Efficient Sub-optimal Temporal Decomposition with Dynamic Weighting of Speech Signals for Coding Applications

Slava Shechtman[1] *and David Malah*

Department of Electrical Engineering
Technion IIT, Haifa 32000, Israel
slava@il.ibm.com, malah@ee.technion.ac.il

Abstract

The Optimized Temporal Decomposition (OTD) technique for Line Spectral Frequencies (LSF) speech envelope representation, under a MMSE criterion, has been shown to be promising for very low bit rate speech coding for storage and broadcast applications. In order to improve perceptual speech quality, a dynamically weighted OTD (DW-OTD) technique is introduced in this work. It extends the OTD by allowing temporally changing weights, so as to improve the perceived speech quality. Use of Gardner's weighted MSE with DW-OTD is found to reduce the Log Spectral Distance (LSD) measure by 0.3 dB, as compared to OTD. The original OTD algorithm delay and complexity requirements make it inappropriate for real-time speech coding. In this paper we also introduce a modification of this technique, which is sub-optimal but suitable for on-line speech coding purposes, with negligible degradation of performance (of only about 0.06 dB in LSD). With the proposed techniques we were able to encode speech spectral envelopes at 300-370 bps at LSD of 2.25-2.1 dB, respectively, with a delay of just 7 frames.

1. Introduction

Fixed-frame-length Linear Predictive Coding (LPC) is a widely used method for spectral envelope representation at low bit rates. Usually, Line Spectral Frequencies (LSF) vectors (referred further as *parameter vectors*) are extracted at a constant rate of 40 – 50 Hz, and are coded using different vector quantization (VQ) techniques. Those techniques may achieve transparent quality of the spectral parameters (i.e. 1dB average LSD, less than 2% frames having spectral distortion greater then 2 dB and no outliers above 4dB) at a coding rate as low as 1000 bps [1].

In order to further reduce the coding rate the inter-frame redundancies in speech spectral parameters are exploited. There are many speech coding schemes based on the fixed-frame LPC spectral representation, which utilize those redundancies to gain an acceptable speech quality at reduced bit rates [3],[4],[6]. Variable or constant bit-rates using joint quantization schemes, such as Segment [4] or Matrix [3] Quantization, represent several LSF vectors by a single codeword. However, those methods require very large codebooks to obtain good speaker-independent performance. Other schemes reduce the bit rate by skipping a number of frames, followed by interpolating at the decoder, or combine skipping and joint quantization paradigms [5].

These schemes inevitably increase the delay of the system, but are still acceptable for a number of applications, such as real-time military/secure voice communications or non-real-time storage and broadcasting applications.

In this work we introduce an improved spectral envelope coding scheme, based on Optimized Temporal Decomposition (OTD) model [6], which is computationally efficient and capable of drastically reducing speech temporal redundancies, resulting in as low as 300 bps for coding the spectral envelope with an acceptable speech quality. Section 2 of this paper introduces the general TD concepts and describes the OTD algorithm, followed by the proposed scheme. Performance evaluation results are reported in section 3. The paper is summarized in section 4.

2. Temporal Decomposition (TD)

2.1. General TD model of speech

The Temporal Decomposition (TD) model of speech was originally introduced by Atal [6], as a useful technique for analyzing the temporal structure of speech. It then proved to be a promising technique for very-low-bit-rate speech coding [7],[9]. TD is a method of modeling a set of consequtive speech *parameter vectors* as a sequence of stable *event parameter vectors* (or *targets*) and an associated set of overlapping *interpolation functions* (*event functions*), centered at the corresponding *event instants*.

Let $\mathbf{Y}$ be the *parameter matrix* of a complete utterance, having consecutive *parameter vectors*, of length p each, as its columns. Then TD aims to approximate it as

$$\mathbf{Y} \cong \mathbf{A}\boldsymbol{\Phi}, \tag{1}$$

where $\mathbf{A}$ is a *target matrix*, consisting of *target* column vectors, and $\boldsymbol{\Phi}$ is an *event function matrix* (usually sparse), containing the event functions as its rows. Each event function is supposed to have a limited support.

