ANU CRAWFORD SCHOOL

Solving GTAP model in parallel using Doubly Bordered Block Diagonal ordering technique*

Pham Van Ha
ha.pham@anu.edu.au

Prof. Tom Kompas
tom.kompas@anu.edu.au

Crawford School of Public Policy
ANU College of Asia & the Pacific

Melbourne, 11 August 2014

*Preliminary, not for citation

Contents

ANU CRAWFORD SCHOOL

INTRODUCTION

ANU CRAWFORD SCHOOL

The Rationale

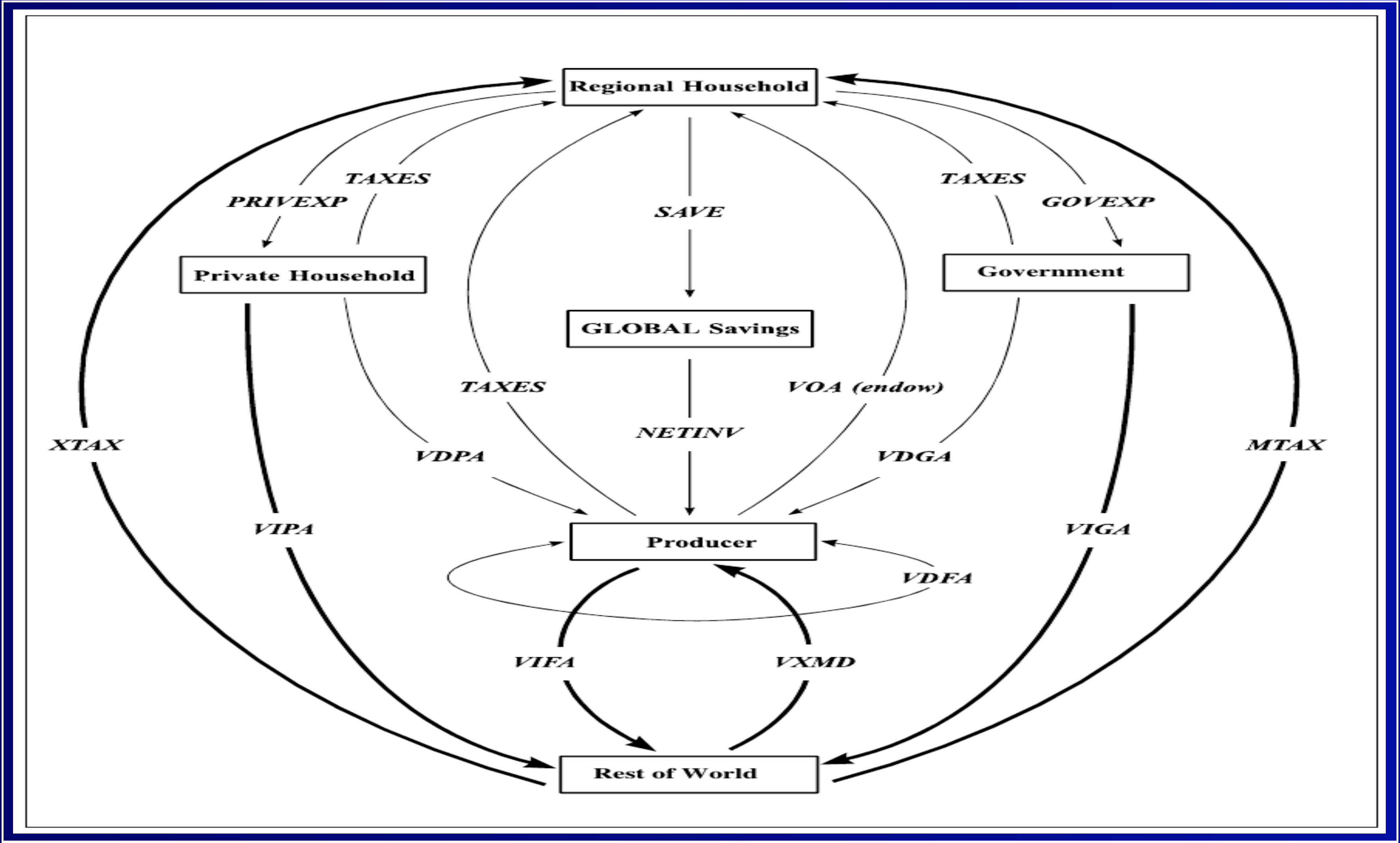
- Modelling regional economies is always a challenge to CGE modelling as they consist of multiple interacting economies (with similar structures).
- There are 2 ways to model regional economies: the 'bottom-up' and 'top-down' approaches. The 'top down' approach models the national economy and solves it first. Regional variables are linked to macro national variables in a one way interface. On the other hand, the 'bottom-up' approach builds a model for all regions in a complete national aggregate model (Klein and Glickman, 1977).
- 'Bottom-up' CGE models are more difficult to solve both in term of data requirement and computing time.
- The purpose of this research is to tackle the computational challenge of bottom-up CGE models to solve a largest bottom-up regional CGE model, the GTAP model (Hertel, 1997).

