Consiglio Nazionale delle Ricerche

ISTITUTO DI ELABORAZIONE DELLA INFORMAZIONE

PISA

AREA-TME TRADEOFFS FOR BILINEAR FORMS COMPUTATIONS IN VLSI

B. Codenotti, G. Lotti

Nota interna B84-23
Ottobre 1984

AREA-TIME TRADEOFFS FOR BILINEAR FORMS COMPUTATIONS IN VLSI

B. CODENOTTI[1] and G. LOTTI[2]

[1]Istituto di Elaborazione dell'Informazione - CNR, via S.Maria 46, PISA, ITALY.

[2]Dipartimento di Scienze dell'Informazione - University of Pisa, Corso Italia 40, PISA, ITALY.

ABSTRACT

An important class of problems in the complexity field is that concerning the computation of bilinear forms. In this work a criterion is given which allows obtaining lower bounds to AT^2 for the bilinear forms computational problem.

Key Words - VLSI Model, Area-Time Complexity, Bilinear Forms, Lower Bound.

1. INTRODUCTION

The area-time tradeoff has been separately provided for several computational problems [3]. The first attempt to obtain lower bounds to AT^2 for different problems by using the same tecnique was made by Vuillmein [5] who considered the class of problems performing the computation of some transitive functions.

Another important class of problems in the complexity field is that concerning the computation of bilinear forms. In this work a criterion is given which allows obtaining lower bounds to AT^2 for the bilinear forms computational problem.

We show also that the product of a Circulant matrix by a vector and of a triangular Toeplitz matrix by a vector are not easier than the usual matrix-vector product.

2. BILINEAR FORMS COMPUTATIONS

Let us consider the set of bilinear forms

$$f_k = x^T A_k y, \qquad k=1,2,\ldots,p,$$

where the mxn matrices A_k are the 3-sections of tensor A.

Let B_i and C_j be respectively the 1-sections and the 2-sections of A. The three matrices A(z), H(x) and C(y) can be defined

$$A(z) = \sum_{k=1}^{p} z_k A_k, \quad H(x) = \sum_{i=1}^{m} x_i B_i^T, \quad C(y) = \sum_{j=1}^{n} y_j C_j .$$

Let H and C be the tensors whose 3-sections are respectively B_i^T and C_j.

The computation of f_k, $k=1,2,\ldots,p$, can be considered as the computation of the matrix-vector product

$$f = B^T(x)y = H(x)y, \qquad (2.1) \text{ or}$$

$$f = C(y)x \qquad (2.2),$$

where the entries of the matrix are variables and the structure of the matrix is fixed.

Let us consider now the computation of (2.1); a bisection of the outputs induces a partition of the inputs, namely

$$|f^1| = |f^2| = p/2, \ |y^1|+|y^2| = n, \ |x^1|+|x^2| = m,$$

where

$$f^1 = H_{11}(x^1)y^1 + H_{11}(x^2)y^1 + H_{12}(x^1)y^2 + H_{12}(x^2)y^2$$

$$f^2 = H_{21}(x^1)y^1 + H_{21}(x^2)y^1 + H_{22}(x^1)y^2 + H_{22}(x^2)y^2.$$

Then tensor H is splitted into the eight tensors H_{rst}, $r,s,t=1,2$ each corresponding to the matrix $H_{rs}(x^t)$.

The information that must be transmitted over the partitioning wires is captured in the terms

$$H_{12}(x^2)y^2, \ H_{21}(x^1)y^1,$$

$$H_{12}(x^1)\,y^2\,,\quad H_{21}(x^2)\,y^1\,,$$

$$H_{11}(x^2)\,y^1\,,\quad H_{22}(x^1)\,y^2\,.$$

By analyzing the term $H_{12}(x^2)\,y^2$, it follows that if $\max_{x^2} \operatorname{rk} H_{12}(x^2) = t$ then at least t words of information must be transferred from part 2 to part 1 of the circuit [4].

Moreover $\dim H_{3\,112}$ words are needed to perform $H_{11}(x^2)y^1$ and $\dim H_{2\,121}$ words are required to perform $H_{12}(x^1)y^2$.