TD is usually evaluated over a finite buffered *block* of N parameter vectors. In that case, a modeled parameter vector at a given instant is given by (2):

$$\hat{\mathbf{y}}(n) = \sum_{k=1}^{M} \mathbf{a}_k \phi_k(n), \qquad 1 \le n \le N, \tag{2}$$

where $\mathbf{a}_k$ is the k-th target vector, $\phi_k(n)$ is the value of the k-th event function at instant n, and $\hat{\mathbf{y}}(n)$ is a TD approximation of the input n-th *parameter vector* $\mathbf{y}(n)$. M is the expected number of events in the *block* of interest. Usually, TD is an iterative process, where each cycle consists of two major steps: event functions calculation (while preserving sparseness of $\boldsymbol{\Phi}$) and then *targets* refinement. Most proposed TD algorithms [6],[7],[8],[9] calculate the *targets* (for a given $\boldsymbol{\Phi}$), so that the total *block* squared error is minimized. This is a classic Least

[1] S. Shechtman is presently with IBM Haifa Research Labs.

Squares (LS) problem, and its solution is:

$$\mathbf{A}^T = (\mathbf{\Phi}\mathbf{\Phi}^T)^{-1}\mathbf{\Phi}\mathbf{Y}^T \quad (3)$$

Since $\mathbf{\Phi}\mathbf{\Phi}^T$ is a positive definite symmetric sparse matrix, the solution in (3) may be efficiently obtained. For example, for the case of a Restricted TD (discussed next), where at most 2 adjacent events may overlap, $\mathbf{\Phi}\mathbf{\Phi}^T$ is actually tri-diagonal, and the solution (3) may be simply found by performing simultaneous (from top and bottom) symmetric Gaussian elimination for each of the p sets of linear equations (one set for each column of $\mathbf{A}^T$) [10].

2.2. Restricted TD model of speech (RTD)

The term Restricted TD [7] (denoted RTD) refers to the overlapping property of event functions and affects the way the *event function* matrix is determined. It assumes that at most two adjacent event functions may overlap, replacing (2) by:

$$\hat{\mathbf{y}}(n) = \mathbf{a}_k\phi_k(n) + \mathbf{a}_{k+1}\phi_{k+1}(n), \;\; n_k \le n < n_{k+1} \quad (4)$$

where n_k and n_{k+1} are the locations of adjacent *event instants* (i.e. column numbers in the *parameter matrix*, associated with the *k*-th and the *(k*+1*)*-th *targets*). The *targets* $\mathbf{a}_k, \mathbf{a}_{k+1}$ are usually initialized by the original input *parameter vectors* at selected *event instances*, i.e.,

$$\mathbf{a}_k = \mathbf{y}(n_k), \quad (5)$$

or fed back from the previous target refinement stage, if the process is iterative.

With RTD, there exists a simple and exact analytic solution for the event functions that minimizes the squared error (SE), defined as $E(n) \triangleq \left\|\mathbf{y}(n) - \hat{\mathbf{y}}(n)\right\|^2$, given event instances and target vectors [8]. It may be readily extended to minimize the weighted squared-error (WSE), re-defined as:

$$E(n) \triangleq \sum_{i=1}^{p} w_i(n)(y_i(n) - \hat{y}_i(n))^2, \quad (6)$$

by,

$$\begin{pmatrix} \phi_k(n) \\ \phi_{k+1}(n) \end{pmatrix} = \begin{pmatrix} \mathbf{a}_k^T\mathbf{W}(n)\mathbf{a}_k & \mathbf{a}_k^T\mathbf{W}(n)\mathbf{a}_{k+1} \\ \mathbf{a}_k^T\mathbf{W}(n)\mathbf{a}_{k+1} & \mathbf{a}_{k+1}^T\mathbf{W}(n)\mathbf{a}_{k+1} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{a}_k^T\mathbf{W}(n)\mathbf{y}(n) \\ \mathbf{a}_{k+1}^T\mathbf{W}(n)\mathbf{y}(n) \end{pmatrix} \quad (7)$$

for $n_k \le n < n_{k+1}$, where $\mathbf{W}(n)$ is a diagonal matrix, containing dynamically varying error weights $\{w_i(n)\}_{i=1}^{p}$ on its main diagonal.