GTAP MODEL AND THE CURRENT SOLUTION METHOD

ANU CRAWFORD SCHOOL

Overview of GTAP model

Figure 1: Graphical representation of GTAP model.



Source: Brockmeier (2001)

ANU CRAWFORD SCHOOL

GTAP model's solution

- Currently, there are two software packages dedicated to solve the GTAP (and CGE models in general) model: GAMS and GEMPACK.
- The two software packages solve GTAP model by direct LU decomposition of its first order differential matrix. GAMS uses an iterative method to solve a CGE model as a system of nonlinear equations (or constraints), meanwhile, GEMPACK uses linear approximations (see Pham and Kompas, under review, for more details).
- In the following sections, we will compare our method with MA48, the core engine behind GEMPACK, the fastest CGE model solver available on the market.

DBBD MATRIX AND DIRECT METHOD FOR SOLVING LINEAR SYSTEM

The DBBD matrix form and its corresponding linear system

$$\begin{pmatrix} A_1 & & & & C_1 \\ & A_2 & & & C_2 \\ & & \dots & & \dots \\ & & & A_K & C_K \\ B_1 & B_2 & \dots & B_K & D \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_K \\ x_d \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_K \\ y_d \end{pmatrix} \qquad (1)$$

where A_i $(i = 1 \dots K)$ and D are rectangular matrices.

Solution algorithm (following Yamazaki and Li, 2011):

1. Solve $A_i u_i = y_i$ problem.
2. Using the same LU decomposition solve the multiple right hand side problem: $A_i v_i = C_i$.
3. Solve the problem: $(D - \sum_i^K B_i v_i) x_d = y_d - \sum_i^K B_i u_i$.
4. Calculate $x_i = u_i - v_i x_d$.

ANU CRAWFORD SCHOOL

DBBD matrix solution (a modified version of the solution method used by Yamazaki and Li (2011))

- The LU decomposition and linear equations solved in Steps 1 and 2 can be done in parallel before the result can be fed back into the leading process in Step 3.
- Step 4 can be done in parallel before the result can be transmitted back to the leading process to assemble the solution vector x.
- The v_i matrix is potentially a dense matrix, hence it should never been directly stored to conserve the memory. Instead of storing v_i we store $B_i v_i$. For the last step, we define:

$$A_i v_i x_d = C_i x_d \tag{2}$$

$$A_i \eta_i = C_i x_d \tag{3}$$

- By solving Equation 3 for η_i as a result of the matrix multiplication $v_i x_d$, we again can avoid storing v_i explicitly.

GTAP MODEL AND DBBD FORM

Bottom-up regional CGE models and DBBD form

- General form of a bottom-up regional CGE model.

$$0 = f[\sum_{j}^{J} x(r,s,i,...), \sum_{j}^{J} y(s,i,...)] \quad \forall r \in R, \forall s \in S, \forall i \in I ... (4)$$

$$0 = g[\sum_{j}^{J}\sum_{r}^{R} x(r,s,i,...), \sum_{j}^{J}\sum_{r}^{R} y(s,i,...)] \quad \forall s \in S, \forall i \in I ... (5)$$

- Equation 4 and Equation 5 represent intra-regional and inter-regional equations, and x, y also represent intra and inter-regional variables.
- Regional set can have subset, but only one regional set will be chosen to classify equations and variables.

