Some Computational Problems in Linear Algebra as Hard as Matrix Multiplication

Peter Bürgisser
International Computer Science Institute
Berkeley, California

Marek Karpinski*
Deptartment of Computer Science
University of Bonn
and
International Computer Science Institute
Berkeley, California

Thomas Lickteig
International Computer Science Institute
Berkeley, California

Abstract

We define the complexity of a computational problem given by a relation using the model of a computation tree with Ostrowski complexity measure. To a sequence of problems we assign an exponent similar as for matrix multiplication. For the complexity of the following computational problems in linear algebra

- KER_n: Compute a basis of the kernel for a given $n \times n$–matrix.
- OGB_n: Find an invertible matrix that transforms a given symmetric $n \times n$–matrix to diagonal form.
- SPR_n: Find a sparse representation of a given $n \times n$–matrix.

we prove relative lower bounds of the form $aM_n - b$ and absolute lower bounds dn^2, where M_n denotes the complexity of matrix multiplication and a, b, d are suitably chosen constants. We show that the exponent of the problem sequences KER, OGB, SPR is the same as the exponent ω of matrix multiplication.

*Supported in part by the Leibniz Center for Research in Computer Science, by the DFG Grant KA 673/2-1 and by the SERC Grant GR-E 68297

1 Introduction

It is well known that matrix multiplication is crucial for many computational problems in linear algebra. Problems like matrix inversion, computation of the determinant or of all coefficients of the characteristic polynomial, LR-decomposition and over the complex numbers also QR-decomposition and unitary transformation to Hessenberg form are all known to be as hard as matrix multiplication.(See [3, 4, 6, 7, 11, 12, 15].) In this paper we study some computational problems in linear algebra that are not specified by a function but by a relation.

Let F denote a field of characteristic zero that may be endowed with an ordering $\leq$. The reader may keep in mind the two important examples $F = \mathbf{C}$ or $F = \mathbf{R}$. A problem is given by a relation

$$P \subset F^m \times F^m.$$

Given an input $x \in F^m$ we are asked to find a $y \in F^n$ such that $(x, y) \in P$. We say that a function

$$f : F^m \longrightarrow F^n$$

solves the problem P if and only if

$$graph(f) \subset P.$$

In order to investigate the complexity of a problem we use the model of a computation tree T using the operation symbols $F \sqcup \{+, -, *, /\}$ and the relation symbol $=$ (or $\leq$ when we are working over an ordered field). We define the cost of a computation tree T as the maximal number of multiplications and divisions that may be performed in a computation of the tree T.(Compare [10, 13, 14, 16].) The complexity $C(f)$ of a function is then defined as the minimal cost of a tree computing f, and finally we put

$$C(P) := min\{C(f) : f \text{ function solving } P\}$$

for the complexity of the problem P. We have chosen the Ostrowski complexity measure because only this gives us enough flexibility to carry through lower bound proofs. However the upper bounds given in this paper also hold when we count all operations and comparisons.

One of the leading problems in computational linear algebra is matrix multiplication. In our formal framework

$$MAMU_{(n,n,n)} := \{((A, B), C) \in (F^{n\times n})^2 \times F^{n\times n} : A_1A_2A_3 = B_1B_2\}.$$

We put for its complexity $M_n := C(MAMU_{(n,n,n)})$. The rate of growth of M_n is measured by the exponent ω of matrix multiplication, which is defined by

$$\omega := \inf\{\tau \in \mathbf{R} : M_n = \mathrm{O}(n^\tau)\}.$$

We will study the following sequences of problems:

(1) 3-COMPRESSION:

$$3{-}CPR_n := \{((A_1, A_2, A_3), (B_1, B_2)) \in (F^{n\times n})^3 \times (F^{n\times n})^2 : A_1A_2A_3 = B_1B_2\}.$$

The investigation of this problem is motivated by the phenomenon that a corresponding problem for the addition of bitnumbers allows savings. The task

given bitnumbers a,b,c, find bitnumbers u,v such that a+b+c=u+v

can be solved more efficiently than by just adding up the numbers a,b,c. This is done with the so-called *carry save adders* which are used in the theory of boolean functions in several places in order to speed up computations. The lower bound we are going to prove shows that an analogue of the carry save adders for matrix multiplication does not exist.(Compare [18].)

(2) KERNEL:

$$KER_n := \{(A,B) \in F^{n\times n} \times \bigsqcup_{i=0}^{n} F^{n\times i} : B \in F^{n\times(n-R(A))}, R(A)+R(B)=n, AB=0\}.$$

This is of course the problem of computing a basis of the kernel for a given matrix.

(3) ORTHOGONAL BASIS:

$$OGB_n := \{(A,S) \in F^{n\times n} \times Gl_n : A \text{ symmetric}, SAS^T \text{diagonal}\}.$$

(4) SPARSE REPRESENTATION:

$$SPR_n := \{(A,(S,T,B)) \in F^{n\times n} \times (Gl_n^2 \times F^{n\times n}) : B = SAT, |supp(B)| \leq cn\}$$

($c > 0$ a fixed constant).

(5) SPARSENESS TRANSFORMATION MATRICES:

$$SPTM_n := \{(A,(S,T)) \in F^{n\times n} \times Gl_n^2 : |supp(SAT)| \leq cn\}$$

($c > 0$ a fixed constant).
In contrast to the SPARSE REPRESENTATION problem only the transformation matrices, but not the sparse representation matrix need to be computed.

The main goal of this paper is to prove lower bounds on the complexity of the problems cited above in terms of the complexity of matrix multiplication. The proofs rest on ideas from [15] and the important Derivation Theorem (see [3]). In the last section we employ the notion of dimension for an affine variety.

