle on pose que les marges du tableau des paramètres sont égales :

$$\sum_{j=1}^{n} \lambda_{ij} = \sum_{j=1}^{n} \lambda_{ji}, \quad i = 1, \ldots, n.$$

L'estimation maximum de vraisemblance de ce modèle se ramène à l'étude du problème d'optimisation suivant :

$$\begin{array}{ll} \text{maximiser} & -\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_{ij} + \sum_{i=1}^{n}\sum_{j=1}^{n} n_{ij}\log(\lambda_{ij}) \\ \text{sous les contraintes :} & \\ \sum_{j=1}^{n}\lambda_{ij} = \sum_{j=1}^{n}\lambda_{ji} & i = 1,\ldots,n\,, \\ \lambda_{ij} > 0 & i = 1,\ldots,n \text{ et } j = 1,\ldots,n. \end{array}$$

Il est clair que l'absence des données diagonales n'influerait pas sur l'analyse. Pour accélérer la convergence on peut prendre comme valeurs initiales des paramètres λ_{ij} celles des effectifs symétrisés $(n_{ij} + n_{ji})/2$.

Il est à noter que l'estimation de ce modèle est délicate mais possible avec le logiciel GLIM (Firth et Treat, 1988). Plus généralement, l'estimation d'un tel modèle requiert un logiciel statistique capable de prendre en compte des contraintes linéaires pour un modèle Poissonien (log-linéaire).

Ce problème s'écrit simplement mais laborieusement dans la syntaxe de **GINO**. Pour faciliter l'interprétation statistique des résultats il est utile d'introduire dans le programme les éléments nécessaires au calcul de la déviance ($-2 \times (VMOD - VSAT)$, où $VSAT$ et $VMOD$ désignent respectivement la log-vraisemblance du modèle saturé et du modèle courant). On cherche alors à minimiser la déviance.

Il est rendu compte de l'application de ce modèle aux données de l'exemple en annexe 1. On y notera l'introduction de la contrainte de positivité des λ_{ij}. Au vu des résultats, il est clair que ce modèle ne s'ajuste pas aux données.

3.3 Modèle Généralisant celui de Quasi-Symétrie

Le modèle de quasi-symétrie suppose que le paramètre λ_{ij} est produit d'un effet ligne, d'un effet colonne et d'une interaction symétrique. Autrement dit :

$$\log(\lambda_{ij}) = \alpha_i + \beta_j + \sigma_{ij} \quad \text{avec} \quad \sigma_{ij} = \sigma_{ji}.$$

L'estimation de ce modèle dans GLIM, par exemple, ne pose aucun problème. Il en est de même pour des extensions simples de ce modèle comme celle consistant, par exemple, à introduire un effet sous diagonal :

$$\log(\lambda_{ij}) = \begin{cases} \alpha_i + \beta_j + \sigma_{ij} & \text{si } i < j \\ \alpha_i + \beta_j + (\kappa + \sigma_{ji}) & \text{sinon.} \end{cases}$$

Ce n'est pas le cas pour la généralisation introduite par Hendrickx et Lammers (1991) qui consiste à introduire plutôt une relation non linéaire entre les interactions :

$$\sigma_{ij} = \kappa\,\sigma_{ji}, \quad i < j, \quad \text{où } \kappa > 0.$$

Il est clair que pour $\kappa = 1$ on retrouve le modèle de quasi-symétrie. On notera que, comme dans ce dernier, la présence ou l'absence de termes diagonaux est sans incidence sur l'estimation et que la formulation du problème garantit la positivité des λ_{ij}.

Ce modèle ne se prête pas directement à une estimation avec GLIM mais **GINO** permet d'obtenir assez simplement des estimations maximum de vraisemblance de ses paramètres. Pour accélérer la convergence on peut encore prendre comme valeurs initiales des λ_{ij} les effectifs symétrisés $(n_{ij} + n_{ji})/2$ et choisir pour k la valeur 1.

