DRD167

DISCUSSION PAPER

Report No. DRD167

BALANCING LARGE SOCIAL ACCOUNTING MATRICES WITH NONLINEAR NETWORK PROGRAMMING

by

Arne Drud
The World Bank

Stavros A. Zenios
John M. Mulvey
Princeton University

Development Research Department
Economics and Research Staff
World Bank

Balancing Large Social Accounting Matrices With Nonlinear Network Programming

Arne Drud *
Stavros A. Zenios **
John M. Mulvey **

Report EES-86-1

*Development Research Department
The World Bank
Washington, D.C.20433

**Engineering-Management Systems
Princeton University
Princeton, NJ08544

ABSTRACT

We formulate the problem of optimally adjusting the components of a large matrix to satisfy consistency requirements as a nonlinear network optimization model. An efficient network optimization algorithm -- NLPNETG -- is incorporated in a user friendly modeling system -- GAMS. The resulting software is used for balancing large Social Accounting Matrices (SAM). We assemble a library of SAM models from developing countries, and report summary computational results.

Keywords: Network optimization; Nonlinear programming; Social Accounting Matrices.

February 15, 1986

This research was completed while the second author was visiting the World Bank, as a consultant with the Development Research Department. Partial support provided by National Science Foundation Grant DCR-8401098.

Balancing Large Social Accounting Matrices With Nonlinear Network Programming

Arne Drud *
Stavros A. Zenios **
John M. Mulvey **

Report EES-86-1

*Development Research Department
The World Bank
Washington, D.C.20433

**Engineering-Management Systems
Princeton University
Princeton, NJ08544

1. Balancing of Social Accounting Matrices

The Social Accounting Matrix (SAM) framework is a convenient way of giving a comprehensive and consistent picture of an economy. It describes all aspects of the circular flow in an economy, and may be viewed as an extension of input/output matrices which look only at the production side. As such, SAM is used as the core data base for many economy-wide models. The presentation of such models based on a SAM framework is discussed in Drud et al[5]. King[7] provides a general introduction to Social Accounting Matrices.

A SAM database can only be used as the basis of a planning model if it is balanced in the sense that it satisfies a set of internal consistency relations. This guaranties that the flows in an economy satisfy the material balance equations. We may not assume that readily available data for a SAM satisfy such consistency relations! Data for the matrix estimation are derived from disparate sources, many of those based on sampling procedures; the resulting SAM is inherently unbalanced. Hence a systematic way is needed to balance the SAM estimate in an efficient and consistent manner. Such a procedure was originally proposed by Stone[8]. Byron[3] recognized the equivalence of Stone's procedure with a nonlinear optimization problem. Cottle et al. [4] used Lagrangian relaxation for the same problem.

In this paper we describe a nonlinear network optimization model for SAM balancing. Our model is equivalent, in a special case, to Byron's formulation. By capitalizing on the intrinsic structure of the underlying network model we are able to solve efficiently very large problems. We develop a modeling system (GAMS/SAMBAL) that incorporates a nonlinear network optimization program(NLPNETG[1]) within a general algebraic modeling system (GAMS[2]).

2. Nonlinear Network Optimization Model

In this section we formulate the SAM balancing problem as a network optimization model. We also outline the main features of the GAMS/SAMBAL system.

2.1. Network Formulation

A SAM is a square matrix in which each row and corresponding column represents an agent in the economy; an agent is an economic entity such as a household, firm, government and so on. Every agent is represented twice: once as an expenditure account (column) and once as an income account (row). Cell (i,j) in the SAM indicates the transaction between agents i and j. The requirement that the SAM satisfies accounting constraints (total income equals total expenditure, or row total equals column total) may be represented by a system of network flow constraints. We use the following notation:

$G = (N,E)$: the graph associated with the SAM.

N : the set of nodes in G; every $i \in N$ corresponds to an account of the SAM. A node exists for every expenditure account (column) and for every income account (row) of the SAM.