Let us summarize our results: We can assign to any sequence $P = (P_n)$ of problems an exponent

$$\omega_P := \inf\{\tau \in \mathbf{R} : C(P_n) = \mathrm{O}(n^\tau)\}.$$

For any of the problem sequences P listed under (1)–(5) we have

$$\omega_P \leq \omega.$$

In section 5 we prove for the sequences P listed under (1)–(4) the lower bound

$$\forall n \;\; C(P_n) \geq aM_n - bn^2$$

for suitably chosen constants $a, b > 0$. This implies immediately

$$\omega_P \geq \omega,$$

provided that $\omega > 2$. For the sequence $SPTM$ this estimate is also shown to be true. The aim of section 6 is to remove this assumption " $\omega > 2$ " by showing absolute lower bounds

$$\forall n \;\; C(P_n) \geq dn^2$$

(d a positive constant).
So for any of the sequences listed under (1)–(5) we have

$$\omega_P = \omega.$$

2 Some terminology

We treat two cases in parallel. In the first case F denotes a field of characteristic zero, in the second F stands for an ordered field. (Think of the two examples $F = \mathbf{C}$ and $F = \mathbf{R}$.)

A *problem* P is defined as being a subset

$$P \subset F^m \times F^n.$$

We say that a function

$$f : F^m \longrightarrow F^n$$

solves the problem P if and only if

$$graph(f) \subset P.$$

In order to investigate these objects from the point of view of computations, we use the model of a computation tree. Let us shortly describe this notion, a detailed discussion can be found in [10, 13, 14, 16].

As the set Ω of operational symbols and the set R of relational symbols (together with arity functions) we take

$$\Omega = F \sqcup \{0, 1, +, -, *, /\}$$

and

$$R = \{=\}$$

$$(\text{or } R = \{\leq\} \text{ when we consider an ordered field } (F, \leq)).$$

Let $s_1, s_2, \ldots$ be variables denoting storage locations in a computer. A *computation tree* T of type (Ω, R) with output list of length n is a binary tree together with a function that assigns

- to any simple vertex an operational instruction of the form

$$s_i := \omega(s_{j_1}, \ldots, s_{j_k}),$$

where $k \geq 0$, $i, j_1, \ldots, j_k > 0$ and $\omega \in \Omega$ k-ary,

- to any branching vertex a test instruction of the form

$$\rho(s_{j_1}, \ldots, s_{j_k}),$$

where $k \geq 0$, $j_1, \ldots, j_k > 0$ and $\rho \in R$ k-ary
and

- to any leaf an output instruction of the form

$$(s_{j_1}, \ldots, s_{j_n}),$$

where $j_1, \ldots, j_n > 0$.

The assumption that all output lists have the same length n is made in order to simplify notation and is not essential. When fixing additionally an input length m such a computation tree T *computes* a partial function

$$f : F^m \supset def(f) \longrightarrow F^n$$

in the following way: given $\xi \in F^m$ we assign to the variables at the root of the tree the values $(\xi_1, \ldots, \xi_m, \infty, \infty, \ldots)$. We say $\xi \in def(f)$ if and only if the directed path starting from the root and defined in an obvious manner by the computation tree T leads to a leaf. If this is the case the values of the output instruction are $(f_1(\xi), \ldots, f_n(\xi))$. It is easy to see that for a directed path π from the root to the leaf the set

$$D_\pi \subset F^m$$

of inputs defining this path π is a locally closed semialgebraic set and that the restriction of f to D_π is restriction of some rational function.

Now we are going to define the complexity of problems and functions. We chose Ostrowski's complexity measure, e.g. we count only the "noncsalar" multiplications and divisions and allow the linear operations and comparisons for free. We do so not only for simplicity and elegance, but mainly because only this measure provides us with enough flexibility to succeed in proving lower bounds. Let be given a computation tree T and a path π from the root to a leaf. The cost of π is defined as the number of vertices of π equipped with an operational instruction $\omega \in \{*, /\}$. By maximizing over all such paths π of T we get the cost $cost(T)$ of the computation tree. As the *complexity* $C(f)$ of a partial function

$$f : F^m \supset def(f) \longrightarrow F^n$$

we then define

$$C(f) := \min\{cost(T) : T \text{ computation tree computing } f\}$$

and finally we call

$$C(P) := \min\{C(f) : f \text{ a function solving } P\}$$

the *complexity* of the problem $P \subset F^m \times F^n$.

The sequence of the matrix multiplication problems

$$MAMU_{(e,h,l)} := \{((A, B), C) \in (F^{e\times h} \times F^{h\times l}) \times F^{e\times l} : AB = C\}$$

is fundamental in linear algebra. We put

$$M_n := C(MAMU_{(n,n,n)}).$$

As a lower bound only the estimate $M_n \geq 2n^2 - 1$ is known ([2, 9]). The asymptotic behaviour of (M_n) is measured by the so-called exponent ω of matrix multiplication

$$\omega := \inf\{\tau \in \mathbf{R} : M_n = \mathrm{O}(n^\tau)\}.$$

The currently best known estimate is $2 \leq \omega < 2.376$ ([5, 17]). Of course we can assign to any sequence $P = (P_n)$ of problems an exponent

$$\omega_P := \inf\{\tau \in \mathrm{R} : C(P_n) = \mathrm{O}(n^\tau)\}.$$

We will show that the exponent for various sequences of computational problems in linear algebra equals the exponent of matrix multiplication.