Il est rendu compte de l'application de ce modèle aux données de mobilité en annexe 2. On y notera l'introduction de contraintes d'identification portant sur les α_i et les β_j ; d'où l'introduction d'un terme constant noté G dans les formules définissant les λ_{ij}.

4 Une Analyse en Composantes Principales de Type L_1

On considère ici l'analyse en composantes principales (ACP) d'une matrice X d'élément courant x_i^j correspondant à l'observation de la variable quantitative X^j, $j = 1, \ldots, p$ sur l'unité statistique i, $i = 1, \ldots, n$. Soit Y la matrice des données centrées et éventuellement standardisées.

L'ACP usuelle de ces données peut être considérée comme la solution du problème d'optimisation :

$$\begin{array}{ll}
\text{minimiser} & \sum_{i=1}^n \sum_{j=1}^p (y_i^j - z_i^j)^2 \\
\text{sous les contraintes :} & \\
z_i^j & = \sum_{k=1}^q l_k^{1/2} u_{ik} v_{jk} \\
\sum_{i=1}^n u_{ik} & = 0 \\
\sum_{i=1}^n u_{ik} u_{ik'} & = \delta_{kk'} \\
\sum_{j=1}^p v_{jk} v_{jk'} & = \delta_{kk'}
\end{array}$$

Ce problème est classiquement résolu par la détermination des éléments propres de la matrice de covariance (resp. de corrélation) $\mathbf{S}$: $SV = V(\frac{1}{n}L)$. On en déduit la décomposition $Y = UL^{\frac{1}{2}}V'$ (avec $U = YVL^{-\frac{1}{2}}$, $V'V = U'U = I$) puis l'approximation d'ordre q qui est recherchée.

Une stratégie de recherche d'ACP robuste des données initiales consiste à calculer d'abord une estimation robuste de la matrice $\mathbf{S}$ et à déduire de ses éléments propres une décomposition robuste de $\mathbf{X}$ (Galpin et Hawkins (1987), Rivest et Plante (1988)).

L'alternative que nous présentons ici consiste à modifier la fonction objectif du problème d'optimisation. Pour une ACP de type L_1 on cherchera par exemple à résoudre le problème :

$$\begin{array}{ll}
\text{minimiser} & \sum_{i=1}^n \sum_{j=1}^p \left| y_i^j - z_i^j \right| \\
\text{sous les contraintes :} & \\
z_i^j & = \sum_{k=1}^q l_k^{1/2} u_{ik} v_{jk} \\
\sum_{i=1}^n u_{ik} & = 0 \\
\sum_{i=1}^n u_{ik} u_{ik'} & = \delta_{kk'} \\
\sum_{j=1}^p v_{jk} v_{jk'} & = \delta_{kk'}
\end{array}$$

On peut montrer que cette ACP est naturelle dans le cadre de l'estimation par maximum de vraisemblance d'un modèle à effet fixe de rang(q) où les erreurs sont i.i.d. Laplace (Baccini, Besse et Falguerolles (1992)). Par ailleurs, on pourra préalablement centrer les variables par soustraction de la valeur d'un indice de tendance centrale robuste (médiane, moyenne élaguée ....) et éventuellement les standardiser dans le même esprit.

Ce type d'ACP peut s'effectuer dans **GINO** et être utilisé pour réaliser, par exemple, l'analyse en dimension 2 des données ($n = 15$, $p = 12$) considérées dans Gower et Harding (1988). Une description du problème à soumettre est reproduite dans l'annexe 3. On notera que, compte tenu des limites de capacité de **GINO**, seuls les paramètres $l_k^{1/2}$, u_{ik} et v_{jk} y sont estimés. Lors de l'exécution, les valeurs absolues présentes dans la fonction objectif semblent ne poser aucun problème.

