Are there faster ways to multiply two matrices ?

Together with Edoardo

Trento

June 21-24, 2017

Giorgio Ottaviani

Università di Firenze

Edoardo's course in 1984 at Genova Nervi

- On Edoardo's web page
  A Brief Introduction to Algebraic Curves (1984 Nervi Lectures translated and updated by C. Fontanari).

Florence 2002

On the Lüroth Quartic Curve.

By Frank Morley.

It has been known since 1870 * that the problem of inscribing a five-line in a planar quartic is poristic; of the ten conditions nine fall on the lines and one on the curve. Thus the quartic is one for which an invariant vanishes, and the degree of this invariant is sought. We use Aronhold's construction of a curve of class 4 from seven given points. And the starting point is the theorem of Prof. Bateman † that the seven points which have the same polar line as to a conic and a cubic give rise to a Lüroth quartic.

For completeness I indicate the proof. A conic and a cubic have the canonical forms (αx^2), (βx^3) where $(x)=0$. The polars of x are (αxy), (βxy^2). Working in a space of three dimensions the line $(y)=0$, $(\alpha xy)=0$ is to touch the quadric (βxy^2). This requires that

$$\Sigma\beta_0\beta_1 x_0 x_1(\alpha_2 x_2 - \alpha_3 x_3)^2 = 0,$$

or

$$(\alpha/\beta)^2/(\alpha^2 x/\beta) = (1/\beta x),$$

and this is a quartic of Lüroth's type. The seven common polar lines are an Aronhold set of double lines of this quartic, and by polarity as to the conic the seven points a_i which have these polar lines are double points of a Lüroth curve of class 4.

§ 1. *The Bateman Conic.*

Take now a conic $(\alpha x)^2$ and a cubic $(\beta x)^3$. The Jacobian of these and a line (ξx)

$$(\alpha x)\,(\beta x)^2\,|\alpha\beta\xi| = 0$$

gives the net of cubics on the seven points a_i. Referred to one of the points and the corresponding line let the conic be $x_0^2 + 2x_1x_2$ and the cubic be

$$x_0^3 + x_0(\gamma x)_0^2 + (\delta x)_0^3.$$

Then for $(\xi x) \equiv x_0$ the Jacobian is

$$(\alpha_1\beta_2 - \alpha_2\beta_1)\,(\alpha x)\,(\beta x)^2 = \beta_2(\beta x)^2 x_2 - \beta_1(\beta x)^2 x_1,$$

so that not only terms in x_0^2 but also the term $x_0x_1x_2$ is missing.

A zoom from Frank Morley paper

On the Lüroth Quartic Curve.

By Frank Morley.

It has been known since 1870 * that the problem of inscribing a five-line in a planar quartic is poristic; of the ten conditions nine fall on the lines and one on the curve. Thus the quartic is one for which an invariant vanishes, and the degree of this invariant is sought. We use Aronhold's construction of a curve of class 4 from seven given points. And the starting point is the theorem of Prof. Bateman † that the seven points which have the same polar line as to a conic and a cubic give rise to a Lüroth quartic.

For completeness I indicate the proof. A conic and a cubic have the canonical forms (αx^2), (βx^3) where $(x)=0$. The polars of x are (αxy), (βxy^2). Working in a space of three dimensions the line $(y)=0$, $(\alpha xy)=0$ is to touch the quadric (βxy^2). This requires that

$$\Sigma\beta_0\beta_1 x_0 x_1(\alpha_2 x_2-\alpha_3 x_3)^2=0,$$

or

$$(\alpha/\beta)^2/(\alpha^2 x/\beta)=(1/\beta x),$$

and this is a quartic of Lüroth's type. The seven common polar lines are an Aronhold set of double lines of this quartic, and by polarity as to the conic the seven points a_i which have these polar lines are double points of a Lüroth

Michigan Math. J. 59 (2010), 365–394

On the Hypersurface of Lüroth Quartics

GIORGIO OTTAVIANI & EDOARDO SERNESI

Introduction

In his celebrated paper [18], Lüroth proved that a nonsingular quartic plane curve containing the ten vertices of a complete pentalateral contains infinitely many such 10-tuples. This implies that such curves, called *Lüroth quartics,* fill an open set of an irreducible, SL(3)-invariant hypersurface $\mathcal{L} \subset \mathbb{P}^{14}$. In his short paper [19], Morley computed the degree of the Lüroth hypersurface $\mathcal{L}$ by introducing some interesting ideas that seem to have been forgotten, maybe because a few arguments are somehow obscure. In this paper we put Morley's result and method on a solid foundation by reconstructing his proof as faithfully as possible. The main result is the following.

