tensors in computational mathematics

Lek-Heng Lim

acknowledgments

- kind colleagues who nominated/supported me:
  - ◇ Shmuel Friedland
  - ◇ Pierre Comon
  - ◇ Ming Gu
  - ◇ Jean Bernard Lasserre
  - ◇ Sayan Mukherjee
  - ◇ Jiawang Nie
  - ◇ Bernd Sturmfels
  - ◇ Charles Van Loan
- postdocs: Ke Ye, Shenglong Hu, Yang Qi, Jose Rodriguez, Anne Shiu
- students: Liwen Zhang, Ken Wong, Greg Naitzat, Kate Turner, many MS students as well
- brilliant collaborators: too many to list
- funding agencies: AFOSR, DARPA (special thanks to Fariba Fahroo), NSF

thank you all from the bottom of my heart

what is a tensor?

what is a tensor?

- for every complex question there is an answer that is clear, simple, and wrong
- clear, simple, and wrong answer:

  "a tensor is a multiway array"

- unfortunately also widely believed — simple answer to complex question has its appeal

what is a tensor?

- indication that answer cannot be so simple: Einstein's letter to Sommerfeld, dated October 29, 1912
- J. Earman, C. Glymour, "Lost in tensors: Einstein's struggles with covariance principles 1912–1916," *Stud. Hist. Phil. Sci.*, **9** (1978), no. 4, pp. 251–278
- fortunately the last century of progress in algebra, geometry, and physics has made tensors a lot easier to explain and understand

why we should get it right

- each example below contains an order-3 tensor
- example 1: multiplication of complex numbers

$$(a + bi)(c + di) = (ac - bd) + i(bc + ad)$$

- example 2: matrix product

$$\begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} \begin{bmatrix} b_1 & b_2 \\ b_3 & b_4 \end{bmatrix} = \begin{bmatrix} a_1 b_1 + a_2 b_3 & a_1 b_2 + a_2 b_4 \\ a_3 b_1 + a_4 b_3 & a_3 b_2 + a_4 b_4 \end{bmatrix}$$

- example 3: Grothendieck inequality

$$\max_{\mathbf{x}_1,\dots,\mathbf{x}_m,\mathbf{y}_1,\dots,\mathbf{y}_n \in \mathbb{S}^{m+n-1}} \sum\nolimits_{i=1}^{m} \sum\nolimits_{j=1}^{n} a_{ij} \langle \mathbf{x}_i, \mathbf{y}_j \rangle \\ \le K_G \max_{\varepsilon_1,\dots,\varepsilon_m,\delta_1,\dots,\delta_n \in \{-1,+1\}} \sum\nolimits_{i=1}^{m} \sum\nolimits_{j=1}^{n} a_{ij} \varepsilon_i \delta_j$$

- no 3-way array anywhere

start from the familiar

- tensors of order 0
  - scalars:

    elements of a field: $\lambda \in \mathbb{R}, \ \mathbb{C}$, etc

  more generally a commutative ring: $\mathbb{Z}$, $C^\infty(M)$, etc
- tensors of order 1
  - vectors:

    elements of a vector space: $\mathbf{v} \in \mathbb{V}$
  - covectors:

    elements of a dual vector space: $\mathbf{v}^* \in \mathbb{V}^*$,

    i.e., $\mathbf{v}^* : \mathbb{V} \to \mathbb{C}$ is a linear functional

  more generally a module

start from the familiar

- tensors of order 2
  - linear operators:

$$\varphi : \mathbb{U} \to \mathbb{V}$$

i.e.,

$$\varphi(\lambda_1 \mathbf{u}_1 + \lambda_2 \mathbf{u}_2) = \lambda_1 \varphi(\mathbf{u}_1) + \lambda_2 \varphi(\mathbf{u}_2)$$