First order partial derivative matrix of the non-linear regional CGE model

$$\begin{pmatrix} f_{x(\text{``}r1\text{''},s,i\ldots)} & \cdots & 0 & f_{y(s,i,\ldots)} \\ 0 & \cdots & f_{x(\text{``}rR\text{''},s,i\ldots)} & f_{y(s,i,\ldots)} \\ g_{x(\text{``}r1\text{''},s,i\ldots)} & \cdots & g_{x(\text{``}rR\text{''},s,i\ldots)} & g_{y(s,i,\ldots)} \end{pmatrix} \begin{matrix} \forall s \in S, \forall i \in I\ldots \\ \forall s \in S, \forall i \in I\ldots \\ \forall s \in S, \forall i \in I\ldots \end{matrix} \qquad (6)$$

GTAP model and DBBD direct matrix ordering technique (1)

- The current version of GTAP model consists of 210 (groups of) equations, indexed by 12 sets in the model.
- GTAP's equation groups can be classified into 4 kinds of equations: those with 2 regional indices (source and destination), with 1 regional index, without regional index and scalar equations:

  1. 11 equations with more than one regional index.
  2. 172 equations with one regional index.
  3. The rest of equation groups are: those with no regional index (15), with no index at all (12).

- Inter-regional equations will be moved to the bottom, intra-regional equations will be reordered by regions.

ANU CRAWFORD SCHOOL

GTAP model and DBBD direct matrix ordering technique (2)

- There are 243 variables in GTAP model. Variables also do have dimensions ranging from a scalar variable (zero dimension) to up to 4 dimensions. The reordering variables also depend on a regional index:

  1. 11 variables with more than 2 regional indices.
  2. 196 variables with one regional index.
  3. The rest of variables, which include 23 variables with no regional index and 13 scalar variables.

- Similarly to the case of equations, inter-regional variables will be moved rightward, intra-regional variables will also be reordered by regions. The combination of reordering of intra-regional equations and variables will result in the block diagonal parts of the DBBD matrix.

Post ordering preparation matrix for parallel solution

- Our algorithm requires the matrices A_i to have full rank and square.
- There is no guarantee that the matrices A_i will have full rank and square.
- Columns or rows of the block diagonal matrices should be dropped and shift to the border to ensure the block diagonal matrices are rectangular and have the full rank. We will employ the MA51 (HSL, 2013) procedure to perform the task.

NUMERICAL ANALYSIS

GTAP model's database

- The paper uses GTAP database version 6 (Dimaranan, 2006). The full database includes 87 regions (countries) and 57 commodities.

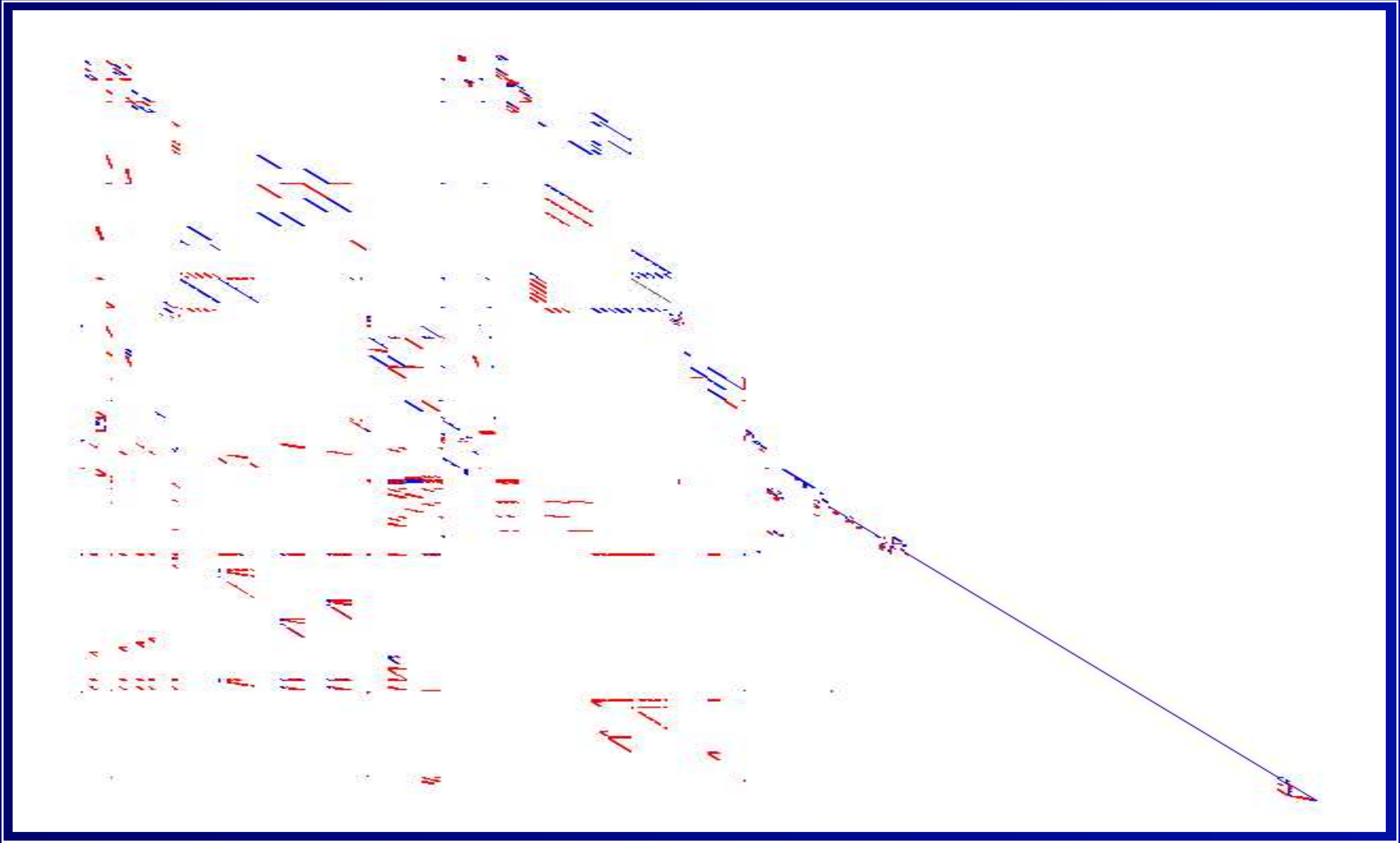
Table 1: GTAP model with different database aggregation levels.