We outline our method for giving lower bounds on the complexity of problems. To do so we need the notion of the nonscalar complexity of a family of rational functions. Let a subring A

$$F[x_1, \ldots, x_m] \subset A \subset F(x_1, \ldots, x_m)$$

of the field of rational functions in the variables $x_1, \ldots, x_m$ be given which contains the polynomial ring $F[x_1, \ldots, x_m]$. Let $f_1, \ldots, f_n \in A$. The *nonscalar complexity* $L_A(f_1, \ldots, f_n)$ of $f_1, \ldots, f_n$ with respect to the subring A is defined as the minimal number of multiplications and divisions by units in the ring A that are sufficient to compute $f_1, \ldots, f_n$ from the input set $F \cup \{x_1, \ldots, x_m\}$ by a straight line program using the operations $F \sqcup \{0, 1, +, -, *, /\}$. Let $P \subset F^m \times F^n$ be a problem, $f : F^m \to F^n$ be an optimal function solving P and T be an optimal computation tree computing f. Then we know

$$C(P) = C(f) = cost(T).$$

We already noticed that we have a finite disjoint union

$$F^m = \bigcup\{D_\pi : \pi \text{ directed path from root to leaf}\},$$

where D_π denotes the set of inputs defining the path π. The sets D_π are locally closed and the restriction of f to D_π is restriction of some rational function. There must be a path π_0 such that D_{π_0} is Zariski-dense in F^m for the simple reason that F^m is irreducible. Let us call such a path π_0 a typical one. If we are considering a field F without ordering, there is exactly one typical path, because in this case a nonempty locally closed set contains a nonempty Zariski-open subset and two of them must intersect. However, if we are working over an ordered field $(F, \leq)$ there might be many typical paths. Let π_0 be a typical one. Then we can consider the $g_i := f_i \mid_{D_{\pi_0}}$ as elements of $F(x_1, \ldots, x_n)$ and we easily see that

$$cost(T) \geq cost(\pi_0) \geq L_{F(\underline{x})}(g_1, \ldots, g_n)$$

and

$$\forall \xi \in D_{\pi_0} \;\; (\xi, (g_1(\xi), \ldots, g_n(\xi)) \in P.$$

We proved

Lemma 1 *Let $P \subset F^m \times F^n$ be a problem. Then there are rational functions $g_1, \ldots, g_n \in F(x_1, \ldots, x_m)$ such that*

$$(\xi, (g_1(\xi), \ldots, g_n(\xi))) \in P \text{ for Zariski-almost all } \xi \in F^m$$

and

$$C(P) \geq L_{F(\underline{x})}(g_1, \ldots, g_n).$$

So we gave a lower bound for the complexity of a problem in terms of the nonscalar complexity of rational functions from which we only know that they satisfy certain relations. For dealing with the nonscalar complexity of rational functions we will use some known techniques that are listed in the next section.

3 Properties of the nonscalar complexity

Let us recall some ideas from [15]. For $\lambda \in F^m$ we consider the local rings

$$\mathcal{O}_\lambda := \{f \in F(x_1, \ldots, x_m) : \ f \text{ defined at } \lambda \ \}.$$

It is well known that $\mathcal{O}_\lambda$ is contained in the ring $F[\![y_1, \ldots, y_m]\!]$ of formal power series via the imbedding

$$\mathcal{O}_\lambda \hookrightarrow F[\![y_1, \ldots, y_m]\!], \ \ x_i - \lambda \mapsto y_i.$$

So the image of an element $f \in \mathcal{O}_\lambda$ is just the taylor expansion of f around λ.

Lemma 2 *For given $f_1, \ldots, f_n \in F(x_1, \ldots, x_m)$ the equality*

$$L_{F(\underline{x})}(f_1, \ldots, f_n) = L_{\mathcal{O}_\lambda}(f_1, \ldots, f_n)$$

holds for Zariski-almost all $\lambda \in F^m$.

We omit the trivial proof.

The next theorem will be used throughout in the paper.

Theorem 1 ("Vermeidung von Divisionen" [15]) *Let $\lambda \in F^m$, $f_1, \ldots, f_n \in \mathcal{O}_\lambda$, $d \in \mathbf{N}$. Then*

$$L_{F[y_1,\ldots,y_m]}(\{f_i^{(k)}(\underline{y}) : 0 \leq k \leq d, \ 1 \leq i \leq m\}) \leq \frac{d(d-1)}{2} L_{\mathcal{O}_\lambda}(f_1, \ldots, f_n),$$

where $\sum_{k=0}^{\infty} f_i^{(k)}(\underline{y})$ denotes the taylor expansion of f_i around the point λ.

Observe that the complexity on the lefthand side is defined with respect to the polynomial ring $F[y_1, \ldots, y_m]$. A well known consequence of the statement above is

$$M_n = C(MAMU_{(n,n,n)}) = L_{F[\underline{X},\underline{Y}]}(\{\sum_{l=1}^{n} X_{il} Y_{lj} : 1 \leq i, j \leq n\}).$$

The proof of Theorem 1 (see [15]) immediately leads to

Corollary 1 *Let F be algebraically closed. Let $\lambda \in F^m, f_1, \ldots, f_n \in \mathcal{O}_\lambda, d \in \mathbf{N}$. Then there exists a closed cone $Z \subset F^m$ with vertex in the origin such that*

$$codim(Z) \leq \frac{d(d-1)}{2} L_{\mathcal{O}_\lambda}(f_1, \ldots, f_n)$$

and

$$Z \subset \{\eta \in F^m : f_i^{(k)}(\eta) = 0 \text{ for all } i \in \{1, \ldots, n\},\ k \in \{2, \ldots, d\}\}$$

Every irreducible component of Z is again a cone with vertex in the origin.

This corollary will be of help in section 6 of this paper.

We will frequently use the simple fact that a linear substitution does not increase the complexity, provided we are working in the polynomial ring. More precisely, let be

$$f_1, \ldots, f_n \in F[x_1, \ldots, x_m],\ A \in F^{m \times m},$$

and put

$$g_i = f_i(A\underline{x}).$$

Then

$$L_{F[\underline{x}]}(g_1, \ldots, g_n) \leq L_{F[\underline{x}]}(f_1, \ldots, f_n).$$

Finally we cite the important

Theorem 2 ("Derivation Theorem" [3]) *Let $f \in F(x_1, \ldots, x_m)$ be a rational function. Then*

$$L_{F(\underline{x})}(f, \frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_m}) \leq 3 L_{F(\underline{x})}(f).$$

4 Relative upper bounds

We recall the definitions of the computational problems we are interested in.

