EVPI–Based Importance Sampling Solution Procedures for Multistage Stochastic Linear Programmes on Parallel MIMD Architectures

M. A. H. Dempster and R. T. Thompson
Judge Institute of Management Studies
Trumpington Street
University of Cambridge, UK

mahd@cam.ac.uk and rt208@cam.ac.uk

Abstract

Multistage stochastic linear programming has many practical applications for problems whose current decisions have to be made under future uncertainty. There are a variety of methods for solving the deterministic equivalent forms of these dynamic problems, including the simplex and interior point methods and nested Benders decomposition – which decomposes the original problem into a set of smaller linear programming problems and has recently been shown to be superior to the alternatives for large problems. The Benders subproblems can be visualised as being attached to the nodes of a tree which is formed from the realisations of the random data vectors determining the uncertainty in the problem. Parallel versions of the nested Benders algorithm involve two obvious techniques for parallelising the associated tree structure for multiprocessors or multicomputers – subtree parallelisation or a nodal parallelisation – both of which utilise a farming approach. The nodal parallelisation technique is presented in this paper, as it balances load more efficiently than its alternative. Differing structures of the test problems cause differing levels of speed–up on a variety of multicomputing platforms: problems with few variables and constraints per node do not gain from this parallelisation. Stage aggregation has been successfully employed for such problems to improve their parallel solution efficiency by increasing the size of the nodes and therefore the time spent calculating relative to the time spent communicating between processors. A parallel version of an importance sampling solution algorithm based on local EVPI information has been developed for extremely large multistage stochastic linear programmes which either have too many data paths to solve directly or a continuous distribution of possible realisations. It utilises the parallel nested Benders algorithm and a parallel version of an algorithm designed to calculate the local EVPI values for the nodes of the tree and achieves near linear speed–up.

Key Words: Linear programming, dynamic stochastic programming, nested Benders decomposition, parallel algorithms, aggregation, MIMD computers

1 Introduction

Stochastic programming is the extension to the field of *mathematical programming* obtained by introducing *uncertainty* into the problem data and decision process. Such models are necessary when the effectiveness of current decisions is dependent on future events. Fields in which stochastic programming models arise include economics, financial planning [7] and telecommunications; many examples can be found in [14, 21]. Given an appropriate *probability space* $(\Omega, \mathcal{F}, \mathrm{P})$ where Ω is a sample space with sample points ω, $\mathcal{F}$ is the σ–field of subsets of Ω and P is a probability measure, and taking the (von Neumann–Morgenstern) *expected utility* of the problem objective, leads to a general formulation for stochastic programming.

Two main approaches to *static* stochastic programming have been developed: the *chance constrained* and *recourse* formulations. The recourse formulation or *two-stage* stochastic programming problem has a later second stage that is dependent on the decision vector from the first stage. This model can be extended to the multistage case, when a set of decisions $\{x_1, x_2, \ldots, x_T\}$ are to be made over time, up to the *horizon* T. The sequence of random vectors $\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \ldots, \boldsymbol{\omega}_T$ yield the *discrete time (data) stochastic process*: $\boldsymbol{\omega} := (\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \ldots, \boldsymbol{\omega}_T)$ and the notation $\boldsymbol{\omega}^t := (\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \ldots, \boldsymbol{\omega}_t)$ is used to represent this stochastic process up to time t. In the multistage problem an observation is made then a decision is taken, followed by another observation and another decision and so on: $\xi_1 \rightsquigarrow x_1, \xi_2 \rightsquigarrow x_2, \ldots, \xi_{T-1} \rightsquigarrow x_{T-1}, \xi_T \rightsquigarrow x_T$. The first observation and decision are usually deterministic in these models. A *filtration* can be used to introduce *nonanticipativity* of the decisions into the recourse model.

If the stochastic process is *discrete* state the expected criterion is a summation and if all the problem functions are linear (and therefore separable) we obtain the *linear multistage recourse problem* model

$$
\begin{array}{lll}
 & \min_{x_1} & \{c_1' x_1 + Q(\mathbf{x}_2) : A_{11} x_1 = b_1\} \\
(\mathbf{LMRP}) & \text{where} & Q(\mathbf{x}_t) := \mathbb{E}_{\boldsymbol{\omega}_t}\{\min_{\mathbf{x}_t} \mathbf{c}'_t \mathbf{x}_t + Q(\mathbf{x}_{t+1}) : \sum_{\tau=1}^{t} \mathbf{A}_{t,\tau} \mathbf{x}_\tau = \mathbf{b}_t \ \ a.s.\} \\
 & \text{and} & Q(\mathbf{x}_{T+1}) \text{ is specified} \\
 & & l_1 \leq x \leq u_1 \\
 & & \mathbf{l}_t \leq \mathbf{x}_t \leq \mathbf{u}_t \ a.s. \quad t = 2, \ldots, T. \qquad (1)
\end{array}
$$

Here all constraints involving random variables hold *almost surely (a.s)*, i.e. with probability one. There are several overviews of the stochastic programming literature which contain information on chance constrained methods, approximating schemes and problems with infinite horizons, see [14, 21, 15, 27]. Dupacova [18] has recently surveyed the multistage stochastic programming literature.

The *deterministic equivalent* of problem (1) can be taken by replacing the expectations by probability weighted summations over the finite number of data paths and writing deterministic constraints for each path. The size of the associated constraint matrix increases exponentially with the number of time stages and the number of outcomes for each random vector $\boldsymbol{\omega}_t$. Each of the data paths $\omega := \{\omega_1, \omega_2, \ldots, \omega_T\}$ is referred to as a *scenario*. The large–scale deterministic equivalent form of problem (1) can be solved by the simplex or interior point method. Alternatively decomposition methods such as *augmented Lagrangian decomposition* [15, 32] or *Benders decomposition* [3, 36, 4, 23] can be employed. Benders decomposition is more suitable for decomposing stochastic linear programmes than Dantzig–Wolfe decomposition [12] and was first developed by Van Slyke and Wets [36] into a cutting plane algorithm for the two–stage problem. This has been extended to the *nested decomposition* method by Birge [4] and Gassmann [23] for linear multistage recourse problems.

Sampling methods are required to reduce the large dimensionality of realistic deterministic equivalent problems, while parallel programming allows problems to be solved more quickly or more detailed problems to be treated. In Section 2 nested Benders decomposition is reviewed and the associated algorithm explained.

Section 3 gives an outline of the MSLiP nested Benders code, while Section 4 describes the parallel version of the code and gives parallel results. Section 5 explains the concept of *expected value of perfect information* (EVPI) [15, 16, 10] and its calculation while Section 6 contains the EVPI–S sampling procedure [9]. The parallelisation of the EVPI algorithms is found in Section 7 which includes numerical results, and conclusions are drawn in Section 8.