E : the set of arcs in the graph G; every transaction between nodes i and j corresponds to an arc $(i,j) \in E$. In addition an arc (i,i') exists between every expenditure account i and the corresponding income account i′. The set is usually sparse.

δ_i^+ (δ_i^-) : set of arcs $(i,j) \in E$ leaving (entering) node i.

$\bar{X}^0 = (x_{ij}^0, x_{ii'}^0 \mid (i,j),(i,i') \in E)$: the level of transaction between i and j, and the total expenditure (income) of account i (i′) in the unbalanced SAM.

$\bar{X} = (x_{ij}, x_{ii'} \mid (i,j),(i,i') \in E)$: the level of transaction between i and j, and the total expenditure (income) of account i (i′) in the balanced SAM.

$\bar{L}$ ($\bar{U}$) : vector of lower (upper) bounds on $\bar{X}$.

i and i′ indicate nodes corresponding to expenditure and income accounts respectively.

The accounting constraints may be written in the form:

$$\sum_{(j,i')\in\delta_{i'}^{-}} x_{ji'} = x_{ii'} \qquad [2.1]$$

$$\sum_{(i,j)\in\delta_{i}^{+}} x_{ij} = x_{ii'} \qquad [2.2]$$

This is the familiar form of a circulation network (without exogenous supply or demand). See Figure 1 for an example.

We want a solution $\bar{X}$ of [2.1]-[2.2] that is close to $\bar{X}^0$ in a suitable norm. This requirement is qualified through the choice of a separable penalty objective function, for all (i,j) ∈ E :

(i) Quadratic penalty : $f_{ij}(x_{ij}) = w_{ij} \cdot (x_{ij} - x_{ij}^0)^2$

(ii) Entropy penalty : $f_{ij}(x_{ij}) = w_{ij} \cdot x_{ij} \cdot (ln\,(x_{ij}/x_{ij}^0) - 1)$

(iii) Piecewise Linear penalty : $f_{ij}(x_{ij}) = w_{ij} \cdot |x_{ij} - x_{ij}^0|$

If we let A denote the constraint matrix corresponding to [2.1]-[2.2] we have the following optimization problem:

[NLN] Minimize $\sum_{(i,j)\in E} f_{ij}(x_{ij}) + \sum_{(i,i')\in E} f_{ii'}(x_{ii'})$

Subject to :

$$A \cdot \bar{X} = \bar{b}$$

$$\bar{L} \le \bar{X} \le \bar{U}$$

Comments:

- The optimization problem with the quadratic penalty is equivalent to Stone's procedure[8], as was shown by Byron[3]; in the special case where row/column totals are fixed it is equivalent to the well known RAS algorithm. The entropy function has been used elsewhere[9], in the context of unconstrained optimization. The linear penalty seems to be a reasonable choice that results, however, in extreme point solutions.

- The objective function weights (w_{ij}) reflect the reliability of the original estimates. Byron[3] is using the prior estimate of the variance matrix $\bar{V}$ as an indicator of the data reliability. The variance matrix is diagonal, and the weights are computed trivially as the inverse of $\bar{V}$.

	LAB	H1	H2	P1	P2	TOT
LAB		*	*	*	*	*
H1	*					*
H2	*					*
P1		*	*		*	*
P2		*	*	*		*
TOT	*	*	*	*	*	

(Blank entries indicate impossible transactions)