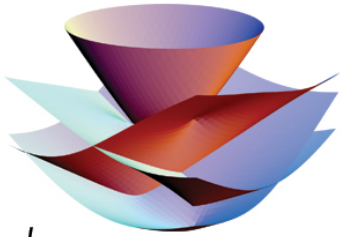
THEOREM 0.1. *The Lüroth hypersurface $\mathcal{L} \subset \mathbb{P}^{14}$ has degree 54.*

Morley's proof uses the description of plane quartics as branch curves of the degree-2 rational self-maps of $\mathbb{P}^2$ called *Geiser involutions.* Every such involution is determined by the linear system of cubics having as base locus a 7-tuple of distinct points $Z = \{P_1, \ldots, P_7\}$; let's denote by $B(Z) \subset \mathbb{P}^2$ the corresponding

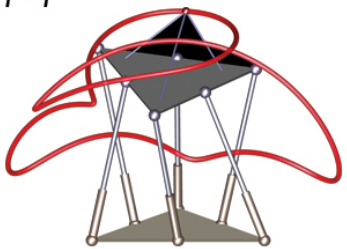
Plan of the talk

- Complexity of Matrix Multiplication and Tensor Rank
  History of recent work.
- The cubic polynomial $\mathrm{trace}(A^3)$ and its Waring rank
  How Algebraic Geometry can be useful to Complexity Theory.

Journal Advertisement

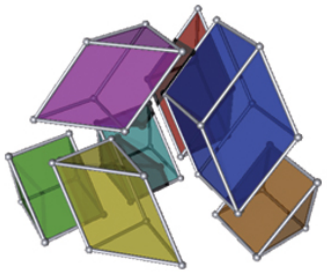


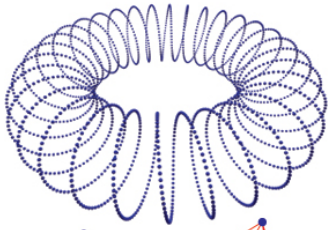
http://www.siam.org/journals/siaga.php

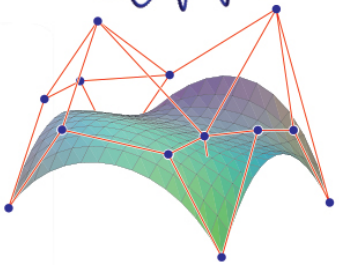


SIAM Journal on

Applied Algebra and Geometry







Relevance of matrix multiplication algorithm, and making it faster

Many numerical algorithms use matrix multiplication. The complexity of matrix multiplication algorithm is crucial in many numerical routines.

$$M_n = \text{space of } n \times n \text{ matrices}$$

Matrix multiplication is a bilinear map

$$\begin{aligned} M_n \times M_n &\to M_n \\ (A, B) &\mapsto A \cdot B \end{aligned}$$

where $A \cdot B = C$ is defined by $c_{ij} = \sum_k a_{ik} b_{kj}$.
This standard way to multiply two matrices requires n multiplications and $(n-1)$ additions for each entry, so $2n^3 - n^2$ ring operations.

Strassen result on 2×2 multiplication

Set

$$\begin{cases} I = (a_{11} + a_{22})(b_{11} + b_{22}) \\ II = (a_{21} + a_{22})b_{11} \\ III = a_{11}(b_{12} - b_{22}) \\ IV = a_{22}(-b_{11} + b_{21}) \\ V = (a_{11} + a_{12})b_{22} \\ VI = (-a_{11} + a_{21})(b_{11} + b_{12}) \\ VII = (a_{12} - a_{22})(b_{21} + b_{22}) \end{cases}$$

Then Strassen showed explicitly in 1969

$$\begin{cases} c_{11} = I + IV - V + VII \\ c_{12} = III + V \\ c_{21} = II + IV \\ c_{22} = I + III - II + VI \end{cases}$$

Notation $A \cdot B = C$.