2 Nested Benders decomposition

In the discrete case, the data sample path (scenario) space is partitioned into sets $A_t \in \mathcal{A}_t \subset \mathcal{A}_T := \Omega$. The sets A_t contain realizations which are identical up to period t and are distinct at the horizon T, so that the stochastic process can be represented by a tree structure. There is a 1-1 mapping between the outcomes of the random variables $\boldsymbol{\omega}^t$ and the *nodes* of the tree at stage t for $t = 1, \ldots T$. The realisation ω^t can thus be used to represent a node of the tree at stage t. The hierarchical structure of the nodes can be described in the same manner as a family, with ancestors, descendants, parents, children, siblings and cousins. The nested Benders method solves the problem in a recursive manner. It can be illustrated by taking a particular node of the decision tree defined by the data path ω^t and its descendants:

$$\begin{aligned} \min \quad & \{c_t(\omega^t)' x_t(\omega^t) + Q(\omega^t, x_t(\omega^t)) \ : \ A_{t,t}(\omega^t) x_t = b_t(\omega^t) - B(\omega^t, x_t(\omega^t))\} = \\ \min \quad & \{c_t(\omega^t)' x_t(\omega^t) + \theta_t(\omega^t) : A_{t,t}(\omega^t) x_t = b_t(\omega^t) - B(\omega^t, x_t(\omega^t)) \\ & \theta_t(\omega^t) \geq Q(\omega^t, x_t(\omega^t))\}, \end{aligned} \tag{2}$$

where

$$B(\omega^t, x_t(\omega^t)) := \sum_{\tau=1}^{t-1} A_{t,\tau}(\omega^\tau) x_\tau(\omega^\tau)$$

and $Q(\omega^t, x_t(\omega^t)) :=$

$$\begin{array}{ll} \min & \sum_{\omega^{t+1}|\omega^t} p(\omega^{t+1}) c_{t+1}(\omega^{t+1}) x_{t+1}(\omega^{t+1}) \\ \text{s.t.} & A_{t+1,t+1}(\omega^{t+1}) x_{t+1}(\omega^{t+1}) \ = b_{t+1}(\omega^{t+1}) - B(\omega^{t+1}, x_{t+1}(\omega^{t+1})) \\ & l_t \le x_t(\omega^t) \le u_t, \quad l_{t+1} \le x_{t+1} \le u_{t+1}. \end{array} \tag{3}$$

Both the corresponding primal and dual problems can be decomposed into sets of smaller subproblems which can be solved using the simplex method. The number of these subproblems is equal to the number of descendants of the node ω^t. The dual problems are all feasible, since the dual decision $y_{t+1}(\omega^{t+1})$ is unconstrained. If the dual problem is unbounded, then the primal problem is infeasible, and a feasibility cut can be placed in the parent node problem. Otherwise, for an optimal solution for (3) either a *single optimality cut* is placed, if weighted sums are taken, or these can be disaggregated to give a set of cuts, known as *multicuts*.

Then having a collection of dual feasible vectors $(y^i_{t+1}(\omega^{t+1})', \lambda^{i'}, \mu^{i'})$; $i = 1, \ldots, I_t(\omega^t)$, and directions $(y^j_{t+1}(\omega^{t+1})', \lambda^{j'}, \mu^{j'})$, $j = 1, \ldots, J_t(\omega^t)$, we obtain the problem

$$\begin{array}{lll} \min\limits_{x_t(\omega^t)} & \{c_t(\omega^t)' x_t(\omega^t) + \theta_t(\omega^t)\} & \\ \text{s.t.} & A_{tt}(\omega^t) x_t(\omega^t) & = b_t(\omega^t) - B(\omega^t, x_t(\omega^t)) \\ & y^j_{t+1}(\omega^{t+1})' A_{t+1,t}(\omega^t) x_t(\omega^t) & \ge y^j_{t+1}(\omega^{t+1})' b_{t+1}(\omega^{t+1}) + \lambda^{j'} l_{t+1} - \mu^{j'} u_{t+1} \\ & & -y^j_{t+1}(\omega^{t+1})' \sum_{\tau=1}^{t-1} A_{t+1,\tau}(\omega^\tau) x_\tau(\omega^\tau) \\ & \sum_{\omega^{t+1}|\omega^t} p(\omega^{t+1}) y^i_{t+1}(\omega^{t+1})' A_{t+1,t}(\omega^t) x_t(\omega^t) & \ge \sum_{\omega^{t+1}|\omega^t} p(\omega^{t+1}) \left\{ y^i_{t+1}(\omega^{t+1})' + \theta_t(\omega^t) \right. \\ & & \left. [b_{t+1}(\omega^{t+1}) - \sum_{\tau=1}^{t-1} A_{t+1,\tau}(\omega^\tau) x_\tau(\omega^\tau)] + \lambda^{i'} l_{t+1} - \mu^{i'} u_{t+1} \right\} \\ & & l_t \le x_t(\omega^t) \le u_t \\ & & l_{t+1} \le x_{t+1}(\omega_t) \le u_{t+1} \\ & & i = 1, ..., I_t(\omega^t), \quad j = 1, ..., J_t(\omega^t). \end{array} \tag{4}$$

The objective $c_t(\omega^t)' x_t(\omega^t) + \theta_t(\omega^t)$ gives a lower bound on the optimal objective value for the nodal problem, while $c_t(\omega^t)' x_t(\omega^t) + Q(\omega^t, x_t(\omega^t))$ gives an upper bound. So if $\theta_t(\omega^t) \ge Q(\omega^t, x_t(\omega^t))$, the nodal problem is fully solved for the right hand side supplied to it. The relaxed problem is solved iteratively, producing optimality and feasibility cuts until this condition is satisfied.

The *nested Benders decomposition* algorithm based on the above ideas can be stated as follows (*cf.* [4, 24]).

Step 0. Solve the root node (first stage problem). At iteration zero, in problem (4), set $\theta_1 := I_1(\omega^1) := J_1(\omega^1) := 0$ and take the first constraint to be $A_{11}x_1 = b_1$. Set $I_t(\omega^t) := J_t(\omega^t) := 0 \ \forall \ t, \ \omega^t$. If problem (4) is infeasible, stop – the original problem is infeasible . If $\theta_1 \geq Q(x_1)$, stop — the optimal solution has been reached. Otherwise set t:=t+1 and let $\tilde{x}_1$ be the current solution vector to the root problem.

Step 1. The formation of the next stage problems can be done by using the decision vectors $x_{t-1}(\omega^{t-1})$ from ancestors to form the right hand sides of the constraints for t for all the outcomes of ω^t whose parents have an optimal solution, to give the appropriate problem (4). Some nodal problems at this time stage will not be solved due to the infeasibility of their parent or parent's siblings problems. If a nodal problem is found to be infeasible, place a feasibility cut in the parent node problem. This branch of the tree is not explored any further until the parent node problem is re-solved to optimality. There is the choice to move either forward to $t+1$ and go to **Step 1**, or backward to $t-1$ and go to **Step 2**. If $t = T$ or all nodal problems in the current time period are infeasible then go to **Step 2**.

Step 2. Set $\theta_t(\omega^t) := -\infty$ for all values which have not been previously been set in this time period. If all the descendant node problems at node ω^t have optimal solutions, a single optimality cut may be placed in the node ω^t problem. Otherwise, a feasibility cut has been previously placed. In the case of multicuts, both feasibility and optimality cuts may be placed in the node ω^t problem. There is the choice to move either forward to $t+1$ and go to **Step 1**, or backward to $t-1$ and go to **Step 2**. If $t = T$ or all nodal problems in the current time period are infeasible, then go to **Step 2**. If $t = 1$, go to **Step 0**.

Problems at any stage need only be re–solved if supplied with a new right hand side or if $\theta_t(\omega^t) < Q(\omega^t, x_t(\omega^t))$, when new cut information is supplied to

the problem at ω^t. The *sequencing protocol* determines whether to descend the tree further in **Step 1** (termed a *forward pass*) or return to the previous stage of the tree in **Step 2** (termed a *backward pass*). The *fast–forward–fast–back method* [37] suggests continuing in the current direction whenever possible and appears to be the most efficient sequencing protocol [24].

3 The MSLiP code

The MSLiP code has been developed over a number of years with the combined efforts of a variety of programmers [4, 23]. The current version (8.3) is an improvement on the version described in Gassmann's working paper [24]. The MSLiP nested Benders decomposition code (see Figure 1) consists of the following modules:-

1) An LP solver (the NORMAL routine). The LP solver found in the code is a modification of the Pfefferkorn–Tomlin LP code [31] which uses the product form of the inverse. The pricing has recently been improved to include steepest edge and random pricing, as well as the default most negative cost. It has no crash–to–feasibility heuristics. Basis, primal, dual and right hand side vectors are needed from the solution of each nodal problem.