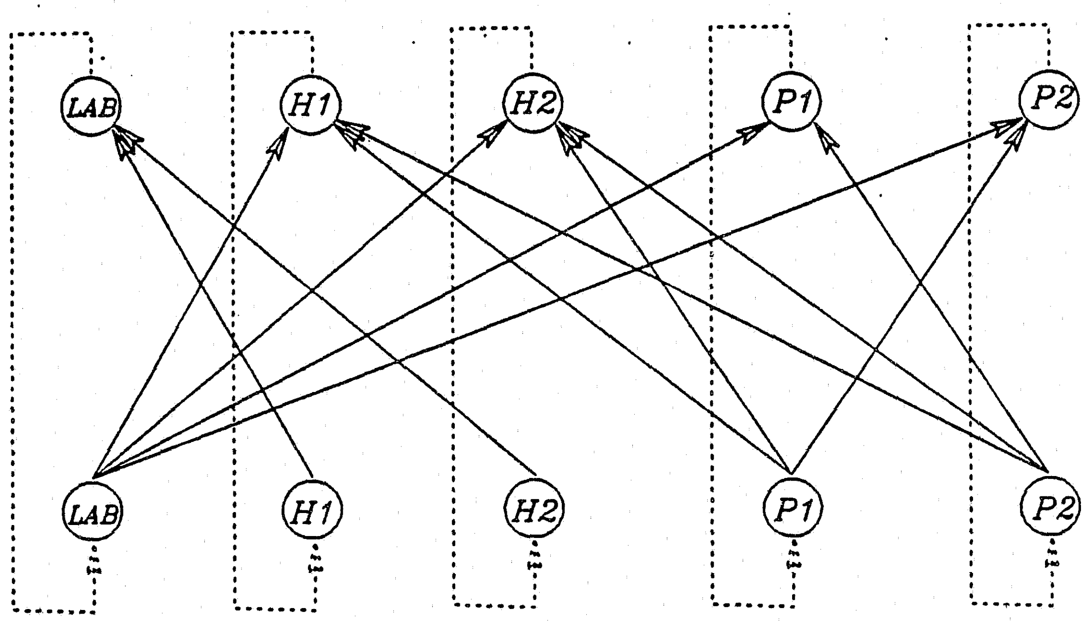

Figure 1: The Structure of a Small SAM and the Corresponding Network Model

This approach is employed in the test problem STONE, were the variance matrix is provided together with the initial SAM estimates.

2.2. The GAMS/SAMBAL System

In order to solve large SAM problems on a routine basis we had to develop a user friendly and reliable software system for generating and solving [NLN] models. An efficient nonlinear network optimization code (NLPNETG) was incorporated within a general algebraic modeling system (GAMS).

In this system the user has flexibility over the choice of suitable penalty functions, and may assign weights to components of the specified objective. The user may adjust bounds imposed on the transactions to eliminate results that seem irrational from an economic perspective; the network model has a built-in advanced start mechanism to facilitate the efficient modification of balanced SAM. Missing data that are known *a priori* to be non-zero, without any estimate on their numerical value, may be treated by specifying a judgmental estimate and assigning a penalty term identically equal to zero. SAM values that are known to be correct may be kept fixed, by specifying the corresponding arc to have constant flow -- i.e. $l_{ij} = u_{ij} = x_{ij}^0$. Finally, GAMS provides tools for report writing and tabular displays.

3. Results and Discussion

The GAMS/SAMBAL system was used in analyzing a wide variety of large SAM's, mostly developed by World Bank staff as part of their planning exercises. These test problems were assembled in a library* of models in GAMS language[10]. In Table 1 we describe the characteristics of the collected problems. All testing was carried out on an IBM 3081-K, at the World Bank Headquarters. The system was coded in Fortran 77. The H compiler with level of optimization OPT2 under OS370 was used. The reported solution times exclude input/output. Table 2 reports some statistics about the balanced SAMs: maximum and average percentage deviation of the balanced SAM values from the original estimates, objective function value and timing information. We observe that even very large problems can be solved within a few seconds on a large mainframe.

Problem	Accounts	Transactions	Network (Nodes/Arcs)	Description
SAM1	5	8	10 / 13	Demonstration test problem
SAM6	18	46	36 / 64	SAM models obtained from
SAM10	25	63	50 / 88	Chapter 6 and 10 of[6].
STONE	6	12	12 / 18	Stone's demonstration SAM
SAMTU	8	19	16 / 27	SAM for Turkey 1973
SAMKE	25	177	50 / 202	SAM for Kenya 1976
SAMBO	63	598	126 / 662	SAM for Botswana 1974-75
SAMMO	232	1664	464 /1896	SAM for Morocco 1980

Table 1 : Test SAM Models

Figure 2 presents a histogram of the errors in the balanced SAMKE obtained with a quadratic and an entropy penalty. Observe a desirable distribution of errors, clustered towards very small values. In addition the deviation between the results obtained with the quadratic and entropy penalties does not exceed 0.3% (average value 0.04%). The histogram of errors for the piecewise linear function does not justify the use of this penalty for any reasonably accurate analysis, except in the case of almost balanced matrices.