Strassen result in tensor notation

Strassen result can be better understood by the following tensor identity

$$M_{\langle 2\rangle} = \mathrm{trace}(ABC) =$$

$$\underbrace{a_{11}\otimes b_{11}\otimes c_{11}}_{1} + \underbrace{a_{12}\otimes b_{21}\otimes c_{11}}_{2} + \underbrace{a_{21}\otimes b_{11}\otimes c_{21}}_{3} + \underbrace{a_{22}\otimes b_{21}\otimes c_{21}}_{4}$$

$$+ \underbrace{a_{11}\otimes b_{12}\otimes c_{12}}_{5} + \underbrace{a_{12}\otimes b_{22}\otimes c_{12}}_{6} + \underbrace{a_{21}\otimes b_{12}\otimes c_{22}}_{7} + \underbrace{a_{22}\otimes b_{22}\otimes c_{22}}_{8} =$$

$$\underbrace{(a_{11}+a_{22})\otimes(b_{11}+b_{22})\otimes(c_{11}+c_{22})}_{1}$$

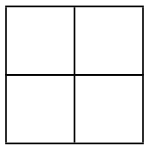
$$+ \underbrace{a_{11}\otimes(b_{12}-b_{22})\otimes(c_{12}+c_{22})}_{2} + \underbrace{(a_{21}+a_{22})\otimes b_{11}\otimes(c_{21}-c_{22})}_{3} + \underbrace{(a_{12}-a_{22})\otimes(b_{21}+b_{22})\otimes c_{11}}_{4}$$

$$+ \underbrace{a_{22}\otimes(-b_{11}+b_{21})\otimes(c_{21}+c_{11})}_{5} + \underbrace{(a_{11}+a_{12})\otimes b_{22}\otimes(-c_{11}+c_{12})}_{6} + \underbrace{(-a_{11}+a_{21})\otimes(b_{11}+b_{12})\otimes c_{22}}_{7}$$

There are elegant proofs of this identity by using group actions ([CILO 2016] Chiantini-Ikenmeyer-Landsberg-O). $M_{\langle n\rangle} = \mathrm{trace}(ABC)$ is called the matrix multiplication tensor.

Iteration of Strassen result, the Strassen algorithm

Dividing a matrix of size $2^k \times 2^k$ into 4 blocks of size $2^{k-1} \times 2^{k-1}$



one gets for
$T(n) := \#\{$ ring operations to multiply two $n \times n$ matrices$\}$

$$T(n) \leq 7T(n/2) + 18(n/2)^2,$$

together with $T(1) = 1$ one shows inductively that

$$T(n) \leq 7n^{\log_2 7} - 6n^2$$

where $\log_2 7 = 2.81\ldots$, which is cheaper than the standard algorithm for $n \geq 718$. Strassen algorithm is currently used for large matrices(roughly $n \geq 10^3$). The number 7 of multiplications needed turns out to be the crucial measure of the complexity.

The exponent of matrix multiplication

The exponent of matrix multiplication ω is defined to be

Definition

$\omega := \underline{\lim}_n \log_n (\#\{ \text{ ring operations to multiply two } n \times n \text{ matrices}\}) = \underline{\lim}_n \log_n (T(n))$

A consequence of Strassen algorithm is that $\omega \leq log_2 7 = 2.81\ldots$, while the standard algorithm gave $\omega \leq 3$.

The world record history for ω

ω=exponent of matrix multiplication of $n \times n$ matrices

- Strassen, $O(n^{2.81})$, 1969
- Bini, Capovani, Romani, Lotti, $O(n^{2.7799})$, 1979
- Strassen, $O(n^{2.48})$, 1987, Laser method
- Coppersmith, Vinogradov, $O(n^{2.375477})$, 1990
- Stothers, $O(n^{2.3736})$, 2010
- Williams, $O(n^{2.37287})$, 2011
- LeGall, $O(n^{2.37286})$, 2014

Basic question:

Compute ω. Is $\omega = 2$?