Then we get

$$I \geq \max_{x^2} \operatorname{rk} H_{12}(x^2) + \max_{x^1} \operatorname{rk} H_{21}(x^1) +$$

$$+ \dim H_{2\,121} + \dim H_{2\,212} + \dim H_{3\,112} + \dim H_{3\,221}\,.$$

Similar considerations for the computation of (2.2) give

$$I \geq \max_{y^2} \operatorname{rk} C_{12}(y^2) + \max_{y^1} \operatorname{rk} C_{21}(y^1) +$$

$$+ \dim C_{2\,121} + \dim C_{2\,212} + \dim C_{3\,112} + \dim C_{3\,221}\,.$$

It is easy to prove the following

LEMMA 2.1

$$\dim_2 A \geq \max_z \text{rk } A(z).$$

Proof.

The proof is trivial in the case $\dim_2 A=n$.

If $\dim_2 A=h<n$ then however we consider $s\geq h+1$ 2-sections of A, say $C_{j_1}, C_{j_2}, \ldots, C_{j_s}$, they are linearly dependent.

This means that it is always possible to find a nonzero s-vector t such that

$$\sum_{l=1}^{s} \left(\sum_{k=1}^{p} z_k a_{k i j_l}\right) t_l = 0, \text{ for any } z \in \mathbb{R}^p \qquad (2.3)$$

Now if $w=\max_z \text{rk } A(z) \geq h+1$ then it would be possible to choose w columns of $A(z)$, say $A'(z)$ such that

$$\max_z \text{rk } A'(z)=w.$$

This would contraddict (2.3). ∎

Let us denote with r and s the quantities

$$r = \dim_2 H_{121} + \dim_2 H_{212} = \dim_3 C_{112} + \dim_3 C_{221},$$

$$s = \dim_3 H_{112} + \dim_3 H_{221} = \dim_2 C_{121} + \dim_2 C_{212}.$$

From Lemma 2.1 we get

$$r \geq \text{rk } H_{12}(u^1) + \text{rk } H_{21}(u^2), \text{ for any } u^1, u^2,$$

$$s \geq \mathrm{rk}\, C_{12}(v^1) + \mathrm{rk}\, C_{21}(v^2), \text{ for any } v^1, v^2.$$

Therefore we have

$$(2.5) \quad I \geq \max_x \mathrm{rk}\, H_{12}(x) + \max_x \mathrm{rk}\, H_{21}(x) + \dim H_{3\,112} + \dim H_{3\,221}$$

$$(2.6) \quad I \geq \max_x \mathrm{rk}\, C_{12}(x) + \max_x \mathrm{rk}\, C_{21}(x) + \dim C_{3\,112} + \dim C_{3\,221}.$$

From relations (2.5), (2.6) it is possible to derive simpler bounds to the minimal information flow, namely

$$(2.7) \qquad I \geq \max_x \mathrm{rk}\, H_{12}(x) + \max_x \mathrm{rk}\, H_{21}(x),$$

$$(2.8) \qquad I \geq \max_x \mathrm{rk}\, C_{12}(x) + \max_x \mathrm{rk}\, C_{21}(x),$$

which are straightforward extension of Thompson criterion [4]. Anyway, in some cases formulas (2.7), (2.8) do not produce significant lower bounds. For example, in the case of the outer product of two n-vectors, it is easy to see that the well known lower bound $AT^2 = \Omega(n^2)$ [3] can be obtained from (2.5) or (2.6) and not from (2.7) or (2.8).

3. TRANSFORMATIONS CONCERNING CIRCULANT AND TOEPLITZ MATRICES

For some special cases of Bilinear Forms, a more immediate approach allows obtaining significant lower bounds, without needing the knowledge of the ranks of certain minors involved in the problem, which can be hard to determine. In the following, we consider the product between two Circulant matrices and the product of a Circulant matrix by a vector.

An $\Omega(n^2)$ lower bound to AT^2 for the product of two nxn Circulant matrices can be proved, by using some results of the previous section. In fact, the bilinear forms associated to this problem are equivalent to the bilinear forms associated to the product of a Circulant matrix by a vector, and in both cases the lower bound to AT^2 is the same.

Moreover from [5] it follows that the lower bound $AT^2=\Omega(n^2)$ holds for the problem of multiplying a Circulant matrix by a vector, since any circuit for this problem has to perform any permutation of the input vector.

Note that Circulant matrices are special cases of Toeplitz matrices; therefore the product of a Toeplitz matrix by a vector and the product between two Toeplitz matrices require $AT^2=\Omega(n^2)$.