- t-COMPRESSION:

$$[t - CPR_n := \{((A_1, \ldots, A_t), (B_1, \ldots, B_{t-1})) \in (F^{n \times n})^t \times (F^{n \times n})^{t-1} : A_1 A_2 \cdots A_t = B_1 B_2 \cdots B_{t-1}\}.$$

  ($t \geq 3$ a natural number).

- KERNEL:

$$KER_n := \{(A, B) \in F^{n \times n} \times \bigsqcup_{i=0}^{n} F^{n \times i} : B \in F^{n \times (n - R(A))}, R(A) + R(B) = n, AB = 0\}.$$

- ORTHOGONAL BASIS:

$$OGB_n := \{(A, S) \in F^{n \times n} \times Gl_n : A \text{ symmetric}, SAS^T \text{diagonal}\}.$$

- SPARSE REPRESENTATION:

$$SPR_n := \{(A,(S,T,B)) \in F^{n\times n} \times (Gl_n^2 \times F^{n\times n}) : B = SAT, |supp(B)| \leq cn\}$$

($c > 0$ a fixed constant).

- SPARSENESS TRANSFORMATION MATRICES:

$$SPTM_n := \{(A,(S,T)) \in F^{n\times n} \times Gl_n^2 : |supp(SAT)| \leq cn\}$$

($c > 0$ a fixed constant).

The following theorem gives an upper bound relative to the complexity of matrix multiplication.

Theorem 3 *The exponent for any of the sequences of problems*

$$t{-}CPR,\ KER,\ OGB,\ SPR,\ SPTM$$

is less or equal the exponent ω of matrix multiplication.

The proof is based on ideas from [4, 11, 12]. See also [1, pages 233–240]. The proceeding is to subdivide the occuring matrices into blocks, to perform a sort of Gaussian Elimination blockwise using a fast hypothetical matrix multiplication algorithm, and then to continue recursively. We leave the details to the reader. For the problem $t{-}CPR$ the statement is of course trivial.

Remark: Theorem 3 remains true when we count all rational operations and tests at unit cost.

5 Relative lower bounds

We are going to prove lower bounds in terms of M_n for the various problems defined above.

Theorem 4

$$C(3{-}CPR_n) \geq \frac{1}{3}M_n - n^2.$$

Proof: Let A, B, C be $n \times n$–matrices whose entries are indeterminates over F and put $K := F(A_{ij}, B_{ij}, C_{ij})$. By Lemma 1 there are $U, V \in K^{n\times n}$ such that

$$UV = ABC$$

and

$$L_K(U,V) \leq C(3{-}CPR_n).$$

If we take into consideration that the trace of the product of two $n \times n$–matrices can be computed with n^2 multiplications, we get

$$L_K(Tr(ABC)) \leq C(3\text{–}CPR_n) + n^2.$$

Furthermore

$$\frac{\partial Tr(ABC)}{\partial A_{ij}} = (BC)_{ji}.$$

Theorem 2 implies now

$$M_n = L_K(BC) \leq 3C(3\text{–}CPR_n) + 3n^2,$$

which completes the proof of the theorem. □

We do not know how to prove a similar lower bound for the problem sequences t–CPR when $t > 3$.

Theorem 5

$$C(KER_n) \geq M_{\lfloor n/4 \rfloor}.$$

Proof: W.l.o.g. we may assume that $n = 4m$, $m \in \mathbf{N}$. Let X, Y denote $2m \times 2m$-matrices whose entries are indeterminates over F. We put $K := F(X_{ij}, Y_{ij})$ and $R := F[X_{ij}, Y_{ij}]$. When we apply Lemma 1 to the restricted problem

$$KER_n \cap \left(\left\{ \begin{pmatrix} \xi & \eta \\ 0 & 0 \end{pmatrix} : \xi, \eta \in F^{2m \times 2m} \right\} \times \bigsqcup_{i=0}^{n} F^{n \times i} \right),$$

we see that there exists a matrix $B \in K^{4m \times 4m}$ with

$$R(B) = 2m,\ (X, Y)B = 0$$

and

$$L_K(B) \leq C(KER_n).$$

There must be a matrix $U \in Gl_{2m}(K)$ such that

$$B = \begin{pmatrix} X^{-1}U \\ Y^{-1}U \end{pmatrix}.$$

We therefore have

$$L_K(X^{-1}U, Y^{-1}U) \leq C(KER_n).$$

¿From Lemma 2 we obtain that there are $\xi, \eta \in Gl_{2m}(F)$ such that $\det U(\xi, \eta) \neq 0$ and

$$L_{\mathcal{O}_{(\xi,\eta)}}(X^{-1}U, Y^{-1}U) = L_K(X^{-1}U, Y^{-1}U).$$

We may replace U by $UU(\xi, \eta)^{-1}$ and therefore assume that $U(\xi, \eta) = E$. Application of the automorphism

$$\varphi \in Aut_F K : \varphi(X) = X\xi^{-1}, \varphi(Y) = Y\eta^{-1}$$

shows that we can assume w.l.o.g. that $\xi = \eta = E$. Furthermore

$$L_{\mathcal{O}_{(0,0)}}((E - X)^{-1}V, (E - Y)^{-1}V) = L_{\mathcal{O}_{(E,E)}}(X^{-1}U, Y^{-1}U),$$

when we put $V := U(E - X, E - Y)$. We use now Theorem 1 with $d = 2$ and get

$$L_R(X^2 + XW^{(1)} + W^{(2)}, Y^2 + YW^{(1)} + W^{(2)}) \leq L_{\mathcal{O}_{(0,0)}}((E - X)^{-1}W, (E - Y)^{-1}W),$$

where $W = E + W^{(1)} + W^{(2)} + \cdots$ denotes the taylor expansion of W around the point $(0, 0)$. The complexity on the lefthand side can be estimated from below by

$$L_R(X^2 - Y^2 + (X - Y)W^{(1)}).$$

We write

$$X = \begin{pmatrix} X^{11} & X^{12} \\ X^{21} & X^{22} \end{pmatrix}, \; Y = \begin{pmatrix} Y^{11} & Y^{12} \\ Y^{21} & Y^{22} \end{pmatrix}$$

where $X^{ij}, Y^{ij} \in K^{m \times m}$ and make the linear substitution ψ

$$\psi(X) := \begin{pmatrix} 0 & X^{12} \\ X^{21} & X^{22} \end{pmatrix}, \psi(Y) := \begin{pmatrix} 0 & X^{12} \\ 0 & 0 \end{pmatrix}.$$