ID	Model's Size	Number of endogenous variables	Number of exogenous variables	Number of non-zeros
1	3 regions, 3 commodities	1910	568	7158
2	87 regions, 3 commodities	295070	103300	1442793
3	87 regions, 5 commodities	490846	172218	2416111
4	87 regions, 26 commodities	2966704	1021920	13455519
5	87 regions, 57 commodities	10612758	3559840	42212180

Source: Author's calculation.

Graphical representation of the un-ordered matrix

Figure 2: The matrix without ordering.

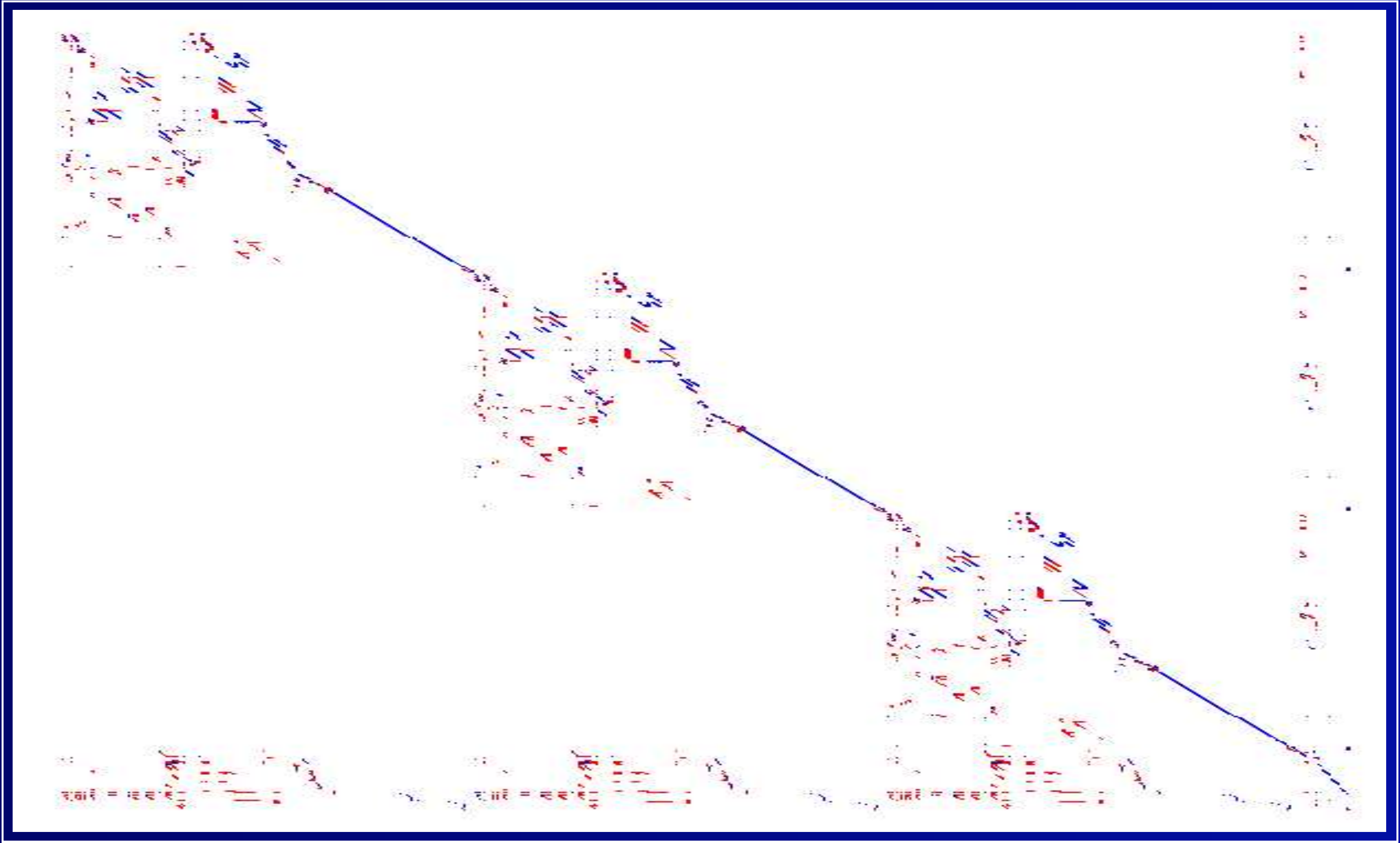


Source: Author's calculation.
The matrix plot function is from PETSC (Balay et al., 1997, 2014, 2013).

ANU CRAWFORD SCHOOL

Graphical representation of the ordered matrix

Figure 3: The reordered matrix.



Source: Author's calculation.
The matrix plot function is from PETSC (Balay et al., 1997, 2014, 2013).

Serial computing performance

Table 2: Calculation time in sec.

ID	MA48	DBBD reordering
1	0.001462	0.005651
2	2.024172	3.398415
3	5.704002	8.592551
4	238.063137	191.910587
5	2503.745974	1800.766207

Source: Author's calculation.
MA48 package from HSL library (HSL, 2013) has been used for matrix solution.
The parallel computing exercises are carried out with 3 Lenovo computers:
Intel Core i7-4770 Processor (8MB Cache, up to 3.90GHz), 32 G ram 128 SSD.
