Improved Construction for Universality of Determinant and Permanent

Hong Liu, Kenneth W. Regan *

University at Buffalo

Abstract

Valiant (1979) proved that every polynomial of formula size e is a projection of the $(e+2) \times (e+2)$ determinant polynomial. We improve "$e+2$" to "$e+1$", also for a definition of formula size that does not count multiplications by constants as gates. Our proof imitates the "$2e+2$" proof of von zur Gathen (1987), but uses different invariants and a tighter set of base cases.

Key words: Computational complexity, algebraic formula size, determinant, permanent.

1 Introduction

Valiant (1979) proved that the determinant polynomials are "universal" for polynomial-sized arithmetical formulas in the following sense: given a formula $\phi(x_1, \dots, x_n)$ with e gates, he constructed an $(e+2) \times (e+2)$ matrix A_ϕ with entries in $\{x_1, \dots, x_n\}$ and the underlying field F, such that $\phi = \det(A_\phi)$. The same is true of the permanent polynomials. The universality of the determinant is a key element of Mulmuley and Sohoni (2001) for proving lower bounds via algebraic geometry. Universality of the permanent is a key step in its VNP-completeness result by Valiant (1979). Hence, it is interesting to optimize this fundamental construction as far as possible. We prove:

Theorem 1.1 *For any arithmetical formula ϕ of size e with at least one $+$ gate, we can build an $(e+1) \times (e+1)$ matrix A_ϕ such that $\phi = \det(A_\phi)$. If ϕ is a formal monomial $c_0 \cdot x_1 \dots x_n$, then we get $\phi = c_0 \cdot \det(A_\phi)$.*

* Corresponding author. Address: Department of CSE, University at Buffalo, 201 Bell Hall, Buffalo, NY 14260-2000; regan@cse.buffalo.edu

Besides moving $e+2$ to $e+1$, there is a second improvement; whereas the formula size measure in Valiant (1979) counts multiplications by constant as gates, ours does not. For example, $2wx+3yz$ has size 5 in Valiant (1979) and is taken to a 7×7 matrix. Here it has formula size 3 and is taken to a 4×4 matrix.

2 Definitions

We expand the formula definition in Valiant (1979) by adding coefficients to the wires in circuits.

Definition 2.1 *A formula ϕ over F is an expression that has one of the following forms:*

(1) a constant $c \in F$ or a variable x_i; or
(2) $c{\cdot}\phi$ where c is a constant and ϕ is a formula; or
(3) $\phi_1 \diamond \phi_2$ where ϕ_1 and ϕ_2 are formulas over F, and $\diamond \in \{+, \cdot\}$.

By applying 2 and 3, we can create $c_1\phi_1 \diamond c_2\phi_2$, where c_1 and c_2 are regarded as constants on the two wires leading into the $\diamond$ gate.

Definition 2.2 *We define formula size inductively as follows,*

(1) If ϕ is a constant c or variable x_i, then $size(\phi) = 0$;
(2) If $\phi = c{\cdot}\phi'$ for constant c and formula ϕ, then $size(\phi) = size(\phi')$;
(3) If $\phi = \phi_1 \diamond \phi_2$ for formulas ϕ_1, ϕ_2, then $size(\phi) = size(\phi_1) + size(\phi_2) + 1$.

Thus, if $size(\phi_1) = e_1$ and $size(\phi_2) = e_2$, then $size(c_1\phi_1 + c_2\phi_2) = e_1 + e_2 + 1$. In Valiant (1979) and von zur Gathen (1987), however, the size would be $e_1 + e_2 + 3$. Note that our size measure is one less than the numbers of leaves in the formula tree of ϕ, ie., the number of occurrences of variables and additive constants.

3 Proof of Main Theorem

Theorem 1.1 follows from the following statement, in which we carry a multiplicative constant c_0 outside in all cases, and preserve additional invariants.