One calculates immediately that

$$\psi(X^2 - Y^2 + (X - Y)W^{(1)}) = \begin{pmatrix} X^{12}X^{21} & X^{12}X^{22} \\ P & Q \end{pmatrix}$$

for some $P, Q \in K^{m \times m}$. Therefore

$$M_m = L_R(X^{12}X^{21}) \leq C(KER_n).$$

□

Theorem 6

$$C(OGB_n) \geq \frac{1}{3}M_{\lfloor n/4 \rfloor} - 4n^2 - n.$$

Proof W.l.o.g. we may assume that $n = 4m$, $m \in \mathbf{N}$. Let A denote a symmetric $n \times n$–matrix whose entries are indeterminates over F. Put $K := F(A_{ij} : i \leq j \leq n)$ and $R = F[A_{ij} : i \leq j \leq n]$. By Lemma 1 there is a matrix $S \in Gl_n(K)$ such that

$$D := SAS^T \text{ is diagonal}$$

and

$$L_K(S) \leq C(OGB_n).$$

By writing

$$\begin{bmatrix} D_{11} \\ \vdots \\ D_{nn} \end{bmatrix} = D \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = S(A(S^T \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix})),$$

we see that D can be computed from A and S with $2n^2$ multiplications. We have

$$Tr(A^{-1}) = Tr(S^T(D^{-1}S)).$$

Therefore

$$L_K(Tr(A^{-1}) \leq L_K(S) + 4n^2 + n.$$

We proceed now similar as in [6]. Let $V \in F^{n\times n}$ be symmetric and ϵ be an indeterminate over K. Then

$$Tr((A+\epsilon V)^{-1}) = Tr(A^{-1}) + \epsilon \sum_{i\leq j} \frac{\partial Tr(A^{-1})}{\partial A_{ij}} V_{ij} + \mathrm{O}(\epsilon^2).$$

On the other hand one easily calculates

$$Tr((A+\epsilon V)^{-1}) = Tr(A^{-1}) - \epsilon\, Tr(A^{-1}VA^{-1}) + \mathrm{O}(\epsilon^2).$$

Comparing the two equations we get

$$\frac{\partial Tr(A^{-1})}{\partial A_{ij}} = \begin{cases} -2(A^{-2})_{ij} & \text{, if } i \neq j \\ -(A^{-2})_{ij} & \text{, otherwise.} \end{cases}$$

¿From the Derivation Theorem 2 we deduce

$$\frac{1}{3} L_K(A^{-2}) \leq L_K(Tr(A^{-1})) \leq L_K(S) + 4n^2 + n.$$

By Lemma 2 there exists a symmetric matrix $\alpha \in Gl_n(F)$ such that

$$L_{\mathcal{O}_\alpha}(A^{-2}) = L_K(A^{-2}).$$

Furthermore

$$L_{\mathcal{O}_\alpha}(A^{-2}) = L_{\mathcal{O}_E}(A^{-2}) = L_{\mathcal{O}_0}((E-A)^{-2}).$$

Using Theorem 1 with $d = 2$ and taking into account that

$$(E-A)^{-2} = E + 2A + 3A^2 + \ldots$$

we get

$$L_R(A^2) \leq L_{\mathcal{O}_0}((E-A)^{-2}).$$

We divide the matrix A into $m \times m$–blocks $A^{ij} \in K^{m\times m}$ and define the substitution ψ by

$$\psi(A) := \begin{pmatrix} 0 & 0 & A^{13} & 0 \\ 0 & 0 & A^{23} & 0 \\ (A^{13})^T & (A^{23})^T & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Obviously

$$\psi(A)^2 = \begin{pmatrix} A^{13}(A^{13})^T & A^{13}(A^{23})^T & 0 & 0 \\ A^{23}(A^{13})^T & A^{23}(A^{23})^T & 0 & 0 \\ 0 & 0 & W & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

where $W = (A^{13})^T A^{13} + (A^{23})^T A^{23}$. Hence

$$M_n = L_R(A^{13}(A^{23})^T) \leq L_R(A^2) \leq 3(C(OGB_n) + 4n^2 + n).$$

□

Let $c > 0$ be fixed. We call a matrix $A \in F^{n\times n}$ *sparse* if and only if

$$|supp(A)| \leq cn.$$

Observe that an arbitrary $n \times n$–matrix can be multiplied with a sparse matrix using only cn^2 multiplications.