Table 3 compares the results of our analysis with Byron's estimates of STONE. We used both a quadratic, and an entropy penalty function; the inverse of the variance matrix was used for

* The collection of test problems may be obtained from the authors.

Problem	Penalty function	Error (%) Maximum	Error (%) Average	Objective value	Optimizer CPU time (sec)
SAM1	QUAD	6.47	1.71	0.54	0.11
	ENTROP	6.10	1.46	- 222.72	0.08
SAM6	QUAD	17.50	5.10	11.75	0.86
	ENTROP	17.50	5.10	-2197.00	0.87
SAM10	QUAD	12.50	3.70	8.41	1.21
	ENTROP	13.50	3.80	-3212.78	1.23
STONE	QUAD	14.90	7.30	14.63	0.31
	ENTROP	16.20	7.50	-1025.99	0.44
SAMTU	QUAD	11.50	4.40	6.19	0.36
	ENTROP	10.20	4.00	-1642.93	0.29
SAMKE	QUAD	8.70	2.10	6.09	2.88
	ENTROP	9.00	2.10	-7768.35	3.84
SAMBO	QUAD	39.00	5.00	10.48	12.57
	ENTROP	40.70	5.40	-1640.85	11.74
SAMMO	QUAD	0.50	0.01	0.001	26.93

Table 2 : Balancing the SAM Models with GAMS/SAMBAL

assigning weights. Note that this table displays some discrepancies between Byron's estimates and our results. This is due to the way that missing values are treated in the two approaches: Byron uses a quadratic penalty with a high variance, while we use a penalty term identically equal to zero.

Network optimization algorithms seem to be a reliable and efficient tool for balancing large SAM models, and a worthy successor of currently used procedures. The developed system can be used as an integral part of the planning process. We do not have enough experiences, as yet, to recommend one penalty function over another. The proper choice of weights and the effect of upper/lower bound on the quality of the balanced SAM can be evaluated only after extended use of the system by its intended clients. Further research is needed to evaluate the relationship of the accuracy in solving (NLN) to the quality of the balanced SAM, since a very accurate solution is not always needed.