Theorem 3.1 *For every formula ϕ of size e, there is an $(e+1)\times(e+1)$ matrix A_ϕ and a constant c_0 such that $\phi \equiv c_0 \det(A_\phi)$, and there exist $A \in \{X \cup F\}^{e \times e}$, $\alpha \in \{X \cup F\}^{1 \times e}$, $\beta \in \{X \cup F\}^{e \times 1}$ such that A is upper triangular with -1 on* *the main diagonal and A_ϕ is given by*

$$A_\phi = \left[\begin{array}{ccc|c} & \alpha & & a \\ \hline -1 & & A & \\ & \ddots & & \beta \\ \mathit{0} & & -1 & \end{array}\right]. \tag{1}$$

The following properties are inductively maintained:

(1) Every entry on the main sub-diagonal is -1.
(2) All entries below the main sub-diagonal are 0.
(3) Whenever ϕ has a $*$ gate, the entry a in the upper right corner is 0.
(4) Whenever ϕ has a $+$ gate, the last row has no variables.
(5) Each column has at most one variable.

Proof. Theorem 1.1 now follows from property 4 since we can multiply the last row by c_0. Our basis comprises all formulas that have no $*$ gates, and all formulas that have no $+$ gates:

(1) For a linear formula $\phi \equiv \sum_{i=1}^{n} a_i x_i$ (the affine linear case is similar with 1 in place of x_n),

$$\phi = a_n \cdot \left|\begin{array}{cccc|c} x_1 & x_2 & \dots & x_{n-1} & x_n \\ \hline -1 & & & 0 & a_1/a_n \\ & -1 & & & a_2/a_n \\ & & \ddots & & \dots \\ 0 & & & -1 & a_{n-1}/a_n \end{array}\right|. \tag{2}$$

(2) For a monomial $\phi \equiv c_0 \prod_{i=1}^{n} x_i$,

$$\phi = c_0 \cdot \left|\begin{array}{cccc} x_1 & & & 0 \\ -1 & x_2 & & \\ & \ddots & \ddots & \\ 0 & & -1 & x_n \end{array}\right|. \tag{3}$$

These formulas have size $n-1$, and the matrices are $n \times n$ as required, with conditions 1–5 holding. To prove equation (2), we need only expand the determinant by the first row; then the right-hand side of (2) equals

$$
\begin{aligned}
& a_n\cdot(-1)^{1+n}x_n\cdot(-1)^{n-1}+ \\
& +\sum_{j=1}^{n-1} a_n\cdot(-1)^{1+j}x_j \left|\begin{array}{cc|cc|c}
-1 & 0 & & & a_1/a_n \\
 & \ddots & & 0 & \dots \\
0 & -1 & & & a_{j-1}/a_n \\ \hline
0\dots 0 & & 0\dots 0 & & a_j/a_n \\ \hline
 & & -1 & 0 & a_{j+1}/a_n \\
 & 0 & & \ddots & \dots \\
 & & 0 & -1 & a_{n-1}/a_n
\end{array}\right| \\
= & \, a_nx_n+\sum_{j=1}^{n-1} a_n\cdot(-1)^{1+j}x_j\cdot(-1)^{j+n-1}(a_j/a_n)\cdot(-1)^{n-2} \\
= & \sum_{j=1}^{n} a_jx_j=\phi.
\end{aligned}
$$

For monomial ϕ, equation (3) is clear. Note that the coefficient cannot be brought inside any $n\times n$ matrix, so the exceptional constant c_0 in Theorem 1.1 cannot be avoided. For example, $c_0=c_0\cdot|1|$, $c_0x=c_0\cdot|x|$, and

$$
a_1x+a_2=a_1\cdot\begin{vmatrix} x & 1 \\ -1 & a_1/a_2 \end{vmatrix},\quad a_1x_1+a_2x_2=a_2\cdot\begin{vmatrix} x_1 & x_2 \\ -1 & a_1/a_2 \end{vmatrix},\quad c_0x_1x_2=c_0\cdot\begin{vmatrix} x_1 & 0 \\ -1 & x_2 \end{vmatrix}.
$$