Theorem 7

$$C(SPR_n) \geq \frac{1}{9} M_{\lfloor n/3 \rfloor} - (2 + 10c/3)n^2.$$

Proof W.l.o.g. we can assume that $n = 3m$, $m \in \mathbf{N}$. Let A denote a $n \times n$–matrix whose entries are indeterminates over F. We set $K := F(A_{ij} : i, j \leq n)$ and $R := F[A_{ij} : i, j \leq n]$. By Lemma 1 there exist $\tilde{S}, \tilde{T} \in Gl_n(K)$ and a sparse matrix $\tilde{B} \in K^{n \times n}$ such that

$$\tilde{B} = \tilde{S} A \tilde{T}$$

and

$$L_K(\tilde{S}, \tilde{T}, \tilde{B}) \leq C(SPR_n).$$

Lemma 2 shows that there is a matrix $\alpha \in Gl_n(F)$ such that

$$L_{\mathcal{O}_\alpha}(\tilde{S}, \tilde{T}, \tilde{B}) = L_K(\tilde{S}, \tilde{T}, \tilde{B})$$

and

$$\tilde{S}(\alpha),\ \tilde{T}(\alpha) \in Gl_n(F).$$

By applying the substitution $A \mapsto \alpha(E - A)$ we see that there exist $S, T \in Gl_n(K)$, $B \in K^{n \times n}$ satisfying

$$B = S(E - A)T,$$

$$S(0), T(0) \in Gl_n(F),\ B \text{ sparse}$$

and

$$L_{\mathcal{O}_0}(S, T, B) = L_{\mathcal{O}_\alpha}(\tilde{S}, \tilde{T}, \tilde{B}).$$

Let $S = S^{(0)} + S^{(1)} + \ldots$, $T = T^{(0)} + T^{(1)} + \ldots$, $B = B^{(0)} + B^{(1)} + \ldots$ denote the taylor expansions around 0 of S, T, B respectively. The matrices $B^{(k)}$ are also sparse. We put $\beta := B^{(0)}$.
Theorem 1 implies

$$L_R(S^{(2)}, S^{(3)}, T^{(2)}, T^{(3)}) \leq 3 L_{\mathcal{O}_0}(S, T, B).$$

By comparing the third order terms in the taylor expansion of both sides of the equation

$$(E - A)^{-1} = T B^{-1} S$$

we get

$$Tr(A^3) = \sum_{i+j+k=3} Tr(T^{(i)}((B^{-1})^{(j)} S^{(k)})).$$

We are going to show now that the products $(B^{-1})^{(j)} S^{(k)}$ $(j + k \leq 3)$ can be computed from $B^{(2)}, B^{(3)}, S^{(2)}, S^{(3)}$ with only $10cn^2$ multiplications:

A short calculation yields (put $\gamma := (B^{(0)})^{-1}$)

$$(B^{-1})^{(1)} = -\gamma B^{(1)} \gamma, \tag{1}$$

$$(B^{-1})^{(2)} = -\gamma B^{(2)} \gamma + \gamma B^{(1)} \gamma B^{(1)} \gamma, \tag{2}$$

$$(B^{-1})^{(3)} = -\gamma B^{(3)} \gamma + \gamma B^{(2)} \gamma B^{(1)} \gamma + \gamma B^{(1)} \gamma B^{(2)} \gamma + \gamma B^{(1)} \gamma B^{(1)} \gamma B^{(1)} \gamma. \tag{3}$$

Observe furthermore that a product

$$\gamma \Sigma_1 \gamma \Sigma_2 \cdots \gamma \Sigma_t \gamma \Gamma,$$

where

$$\gamma \in Gl_n(F),\ \Gamma \in K^{n\times n}$$

and

$$\Sigma_i \in K^{n\times n} \text{ sparse } (i = 1, \ldots, t),$$

can be computed from the matrices Σ_i, Γ with only ctn^2 nonscalar multiplications. (Compute from the righthand side to the left.) Taking this into account the upper bound $10cn^2$ follows now easily.

The result of this intermediate reasoning gives the upper bound

$$L_R(Tr(A^3)) \leq 3C(SPR_n) + 6n^2 + 10cn^2.$$

We subdivide A into $m \times m$–blocks $A^{ij} \in K^{m\times m}$ and make the substitution ψ defined by

$$\psi(A) := \begin{pmatrix} 0 & A^{12} & 0 \\ 0 & 0 & A^{23} \\ A^{13} & 0 & 0 \end{pmatrix}.$$

One easily verifies

$$\psi(A)^3 = \begin{pmatrix} A^{12}A^{23}A^{31} & 0 & 0 \\ 0 & A^{23}A^{31}A^{12} & 0 \\ 0 & 0 & A^{31}A^{12}A^{23} \end{pmatrix}$$

and hence

$$\psi(Tr(A^3)) = Tr(\psi(A)^3) = 3Tr(A^{12}A^{23}A^{31}).$$

So we showed that

$$L_R(A^{12}A^{23}A^{31}) \leq L_R(Tr(A^3)).$$

When we apply the Derivation Theorem 2 we finally get

$$\frac{1}{3}M_m \leq L_R(A^{12}A^{23}A^{31})$$

which implies the desired bound

$$\frac{1}{9}M_m - (2 + 10c/3)n^2 \leq C(SPR_n).$$

□

As an immediate consequence of Theorem 3 and Theorems 4 - 7 we get the following

Corollary 2 *Any of the sequences of problems*

$$3{-}CPR,\ KER,\ OGB,\ SPR$$

has as exponent the exponent ω of matrix multiplication, provided that $\omega > 2$.

For the sequence $SPTM$ we will only make a statement about the exponent. We need the following

Lemma 3 *The sequence of problems $MAMU_{(n,n,\lfloor\sqrt{n}\rfloor)}$ has an exponent strictly smaller than the exponent ω of matrix multiplication, provided that $\omega > 2$.*

Proof: We assume $\omega > 2$ and chose ϵ satisfying $0 < \epsilon < \omega/2 - 1$. For a suitable constant $d > 0$ and all squares n we have

$$C(MAMU_{(n,n,\sqrt{n})}) \leq C(MAMU_{(\sqrt{n},\sqrt{n},\sqrt{n})})C(MAMU_{(\sqrt{n},\sqrt{n},1)}) \leq d(\sqrt{n})^{\omega+2\epsilon}n.$$

Therefore

$$C(MAMU_{(n,n,\sqrt{n})}) = \mathrm{O}(n^{\omega/2+\epsilon+1}).$$

But $\omega/2 + \epsilon + 1 < \omega$ and the statement follows. □

Theorem 8 *The exponent for the sequence $SPTM$ equals the exponent ω of matrix multiplication, provided that $\omega > 2$.*

Proof: Suppose $\omega > 2$. Since we already proved Theorem 7 it is sufficient to show the following:

Given $(A, S, T) \in F^{n\times n} \times Gl_n^2$ and the information that $B := SAT$ is sparse, then we can compute the matrix B with cost $\mathrm{O}(n^\tau)$, where $\tau < \omega$.