5 Conclusion

En présentant ces exemples nous avons voulu montrer l'intérêt qu'un logiciel d'opti-misation peut présenter pour un statisticien. Une étude comparative des différents logiciels existants et notamment de leur intégration à des logiciels statistiques serait à faire. Limité à des problèmes de taille peu importante et raisonablement lent, le logiciel **GINO** reste néanmoins un bon candidat du fait de sa simplicité, sa généralité et son efficacité. Il peut donc rendre de réels services sur des données de faible volume. Bien que cela ne soit pas essentiel il faut hélas souligner que la saisie des problèmes est assez laborieuse ; l'emploi d'un traitement de texte facilite cependant la génération du code.

Références

Abadie, J. (1978) "The GRC Method for Nonlinear Programming" in *Design and Implementation of Optimization Software*, H.J. Greenberg (ed.), Siijthoff and Noordhof, 335-363.

Baccini, A., Besse, Ph., and Falguerolles, A. de (1992) "A PCA Based on GINI's Mean Absolute Difference", communication présentée lors de la SECOND INTERNATIONAL CONFERENCE ON STATISTICAL ANALYSIS BASED ON THE L_1-NORM AND RELATED METHODS", *Neuchâtel, August 17-20, 1992.*

Firth, D., and Treat, B.R. (1988) "Square Contingency Tables and GLIM" ,*The GLIM Newsletter*, No. 16, 16-20.

Galpin, J.S., and Hawkins, D.M. (1987) "Methods of L_1 Estimation of a Covariance Matrix", *Computational Statistics & Data Analysis*, 5, 305-319.

Goodman, L. (1991) "Measures, Models, and Graphical Displays in the Analysis of Cross-Classified Data", *Journal of the American Statistical Association*, Vol. 86, 1085-1138.

Gower, J.C., and Harding, S.A. (1988) "Non Linear Biplot", *Biometrika*, 75, 445-455.

Hendrickx, J., and Lammers, J. (1991) "Design Techniques for Equal/Unequal Main Effects and Symmetrical/Asymmetrical Interactions", communication présentée lors du SIXTH INTERNATIONAL WORKSHOP ON STATISTICAL MODELLING, *Utrecht, July 15-19, 1991.*

Hout , M. (1983) "Mobility Tables", *Sage University Paper 07-031*, Beverly Hill and London : Sage Publication.

Lasdon, L., and Waren, A. (1978) "Generalized Reduced Gradient Software for Linearly and Nonlinearly Constrained Problems " in *Design and Implementation of Optimization Software*, H.J. Greenberg (ed.), Siijthoff and Noordhof, 363-397.

Liebman, J., Lasdon, L., Schrage, L., and Waren, A. (1986) *Modeling and Optimization with GINO*, The Scientific Press.

Rao, C.R. and Galigiuri, M. (1992) "Scoring of Ordinal Data", communication présentée lors de la SEVENTH INTERNATIONAL CONFERENCE ON MULTIVARIATE ANALYSIS, *Barcelona, September 21-24, 1992.*

Rivest, L.-P. et Plante, N. (1988) "L'Analyse en Composantes Principales Robuste", *Revue de Statistique Appliquée*, Vol. 36, 55-66.