For the induction, note that if Theorem 3.1 holds for a formula ϕ, then for any constant c, it holds for $c\cdot\phi$. So given ϕ of size e, assume the induction hypothesis (IH) that, for any ϕ' of size $e'<e$, there exists an $s'\times s'$ matrix A'_ϕ such that for all clauses of Theorem 3.1 hold for ϕ' and A'_ϕ. There are two top-level cases:

(1) Case 1: $\phi=\phi_1\cdot\phi_2$. From IH, we get $\phi_1=c_1\cdot\det(A_{\phi_1})$ and $\phi_2=c_2\cdot\det(A_{\phi_2})$. Similar to von zur Gathen (1987), we build the matrix A_ϕ by

$$
A_\phi=\left[\begin{array}{c|c}
A_{\phi_1} & 0 \\ \hline
\begin{matrix} & -1 \\ & \\ 0 & \end{matrix} & A_{\phi_2}
\end{array}\right].
$$

Then

$$\phi = \phi_1 \cdot \phi_2 = c_1 \det(A_{\phi_1}) \cdot c_2 \det(A_{\phi_2}) = c_1 \cdot c_2 \cdot \det(A_\phi).$$

The size s equals $s_1 + s_2 = e_1 + 1 + e_2 + 1 = e + 1$. If ϕ has a $+$ gate, we may wlog. suppose ϕ_2 has a $+$ gate, so that last row of A_{ϕ_2} and hence the last row of A_ϕ has no variables. Then the properties 1–5 of Theorem 3.1 are clear.

(2) Case 2: $\phi = \phi_1 + \phi_2$. We will consider the following subcases,

(a) If ϕ_1 and ϕ_2 both have $*$ gates, we modify von zur Gathen (1987)'s method as follows. According to IH, A_{ϕ_k} is as shown in equation (4) for $k = 1, 2$.

$$A_{\phi_k} = c_k \cdot \left[\begin{array}{ccc|c} \multicolumn{3}{c|}{\alpha_k} & 0 \\ \hline -1 & & A_k & \\ & \ddots & & \beta_k \\ 0 & & -1 & \end{array}\right]. \tag{4}$$

We build the $s \times s$ matrix A_ϕ by

$$A_\phi = \left[\begin{array}{ccc|ccc|c|c} \multicolumn{3}{c|}{\alpha_1} & \multicolumn{3}{c|}{\alpha_2} & 0 & 0 \\ \hline -1 & & A_1 & & & & & \\ & \ddots & & & 0 & & \beta_1 & 0 \\ 0 & & -1 & & & & & \\ \hline & & & -1 & & A_2 & & \\ & 0 & & & \ddots & & 0 & \beta_2 \\ & & & 0 & & -1 & & \\ \hline \multicolumn{3}{c|}{0} & \multicolumn{3}{c|}{0} & -c_2 & c_1 \end{array}\right]. \tag{5}$$

Besides having $-1s$ not $+1s$ on the main subdiagonal, the chief difference from von zur Gathen (1987) is staggering β_1 and β_2 rather than having one atop of the other. We show that this staggering introduces no unwanted nonzero diagonal products, and that the signs of all products come out right. Now we prove that $\phi = \det(A_\phi)$. In (5), we expand the determinant by the last row, getting

$$\det(A_\phi) = c_1 \cdot \det(R_1) + (-1)(-c_2) \cdot \det(R_2),$$

where

$$R_1 = \left[\begin{array}{c|c|c}
\alpha_1 & \alpha_2 & 0 \\ \hline
\begin{matrix} -1 & & A_1 \\ & \ddots & \\ 0 & & -1 \end{matrix} & 0 & \beta_1 \\ \hline
0 & \begin{matrix} -1 & & A_2 \\ & \ddots & \\ 0 & & -1 \end{matrix} & 0
\end{array}\right] \tag{6}$$

and

$$R_2 = \left[\begin{array}{c|c|c}
\alpha_1 & \alpha_2 & 0 \\ \hline
\begin{matrix} -1 & & A_1 \\ & \ddots & \\ 0 & & -1 \end{matrix} & 0 & 0 \\ \hline
0 & \begin{matrix} -1 & & A_2 \\ & \ddots & \\ 0 & & -1 \end{matrix} & \beta_2
\end{array}\right]. \tag{7}$$