Perhaps you think that I should end my talk with a conjecture about ω. But this is dangerous. V. Strassen, 2010

How ω depends on the field

ω a priori depends on the field K.

Theorem (Schönhage)

ω depends only on the characteristic of the field.

Sketch of proof

By NullStellenSatz, ω is invariant by field extensions.

ω can be computed from tensor rank

A *tensor* $t \in U \otimes V \otimes W$ has *rank* r if $t = \sum_{i=1}^{r} u_i v_i w_i$ and r is minimal.
We write $\mathrm{rk}(t) = r$.
Let $End(V) = M_n$ be the vector space of $n \times n$ matrices over $\mathbb{C}$.
The matrix multiplication tensor $M_{\langle n \rangle} \in M_n^\vee \otimes M_n^\vee \otimes M_n^\vee$ is defined by $M_{\langle n \rangle}(A, B, C) = tr(ABC)$.

Theorem (Strassen)

$\omega = \limsup_n \left[\log_n \mathrm{rk}(M_{\langle n \rangle})\right]$

Excellent references are [Buergisser, Clausen, Shokrollai] and Landsberg notes.

Rank and complexity

Matrix multiplication can be seen as a tensor
$M_{\langle n \rangle} \in M_n \otimes M_n \otimes M_n$
$M_{\langle n \rangle}(A \otimes B \otimes C) = \sum_{i,j,k} a_{ik} b_{kj} c_{ji} = tr(ABC)$
and the number of multiplications needed coincides asymptotically with the rank of $M_{\langle n \rangle}$.

Allowing approximations, the border rank of t is a good measure of the complexity of matrix multiplication algorithm (Strassen, Bürgisser, Bini).
Border rank$(t) = \underline{\mathrm{rk}}(t) := \min\{r | \exists t_n \to t \text{ with } \mathrm{rk} t_n = r\}$
$\underline{\mathrm{rk}}(t) \leq \mathrm{rk}(t)$, for d-way tensors with $d \geq 3$ there are examples where strict inequality holds.

Geometry in tensor space

In the tensor space $A \otimes B \otimes C$ there is the Segre variety of decomposable (rank 1) tensors $X = \mathbb{P}(A) \times \mathbb{P}(B) \times \mathbb{P}(C)$.
Tensors of rank 2 like $a_0 b_0 c_0 + a_1 b_1 c_1$ lie in the line joining $a_0 b_0 c_0$ and $a_1 b_1 c_1$.
Tensors of rank k lie in the span of k points on the Segre variety.

The k-secant variety $\sigma_k(X)$ is the Zariski closure

$$\sigma_k(X) = \overline{\bigcup_{x_1,\dots x_k \in X} \langle x_1, \dots, x_k \rangle}$$

The seven summands in Strassen identity

The seven summands in Strassen tensor identity can be understood by a naive parameter count.
Expected dimension of $\sigma_k X \subset \mathbb{P}^M$ is $\min\left(k(\dim X + 1) - 1, M\right)$.

In the case of
$X = \mathbb{P}(M_2) \times \mathbb{P}(M_2) \times \mathbb{P}(M_2) \subset \mathbb{P}(M_2 \otimes M_2 \otimes M_2) = \mathbb{P}^{63}$ we have $\dim X = 9$ and indeed

$$dim\sigma_6 X = \min(59, 63) = 59 \qquad \dim \sigma_7 X = \min(69, 63) = 63.$$

By Terracini Lemma it is easy to compute that expected dimensions are actually attained.

rank of 2×2 multiplication tensor

Theorem

Rank and border rank of 2×2 *multiplication tensor are both* 7.

Theoretical proof by Landsberg (2006) with representation theory techniques.
Computational proof by Hauenstein, Ikenmeyer, Landsberg (2013).
In this case the rank of general tensor of the same size is again 7.

In the 3×3 case, the rank and the border rank of the matrix multiplication tensor are not yet known.

Bounds on border rank

Theorem (O-Landsberg 2012)

$$brk(M_{<n>}) \geq 2n^2 - n.$$

Theorem (Landsberg-Michałek 2016)

$$brk(M_{<n>}) \geq 2n^2 - \log_2 n + 1.$$

For 3×3 matrices, the state of the art is $16 \leq \mathrm{brk}(M_{\langle 3 \rangle}) \leq 21$, the upper bound is due to Schönhage .