So let $(A, S, T) \in F^{n\times n} \times Gl_n^2$ be given. We put $B := SAT$ and assume that B is sparse. For $i \in \{1, \ldots, n\}$ we define

$$I_i := \{j \in \{1, \ldots, n\} : B_{ij} \neq 0\}.$$

In order to simplify notation we assume that $(|I_i|)_{j=1,\ldots,n}$ is a decreasing sequence. We set

$$M := \max\{i : |I_i| \leq \lfloor\sqrt{n}\rfloor\}.$$

Then

$$n - M \leq c\sqrt{n}.$$

We now chose a matrix $U \in F^{n\times\lfloor\sqrt{n}\rfloor}$ with the property that all its subdeterminants are different from zero. According to the preceeding remark we can compute the product C

$$C := BV = S(A(TU))$$

with cost $\mathrm{O}(n^\tau)$, where $\tau < \omega$. However, the first M rows of B can be computed from C without any nonscalar operations: for a fixed $i \leq \lfloor\sqrt{n}\rfloor$ we have

$$\forall\ k \leq \lfloor\sqrt{n}\rfloor \quad \sum_{j\in I_i} B_{ij}U_{jk} = C_{ik}.$$

Now observe that $(U_{jk})_{j\in I_i,\, k=1,\ldots,|I_i|}$ is an invertible matrix with entries in F.

In order to get the remaining rows of B we do simply the following. We chose a matrix $V \in Gl_{n-M}(F)$ and put

$$W := (0, V) \in F^{(n-M)\times n}.$$

The product

$$WB = ((WS)A)T$$

can be computed with only $\mathrm{O}(n^\tau)$ nonscalar operations.
But from WB we can obtain $(B_{ij})_{i=M+1,\ldots,n,\ j=1,\ldots,n}$ without using any nonscalar operations. □

6 Absolute lower bounds

The aim of this section is to show that the assumption "$\omega > 2$" in Corollary 2 and Theorem 8 is unnecessary. We will do so by proving lower bounds of the type

$$\text{constant} \cdot n^2$$

for the various computational problems in linear algebra we considered before. There is no harm in assuming that F is algebraically closed.

We need the following

Lemma 4 *The set*

$$\Delta_n := \{\alpha \in F^{n\times n} : \alpha^3 = 0\}$$

is a closed subvariety of the affine variety $F^{n\times n}$. For its dimension we have the estimate

$$\dim(\Delta_n) \leq \frac{2}{3}n^2.$$

This inequality is sharp if n is a multiple of 3.

This lemma is of course classical and can be proved by standard techniques. For the reader's convenience we add a proof.

Proof: Let β be a matrix in Jordan normal form, i.e.

$$\beta = diag(J(n_1, \lambda_1), \ldots, J(n_t, \lambda_t)),$$

where $(n_1, \ldots, n_t)$ is a partition of n and $J(n_i, \lambda_i)$ is defined as

$$J(n_i, \lambda_i) = \begin{pmatrix} \lambda & 1 & 0 & 0 \\ 0 & \ddots & \ddots & 0 \\ 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & \lambda \end{pmatrix} \in F^{n_i \times n_i}.$$

We denote the dimension of the conjugacy class of β by $d_\lambda(n_1, \ldots, n_t)$. It is easy to see that

$$\beta^3 = 0 \iff \forall i\ \lambda_i = 0, n_i \leq 3.$$

This implies at once

$$\dim(\Delta_n) = \max\{d_0(n_1, \ldots, n_t) : (n_1, \ldots, n_t) \text{ partition of } n \text{ with } n_i \leq 3 \text{ for all } i\}.$$

The value of $d_0(n_1, \ldots, n_t)$ can be exactly determined, namely

$$d_0(n_1, \ldots, n_t) = n^2 - \sum_{j=1}^{t'} (n'_j)^2,$$

where $(n'_1, \ldots, n'_{t'})$ denotes the partition dual to $(n_1, \ldots, n_t)$. (See [8, page 192].) Using this fact we conclude

$$\dim(\Delta_n) = \max\{n^2 - (a+b+c)^2 - (a+b)^2 - a^2 : n = 3a + 2b + c,\ a, b, c \in \mathbf{N}\} \leq 2n^2/3,$$

with equality when n is a multiple of 3. $\square$

Remark: We sketch here a more elementary proof method, which, however, only yields the bound $8n^2/9$.

By thinking of the Jordan normal form of a matrix we see that

$$\Delta_n \subset \{\alpha \in F^{n\times n} : R(\alpha) \leq 2n/3\}.$$

Employing the well known fact

$$\dim(\{\alpha \in F^{n\times n} : R(\alpha) \leq r\}) = r(2n-r)$$

we obtain

$$\dim(\{\alpha \in F^{n\times n} : R(\alpha) \leq 2n/3\}) \leq 8n^2/9.$$

Theorem 9

$$\liminf_{n\to\infty} C(SPTM_n)n^{-2} \geq \frac{1}{9}.$$

Proof: Let A denote a $n \times n$–matrix whose entries are indeterminates over F.
Put $K := F(A_{ij} : i,j \leq n)$. The same reasonings as in the proof of Theorem 8 show that there are
$S,T \in Gl_n(K), B \in K^{n\times n}$ such that

$$B = S(E-A)T,$$

$$S(0),T(0) \in Gl_n(F),\ B \text{ sparse}$$

and

$$L_{\mathcal{O}_0}(S,T) \leq C(SPTM_n).$$

Suppose that $L_{\mathcal{O}_0}(S,T) \leq \epsilon n^2$ for some $\epsilon > 0$. Corollary 1 says that there is a closed cone $Z \subset F^{n\times n}$ with vertex in the origin satisfying

$$codim(Z) \leq 3\epsilon n^2$$

and

$$Z \subset \{\alpha \in F^{n\times n} : S^{(k)} = T^{(k)} = 0 \text{ for } k = 2,3\}.$$