This solution may result in irregular event functions shapes, which could be difficult to quantize using vector quantization. It has been proposed [7],[8] to regularize these functions, by imposing constraints on the solution (7), such as the one's complementary property of each event function pair at each time instant *n* in addition to non-negativity [7], and monotonicity [8], of each event function throughout a *segment*. A segment is defined as the time interval between adjacent *event instants*. A solution that satisfies all of those constraints was proposed in [8]: (8)

$$\tilde{\phi}_k(n) = \begin{cases} 1 - \tilde{\phi}_{k-1}(n), & n_{k-1} \le n < n_k \\ 1, & n = n_k \\ \min\left(\tilde{\phi}_k(n-1), \min\left(1, \max\left(0, \bar{\phi}_k(n)\right)\right)\right) & n_k < n < n_{k+1} \\ 0, & else \end{cases},$$

where, $\bar{\phi}_k(n) = \dfrac{\left(\mathbf{y}(n) - \mathbf{a}_{k+1}\right)^T \mathbf{W}(n)\left(\mathbf{a}_k - \mathbf{a}_{k+1}\right)}{\left(\mathbf{a}_k - \mathbf{a}_{k+1}\right)^T \mathbf{W}(n)\left(\mathbf{a}_k - \mathbf{a}_{k+1}\right)}$.

2.3. Optimized TD (OTD)

Due to the simplicity of determining the instant model errors in RTD, it is possible to define a full search procedure to find the optimal *event-functions-matrix* in terms of the total SE of a block of parameter vectors, given a desired number of events, *M*, per *a* block of length *N* [8].

Let $E(n)$ be the optimal instant error, given an initial segmentation $\{n_k\}_{k=1}^{M}$ and target vectors - either according to (5) or some other explicit values. Define also the total block error for a given segmentation $\{n_0 \triangleq 0, n_1, \ldots n_M, n_{M+1} \triangleq N+1\}$ (events at the first and the last event instants are dummy zero events) as: $E_{block}(0, n_1, \ldots n_M, N+1) \triangleq \sum_{n=1}^{N} E(n)$. Then,

$$\begin{cases} D(n_k) = \min\limits_{n_{k-2} < n_{k-1} < n_k} \left(D(n_{k-1}) + E_{seg}(n_{k-1}, n_k)\right) \\ n_{k-1}^* = \arg\min\limits_{n_{k-2} < n_{k-1} < n_k} \left(D(n_{k-1}) + E_{seg}(n_{k-1}, n_k)\right) \end{cases}, \; k = 2, \ldots M$$

where,

$$E_{seg}(n_k, n_{k+1}) = \sum_{n=n_k}^{n_{k+1}-1} E(n) \text{ and } D(n_k) \triangleq \sum_{n=1}^{n_k-1} E(n) \quad (9)$$

This formulation can be implemented with the Viterbi algorithm, to find the optimal segmentation and event functions. In order to minimize large errors at block edges, a block overlapping technique is applied, so that the last event of the previous block is the first one of the current block [8]. This full search procedure is followed by target refinement stage resulting in (3), and the refined targets are fed back to the full search stage for the next iteration. Usually 3-5 iterations is enough to converge [8]. When the full search stage is entered for the first time, (5) is assumed, but in the next iterations, following each target-refinement step, the targets are kept during the search algorithm.

Let's estimate the number of instant error calculations for the full search algorithm. For the first run, each parameter vector (out of N) is contained in about $N^2/2$ segments with different surrounding targets, so there are about $N^3/2$ instant error calculations for a N-length block. For the next runs of the algorithm, there exist only about $M^2/2$ different error calculations for each parameter vector, so there are nearly $M^2N/2$ elementary error calculations for the full block.

2.4. Dynamically weighted Optimized TD (DW-OTD)

2.4.1. Motivation

It is known that the SE measure, applied to spectral parameters, may fail to describe perceptual distance between the original and modeled spectra. The most widely used criterion for perceptual model/quantization error evaluation is the Log Spectral Distance (LSD) measure, given by

$$E_{LSD}[dB] = \sqrt{\tfrac{1}{\pi}\int_{\theta_1}^{\theta_2} (10\log(S_n(e^{j\omega})) - 10\log(\hat{S}_n(e^{j\omega}))^2 d\omega},$$

where $S_n(e^{j\omega})$ and $\hat{S}_n(e^{j\omega})$ are the LPC power spectra corresponding to the original and modeled spectral parameter vectors at instant *n,* and $0 \le \theta_1 < \theta_2 \le \pi$ define the bandwidth of interest. It is hard to incorporate this criterion in the search procedure due to its complexity, but there exist a number of dynamically weighted squared error measures, applied directly