Basic book references

Bürgisser, Clausen, Shokrollai, *Algebraic Complexity Theory*, **Springer, 1997**

J.M. Landsberg, *Tensors, Geometry and Applications*, **AMS, 2012**

J.M. Landsberg, *Geometry and Complexity Theory*, **Cambridge, 2017**

Not uniqueness

Matrix multiplication tensor is quite special.
One cannot expect a unique honest decomposition.
Indeed it is invariant by a big isotropy group, because

$$tr(ABC) = tr\left((G^{-1}AH)(H^{-1}BK)(K^{-1}CG)\right)$$

Symmetric tensors and symmetric rank

A cubic *symmetric tensor* is $t \in Sym^3 M_n$. It is a homogeneous cubic polynomial. t has *symmetric rank* r if $t = \sum_{i=1}^{r} l_i^3$ and r is minimal.
We write $\mathrm{symrk}(t) = r$.

In the case of symmetric rank, the Segre variety X is replaced by the Veronese variety.
Example $\mathrm{symrk}(2x_0^3 + 6x_0x_1^2)$
Comon conjecture claims that for symmetric tensors, the rank equals the symmetric rank.

Comon Conjecture

If t is a symmetric tensor, then

$$\mathrm{rk}(t) = \mathrm{symrk}(t)$$

Shitov announced a counterexample, arXiv May 24, 2017.

Symmetric version of matrix multiplication tensor.

Symmetrize $M_{\langle n \rangle}$ as $sM_{\langle n \rangle} \in \mathrm{Sym}^3 M_n^\vee$ defined by $sM_{\langle n \rangle}(A) = tr(A^3)$. It is a cubic polynomial in n^2 indeterminates.

Theorem (CHILO 2017)

(Chiantini-Hauenstein-Ikenmeyer-Landsberg-O)[Asymptotically, the symmetric version works as well.]

$$\omega = \limsup_n \left[\log_n \mathrm{symrk}(sM_{\langle n \rangle})\right]$$

$$\omega = \limsup_n \left[\log_n \mathrm{bsymrk}(sM_{\langle n \rangle})\right]$$

Sketch of proof

Sketch of Proof

$\geq$ is easy

$\leq$ For $n \times n$ matrices A, B, C consider the $3n \times 3n$ matrix $X = \begin{pmatrix} 0 & 0 & A \\ C & 0 & 0 \\ 0 & B & 0 \end{pmatrix}$. Then, $X^3 = \begin{pmatrix} ABC & 0 & 0 \\ 0 & CAB & 0 \\ 0 & 0 & BCA \end{pmatrix}$ and $\mathrm{tr}(X^3) = 3\mathrm{tr}(ABC)$.

It follows $\mathrm{rk}(M_{\langle n \rangle}) \leq \mathrm{rk}(sM_{\langle 3n \rangle})$, hence the inequality $\leq$.

Examples

For $n = 2$,

$$sM_{\langle 2\rangle} = \mathrm{tr}(A^3) = a_{0,0}^3 + 3a_{0,0}a_{0,1}a_{1,0} + 3a_{0,1}a_{1,0}a_{1,1} + a_{1,1}^3$$
$$= \underbrace{[\mathrm{trace}(A)]}_{\text{non tg hyperp.}} \cdot \underbrace{\left[\mathrm{trace}^2(A) - 3\det(A)\right]}_{\text{smooth quadric}}$$

has $\mathrm{rk}(sM_{\langle 2\rangle}) = 6$, $\mathrm{bordrk}(sM_{\langle 2\rangle}) = 5$ (B. Segre).

For $n = 3$,

$$sM_{\langle 3\rangle} = \mathrm{tr}(A^3) = a_{0,0}^3 + 3a_{0,0}a_{0,1}a_{1,0} + 3a_{0,1}a_{1,0}a_{1,1} + a_{1,1}^3 + 3a_{0,0}a_{0,2}a_{2,0} +$$
$$3a_{0,1}a_{1,2}a_{2,0} + 3a_{0,2}a_{1,0}a_{2,1} + 3a_{1,1}a_{1,2}a_{2,1} + 3a_{0,2}a_{2,0}a_{2,2} + 3a_{1,2}a_{2,1}a_{2,2} + a_{2,2}^3$$

is irreducible, $\mathrm{rk}(sM_{\langle 3\rangle}) \leq 18$, found numerically with *Bertini*.