¿From $(E-A)^{-1} = TB^{-1}S$ we get

$$A^3 = \sum_{i+j+k=3} T^{(i)}(B^{-1})^{(j)}S^{(k)}.$$

We define

$$Z' := Z \cap \{\alpha \in F^{n\times n} : B^{(k)}(\alpha) = 0 \text{ for } k = 1,2,3\ \}.$$

Z' is again a closed cone with vertex in the origin and we have

$$codim_{F^{n\times n}}(Z') \leq 3\epsilon n^2 + 3cn,$$

since the matrices $B^{(k)}$ are sparse. From the equations (1)– (3) of section 5 follows

$$Z' \subset \{\alpha \in F^{n\times n} : (B^{-1})^{(k)}(\alpha) = 0 \text{ for } k = 1,2,3\}.$$

We have therefore shown that

$$Z' \subset \{\alpha \in F^{n\times n} : \alpha^3 = 0\}.$$

Lemma 4 and a comparison of dimensions lead to the inequality

$$3\epsilon n^2 + 3cn \geq n^2/3$$

or

$$\epsilon + c/n \geq 1/9.$$

If $\epsilon < 1/9$ we get a contradiction for sufficiently large n. So we have proved that

$$\forall \epsilon < 1/9 \;\; \exists n_0 \;\; \forall n \geq n_0 \;\; C(SPTM_n) > \epsilon n^2,$$

which proves the theorem. □

Theorem 10

$$\liminf_{n\to\infty} C(3{-}CPR_n)n^{-2} \geq \frac{1}{9},$$

$$\liminf_{n\to\infty} C(OGB_n)n^{-2} \geq \frac{1}{54},$$

$$\liminf_{n\to\infty} C(SPR_n)n^{-2} \geq \frac{1}{9}.$$

Since the proof is very similar to the previous one, we only give some hints.
$3-CPR$: Consider the restricted problem

$$\{(A,(U,V)) \in F^{n\times n} \times (F^{n\times n})^2 : A^3 = UV\}.$$

OGB: Use

$$\dim(\{\alpha \in F^{n\times n} : \alpha \text{ symmetric}, R(\alpha) \leq r\}) = rn - r(r-1)/2$$

in order to show that

$$\dim(\{\alpha \in F^{n\times n} : \alpha \text{ symmetric}, \alpha^3 = 0\}) \leq 4n^2/9 + n/3.$$

SPR: The statement follows trivially from $C(SPR_n) \geq C(SPTM_n)$.

Putting all our information together we get the final result

Theorem 11 *Any of the sequences of problems*

$$3{-}CPR,\ KER,\ OGB,\ SPR,\ SPTM$$

has as exponent the exponent ω of matrix multiplication.

References

[1] *A.V. Aho* and *J.E. Hopcroft* and *J.D. Ullman*, The design and analysis of computer algorithms, Reading MA: Addison–Wesley, 1974.

[2] *A. Alder* and *V. Strassen*, On the algorithmic complexity of associative algebras, Theor. Computer Science **15**(1981), 201–211.

[3] *W. Baur* and *V. Strassen*, The complexity of partial derivatives, Theor. Computer Science **22**(1982), 317–330.

[4] *J. Bunch* and *J. Hopcroft*, Triangular factorizationand inversion by fast matrix multiplication, Math. Comp. **28**(1974), 231–236.

[5] *D. Coppersmith* and *S. Winograd*, Matrix multiplication via arithmetic progressions, Proc. 19th ACM STOC, New York(1987) , 1–6.

[6] *K. Kalorkoti*, The trace invariant and matrix inversion, Theor. Computer Science **59**(1988), 277–286.

[7] *W. Keller–Gehrig*, Fast algorithms for the characteristic polynomial, Theor. Computer Science **36**(1985), 309–317.

[8] *H. Kraft*, Geometric methods in representation theory, in: Representations of Algebras, Workshop Proc., Puebla, Mexico 1980, LNM **944**, Berlin–Heidelberg–New York 1982.

[9] *J.C. Lafon* and *W. Winograd*, A lower bound for the multiplicative complexity of the product of two matrices, (unpublished) manuscript, 1978.

[10] *T. Lickteig*, On semialgebraic decision complexity, Habilitationsschrift, to appear.

[11] *A. Schönhage*, Unitäre Transformationen grosser Matrizen, Num. Math. **20**(1973), 409–417.

[12] *V. Strassen*, Gaussian elimination is not optimal, Numer. Mathematik **13**(1969), 354–356.

[13] *V. Strassen*, Berechnung und Programm I, Acta Informatica **1**(1973), 320–335.

[14] *V. Strassen*, Berechnung und Programm II, Acta Informatica **2**(1973), 64–79.

[15] *V.Strassen*, Vermeidung von Divisionen, Crelles Journal für die reine und angewandte Mathematik **264**(1973), 184–202.

[16] *V. Strassen*, The complexity of continued fraction, SIAM J. Comp. **12/1**(1983), 1–27.

[17] *V. Strassen*, Relative bilinear complexity and matrix multiplication, J. für die reine und angewandte Mathematik **375/376**(1987), 406–443.

[18] *I. Wegener*, The complexity of Boolean functions, Wiley–Teubner, 1987.