Annexe 1: homogénéité des marges

fichier "marghom"

```
MODEL:

1) MIN=-2*(VMOD-VSAT);

2) VSAT=  - 1414         - 521         - 302         - 643         -   40
          -  724         - 524         - 254         -  703        -   48
          -  798         - 648         - 856         - 1676        -  108
          -  756         - 914         - 771         - 3325        -  237
          -  409         - 357         - 441         - 1611        - 1832
  +1414*LOG(1414)+ 521*LOG( 521)+ 302*LOG( 302)+ 643*LOG( 643)+  40*LOG(  40)
  + 724*LOG( 724)+ 524*LOG( 524)+ 254*LOG( 254)+ 703*LOG( 703)+  48*LOG(  48)
  + 798*LOG( 798)+ 648*LOG( 648)+ 856*LOG( 856)+1676*LOG(1676)+ 108*LOG( 108)
  + 756*LOG( 756)+ 914*LOG( 914)+ 771*LOG( 771)+3325*LOG(3325)+ 237*LOG( 237)
  + 409*LOG( 409)+ 357*LOG( 357)+ 441*LOG( 441)+1611*LOG(1611)+1832*LOG(1832);

3) VMOD=  -  L11         - L12         - L13         - L14         - L15
          -  L21         - L22         - L23         - L24         - L25
          -  L31         - L32         - L33         - L34         - L35
          -  L41         - L42         - L43         - L44         - L45
          -  L51         - L52         - L53         - L54         - L55
  +1414*LOG( L11)+ 521*LOG( L12)+ 302*LOG( L13)+ 643*LOG( L14)+  40*LOG( L15)
  + 724*LOG( L21)+ 524*LOG( L22)+ 254*LOG( L23)+ 703*LOG( L24)+  48*LOG( L25)
  + 798*LOG( L31)+ 648*LOG( L32)+ 856*LOG( L33)+1676*LOG( L34)+ 108*LOG( L35)
  + 756*LOG( L41)+ 914*LOG( L42)+ 771*LOG( L43)+3325*LOG( L44)+ 237*LOG( L45)
  + 409*LOG( L51)+ 357*LOG( L52)+ 441*LOG( L53)+1611*LOG( L54)+1832*LOG( L55);

4)       L12 + L13 + L14 + L15       - L21 - L31 - L41 - L51 =0;
5) L21       + L23 + L24 + L25 - L12       - L32 - L42 - L52 =0;
6) L31 + L32       + L34 + L35 - L13 - L23       - L43 - L53 =0;
7) L41 + L42 + L43       + L45 - L14 - L24 - L34       - L54 =0;
8) L51 + L52 + L53 + L54       - L15 - L25 - L35 - L45       =0;

 9) L11=1414;
10) L22= 524;
11) L33= 856;
12) L44=3325;
13) L55=1832;

 END

GUESS L11     1414.0
GUESS L12      622.5
GUESS L13      550.0
GUESS L14      699.5
GUESS L15      224.5
GUESS L21      622.5
GUESS L22      524.0
GUESS L23      451.0
GUESS L24      808.5
GUESS L25      202.5
```

```
GUESS L31      550.0
GUESS L32      451.0
GUESS L33      856.0
GUESS L34     1223.5
GUESS L35      274.5
GUESS L41      699.5
GUESS L42      808.5
GUESS L43     1223.5
GUESS L44     3325.0
GUESS L45      924.0
GUESS L51      224.5
GUESS L52      202.5
GUESS L53      274.5
GUESS L54      924.0
GUESS L55     1832.0

SLB L11 0.00005
SLB L12 0.00005
SLB L13 0.00005
SLB L14 0.00005
SLB L15 0.00005
SLB L21 0.00005
SLB L22 0.00005
SLB L23 0.00005
SLB L24 0.00005
SLB L25 0.00005
SLB L31 0.00005
SLB L32 0.00005
SLB L33 0.00005
SLB L34 0.00005
SLB L35 0.00005
SLB L41 0.00005
SLB L42 0.00005
SLB L43 0.00005
SLB L44 0.00005
SLB L45 0.00005
SLB L51 0.00005
SLB L52 0.00005
SLB L53 0.00005
SLB L54 0.00005
SLB L55 0.00005
```

session GINO

```
:TAKE marghom
:GO

 SOLUTION STATUS:  OPTIMAL TO TOLERANCES.  DUAL CONDITIONS:  SATISFIED.

              OBJECTIVE FUNCTION VALUE

           1)      2743.562185
```

VARIABLE	VALUE	REDUCED COST
VMOD	118298.473151	.000000
VSAT	119670.254244	.000000
L11	1414.000000	.000000
L12	555.168426	.000001
L13	553.257438	.000004
L14	727.398924	.000000
L15	291.918009	-.000007
L21	682.023419	-.000001
L22	524.000000	.000000
L23	418.171856	.000000
L24	743.507780	-.000001
L25	241.734159	.000003
L31	548.777425	-.000001
L32	465.318285	.000000
L33	856.000000	.000000
L34	1252.510718	.000000
L35	182.691217	.000001
L41	677.401248	-.000001
L42	866.775064	.000000
L43	1164.844748	-.000004
L44	3325.000000	.000000
L45	936.574235	.000000
L51	219.540705	-.000001
L52	198.175438	.000000
L53	313.023603	.000000
L54	922.177873	-.000001
L55	1832.000000	.000000