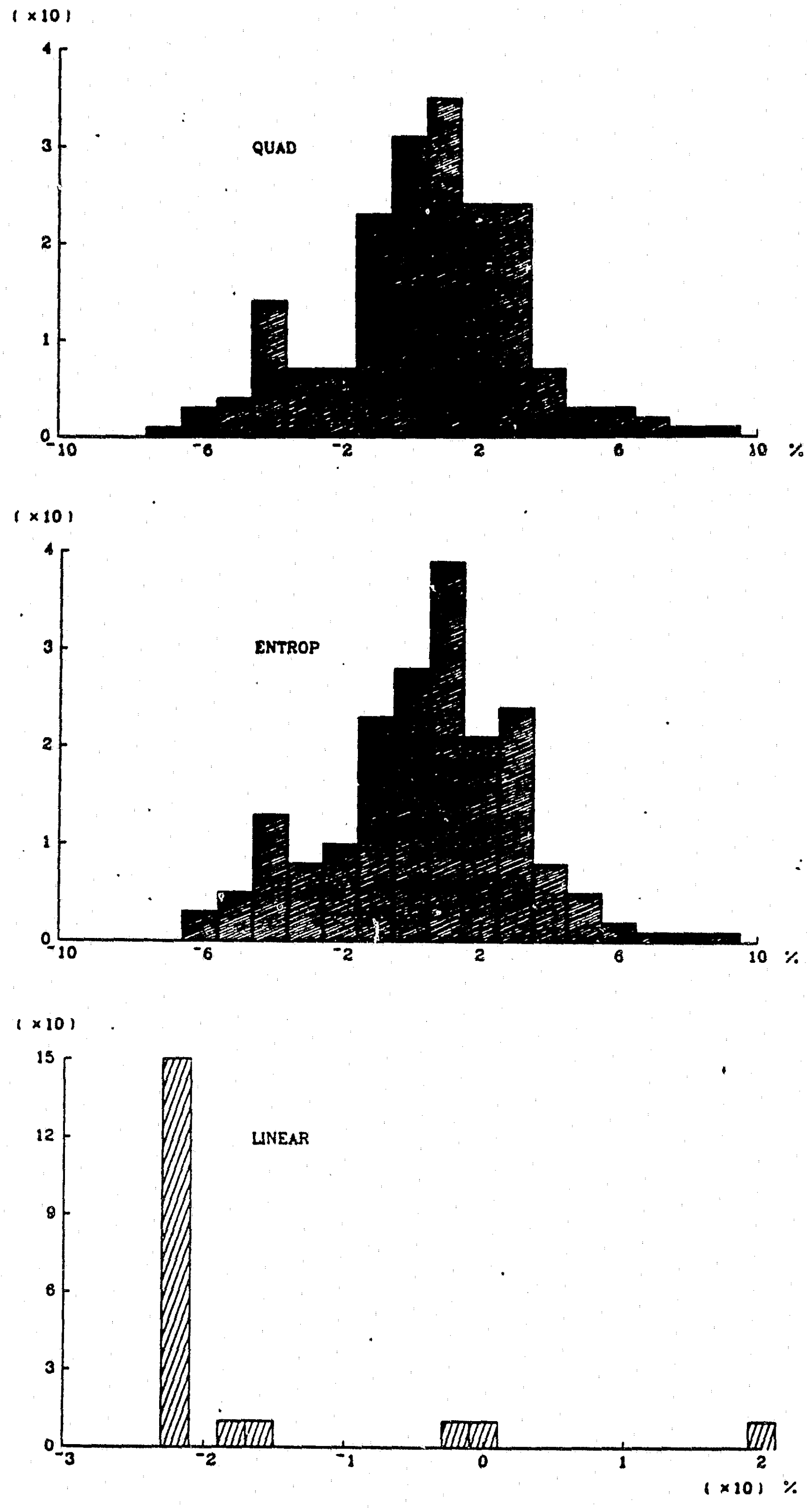

Figure 2: Error Distribution for the Balanced SAMKE

Cell	Original estimate	Byron's estimate	QUAD penalty	ENTROP penalty	QUAD with weights	ENTROP with weights
LAB.H1	15	14.45	14.53	13.93	13.56	14.81
LAB.H2	3	2.98	2.91	2.89	2.71	2.99
LAB.P1	130	124.12	126.80	127.08	127.65	125.76
LAB.P2	80	85.84	84.24	84.42	81.42	86.76
H1 .LAB	*	54.35	52.57	52.09	53.17	54.36
H2 .LAB	*	173.04	175.91	176.23	172.16	175.98
P1 .H1	15	15.13	15.87	15.89	15.78	15.11
P1 .H2	130	139.65	137.51	137.71	136.89	138.36
P1 .P2	20	20.60	22.98	23.20	26.48	22.50
P2. H1	25	24.77	22.18	22.27	23.82	24.44
P2. H2	40	30.42	35.49	35.64	32.56	34.62
P2 .P1	55	51.26	49.56	49.71	51.51	50.21
TOT-LAB	220	227.39	228.48	228.32	225.33	230.34
TOT-H1	*	54.35	52.75	52.09	53.17	54.36
TOT-H2	*	173.04	175.91	176.23	172.16	175.98
TOT-P1	190	175.38	176.37	176.80	179.16	175.98
TOT-P2	105	106.44	107.23	107.62	107.90	109.27

* Estimate for this entry not available; penalty term identically equal to zero.

Table 3: Solutions to the STONE problem

References

1. D. P. Ahlfeld, R. S. Dembo, J. M. Mulvey, and S. A. Zenios, "Nonlinear Programming on Generalized Networks," Report EES-85-7, submitted for publication to Transactions on Mathematical Software, Princeton University, June 1985.