To prove $\det(R_1) = \det(A_{\phi_1})$, we just flip the last column with the second last column, then flip the second last with the third last, so on and so forth. After $s_2 - 1$ flips, we get a new matrix R_1 as follows,

$$R_1' = \left[\begin{array}{c|c|c}
\alpha_1 & 0 & \alpha_2 \\ \hline
\begin{matrix} -1 & & A_1 \\ & \ddots & \\ 0 & & -1 \end{matrix} & \beta_1 & 0 \\ \hline
0 & 0 & \begin{matrix} -1 & & A_2 \\ & \ddots & \\ 0 & & -1 \end{matrix}
\end{array}\right]. \tag{8}$$

Therefore

$$\det(R_1) = (-1)^{s_2-1}\det(R_1') = (-1)^{s_2-1}(-1)^{s_2-1}\det(A_{\phi_1}) = \det(A_{\phi_1}).$$

Similarly, we can prove that $\det(R_2) = \det(A_{\phi_2})$, and so

$$\det(A_\phi) = \det(A_{\phi_1}) + \det(A_{\phi_2}).$$

At last, if we take outside the constant c_2 from the last row of A_ϕ, we prove that properties 1–5 of Theorem 3.1 hold in this case.

(b) If one of ϕ_1 and ϕ_2 does not have a $*$ gate, wlog. suppose ϕ_2 does not have one, then ϕ_2 is an affine formula, and we can express as $\phi_2 = \phi_2' + aw$ (where aw is an atom). Then

$$\phi = \phi_1 + \phi_2 = \phi_1 + \phi_2' + aw.$$

So if we set $\phi' = \phi_1 + \phi_2'$, then $\phi = \phi' + aw$. From IH, there exists a matrix $A_{\phi'}$ with submatrix A' as in Theorem 3.1, giving

$$\phi' = c_0 \cdot \det \left[\begin{array}{ccc|c} & \alpha & & 0 \\ \hline -1 & & A' & \\ & \ddots & & \beta \\ 0 & & -1 & \end{array}\right]. \tag{9}$$

Thus

$$\phi = a \cdot \det \left[\begin{array}{ccc|c|c} & \alpha & & 0 & w \\ \hline -1 & & A' & & \\ & \ddots & & \beta & 0 \\ 0 & & -1 & & \\ \hline & 0 & & -1 & c_0/a \end{array}\right]. \tag{10}$$

In equation (10), we flip the last two columns, and take out constant $-c_0/a$ from the last row, and we get:

$$\phi = c_0 \cdot \det \left[\begin{array}{ccc|c|c} & \alpha & & w & 0 \\ \hline -1 & & A' & & \\ & \ddots & & 0 & \beta \\ 0 & & -1 & & \\ \hline & 0 & & -1 & a/c_0 \end{array}\right]. \tag{11}$$

From (11), we have established properties 1–5 of Theorem 3.1. As remarked above, this also finishes the proof of Theorem 1.1. □

For the universality of the permanent polynomials, we have the following corollary.