ROW	SLACK OR SURPLUS	PRICE
2)	.000000	-2.000000
3)	.000000	2.000000
4)	.000000	-.232058
5)	.000000	-.108966
6)	.000000	.676223
7)	.000000	.000000
8)	.000000	1.493900
9)	.000000	.000000
10)	.000000	.000000
11)	.000000	.000000
12)	.000000	.000000
13)	.000000	.000000

Annexe 2: une généralisation de la quasi-symétrie

fichier "genqs"

```
MODEL:

1)  MIN=-2*(VMOD-VSAT);

2)  VSAT= - 1414        - 521        - 302        - 643        -  40
          -  724        - 524        - 254        - 703        -  48
          -  798        - 648        - 856        - 1676       - 108
          -  756        - 914        - 771        - 3325       - 237
          -  409        - 357        - 441        - 1611       - 1832
  +1414*LOG(1414)+ 521*LOG( 521)+ 302*LOG( 302)+ 643*LOG( 643)+  40*LOG(  40)
  + 724*LOG( 724)+ 524*LOG( 524)+ 254*LOG( 254)+ 703*LOG( 703)+  48*LOG(  48)
  + 798*LOG( 798)+ 648*LOG( 648)+ 856*LOG( 856)+1676*LOG(1676)+ 108*LOG( 108)
  + 756*LOG( 756)+ 914*LOG( 914)+ 771*LOG( 771)+3325*LOG(3325)+ 237*LOG( 237)
  + 409*LOG( 409)+ 357*LOG( 357)+ 441*LOG( 441)+1611*LOG(1611)+1832*LOG(1832);

3)  VMOD= -  L11        - L12        - L13        - L14        - L15
          -  L21        - L22        - L23        - L24        - L25
          -  L31        - L32        - L33        - L34        - L35
          -  L41        - L42        - L43        - L44        - L45
          -  L51        - L52        - L53        - L54        - L55
  +1414*LOG( L11)+ 521*LOG( L12)+ 302*LOG( L13)+ 643*LOG( L14)+  40*LOG( L15)
  + 724*LOG( L21)+ 524*LOG( L22)+ 254*LOG( L23)+ 703*LOG( L24)+  48*LOG( L25)
  + 798*LOG( L31)+ 648*LOG( L32)+ 856*LOG( L33)+1676*LOG( L34)+ 108*LOG( L35)
  + 756*LOG( L41)+ 914*LOG( L42)+ 771*LOG( L43)+3325*LOG( L44)+ 237*LOG( L45)
  + 409*LOG( L51)+ 357*LOG( L52)+ 441*LOG( L53)+1611*LOG( L54)+1832*LOG( L55);

 4)  L11=EXP(G+A1+B1+  S11);

 5)  L12=EXP(G+A1+B2+  S12);
 6)  L21=EXP(G+A2+B1+K*S12);

 7)  L13=EXP(G+A1+B3+  S13);
 8)  L31=EXP(G+A3+B1+K*S13);

 9)  L14=EXP(G+A1+B4+  S14);
10)  L41=EXP(G+A4+B1+K*S14);

11)  L15=EXP(G+A1+B5+  S15);
12)  L51=EXP(G+A5+B1+K*S15);

13)  L22=EXP(G+A2+B2+  S22);

14)  L23=EXP(G+A2+B3+  S23);
15)  L32=EXP(G+A3+B2+K*S23);

16)  L24=EXP(G+A2+B4+  S24);
17)  L42=EXP(G+A4+B2+K*S24);

18)  L25=EXP(G+A2+B5+  S25);
19)  L52=EXP(G+A5+B2+K*S25);
```

```
20)  L33=EXP(G+A3+B3+  S33);

21)  L34=EXP(G+A3+B4+  S34);
22)  L43=EXP(G+A4+B3+K*S34);

23)  L35=EXP(G+A3+B5+  S35);
24)  L53=EXP(G+A5+B3+K*S35);

25)  L44=EXP(G+A4+B4+  S44);

26)  L45=EXP(G+A4+B5+  S45);
27)  L54=EXP(G+A5+B4+K*S45);

28)  L55=EXP(G+A5+B5+  S55);

29)  L11=1414;
30)  L22= 524;
31)  L33= 856;
32)  L44=3325;
33)  L55=1832;

34)  A1+A2+A3+A4+A5=0;
35)  B1+B2+B3+B4+B5=0;

36)  K>0;

 END

GUESS K 1
GUESS L11      1414
GUESS L12       622.5
GUESS L13       550.0
GUESS L14       699.5
GUESS L15       224.5
GUESS L21       622.5
GUESS L22       524
GUESS L23       451.0
GUESS L24       808.5
GUESS L25       202.5
GUESS L31       550.0
GUESS L32       451.0
GUESS L33       856
GUESS L34      1223.5
GUESS L35       274.5
GUESS L41       699.5
GUESS L42       808.5
GUESS L43      1223.5
GUESS L44      3325
GUESS L45       924.0
GUESS L51       224.5
GUESS L52       202.5
GUESS L53       274.5
GUESS L54       924.0
GUESS L55      1832
```