All numerical experiments are carrying out with one step Johansen method (see Pearson, 1991; Dixon et al., 1992, for clarification of the method)). The time is counted for linear system (matrix) solution only.

Parallel computing performance (1)

Table 3: Parallel computing performance shared vs distributed memory environment (in sec.).

	Shared memory		Distributed memory	
	2 processes on machine 1	3 processes on machine 1	2 processes on 2 machines	3 processes on 3 machines
1	0.004633	0.003049	0.012293	0.020846
2	1.837131	1.356841	2.106944	2.008004
3	4.396639	4.587221	4.983047	4.482060
4	183.761794	164.795130	134.642178	104.518418
5	1265.381280	1314.069277	1084.619550	863.920541

Source: Author's calculation.
MA48 package from HSL library (HSL, 2013) has been used for matrix solution.
The parallel computing exercises are carried out with 3 Lenovo computers:
Intel Core i7-4770 Processor (8MB Cache, up to 3.90GHz), 32 G ram 128 SSD.
All numerical experiments are carrying out with one step Johansen method (see Pearson, 1991; Dixon et al., 1992, for clarification of the method)). The time is counted for linear system (matrix) solution only.

ANU CRAWFORD SCHOOL

Parallel computing performance (2)

Table 4: Parallel computing performance in a mixed shared and distributed memory environment.