Corollary 3.2 *For any arithmetical formula ϕ of size e with at least one $+$ gate, we can build an $(e+1) \times (e+1)$ matrix A_ϕ such that $s = e+1$ and $\phi = \mathrm{per}(A_\phi)$. If ϕ is a formal monomial $a \cdot x_i \ldots x_n$, then we get $\phi = a \cdot \mathrm{per}(A_\phi)$.*

Proof. The only change to the proof of Theorem 1.1 is that property *(1)* now reads, "*Every entry on the main sub-diagonal is* +1"—instead of −1 as for the determinant. This works because the proof Theorem 3.1 does not rely on cancellations—the −1s are solely to make odd permutations contribute terms with the correct sign. □

4 Examples and Conclusions

According to the method in von zur Gathen (1987), the matrix size s_e for circuit size e obeys the recursive relation

$$s_e = s_i + s_j, \quad \text{for any} e = i + j + 1.$$

von zur Gathen (1987) takes $s_0 = 2$ as basis, and hence gets $s_e = e + 2$, but our more-extensive treatment of base-case formulas allows $s_0 = 1$ and gives us $s_e = e + 1$. For example, Valiant (1979) gets a 3 × 3 matrix for the formula "$5x$", while we get a 1 × 1 matrix with outside constant 5.

Clearly, "e+1" is optimal, if one requires each column to have at most one variable. Without this requirement, can one do better with a still more extensive set of base cases? The answer is no in general, because when $\phi = x_1 \cdot x_2 \cdot \cdots \cdot x_n$ or $\phi = x_1 + x_2 + \cdots + x_n$, then $e + 1$ is absolutely optimal. However, for $\phi = 2wx + 3yz$, we can get the 3 × 3 matrix

$$\begin{bmatrix} 2x & 3y & 0 \\ -1 & 0 & w \\ 0 & -1 & z \end{bmatrix}.$$

This "cheats" a little by allowing "atoms" such as $2x$ and $3y$ as individual entries, but significantly, it has fewer columns than variables. Note that some formulas allow colossal savings in row/column size, such as the standard formulas for the permanent and determinant themselves. Finding general cases allowing $s \times s$ matrices with $s = o(e)$ is a step for further research. Two such cases are suggestive: For a family of formulas $\phi_n = x_1 y_n + x_2 y_{n-1} + x_3 y_{n-2} + \cdots + x_n y_1$, ϕ_n has size $e_n = 2n - 1$, and we build a matrix A_{ϕ_n} as follows,

$$A_{\phi_n} = \begin{bmatrix} x_1 & x_2 & \dots & x_n & 0 \\ -1 & & & \text{\Large 0} & y_n \\ & -1 & & & \dots \\ & & \ddots & & y_2 \\ \text{\Large 0} & & & -1 & y_1 \end{bmatrix}. \tag{12}$$

We have $\phi_n = \det(A_{\phi_n})$, and matrix size $s_n = n + 1$. Hence, $\lim_{n\to\infty} s_n/e_n = 1/2$. We can find a more compressible family of formulas ϕ_n with determinant polynomials A_{ϕ_n} as follows,

$$A_{\phi_n} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1,n-1} & x_{1n} \\ -1 & x_{22} & \dots & x_{2,n-1} & x_{2n} \\ & -1 & \dots & x_{3,n-1} & x_{3n} \\ & & \ddots & \dots & \dots \\ & & & x_{n-1,n-1} & x_{n-1,n} \\ \text{\Large 0} & & & -1 & x_{nn} \end{bmatrix}. \tag{13}$$

With $\phi_n = \det(A_{\phi_n})$, we have formula size $e_n \geq n(n+1)/2$, and matrix size $s_n = n$. Therefore,

$$\lim_{n\to\infty} s_n \leq \sqrt{2e_n}.$$

References

Mulmuley, K., Sohoni, M., 2001. Geometric complexity theory, P vs. NP, and explicit obstructions. In: Proceedings, International Conference on Algebra and Geometry, Hyderabad, 2001.

Valiant, L., 1979. Completeness classes in algebra. In: Proc. 11th Annual ACM Symposium on the Theory of Computing. Atlanta GA, pp. 249–261.

von zur Gathen, J., 1987. Feasible arithmetic computations: Valiant's hypothesis. Journal of Symbolic Computation 4, 137–172.