to LSF parameters, that approximate and track it better then SE, such as Paliwal-Atal WSE [1] (referred here as PA-WSE), or Gardner WSE [2] (referred here as G-WSE). A modified G-WSE measure, referred as G-WSE(2), is proposed below to further improve LSD tracking ability. It is defined by a fixed weighting of G-WSE to reduce high band importance, relative to the low band:

$$\begin{aligned}&\tilde{w}_i(n) = (c_i)^2 w_i(n),\\ &\mathbf{c} = [1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 0.9 \quad 0.8 \quad 0.7 \quad 0.1 \quad 0.01],\end{aligned} \tag{10}$$

where $w_i(n)$ is G-WSE weight for the *i*-th component at instant *n*. Such a drastic attenuation of high band components may be unacceptable for targets or parameter vectors quantization purposes, but is found to serve well the DW-OTD algorithm. Experimentally, it was found that it performs better then SE, PA-WSE and G-WSE, in the sense of LSD tracking and sensitivity to gross LSD errors, resulting, therefore, in the best performance when used as a TD minimization criterion.

2.4.2. OTD modification

The event-function-matrix determination step of OTD is readily modified to use Dynamic WSE, by using the instant-error criterion in (6) and event functions (7) or (8). However, the target refinement step should be reviewed. We assume that the overlapping technique is used, so that the zero target, which is placed at zero instant, is not modified by the refinement process. To find optimal targets, given the matrix $\mathbf{\Phi}$, the following total block-error has to be minimized (for each speech parameter, out of *p*):

$$E_{block}^{(i)} = \sum_k \sum_{n=n_k}^{n_{k+1}-1} w_i(n)(y_i(n) - a_{i,k}\phi_k(n) - a_{i,k+1}\phi_{k+1}(n))^2,\ 1 \le i \le p \tag{11}$$

The minimization with respect to $\{a_{i,k}\}_{k=1}^{M}$ results in the following symmetric tri-diagonal set of linear equations:

$$\begin{pmatrix} d_1 & x_1 & 0 & 0 \\ x_1 & \ddots & \ddots & 0 \\ 0 & \ddots & d_{M-1} & x_{M-1} \\ 0 & 0 & x_{M-1} & d_M \end{pmatrix} \begin{pmatrix} a_{i,1} \\ \vdots \\ a_{i,M-1} \\ a_{i,M} \end{pmatrix} = \begin{pmatrix} b_1 - x_0 a_{i,0} \\ \vdots \\ b_{M-1} \\ b_M \end{pmatrix}, \tag{12}$$

where $d_k = \sum_n \phi_k^2(n) w_i(n)$, $x_k = \sum_n \phi_k(n)\phi_{k+1}(n) w_i(n)$ and $b_k = \sum_n \phi_k(n) y_i(n) w_i(n)$. There exist *p* such sets, and each one can be solved as in (3).

2.5. Sub-optimal TD algorithm (SOTD)

2.5.1. Sub-optimal search

To reduce the computational burden of event-functions determination in OTD, we propose a sub-optimal algorithm. In [7] no search is done at all. Rather, the event centers are set in the block according to a local spectral stability criterion, called Spectral Feature Transition Rate (SFTR). However, this scheme results in 18-20 events/sec [7], which are too many for very low bit rate coding. So, we choose to refine the initial segmentation by a partial search scheme, as follows. Let $\mathbf{Y}_k$ be a sub-matrix of **Y,** consisting of the parameter vectors in the interval $[n_{k-1}, n_{k+1})$. Assuming that its boundary events are fixed, the best placement of the *k*-th event instant n_k may be found so that the total modeling error of the $\mathbf{Y}_k$'s is minimized:

$$\begin{cases} E_{\min}(\mathbf{Y}_k) = \min\limits_{n_k \in (n_{k-1}, n_{k+1})} \left(E_{seg}(n_{k-1}, n_k) + E_{seg}(n_k, n_{k+1}) \right) \\ n_k^* = \arg\min\limits_{n_k \in (n_{k-1}, n_{k+1})} \left(E_{seg}(n_{k-1}, n_k) + E_{seg}(n_k, n_{k+1}) \right) \end{cases}, k = 1, .., M \tag{13}$$

Applying this operation sequentially on all the events in the block, in increasing order, results in the desired suboptimal solution. It was found that the algorithm performs better, when the effective block end is chosen to be one of the last N/M parameter vectors, which minimizes the SFTR criterion [7], while all initial *event instants* are distributed uniformly.