	4 processes on 2 machines	6 processes on 2 machines	6 processes on 3 machines	9 processes on 3 machines
2	1.772112	1.635990	1.991821	1.977537
3	3.190637	2.943495	3.533839	4.129342
4	128.249460	85.860631	95.069904	93.379519
5	801.508666	894.805971	704.188663	736.697682

Source: Author's calculation.
MA48 package from HSL library (HSL, 2013) has been used for matrix solution.
The parallel computing exercises are carried out with 3 Lenovo computers:
Intel Core i7-4770 Processor (8MB Cache, up to 3.90GHz), 32 G ram 128 SSD.
All numerical experiments are carrying out with one step Johansen method (see Pearson, 1991; Dixon et al., 1992, for clarification of the method)). The time is counted for linear system (matrix) solution only.

CONCLUSION

Some findings

- The numerical experiment shows a clear advantage of our direct ordering method to solve GTAP model in parallel.
- Together with our other work (Pham and Kompas, under review), we have proved that parallel computing has a clear advantage in solution of CGE models.
- Our works show the need for a new CGE model solver, which can recognise the special structure of CGE model to solve it more efficiently.

REFERENCES

ANU CRAWFORD SCHOOL

Cited References

Betina V. Dimaranan, editor. *Global Trade, Assistance, and Production: The GTAP 6 Data Base*. Center for Global Trade Analysis, Purdue University, 2006

HSL. A collection of fortran codes for large scale scientific computation. http://www.hsl.rl.ac.uk, 2013

Ichitaro Yamazaki and XiaoyeS. Li. On techniques to improve robustness and scalability of a parallel hybrid linear solver. In JosM.LaginhaM. Palma, Michel Dayd, Osni Marques, and JooCorreia Lopes, editors, *High Performance Computing for Computational Science VECPAR 2010*, volume 6449 of *Lecture Notes in Computer Science*, pages 421–434. Springer Berlin Heidelberg, 2011. ISBN 978-3-642-19327-9. doi: 10.1007/978-3-642-19328-6_38. URL `http://dx.doi.org/10.1007/978-3-642-19328-6_38`

K.R. Pearson. Solving nonlinear economic models accurately via a linear representation. Centre of Policy Studies/IMPACT Centre Working Papers ip-55, Monash University, Centre of Policy Studies/IMPACT Centre, July 1991. URL `http://ideas.repec.org/p/cop/wpaper/ip-55.html`

Lawrence R. Klein and Norman J. Glickman. Econometric model-building at regional level. *Regional Science and Urban Economics*, 7(12):3 – 23, 1977. ISSN 0166-0462. doi: http://dx.doi.org/10.1016/0166-0462(77)90016-3. URL `http://www.sciencedirect.com/science/article/pii/0166046277900163`

Martina Brockmeier. A graphical exposition of the gtap model. GTAP Technical Paper, No. 8, 2001

Satish Balay, William D. Gropp, Lois Curfman McInnes, and Barry F. Smith. Efficient management of parallelism in object oriented numerical software libraries. In E. Arge, A. M. Bruaset, and H. P. Langtangen, editors, *Modern Software Tools in Scientific Computing*, pages 163–202. Birkhäuser Press, 1997

Satish Balay, Mark F. Adams, Jed Brown, Peter Brune, Kris Buschelman, Victor Eijkhout, William D. Gropp, Dinesh Kaushik, Matthew G. Knepley, Lois Curfman McInnes, Karl Rupp, Barry F. Smith, and Hong Zhang. PETSc Web page. `http://www.mcs.anl.gov/petsc`, 2014. URL `http://www.mcs.anl.gov/petsc`

Satish Balay, Mark F. Adams, Jed Brown, Peter Brune, Kris Buschelman, Victor Eijkhout, William D. Gropp, Dinesh Kaushik, Matthew G. Knepley, Lois Curfman McInnes, Karl Rupp, Barry F. Smith, and Hong Zhang. PETSc users manual. Technical Report ANL-95/11 - Revision 3.4, Argonne National Laboratory, 2013. URL `http://www.mcs.anl.gov/petsc`

Peter B. Dixon, B.R. Parmenter, Alan A. Powell, and Peter J. Wilcoxen. Notes and problems in applied general equilibrium economics. In C.J. Bliss and M.D. Intriligator, editors, Advanced Textbooks in Economics, volume 32. North-Holland, Amsterdam, London, New York, Tokyo, 1992

Thomas W. Hertel, editor. *Global trade analysis: modeling and applications*. Cambridge University Press, Cambridge, New York, 1997

Van Ha Pham and Tom Kompas. Solving intertemporal cge models in parallel using a singly bordered block diagonal ordering technique. *Economic Modelling*, under review