session GINO

```
:TAKE genqs
:GO

 SOLUTION STATUS:  OPTIMAL TO TOLERANCES.  DUAL CONDITIONS:  SATISFIED.

              OBJECTIVE FUNCTION VALUE

        1)          2.539059

   VARIABLE          VALUE          REDUCED COST
       VMOD     119668.984714          .000000
       VSAT     119670.254244          .000000
        L11       1414.000000          .000000
        L12        520.999957          .000000
        L13        295.135570          .000000
        L14        647.004730          .000000
        L15         42.859742          .000000
        L21        724.000063          .000000
        L22        524.000000          .000000
        L23        260.864442          .000000
        L24        704.333933          .000000
        L25         39.801670          .000000
        L31        801.490238          .000000
        L32        644.509645          .000000
        L33        856.000000          .000000
        L34       1670.661436          .000000
        L35        113.338625          .000000
        L41        753.963682          .000000
        L42        913.321670          .000000
        L43        773.714567          .000000
        L44       3325.000000          .000000
        L45        236.999974          .000000
        L51        407.545943          .000000
        L52        361.168509          .000000
        L53        438.285479          .000000
        L54       1611.000009          .000000
        L55       1832.000000          .000000
          G          2.436241         -.000226
         A1         -4.880751          .000111
         B1         -8.640930         -.000173
        S11         18.339618          .000000
         B2          2.620685         -.000461
        S12          6.079575          .000160
         A2           .832819          .000329
          K          1.966694         -.000771
         B3          2.434627          .000093
        S13          5.697318         -.000402
         A3          1.686285          .000003
         B4          3.587131          .000191
        S14          5.329733         -.000169
         A4          2.348083         -.000099
         B5          -.001513          .000000
```

S15	6.203956	.000116
A5	.013564	.000000
S22	.371747	.000000
S23	-.139686	-.000032
S24	-.298938	-.000268
S25	.416362	-.000301
S33	.195118	.000000
S34	-.288681	.000316
S35	.609367	-.000012
S44	-.262230	.000000
S45	.685249	-.000016
S55	5.064872	.000000