2) The subproblem manager routine (NDCOM) decides which direction in the tree to move, which is dependent on the sequencing protocol. After a problem is solved it is marked infeasible or optimal. NDCOM will then decide to place a cut in the previous time stage and re–solve or form the right hand side for the next time stage. Each nodal problem is marked if it needs to be re–solved.

3) The cut routines form the optimality (LKHDCT) or feasibility (FEASCT) cuts to be placed in the appropriate nodal problem. The optimality cuts depend on whether the user has defined single or multicuts in the specification input file. The default in MSLiP is single cuts for the two–stage problem and multicuts for

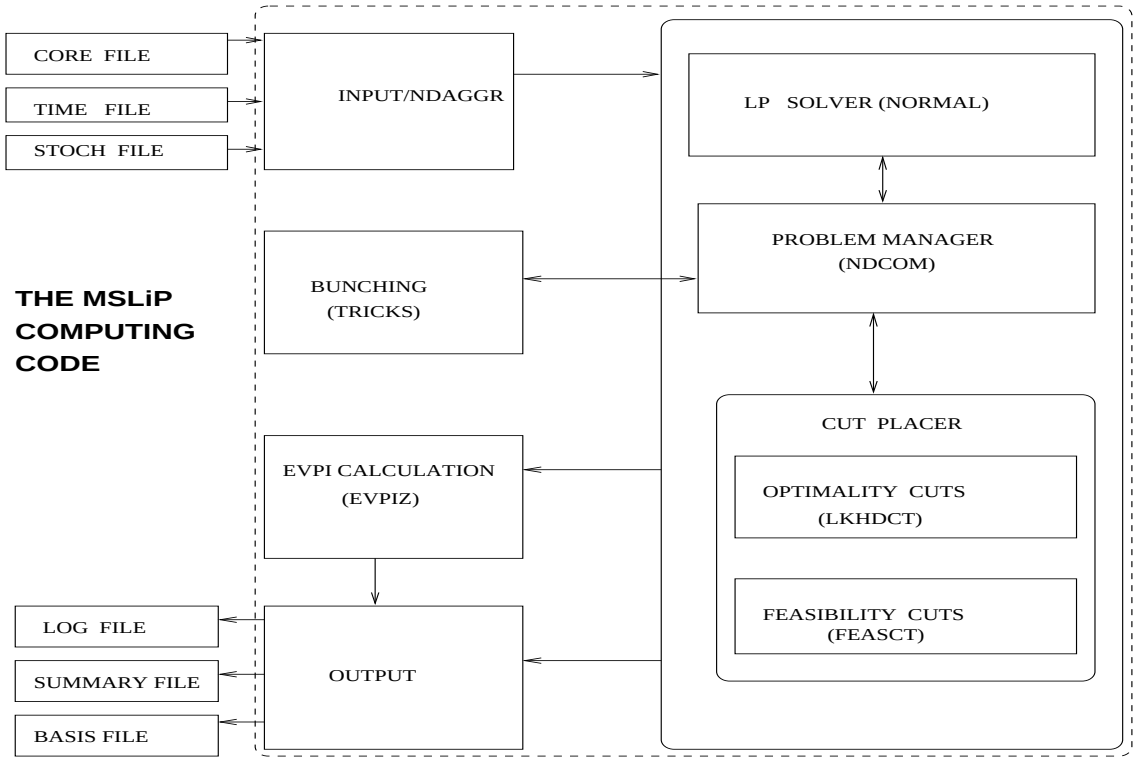


Figure 1: The MSLiP System

multistage problems.

4) The bunching routine (TRICKS) tries to reduce the amount of calculation done in solving a set of LP problems which differ only in their right hand sides. These problems may share bases and pivots, and may even have the same optimal basis (see [23]).

5) After the solution has been reached there is a routine (EVPIZ) for calculating the EVPI at each node of the scenario tree. This is an indicator useful in determining how 'stochastic' the problem actually is. Calculation of the EVPI process is explained in Section 5.

There are two other Benders based codes which have been developed for solving multistage SLPs: ND–UM [5] and SP/OSL [28]. Both are based on the IBM subroutine library OSL [26] which performs the simplex calculations within the Benders algorithm. The hybrid code MSLiP–OSL [17, 34] has been developed to

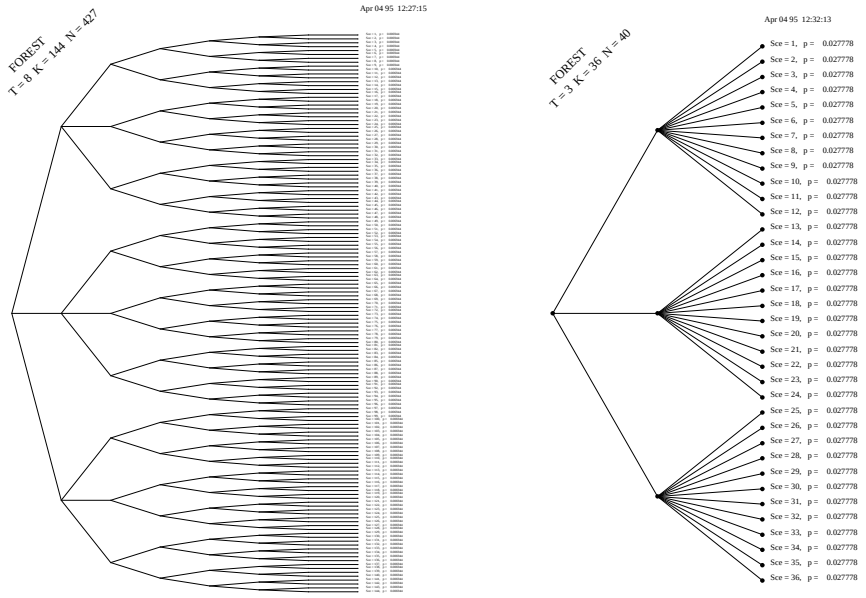


Figure 2: Aggregation of FOREST.8.144

give a more stable and accurate solution to each nodal LP problem, as well as to allow the possibility of crash–to–feasibility and presolve procedures.

An additional routine NDAGGR has been added to the MSLiP and MSLiP–OSL codes, which allows stages to be aggregated, see [17, 34, 35] for more details. An illustration of stage aggregation can be seen in Figure 2. The aggregation scheme for this example combined the first three stages, then the next three and finally the last two. This takes a 144 scenario problem down to a 36 scenario problem.

CPU time reduction can be obtained through stage aggregation. This may not be the case with MSLiP, where larger LP problems take an extremely long time to solve. With MSLiP–OSL however, restructuring so that the earlier stages have larger nodal dimensions, especially the first time stage, can both reduce the computational time considerably and increase the stability [17]. The nodal dimensions of the subproblems at other stages is dependent on the stochasticity; with higher stochasticity more information is required from iteration to iteration and the nodal dimensions can be larger, with lower stochasticity the nodal di-

mensions can be lower. Stage aggregation also reduces the number of iterations and cuts needed to reach a solution.

It has been shown [8, 34] that MSLiP outperforms the simplex and interior point methods for large sparse problems. This , as well as the inherent parallel properties of nested Benders decomposition and the associated sampling scheme, lead the authors to implement a parallel version of the MSLiP code.

4 Parallel MSLiP

Before explaining the parallel nested Benders decomposition method employed in this paper, a summary of other parallel methods is given. Parallelisation of the interior point method in effect means parallelisation of the Cholesky factorisation of the compound matrix used in calculating the iterative step. This has been done in a number of ways, by vectorising the procedure, by making use of the elimination tree to distribute the columns of the matrix to processors [29], by using a block Cholesky factorisation and distributing the blocks [11] and by using the dual Sherman–Morrison decomposition formulation for multistage recourse problems of stochastic programming due to Birge–Qi [38]. Parallelisation of the simplex method on the other hand is extremely difficult [22].