2. J. Bisschop and A. Meeraus, "On the Development of a General Algebraic Modeling System in a Strategic Planning Environment," *Mathematical Programming Study*, vol. 20, pp. 1-29, 1982.

3. R. P. Byron, "The Estimation of Large Social Account Matrices," *J. R. Statistical Society A*, vol. 141, Part 3, pp. 359-367, 1978.

4. R. W. Cottle, S. G. Duvall, and K. Zikan, A Lagrangian Relaxation Algorithm for the Constrained Matrix Problem , Working Paper, Operations Research Department, Stanford University, 1985.

5. A. Drud, W. Grais, and G. Pyatt, "The Transaction Value Approach: A Systematic Method of Defining Econolywide Models Based on Social Accounting Matrices," in *Dynamic Modelling and Control of National Economies 1983*, ed. L. F. Piu, Pergamon Press, New York, 1983 .

6. D. Kendrick and A. Drud, *SIMS: The SAM Integrated Modeling System* , The World Bank, 1985. (in preparation)

7. B. B. King, "What is a SAM?," in *Social Accouting Matrices: A Basis for Planning*, ed. G. Pyatt. J.I. Round, Washington, D.C. , 1985.

8. R. Stone, "The Development of Economic Data Systems," in *Social Accounting for Development Planning with Special Reference to Sri Lanka*, ed. G. Pyatt et al., Cambridge University Press, U.K., 1976.

9. H. Theil, *Economics and Information Theory*, North-Holland, Amsterdam, 1967.

10. S.A. Zenios, "A Library of SAM Balancing Models," Report EES-86-2 , Engineering-Management Systems, Civil Engineering Department , Princeton University, January, 1986.

SOME RECENT DRD DISCUSSION PAPERS

153. An Overview of the Labor Market in India by Robert E. B. Lucas, January 1986.

154. A Profile of the Labor Market in Israel by Yoram Ben-Porath, The Hebrew University of Jerusalem and The Maurice Falk Institute for Economic Research in Israel, August 1985.

155. An Analysis of the Nigerian Labor Market by Paul Collier, Institute of Economics and Statistics, Oxford, February 1986.

156. The Labor Market in Kenya: Recent Evidence by Peter R. Fallon, October 1985.

157. The Labor Market in Colombia: An Overview of Its Evolution Over the Past Three Decades by Francois Bourguigon, Ecole des Hautes Etudes en Sciences Sociales et Ecole Normale Supériere-Ulm, Paris, January 1986.

158. Employment and Wages in Pakistan by Ehtisham Ahmad and Nicholas Stern, December 1985.

159. Export Incentives and Export Growth in Developing Countries: An Econometric Investigation by Bela Balassa, Evangelos Voloudakis, Panagiotis Fylaktos, and Suk Tai Suh, February 1986.

160. The Egyptian Labor Market - An Overview, by Bent Hansen, Department of Economics, University of California, Berkeley, June 1985.

161. Public Policy and the Evolution of the Labor Market in Sri Lanka by Gary S. Fields, Consulting Economist, Ithaca, New York, March, 1986.

162. The Structure of Labor Markets in LDCs: Overview for South Korea, 1953-1984 by Ray Richardson and Byung Whan Kim, London School of Economics, (revised draft) March 1986.

163. Efficiency of Tax-Contract Policies for Natural Resource Rent Collection by Arvind Virmani, April 1986.

164. Efficiency of Practical Resource Rent Tax System: Threshold Rates and Income Taxes by Arvind Virmani, April 1986.

165. Implications of a Vat in India: The Deductability System by Arvind Virmani, April 1986.

166. The Effects of Financial Liberalization on the Development of the Financial Market and the Allocation of Credit to Corporate Sectors: The Korean Case by Yoon Je Cho, March 1986.

167. Balancing Large Social Accounting Matrices with Nonlinear Network Programming by Arne Drud, S. Zenios and J. Mulvey, February 1986.