ROW	SLACK OR SURPLUS	PRICE
2)	.000000	-2.000000
3)	.000000	2.000000
4)	.000000	.000000
5)	.000000	.000000
6)	.000000	.000000
7)	.000000	.046517
8)	.000000	-.008709
9)	.000000	-.012379
10)	.000000	.005402
11)	.000000	-.133447
12)	.000000	.007136
13)	.000000	.000000
14)	.000000	-.052628
15)	.000000	.010831
16)	.000000	-.003788
17)	.000000	.001485
18)	.000000	.411959
19)	.000000	-.023083
20)	.000000	.000000
21)	.000000	.006391
22)	.000000	-.007017
23)	.000000	-.094207
24)	.000000	.012387
25)	.000000	.000000
26)	.000000	.000000
27)	.000000	.000000
28)	.000000	.000000
29)	.000000	.000000
30)	.000000	.000000
31)	.000000	.000000
32)	.000000	.000000
33)	.000000	.000000
34)	.000000	.000114
35)	.000000	-.000025
36)	1.966694	.000000

Annexe 3 : une ACP de type L_1

```
MODEL:
! UNE ACP DE TYPE L1 POUR LES DONNEES ILLUSTRANT L'ARTICLE
! DE J.C. GOWER AND S.A. HARDING (1988).
! LA DIMENSION EST 2.

! VALEURS SINGULIERES (RACINES CARREES DES VALEURS PROPRES) Sk1 >= Sk2
   Sk1-Sk2 > 0;

! SCORES LIGNES CENTRES (15 SITES)
   U01k1 +U02k1 +U03k1 +U04k1 +U05k1 +U06k1 +U07k1 +U08k1
         +U09k1 +U10k1 +U11k1 +U12k1 +U13k1 +U14k1 +U15k1 =0;

   U01k2 +U02k2 +U03k2 +U04k2 +U05k2 +U06k2 +U07k2 +U08k2
         +U09k2 +U10k2 +U11k2 +U12k2 +U13k2 +U14k2 +U15k2 =0;


! ORTHOGONALITE DES SCORES LIGNES:
  U01k1 * U01k1 + U02k1 * U02k1 + U03k1 * U03k1
+ U04k1 * U04k1 + U05k1 * U05k1 + U06k1 * U06k1
+ U07k1 * U07k1 + U08k1 * U08k1 + U09k1 * U09k1
+ U10k1 * U10k1 + U11k1 * U11k1 + U12k1 * U12k1
+ U13k1 * U13k1 + U14k1 * U14k1 + U15k1 * U15k1=1;


  U01k2 * U01k2 + U02k2 * U02k2 + U03k2 * U03k2
+ U04k2 * U04k2 + U05k2 * U05k2 + U06k2 * U06k2
+ U07k2 * U07k2 + U08k2 * U08k2 + U09k2 * U09k2
+ U10k2 * U10k2 + U11k2 * U11k2 + U12k2 * U12k2
+ U13k2 * U13k2 + U14k2 * U14k2 + U15k2 * U15k2=1;

  U01k2 * U01k1 + U02k2 * U02k1 + U03k2 * U03k1
+ U04k2 * U04k1 + U05k2 * U05k1 + U06k2 * U06k1
+ U07k2 * U07k1 + U08k2 * U08k1 + U09k2 * U09k1
+ U10k2 * U10k1 + U11k2 * U11k1 + U12k2 * U12k1
+ U13k2 * U13k1 + U14k2 * U14k1 + U15k2 * U15k1=0;

! ORTHOGONALITE DES SCORES COLONNES (12 COLONNES):
  V01k1 * V01k1 + V02k1 * V02k1 + V03k1 * V03k1
+ V04k1 * V04k1 + V05k1 * V05k1 + V06k1 * V06k1
+ V07k1 * V07k1 + V08k1 * V08k1 + V09k1 * V09k1
+ V10k1 * V10k1 + V11k1 * V11k1 + V12k1 * V12k1=1;