The augmented Lagrangian decomposition method has been parallelised by Ruszczynski [33], while the similar row-action equivalent to this method used by Nielsen and Zenios [30] again decomposes by scenarios. Speed–up has been achieved for the two-stage problem on Thinking Machine's CM-2. Dantzig *et al* [13] have a parallel implementation of their importance sampling based Benders decomposition code, while Ariyawansa and Hudson [2] parallelised the two-stage problem in a similar manner to the Dantzig method – the main difference being that each processor is assigned a subset of the second stage nodal problems. Hiller and Eckstein [19] have produced a Benders-based decomposition scheme as

a subsystem to a portfolio selection algorithm. Speed–up by a factor of 32 was obtained for the Benders-based portion of this algorithm on a CM-2 with 16,000 processing elements.

A parallel nested Benders decomposition which uses a farming strategy has been formed from the ND–UM code of Birge *et al* [6]. The master processor starts the computation and initialises the slave processors. A subtree is solved on each slave processor, with the root node of the subtree lying in a stage (not the first) of the whole tree structure. The master solves the portion of the tree up to the root node of the subtree. Slave processors do not branch from other slave processors.

A parallel version of MSLiP has been implemented on Fujitsu's AP1000 and IBM's SP2. See [17] for a description of the AP1000. The SP2 had 16 IBM RS6000 390 Power 2 processors, with 128 Mbytes DRAM and double precision peak performance of 250 Mflops. Any speed–up achieved is from a level which is faster than most sequential computers. The host workstation was an RS6000 PowerPC processor running under MPI and the theoretical peak for communication bandwidth was 40 Mb/s.

The *parallel algorithm* described below is more similar to the methods of [2, 20, 13] than to that of [6]. The idea is to achieve load balancing which can not be achieved by splitting the tree into a set of subtrees at the possible expense of communication overhead. It uses the *farming* technique, with a *master* and a number of *slaves* decided by the user. The master sends LP information to a slave processor when that slave becomes available. The method can be described for a multicomputer attached to a host processor as follows.

Host

Step H: 0 Send master module to processor 0. Send slave module to processors $s = 1, 2, \ldots, S$. Send data to processor 0.

Step H: 4 Receive optimal solution from processor 0.

Master processor

Step M: 0 Receive module from host. Receive data from host.

Step M: 1 Solve master problem, obtaining trial solution.

Step M: 2 Assuming there are $m \geq S$ nodes to be solved in the current time stage, send information for the first S subproblems to slave processors $s = 1, 2, \ldots, S$.

Step M: 3 When an optimal solution is returned from a processor $s\epsilon\{1, 2, \ldots, S\}$, send the information to solve another nodal problem to processor s unless there are no problems left to be sent. If there are no nodal problems remaining to be sent, go to **Step M: 4**. If an infeasible solution is returned from a processor go to **Step M: 4**. If there is only one nodal problem left to be sent and there is only one slave, this problem is solved on the master processor.

Step M: 4

When all nodal problems have been solved or a problem is found infeasible, the master determines whether to move forward or backward in the tree. The master forms the cuts for all nodes.

On reaching the first stage in the tree and finding there are no more subproblems to be solved in the tree, the optimal solution has been reached. It is returned to the host, and the slave processors are informed to stop.

Otherwise return to **Step M: 2**.

Slave processor

Step S: 0 Receive module from host.

Step S: 2 Receive and solve LP problem sent from master.

Step S: 3 Return solution to master and return to **Step S: 2**.

Step S: 4 Receive message from master to stop and stop.

A slave processor requires new constraints (cuts), new right hand sides and basis information to be sent from the master processor, while the master processor

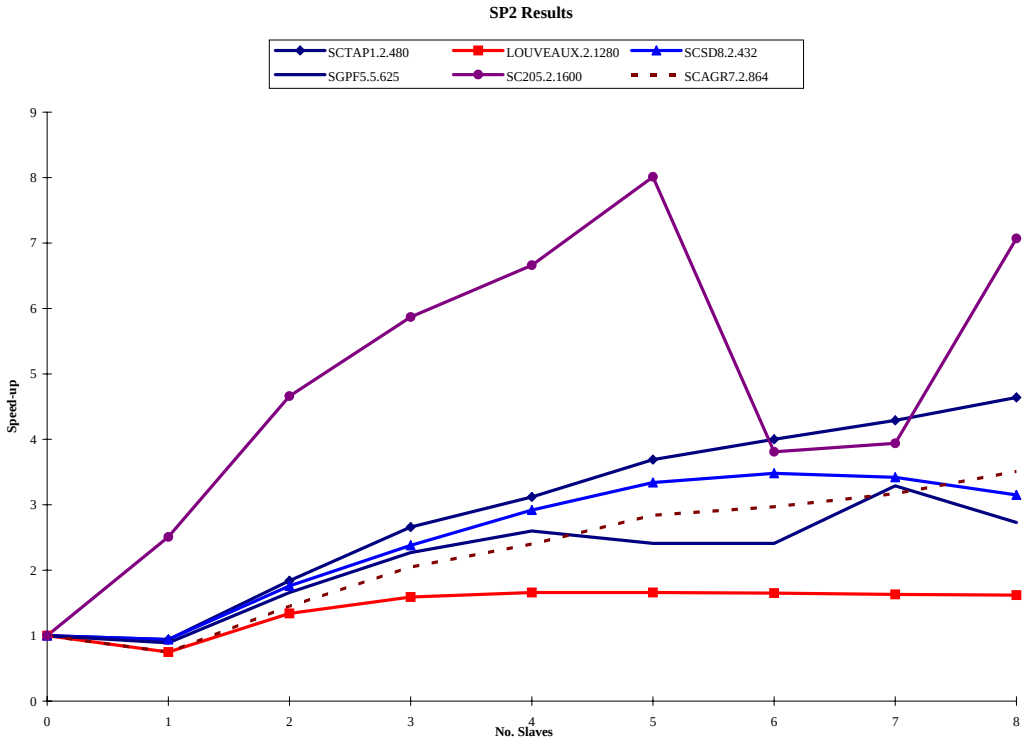
Name	Rows	Cols	Non	Obj.val.	Tree	EVPI	MSLiP CPU
SC205.2.1600	35212	35214	96030	0.0	1600	-	31.25
SCSD8.2.432	8650	60550	190210	25.80	432	0.24	13.73
SCAGR7.2.864	32847	34580	108906	-834446	864	0.01	29.67
SCTAP1.2.480	28830	46128	169068	248.50	480	0.20	25.59
LOUV.2.1280	8962	20484	40968	482.64	1280	9.35	60.34
FOREST.8.384	15667	14927	62795	-43759	$4^2.3.2^3.1$	2.75	32.86
FOREST.8.512	20771	19791	83307	-43869	$4^3.2^3.1$	2.51	40.64
SGPF5.5.625	49202	61759	165570	-5201.282	5^4	10.84	41.07
WATSON.10.512.I	67069	128001	350728	-1959.63	2^9	44.63	201.96
WATSON.10.1024.I	134127	255987	701428	-1926.79	4.2^8	46.53	409.97

Table 1: Problem characteristics - A summary

requires a new primal vector, a new dual vector, basis information and the status of the LP problem — feasible or optimal – from a slave. The slave processors operate as LP solvers in an asynchronous manner. A synchronising step occurs when the master processor changes the stage and forms the new cut information.

If the nodal subproblems are large enough this algorithm should have reasonable load balancing properties, though there may be a communications bottleneck while slave processors try to communicate with the master.