The number of instant error calculations of this algorithm may be evaluated by noting that each sub-matrix $\mathbf{Y}_k$ has an average length of $2N/M$ frames. At the initial run of the algorithm, there are $(2N/M)^2 M = 4N^2/M$ instant error calculations for each pass over the block, compared to $N^3/2$ in OTD. For the next runs, only two instant errors have to be recalculated for each search step. The estimated number of error calculations for those runs is, therefore, $N + 2(2N/M)M = 5N$, compared to $M^2 N/2$ in OTD.

2.5.2. Constant event rate

The OTD algorithm assumes a constant block length, with variable overlap length [8]. This imposes that the system has a variable event rate, which may complicate the implementation. Our proposed algorithm forces a constant event rate, and therefore *constant bit rate*, by retrieving *N* new frames for each analysis block (not including the overlap with previous block). Of course, we need to limit the overlap range to prevent memory buffer overflow.

2.6. Quantization of TD parameters

The parameters to be quantized are the target vectors and the event functions. For the constrained event functions, given in (8), only the decreasing branch shape and its length have to be quantized. Shape quantization, that fits the very-low-bit-rate paradigm, is performed by a small codebook VQ of decreasing event functions branches. For each possible length a different codebook is stored, while the allowable distance between adjacent event instances is limited to 8 (3 bit length coding). The targets are quantized, as common, by the Split-VQ technique of the lower 4 and upper 6 LSF parameters [1], using the G-WSE criterion. Small event-function shape codebooks (2-4 bits) are trained with event functions created by the *constrained* Dynamically Weighted SOTD (DW-SOTD) algorithm. In order to reduce the quantization distortion, both targets and event functions shape quantization are incorporated in the TD process (i.e. search in shape codebooks is done instead of using the analytic solution for event functions (8), and quantized parameter vectors are used as a targets).

3. Experimental results

A speech dataset consisting of 20 (10 male and 10 female) phonetically diverse sentences from the TIMIT speech corpus was used as the test data set for TD performance evaluation. 10^{th} order LSF parameters were extracted, according to the

MELP-2400 federal standard (MIL-STD-3005), at a 44.44Hz frame rate.

LSD performance of the different algorithms examined is shown in the Table 1. These runs use the constant rate scheme, with N=15 and M=4,5,6 targets. The weights used in DW-OTD and DW-SOTD are G-WSE(2). It is concluded from the results that DW-OTD outperforms OTD by 0.3 dB (LSD), while the degradation caused by sub-optimality is as low as 0.05-0.07 dB.

Table 1: Spectral distortions, obtained for different TD models

Algorithm	Ev./ sec	LSD		
		Avg., dB	2-4dB,%	>4dB,%
OTD	11.85	1.77	31.48	4.02
DW-OTD		1.43	21.28	1.36
DW-SOTD		1.48	20.96	2.15
OTD	14.81	1.49	22.57	1.93
DW-OTD		1.20	13.33	0.64
DW-SOTD		1.26	14.81	1.08
OTD	17.78	1.25	16.13	0.79
DW-OTD		0.99	7.49	0.25
DW-SOTD		1.06	9.54	0.50

In Table 2, the performance of the proposed DW-SOTD algorithm is presented for different bit rates and delays. The proposed algorithm was embedded into the MELP vocoder and the resultant speech perceptual quality was evaluated by the PESQ (ITU P.862) standard software. The last entry in the table includes standard MELP results. The triplets in the 2nd column of the table indicate target Split-VQ codebooks' (low 4 LSFs, then high 6 LSFs) and event function shape VQ codebook's sizes in bits. Three additional bits for each event length are added in the rate calculation. It is observed, that the DW-SOTD algorithm reduced the bit rate for the spectral envelope parameters by about 70%, while causing a 1.1-1.25 dB increase of the average LSD and a reduction of about 0.2 in Mean Opinion Score estimation (PESQ). It is seen that the performance of the algorithm deteriorates when reducing the *block* size (N), but it is still acceptable for as low as 7 frame lengths (157.5 ms of delay).