For the latter problem an almost optimal circuit can be obtain by using the well known circuits for convolution [1] to perform the product between a lower and an upper triangular Toeplitz matrix. Indeed, a Toeplitz matrix T can be splitted as the sum of a lower triangular Toeplitz matrix L and an upper triangular Toeplitz matrix U, so that 2 convolutions and 2 products between lower (upper) triangular Toeplitz matrices suffice to solve the problem.

4. TRIANGULAR MATRIX BY VECTOR PRODUCT

An $\Omega(n^2)$ lower bound to AT^2 for the product of a 2nx2n lower triangular matrix A by a 2n-vector b can be easily proved by the following argument.

Let $A = \begin{bmatrix} A_1 & 0 \\ A_3 & A_2 \end{bmatrix}$, where A_1 and A_2 are nxn lower triangular matrices and let $b^T = [\ b_1^T \ \ b_2^T\ ]$, where b_1 and b_2 are n-vectors.

Since any VLSI network performing the product Ab has also to perform the product $A_3 b_1$, where A_3 is a full nxn matrix, the result follows from the well known lower bound for matrix vector product [3].

An important special case of triangular matrices is given by triangular Toeplitz matrices, which are of great interest in many applicative fields.

For the product of a triangular Toeplitz matrix by a vector, in the sequential model, there exist several "ad hoc" algorithms of complexity lower than the complexity required by the generic matrix-vector product.

In the VLSI model, the following arguments show how the lower bound to AT^2 for this special linear transformation is the same of the one for generic matrix vector product. Indeed, let L be a 2nx2n triangular Toeplitz matrix, partitioned as

$$L = \begin{bmatrix} L_1 & 0 \\ T & L_2 \end{bmatrix}$$

and let b a 2n-vector partitioned as $b^T = [\ b_1^T \ \ b_2^T\ ]$, where

b_1 and b_2 are n-vectors.

It is trivial to see that any VLSI circuit performing the matrix vector product Lb, has also to perform the computation Lb, when $b_2 = 0$. This means that the circuit has to compute:

$$(3.1) \qquad \begin{bmatrix} L_1 \\ \hline T \end{bmatrix} b_1 .$$

This result leads to the lower bound $\Omega(n^2)$, since the outputs of the circuit performing (3.1) contain the n-vector Tb_1, whose computation requires $AT^2 = \Omega(n^2)$.

5. CONCLUSIONS AND OPEN PROBLEMS

In this paper, we have presented a criterion to give lower bounds to the Bilinear Forms computational problem; this approach is related to the minimal "information flow" [4], which seems to be not suitable to obtain not trivial lower bounds, in some particular cases.

For example, it is easy to see that the product of a generic nxn matrix by an nxn Circulant matrix can be performed in VLSI with $AT^2 = O(n^3\log^2 n)$ by using $O(n)$ FFT modules to perform the products of the Circulant by all the columns of the generic matrix. Indeed any Circulant matrix can be factorized as $F D F^H$, where F is the Fourier matrix and D is a diagonal matrix whose entries are the elements of the vector obtained by applying the DFT to the first row of the Circulant matrix [2].

It is possible to show a lower bound $\Omega(n^2)$, for this problem, either by using the criterion of section 1 or by using Savage's technique [3], while an $AT^2=\Omega(n^3)$ lower bound is supposed to hold.

An analogous "gap" between lower and upper bound exists for the product of a Toeplitz matrix by a generic matrix.

It is an open question whether or not the lower bounds for the above mentioned problems can be improved.

REFERENCES

[1] G.M.BAUDET, F.P.PREPARATA and J.E.VUILLEMIN, Area Time Optimal VLSI Circuits for Convolution, IEEE Trans. Comput., 7, (1980) pp.684-688.

[2] P.DAVIS, Circulant Matrices. Wiley Interscience, New Jork (1979).

[3] J.E.SAVAGE, Area-Time Tradeoffs for Matrix Multiplication and related problems in VLSI models, J. Comput. and System Sci., vol. 22 ,pp. 230-242, 1981.

[4] C.D. THOMPSON, Area-Time Complexity for VLSI. Proc. 11th Annual ACM Symp. on the Theory of Computing (SIGACT) (1979) 81-88.

[5] J.E.VUILLEMIN, A Combinatorial Limit to the Computing Power of VLSI Circuits, IEEE Trans. Comput. 3 (1983) 294-300.