Figure 3 plots speed–up against the number of slave processors. Problems throughout this paper are labelled in the following way: *Prob.Per.Scen* where *Prob* gives the problem name, *Per* gives the number of periods and *Scen* gives the number of scenarios in the problem. Table 1 gives the deterministic equivalent dimensions of the problems presented here, the objective value, the associated tree structure, the *stochasticity* (this is defined in Section 5) and their sequential solution times on the SP2. Only a subset of the problems tested are given in



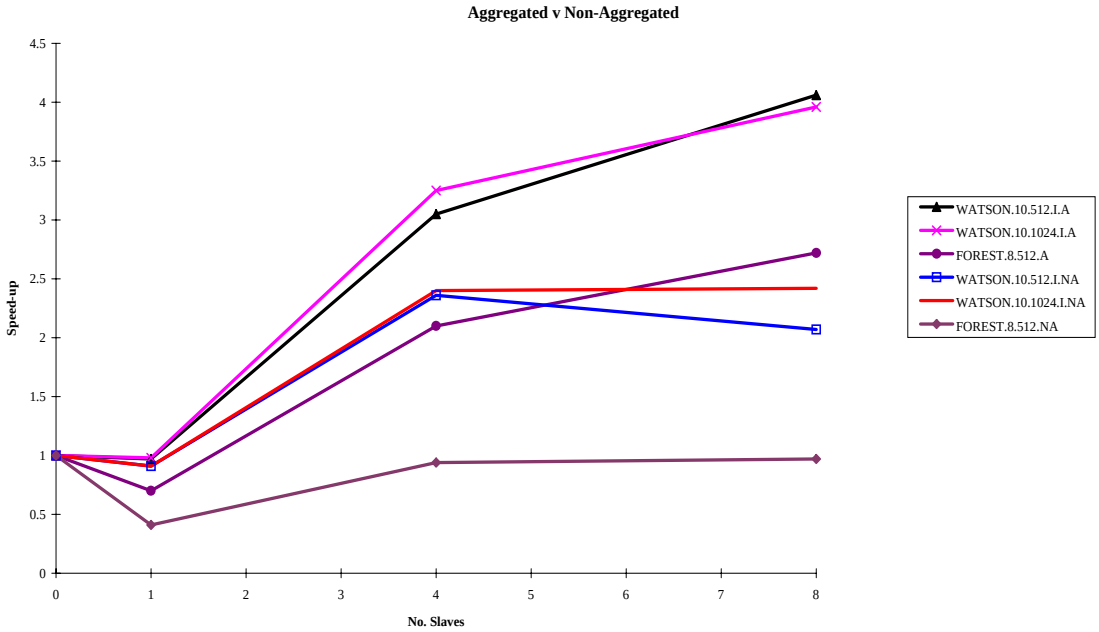


Figure 3: SP2 – various problem sets

this paper, but they show the general pattern of parallel performance [35]. It is noticeable that the parallel algorithm generally performs better on two stage problems than on multistage problems. Near linear speed-up is obtained on the two–stage SCSD8 and SCTAP1 problem sets, while speed–up was achieved on both the multistage SGPF and WATSON problem sets. The nodal problem sizes for the SGPF and WATSON problems are larger than for the FOREST multistage problems which have poorer parallel performance. It should be noted that for two–stage problems our algorithm is equivalent to both the Birge *et al* and Dantzig *et al* parallelisation techniques [6, 13].

While Amdahl's Law would indicate that the speed-up obtainable is less for multistage problems than for two stage problems, the theoretical speed-ups are not achieved. This is partly due to the tree structure; multistage problems can not be totally parallelised in later stages as there are not enough processors, while in the earlier stages there may be more processors than nodal problems to be solved, resulting in some processors remaining idle. In the multistage problems, cuts have to be passed from master to slave, and on large problems where many cuts may be created, this increases the communication time significantly. If the communication to calculation ratio becomes too high, a bottleneck develops, with too many slave processors attempting to communicate with the master processor.

To reduce the communication to calculation ratio for multistage problems, stage aggregation has been employed. Bigger nodal subproblems will lead to larger intervals between communication. A reduction in the number of cuts formed by the master should also ensue from this aggregation, reducing the number of sequential operations. The second graph in Figure 3 compares aggregated speed–ups with the non–aggregated equivalent. The aggregated problems always perform better and it can be seen that for both the FOREST and WATSON problem sets, greater speed–ups can be achieved. The parallel algorithm is reasonably efficient on up to 8 processors for the SP2.

Anomalous effects have occurred with certain problem sets where feasibility cuts have been placed. With more than one slave, a different ordering in the return of nodal subproblems may lead to a different solution path. Possible anomalies are classified as acceleration anomalies, deceleration anomalies and detrimental anomalies. *Acceleration anomalies* occur when the solution time achieved is quicker than expected, while *deceleration anomalies* are exactly the opposite. *Detrimental anomalies* are severe deceleration anomalies where the solution time achieved is slower than the sequential solution time. A similar observation has been made by Hajian *et al* [25]. (FOREST.8.512 does not experience a detrimental anomaly — there is no speed–up because the communication to calculation ratio is too high.)

5 EVPI

For a multistage recourse problem each node of the tree has an associated *expected value of perfect information* value, called the *local* EVPI, defined as:

$$\text{EVPI}_t(\omega^t) := \pi_t(\omega^t) - \phi_t(\omega^t)$$

where $\pi_t(\omega^t)$ denotes the optimal value of the problem with root node defined by ω^t and

$$\phi_t(\omega^t) := \mathbb{E}_{\boldsymbol{\omega}|\omega^t}\pi_t(\boldsymbol{\omega})$$

is the expected value of having perfect information about the future – i.e the optimal value $\pi_t(\omega)$ along each data path ω from time t onwards from the data history path ω^t. The local EVPI values form a stochastic process $\{\text{EVPI}(\boldsymbol{\omega}^t) : t = 1, \ldots, T\}$ which is a supermartingale. This leads to the following observations about the EVPI process:–

1) If the EVPI at a node is zero, all its descendant nodes have zero EVPI.

2) A node in the tree whose descendants have no branching has zero EVPI.

3) At the root node, the deterministic process value is the EVPI problem.

4) The *local stochasticity* of a node of the tree is the local EVPI normalised by the optimal value of the whole problem.

For each path which branches at or after the node ω^t the optimal value is calculated with all nonanticipativity conditions ignored. These paths are all in the set $A_t(\omega^t) \in \mathcal{A}_t \subset \mathcal{P}(\Omega)$, the set of all subsets of scenarios, at time t. The local EVPI value at a node of the tree at stage t is calculated by solving to find $\pi_t(\omega)$ for all $\omega \in A_t(\omega^t)$. Each $\pi_t(\omega)$ is an LP problem conditional on ω^{t-1} with all the data from ω^t along the path ω. This is multiplied by the conditional probability at ω^t for each path before the sum over $\omega \in A_t(\omega^t)$ is taken to obtain ϕ_t. This value is then subtracted from the actual optimal value $\pi_t(\omega^t)$ at the node to give the local EVPI.

To calculate the values of the EVPI process, this procedure has to be repeated for every node of the tree except for those in the last period.

This procedure is performed in MSLiP as follows:–

Step 1 Define an array, called *EVPI* say, with length equal to (or exceeding) the number of nodes in the tree. Set all values in the array to zero.

Step 2 Take the first scenario in the tree and set $t := T - 1$.

Step 3 Calculate the optimal value of the current path by a call to the LP solver and multiply it by the conditional probability of the occurrence of this path at the current node. Add this value to the corresponding position for this node in the *EVPI* array. Set $t := t - 1$. If $t < 1$ go to **Step 4**, otherwise repeat **Step 3**.

Step 4 Find the next path of the tree, set $t := T - 1$ and go to **Step 3**. If all paths of the tree have been traversed, go to **Step 5**.