  V01k2 * V01k2 + V02k2 * V02k2 + V03k2 * V03k2
+ V04k2 * V04k2 + V05k2 * V05k2 + V06k2 * V06k2
+ V07k2 * V07k2 + V08k2 * V08k2 + V09k2 * V09k2
+ V10k2 * V10k2 + V11k2 * V11k2 + V12k2 * V12k2=1;

  V01k2 * V01k1 + V02k2 * V02k1 + V03k2 * V03k1
+ V04k2 * V04k1 + V05k2 * V05k1 + V06k2 * V06k1
+ V07k2 * V07k1 + V08k2 * V08k1 + V09k2 * V09k1
+ V10k2 * V10k1 + V11k2 * V11k1 + V12k2 * V12k1 =0;
```

```
! FONCTION OBJECTIF:
MIN=ABS( Sk1 * U01k1 * V01k1 + Sk2 * U01k2 * V01k2 - 0.1174 )
  + ABS( Sk1 * U01k1 * V02k1 + Sk2 * U01k2 * V02k2 - 0.2958 )
  + ABS( Sk1 * U01k1 * V03k1 + Sk2 * U01k2 * V03k2 - 0.0137 )
  + ABS( Sk1 * U01k1 * V04k1 + Sk2 * U01k2 * V04k2 - 0.2558 )
  + ABS( Sk1 * U01k1 * V05k1 + Sk2 * U01k2 * V05k2 - 1.2017 )
  + ABS( Sk1 * U01k1 * V06k1 + Sk2 * U01k2 * V06k2 - 0.2954 )
  + ABS( Sk1 * U01k1 * V07k1 + Sk2 * U01k2 * V07k2 + 0.2065 )
  + ABS( Sk1 * U01k1 * V08k1 + Sk2 * U01k2 * V08k2 + 0.0937 )
  + ABS( Sk1 * U01k1 * V09k1 + Sk2 * U01k2 * V09k2 + 0.2511 )
  + ABS( Sk1 * U01k1 * V10k1 + Sk2 * U01k2 * V10k2 + 0.1058 )
  + ABS( Sk1 * U01k1 * V11k1 + Sk2 * U01k2 * V11k2 - 0.0891 )
  + ABS( Sk1 * U01k1 * V12k1 + Sk2 * U01k2 * V12k2 - 0.5388 )
  + ABS( Sk1 * U02k1 * V01k1 + Sk2 * U02k2 * V01k2 - 0.0845 )
  + ABS( Sk1 * U02k1 * V02k1 + Sk2 * U02k2 * V02k2 + 0.0898 )
  + ABS( Sk1 * U02k1 * V03k1 + Sk2 * U02k2 * V03k2 - 0.133 )
  ..........................................................
  etc
  ..........................................................
  + ABS( Sk1 * U14k1 * V12k1 + Sk2 * U14k2 * V12k2 + 1.0048 )
  + ABS( Sk1 * U15k1 * V01k1 + Sk2 * U15k2 * V01k2 + 0.0172 )
  + ABS( Sk1 * U15k1 * V02k1 + Sk2 * U15k2 * V02k2 + 0.0327 )
  + ABS( Sk1 * U15k1 * V03k1 + Sk2 * U15k2 * V03k2 + 0.2417 )
  + ABS( Sk1 * U15k1 * V04k1 + Sk2 * U15k2 * V04k2 - 0.365 )
  + ABS( Sk1 * U15k1 * V05k1 + Sk2 * U15k2 * V05k2 + 0.112 )
  + ABS( Sk1 * U15k1 * V06k1 + Sk2 * U15k2 * V06k2 - 0.097 )
  + ABS( Sk1 * U15k1 * V07k1 + Sk2 * U15k2 * V07k2 - 0.6264 )
  + ABS( Sk1 * U15k1 * V08k1 + Sk2 * U15k2 * V08k2 - 0.1256 )
  + ABS( Sk1 * U15k1 * V09k1 + Sk2 * U15k2 * V09k2 + 0.0475 )
  + ABS( Sk1 * U15k1 * V10k1 + Sk2 * U15k2 * V10k2 - 0.022 )
  + ABS( Sk1 * U15k1 * V11k1 + Sk2 * U15k2 * V11k2 - 0.259 )
  + ABS( Sk1 * U15k1 * V12k1 + Sk2 * U15k2 * V12k2 - 0.3386 );
END
```