Table 2: DW-SOTD with quantization evaluation

M/N	Codebook, Bits	Rate, Bps	LSD			PESQ
			Avg., dB	2-4dB, %	>4dB, %	
4/15	12,12,4	367	2.09	38.30	6.56	2.82
	12,12,2	343	2.16	38.98	7.70	2.81
	11,9,3	308	2.20	42.28	6.99	2.78
3/11	12,12,4	376	2.08	36.80	7.45	2.83
	11,9,4	327	2.17	39.23	7.20	2.79
	10,8,4	303	2.24	41.28	8.56	2.78
2/7	12,12,3	381	2.14	35.99	9.57	2.77
	11,9,4	343	2.19	38.03	8.96	2.76
	10,8,3	305	2.27	41.18	9.53	2.73
MELP-2400		1111	0.94	1.50	0.00	2.99

4. Summary

An extension for the OTD technique, called DW-OTD, was proposed. When applied to Line Spectral Frequencies (LSF) speech envelope representation, it reduced the average modeling error (LSD) by 0.3 dB, compared to OTD. In addition, a computationally efficient (and suitable for real-time applications) sub-optimal modification of OTD/DW-OTD denoted DW-SOTD, was presented, featured by negligible degradation of performance (of only about 0.06 dB in LSD), compared to the optimal algorithms (OTD/DW-OTD).

The DW-SOTD scheme, which exploits both modifications on OTD, was combined with quantization, resulting in an efficient constant rate algorithm for very low bit rate speech coding. It allows representation of the speech envelope by as low as 300 bps, still preserving acceptable speech quality and possesses moderate computational requirements, allowing its real-time implementation.

To assess the spectral envelope coding subjective quality, the DW-SOTD was incorporated into MELP-2400 standard codec (thus presenting a 1.6 Kbps codec). Its objective quality was estimated with the PESQ standard software and a degradation of 0.18 – 0.25 of MOS score was reported, as compared to MELP-2400 spectral quantization. With the low rate needed for coding the spectral envelope (300bps), the technique has potential for developing a MELP-based coder operating at 600bps.

5. References

[1] K. K. Paliwal, B. S. Atal, "Efficient Vector Quantization of LPC Parameters at 24 Bits/Frame", *IEEE Tran. On Speech and Audio Processing*, Vol. 1, 1993, pp. 3-14.

[2] W.R. Gardner and B.D. Rao, "Theoretical Analysis of the High Rate Vector Quantization of LPC Parameters", *IEEE Trans. Speech, Audio Processing*, Vol. 3, No. 5, pp. 367-381, 1995.

[3] C. Tsao, R.M. Gray, "Matrix Quantizer Design for LPC Speech Using the Generalized Lloyd Algorithm", *IEEE Trans. on ASSP,* Vol. 32, 1985, No. 3, pp. 537-545.

[4] M. Honda, Y. Shiraki, "Very low-bit-rate speech coding", in *Advances in Speech Signal Processing*, S.Furui and M.M. Sondhi, Eds. , Marcel Dekker, 1992, pp. 209-230

[5] R. Mayrench, D. Malah, "Low-bit-rate Speech Coding Using Quantization of Variable Length Segments", *Eurospeech-1999*.

[6] B. Atal, "Efficient coding of LPC parameters by temporal decomposition", *IEEE ICASSP-83*, Boston, pp 81-84.

[7] S.J. Kim, Y.H. Oh "Efficient Quantization Method for LSF Parameters Based on Restricted Temporal Decomposition", *Elec..Letters*, 1999, Vol. 35, pp 962-964

[8] P.C. Nguyen, M. Akagi, "Imrovement of the Restricted Temporal Decomposition Method for Line Spectral Frequency Parameters, Proc. ICASSP-2002. pp 265-268.

[9] C.N. Athaudage, A.B. Bradley, M. Lech, "Model-Based Speech Signal Coding Using Optimized Temporal Decomposition for Storage and Broadcasting Applications", *EURASIP Journal On Applied Signal Processing,* 2003, Oct., pp 1016-1026.

[10] *NAG Fortran Library Manual*, Mark 20, The Numerical Algorithm Group Limited, 2002