Step 5 These calculations lead to the position where for each node ω^t, $t < T$, the value $\phi_t(\omega^t)$ has been calculated in the array *EVPI*, i.e. $EVPI_t(\omega^t) = \phi_t(\omega^t)$. For all nodes reset the value in the *EVPI* array by subtracting the calculated

value from the nodal optimal value, i.e. in programming notation

$$EVPI_t(\omega^t) := \pi_t(\omega^t) - EVPI_t(\omega^t).$$

The values $\phi_t(\omega^t)$ are stored in the same vector as the $EVPI_t(\omega^t)$ values to save memory. This implementation applies to both the MSLiP and MSLiP–OSL codes.

6 EVPI–S sampling

Corvera–Poiré and Dempster [15, 9, 10] developed a method which is largely based on artificial intelligence ideas to approximate the true EVPI and therefore the optimal solution of the problem, using *sequential importance sampling heuristics* which push the approximations towards the true optimal value. The heuristics can be used in a variety of ways to give a whole family of *EVPI–sampling* algorithms, which are termed EVPI–S.

The sampling method employed by the EVPI–S algorithms tries to represent the data path tree $\mathcal{T}$ associated with the problem by choosing a sequence of trees $\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_n$, where $\mathcal{T}_n$ is always more representative of $\mathcal{T}$ than $\mathcal{T}_{n-1}$.

The data scenario–subtree is formed and the corresponding problem is converted to the standard format SMPS using STOCHGEN (see [9]). It can then be solved with MSLiP or MSLiP–OSL and its EVPI values $\mathrm{EVPI_t}(\omega^{\mathrm{t}})$ computed. Given that the problem corresponding to $\mathcal{T}_{n-1}$ has been solved and its local EVPI values computed, one of the following may occur at each node of the tree $\mathcal{T}_{n-1}$ to form $\mathcal{T}_n$:–

1) Where the nodal EVPI is above a certain zero tolerance level and its descendents have not been considered previously (there was no branching from them in the tree $\mathcal{T}_{n-1}$), the tree is *expanded* – to a depth of one or more stages – by independent sampling conditional on the appropriate nodal data so that more multiple branches to descendent nodes are considered in $\mathcal{T}_n$.

2) At nodes where the local EVPI value is below the zero tolerance, a new path is taken by *resampling* the set of paths conditional on the path to the particular node, with or without replacement. The true local EVPI value may not be zero and so the local EVPI is thus re–estimated.

3) At nodes where the local EVPI value is found to be repeatedly below the zero tolerance over a sequence of iterations the tree is *collapsed*. It becomes more certain with each iteration that this local value is indeed zero or very small. This collapsing is in fact resampling of a single path done from the parent node of the node found to have near zero local EVPI. An iteration number increment before which EVPI values below the zero tolerance are not allowed to lead to collapsing is required. The increment can be changed at each iteration.

In [9] the ratio of local EVPI to root optimal value – which corresponds to the *local stochasticity* – is the tolerance criterion used (1 or 5 percent). The algorithm can be stated as follows:–

Step 1 Set the number of iterations to be performed, N say. (Alternatively a termination condition could be used, such as when the rate of increase in EVPI decreases below a certain threshold.) Set $n := 1$.

Step 2 Sample a scenario–subtree $\mathcal{T}_n$ based on the current EVPI information carried by $\mathcal{T}_{n-1}$.

Step 3 Generate the corresponding SMPS files for the problem corresponding to the tree $\mathcal{T}_n$, using STOCHGEN.

Step 4 Solve the multistage recourse problem defined by the SMPS files with MSLiP or MSLiP–OSL and obtain the nodal EVPI values.

Step 5 If $n < N$, set $n := n + 1$. (In the finite scenario case, transfer the sample EVPI information from $\mathcal{T}_n$ to the full tree $\mathcal{T}$). Otherwise stop, the sampling has terminated and an estimate of the optimal solution obtained.

The set of sampling procedures which are defined as above can greatly reduce the size of the problem to be solved so that more detailed problems, or problems

with a continuous distribution of data paths, may be solved.

7 Parallel EVPI

In producing a fast EVPI–S sampling system unbiased in the small sample the solution times of the nested Benders decomposition and EVPI calculations are highly important. Therefore a parallel EVPI method has been developed for MSLiP. As the parallel EVPI is within the parallel MSLiP system, the algorithm is presented below as a continuation of the nested Benders parallel algorithm.

The host processor is as before with **Step H: 4** relabelled as **Step H: 7**. It receives all the EVPI information.

Master processor

Step M: 0–Step M: 3 As before.

Step M: 4 When all nodes have been solved, the master determines whether to move forward or backward in the tree. The master forms the cuts for all nodes. On reaching the first stage in the tree and finding there are no more subproblems to be solved in the tree, the optimal solution has been reached. Send message to slaves to perform EVPI calculation and go to **Step M: 5**. Otherwise return to **Step M: 2**.

Step M: 5 Take the first scenario in the tree and set $t := T - 1$. Assuming there are $m \geq S$ paths in the tree, send information defining the first S paths to slave processors $s = 1, 2, \ldots, S$.

Step M: 6 When a solution for each LP defined along the path has been returned from a processor $s \in \{1, 2, \ldots, S\}$, add the value of each node in the path to the corresponding position for each node in the *EVPI* array. Send the information defining the next path to processor s unless there are no paths left to be sent. If there are no paths left to be sent go to **Step M: 7**.

Step M: 7 The EVPI for each node in the tree is calculated by subtracting

Name	Slaves								
	Serial	1	2	3	4	5	6	7	8
WATSON.10.512.I	1.00	0.97	1.95	2.92	3.81	4.84	5.80	6.77	
WATSON.10.1024.I	1.00	1.00	2.02	3.03	4.01	5.03	6.04	6.99	7.97

Table 2: SP2 EVPI results

the appropriate *EVPI* array value from the nodal optimal value.

The optimal solution has been reached and the EVPI values calculated for nodes of the tree. This information is returned to the host, and the slave processors are informed to stop.

Slave processor

Step S: 0–Step S: 3 As before.

Step S: 4 Receive message from master to perform EVPI calculation.

Step S: 5 Receive path information or termination condition from the master. If the termination condition has been received then stop. Otherwise calculate the optimal value of this path, multiply by the conditional probability of the occurrence of this path at the current node and store the value. Let $t = t - 1$ and repeat **Step S: 5**. Otherwise go to **Step S: 6**.

Step S: 6 If $t < 1$ the set of optimal values, one for each node (apart from the leaf node) along the path are sent to the master processor. Return to **Step S: 5**.

The communication required between the master and slave processors when there are more than two stages is much less for this algorithm than for the nested Benders algorithm. The load balancing property observed in the nested Benders parallel code is maintained with each newly available processor receiving the latest information. The results reported here are for the SP2 implementation running under MPI. Large problems were again run on the SP2 on up to 8 slaves. Solution of two-stage problems does not gain from the addition of extra processors and, as EVPI is appropriate only for multistage problems, results are not shown here.