Theorem (B. Segre)

$\mathrm{brk}(sM_{\langle 2\rangle}) = 5, \mathrm{rk}(sM_{\langle 2\rangle}) = 6$

A minimal Waring decomposition is given by $8sM_{\langle 2\rangle} = 8trace(A^3) = \sum_{i=1}^{6}(trace(A \cdot L_i))^3$ with

$$L_1 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \; L_2 = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \; L_3 = \begin{pmatrix} -1 & 1 \\ -1 & -1 \end{pmatrix} \; L_4 = \begin{pmatrix} -1 & -1 \\ 1 & -1 \end{pmatrix}$$

$$L_5 = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \; L_6 = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}$$

Sylvester Pentahedral in 2×2 case

2×2 case is instructive because general cubic in 4 variables has a UNIQUE Waring decomposition with 5 summands, by *Sylvester Pentahedral Theorem*.

Consider a family $sM_{\langle 3,\epsilon \rangle}$ such that $sM_{\langle 3,0 \rangle} = sM_{\langle 3 \rangle}$. For $\epsilon \neq 0$ we have five hyperplanes. When $\epsilon \to 0$, the five hyperplanes converge to the same hyperplane $a_{00} + a_{11} = 0$, corresponding to the identity matrix.

Group action

Appearance of identity matrix is not suprising because $SL(n)$ acts by conjugation
$(G, A) \mapsto G^{-1}AG$ and $sM_{\langle n \rangle} = \mathrm{tr}(A^3)$ is $SL(n)$-invariant polynomial.

The identity is the unique zero dimensional orbit in $\mathbb{P}(M_n)$.

3×3 case

Theorem (CHILO, found numerically by Jon Hauenstein)

- *There is a Waring decomposition of* $sM_{\langle 3 \rangle}$ *with* 19 *summands, found numerically, such that* 15 *of them have rank* 2*, the remaining* 4 *of them are traceless. So* $\mathrm{rk}(sM_{\langle 3 \rangle}) \leq 19$.
- *There is a Waring decomposition of* $sM_{\langle 3 \rangle}$ *with* 18 *summands, found numerically, all* 18 *summands have rank* 3.

Question

Is $\mathrm{rk}(sM_{\langle 3 \rangle}) = 18$?

Apolarity Lemma

$M_n = End(V)$ has coordinates a_{ij},

we have the *ring* $\mathrm{Sym}^* M_n^\vee = \mathbb{C}[a_{0,0}, \ldots, a_{n-1,n-1}]$,

and the *dual ring* $\mathrm{Sym}^* M_n = \mathbb{C}[\frac{\partial}{\partial a_{0,0}}, \ldots, \frac{\partial}{\partial a_{n-1,n-1}}]$.

Given $f \in \mathrm{Sym}^* M_n^\vee$, the *apolar ideal* is

$$f^\perp := \{D \in \mathrm{Sym}^* M_n | D \cdot f = 0\}.$$

Lemma (Apolarity Lemma)

$$\begin{cases} f = \sum_i^r l_i^d \\ with\ Z = \{l_1, \ldots, l_r\} \end{cases} \iff I_Z \subset f^\perp$$

The role of A^2

Theorem

Singular locus of hypersurface $sM_{\langle n \rangle}(A) = trace(A^3)$ *is given by* $\{A | A^2 = 0\}$.

Proof

$(A, B) \mapsto \mathrm{tr}(AB^t)$ is a nondegenerate pairing. $\mathrm{tr}A^3 = \mathrm{tr}(A \cdot A^2)$. The coefficients of A^2 are the partial derivatives of $sM_{\langle n \rangle}$. □

Basic question reformulated

Basic question reformulated:

How grows the Waring rank of the cubic polynomial $sM_{\langle n \rangle} = \mathrm{trace}(A^3)$ with n, where A is a $n \times n$ matrix ?

Thanks Edoardo.

Thanks !!