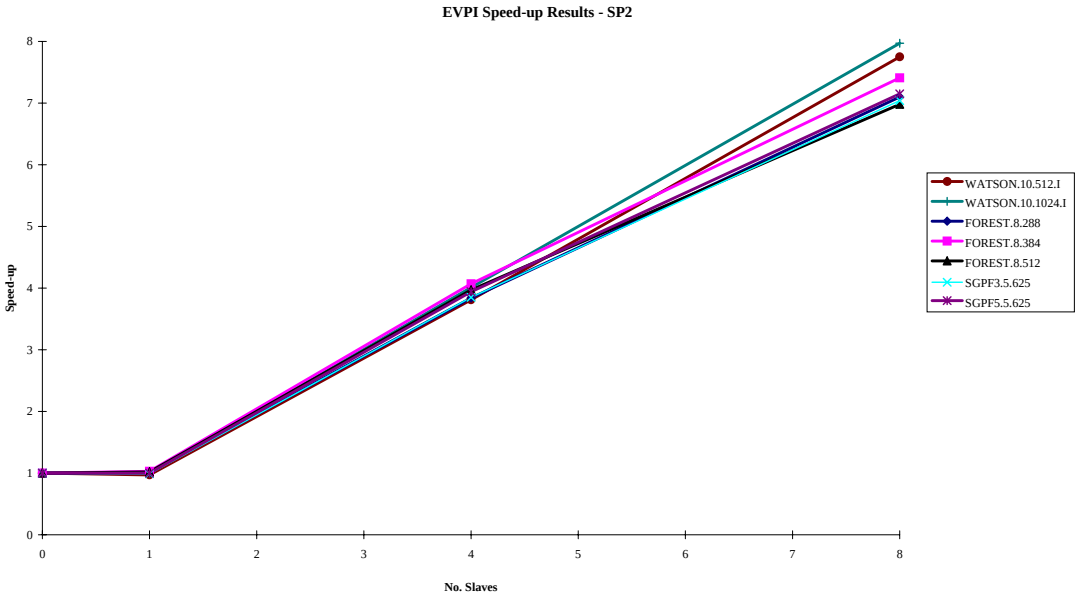


Figure 4: EVPI speed–up on the SP2

The algorithm appears to perform reasonably on three stage problems and is approximately 50% efficient on 8 slave processors. For problems of more than three stages its efficiency and speed–up are high (>90% efficiency) especially for the FOREST, WATSON.I and SGPF problems sets, see Figure 4. Table 2 shows the speed-up on 1 to 8 slaves for the problems WATSON.10.512.I and WATSON.10.1024.I. It can be seen that they both nearly achieve linear speed–up. Amdahl's effect can be seen in the efficiency increase for the larger problem.

On the AP1000 up to 35 processors were used to test the performance of the parallel EVPI implementation. FOREST.8.288 achieved a speed–up of 29 on 35 slaves, an efficiency of 83%. A large SGPF problem achieved speed–up of 18 on 24 slaves, an efficiency of 75%. The WATSON problem set required too much DRAM to be solved on the Fujitsu machine.

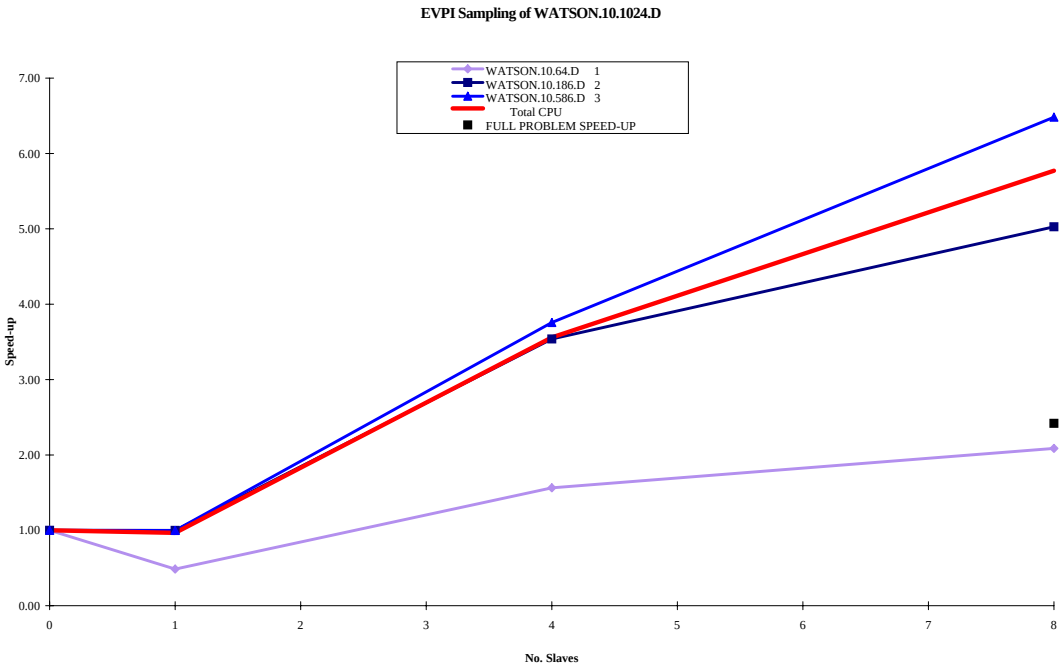


Figure 5: EVPI–S speed–up on the SP2

The efficiency of the EVPI calculation suggests the use of parallelisation for the EVPI–S sampling technique. The EVPI takes a significant proportion of the computational time for the multistage problems for which the sampling technique was designed. An example is shown in Figure 5 for the parallel MSLiP code. The three iterations of the parallel EVPI–S procedure have different parallel performances; as the size of the problem increases ($\mathcal{T}_n > \mathcal{T}_{n-1}$) the parallelisation improves. Since MSLiP–OSL performs the EVPI calculation much faster than MSLiP, it is to be preferred in a sequential sampling algorithm. The parallelisation of the MSLiP–OSL code would lead to lower speed–up for the WATSON.10.1024.I problem shown in Figure 5, since the solution time of the EVPI calculation would play a smaller part in the overall solution time. (Unfortunately this has not been performed as the OSL software has not been installed on either parallel machine.) However, as the solution of a whole set of problems is needed when sampling, which can require a large amount of CPU time, even lower speed–ups can still be extremely useful. The problem shown can be solved in parallel by EVPI–S on the

SP2 with 8 processors in $\sim$ 9.1 min versus solution in parallel of the full problem (with EVPI calculation) in $\sim$ 11.5 min. The full problem is solved serially in $\sim$ 75 min (parallel speed–up 6.5/8). For much larger problems the advantage of the parallel EVPI–S algorithm will be still further highlighted.

8 Conclusions

As nested Benders decomposition for linear multistage recourse stochastic programming problems compares favourably with both interior point and simplex methods and can be stabilised for difficult problems by the use of stage aggregation, it appears to be a sensible method to parallelise, especially as there is a sampling procedure which utilises this method. This paper has mainly be concerned with the parallelisation of nested Benders decomposition and the EVPI calculation within the MSLiP code. These algorithms are integrated into the EVPI–S sampling algorithm for solving linear multistage recourse problems with large, unknown, stochastic structures.

Along with most coarse grain parallel algorithms for solving multistage recourse problems this parallel nested Benders algorithm uses a farming strategy, as does the parallel EVPI algorithm. Each slave processor solves an LP problem, with the master processor directing operations. This type of algorithm has good load–balancing properties. Results on a variety of multicomputing platforms show that as with other parallel decomposition strategies, the performance of the algorithm is problem dependent. It is particularly dependent on the communication to calculation ratio of the particular platform.

The problem sets on which the parallel algorithm performs best – with near linear speed–up on up to 8 processors – had larger nodal LP problems to solve. The length of time spent communicating between processors, is critical to the performance of the algorithm, as is the tree structure and the percentage of the

total time spent forming cuts. Through the use of aggregation, a reduction in the time spent forming cuts and an increase in nodal problem dimensions can be achieved so that higher levels of speed–up can be achieved. The algorithm may be prone to both detrimental and acceleration anomalies due to different solution paths occurring in the parallel algorithm.

As the communication to calculation ratio is critical to performance, and the data needs to be replicated on each processor, a further improvement to the algorithm may be made by converting the algorithm to the multiprocessing environment. Shared memory machines will generally have large amounts of centralised RAM and communication will generally be much faster, improving both the parallel nested Benders and parallel EVPI algorithm.

Further methods which could improve the parallel algorithm are:–

1) A hybrid parallel code could be attempted which is a cross between the parallel algorithm presented here and the Birge *et al* [6] parallel algorithm, with a slave processor solving a node and its children, say, before returning a solution to the master processor. There may however be a problem with transferring cut information to the appropriate processor in this case.

2) A slave processor could solve a set of LP problems at the same stage before returning solutions to the master. This resembles the parallel EVPI calculation. The load–balancing property of the algorithm can be maintained by not allocating all the LPs to be solved to the slave processors initially.

4) A semi–distributed load balancer could be employed with more than one processor forming cuts. Depending on the node, a particular processor may form the cut and broadcast the information to other slave processors.

The penultimate section of this paper concentrates on the EVPI–S sampling algorithm. For a further reduction in solution time and to make the whole system viable in a practical setting, the parallelisation of the EVPI calculation was needed. This is a much simpler operation than that of the parallelisation of nested

Benders decomposition and has very small percentage of sequential calculations in the multistage case. Near linear speed–up was achieved on most problems with more than three stages. The parallel EVPI–S algorithm shows reasonable speed–up on up to 8 processors.

9 Acknowledgements

We are grateful to Professor Tony Hey and the Southampton High Performance Computing Consortium, and Fujitsu (UK) Limited and the Imperial College Fujitsu Parallel Computing Research Centre for access to and support for the parallel computing facilities used in the experiments reported here.

References

[1] G. M. Amdahl. Validity of the single processor approach to achieving large scale computing abilities. *AFIPS Conference Proceeding* **30** (1967) 483–485.

[2] K. A. Ariyawansa & D. D. Hudson. Performance of a benchmark parallel implementation of the Van-Slyke and Wets algorithm. *Concurrency: Practice and Experience* **3.2** (1991) 109-128.

[3] J. F. Benders. Partitioning procedures for solving mixed-variable programming problems. *Numerische Mathematik* **4** (1962) 238-252.

[4] J. R. Birge. Decomposition and partitioning methods for multistage stochastic linear programs. *Operations Research* **33** (1985) 989-1007.

[5] J. R. Birge, C. J. Donohue, D. F. Holmes & O. G. Svintsitski. ND-UM version 1.0 computer code for the nested decomposition algorithm. Department of Industrial and Operations Engineering, University of Michigan (August 1994).

[6] J. R. Birge, C. J. Donohue, D. F .Holmes & O. G. Svintsitski. Tech Report 94-1. A parallel implementation of the nested decomposition algorithm for multistage stochastic linear programs. Department of Industrial and Operations Engineering, University of Michigan (1994).

[7] G. Consigli & M. A. H. Dempster. Dynamic stochastic programming for asset–liability management. Submitted to: *Annals of Operations Research* (1996).

[8] G. Consigli & M. A. H. Dempster. Comparative evaluation of algorithms for dynamic stochastic programming. Submitted to: *European J. of OR* (1996).

[9] X. Corvera Poiré. Model Generation and Sampling Algorithms for Dynamic Stochastic Programming. PhD Thesis, Department of Mathematics, University of Essex (September 1995).

[10] X. Corvera Poiré & M. A. H. Dempster. Simplifying multistage stochastic programming problems. To appear in: *IMA Journal on Mathematics in Industry* (1995).

[11] P. D'Ambra & M. I. Sales. Concurrent banded Cholesky factorization on workstation networks. Working Paper, Department of Mathematics and Applications, University of Naples, and IBM ECSEC, Rome (1993).

[12] G. B. Dantzig & P. Wolfe. Decomposition principle for linear programs. *Operations Research* **8** (1960) 101-111.

[13] G. B. Dantzig, J. K. Ho & G. Infanger. Solving stochastic linear programs on a hypercube multicomputer, Technical Report SOL 91-10, Department of Operations Research, Stanford University (1991).

[14] M. A. H. Dempster (ed.). *Stochastic Programming.* Academic Press, London (1980).

[15] M. A. H. Dempster. On stochastic programming II: Dynamic problems under risk. *Stochastics* **25.1** (1988) 15-42.

[16] M. A. H. Dempster. & H.I. Gassmann. Using the expected value of perfect information to simplify the decision tree. *Abstracts* 15^{th} *IFIP Conference on System Modelling and Optimization.* IFIP, Zurich (1991) 301–303.

[17] M. A. H. Dempster & R. T. Thompson. Parallelization and aggregation of nested Benders decomposition. Submitted to *Annals of Operations Research* (1995).

[18] J. Dupačová. Multistage stochastic programs: The state-of-the-art and selected bibliography. *Kybernetika* **31.2** (1995) 151-174.

[19] J. Eckstein & R. S. Hiller. Stochastic dedication: Designing fixed income portfolios using massively parallel Benders decomposition. Working Paper 91-02, Harvard Business School (1991).

[20] R. Entriken. The parallel decomposition of linear programs. Technical Report SOL 89-17, Department of Operations Research, Stanford University (1989).

[21] Yu. Ermoliev & R. J-B. Wets (eds.). *Numerical Techniques for Stochastic Optimization.* Springer-Verlag, Berlin (1988).

[22] J. J. H. Forrest & J. A. Tomlin. Vector processing in simplex and interior methods. *Annals of Operations Research* **22** (1990) 71–100.

[23] H. I. Gassmann. MSLiP: A computer code for the multistage stochastic linear programming problem. *Mathematical Programming* **47** (1990) 407–423.

[24] H. I. Gassmann. MSLiP 8.2 user's guide. School of Business Administration, Dalhousie University (1992).

[25] M. T. Hajian, I. Hai & G. Mitra. A distributed processing algorithm for solving integer programs using a cluster of workstations. Research Report, Department of Mathematics, University of Brunel (September 1994).

[26] IBM Optimization Subroutine Library (OSL). Fourth Edition. IBM Corporation (1992).

[27] P. Kall & S. W. Wallace. *Stochastic Programming.* John Wiley and Sons (1994).

[28] A. King. SP/OSL Version 1.0. Stochastic Programming Interface User's Guide. IBM Research Division, Yorktown Heights, New York (1994).

[29] R. Levkovitz & G. Mitra. Solution of large–scale linear programs: a review of hardware, software and algorithmic issues. In: T. A. Ciriani & R. C. Leachman (eds.). *Optimization in Industry.* John Wiley & Sons Ltd (1993).

[30] S. S. Nielsen & S. A. Zenios. Solving multistage stochastic network programs. Report 92–08–04. Decision Sciences Department, The Wharton School, University of Pennsylvania, PA (August 1992).

[31] C. E. Pfefferkorn & T. A. Tomlin. Design of a linear programming system for ILLIAC —V. Technical Report SOL 76-8, Department of Operations Research, Stanford University and Technical Report 5487, NASA-Ames Institute for Advanced Computation, Sunnyvale, CA. (1976).

[32] R. T. Rockafellar & R. J-B. Wets. Scenarios and policy aggregation in optimization under uncertainty. *Mathematics of Operations Research* **16** (1991) 309-333.

[33] A. Ruszczyński. Parallel decomposition of multistage stochastic programming problems. Institute of Automatic Control, Warsaw University of Technology, 00665, Warsaw, Poland. August(1991).

[34] R. T. Thompson. MSLIP-OSL 8.3. User's Guide. Judge Institute of Management Studies, University of Cambridge (1996), forthcoming.

[35] R. T. Thompson. Fast Sequential and Parallel Methods for Solving Multistage Stochastic Linear Programmes. PhD Thesis, Department of Mathematics, University of Essex (1996), forthcoming.

[36] R. Van Slyke & R. J-B Wets. L-shaped linear programs with applications to optimal control and stochastic programming. *SIAM Journal of Applied Mathematics* **17** (1969) 638-663.

[37] R. Wittrock. Dual nested decomposition of staircase linear programs. *Mathematical Programming Study* **24** (1985) 65-86.

[38] D. Yang & S. A. Zenios. A scalable parallel interior point algorithm for stochastic linear programming and robust optimization. Report 95–07, Department of Public and Business Administration, University of Cyprus, Cyprus (February 1995).