  - bilinear functionals:

$$\beta : \mathbb{U} \times \mathbb{V} \to \mathbb{C}$$

i.e.,

$$\beta(\lambda_1 \mathbf{u}_1 + \lambda_2 \mathbf{u}_2, \mathbf{v}) = \lambda_1 \beta(\mathbf{u}_1, \mathbf{v}) + \lambda_2 \beta(\mathbf{u}_2, \mathbf{v}),$$
$$\beta(\mathbf{u}, \lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2) = \lambda_1 \beta(\mathbf{u}, \mathbf{v}_1) + \lambda_2 \beta(\mathbf{u}, \mathbf{v}_2)$$

  - other possibilities: linear operators

$$\varphi : \mathbb{U}^* \to \mathbb{V}, \quad \varphi : \mathbb{U} \to \mathbb{V}^*, \quad \varphi : \mathbb{U}^* \to \mathbb{V}^*$$

and bilinear functionals

$$\beta : \mathbb{U}^* \times \mathbb{V} \to \mathbb{C}, \quad \beta : \mathbb{U} \times \mathbb{V}^* \to \mathbb{C}, \quad \beta : \mathbb{U}^* \times \mathbb{V}^* \to \mathbb{C}$$

first (and only) unfamiliar case

- tensors of order 3
  - bilinear operators:
$$\beta : \mathbb{U} \times \mathbb{V} \to \mathbb{W}$$
i.e.,
$$\begin{aligned}\beta(\lambda_1 \mathbf{u}_1 + \lambda_2 \mathbf{u}_2, \mathbf{v}) &= \lambda_1 \beta(\mathbf{u}_1, \mathbf{v}) + \lambda_2 \beta(\mathbf{u}_2, \mathbf{v}),\\ \beta(\mathbf{u}, \lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2) &= \lambda_1 \beta(\mathbf{u}, \mathbf{v}_1) + \lambda_2 \beta(\mathbf{u}, \mathbf{v}_2)\end{aligned}$$
  - trilinear functionals:
$$\tau : \mathbb{U} \times \mathbb{V} \times \mathbb{W} \to \mathbb{C}$$
i.e.,
$$\begin{aligned}\tau(\lambda_1 \mathbf{u}_1 + \lambda_2 \mathbf{u}_2, \mathbf{v}, \mathbf{w}) &= \lambda_1 \tau(\mathbf{u}_1, \mathbf{v}, \mathbf{w}) + \lambda_2 \tau(\mathbf{u}_2, \mathbf{v}, \mathbf{w}),\\ \tau(\mathbf{u}, \lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2, \mathbf{w}) &= \lambda_1 \tau(\mathbf{u}, \mathbf{v}_1, \mathbf{w}) + \lambda_2 \tau(\mathbf{u}, \mathbf{v}_2, \mathbf{w}),\\ \tau(\mathbf{u}, \mathbf{v}, \lambda_1 \mathbf{w}_1 + \lambda_2 \mathbf{w}_2) &= \lambda_1 \tau(\mathbf{u}, \mathbf{v}, \mathbf{w}_1) + \lambda_2 \tau(\mathbf{u}, \mathbf{v}, \mathbf{w}_2)\end{aligned}$$

first (and only) unfamiliar case

- tensors of order 3
  - other possibilities: bilinear operators

$$\beta : \mathbb{U}^* \times \mathbb{V} \to \mathbb{W},\ \beta : \mathbb{U} \times \mathbb{V}^* \to \mathbb{W}, \dots, \beta : \mathbb{U}^* \times \mathbb{V}^* \to \mathbb{W}^*$$

  and trilinear functionals

$$\tau : \mathbb{U}^* \times \mathbb{V} \times \mathbb{W} \to \mathbb{C},\ \tau : \mathbb{U} \times \mathbb{V}^* \times \mathbb{W} \to \mathbb{C}, \dots, \tau : \mathbb{U}^* \times \mathbb{V}^* \times \mathbb{W}^* \to \mathbb{C}$$

- but they are all the same up to covariance and contravariance
- notation:

$$\mathbb{U} \otimes \mathbb{V} = \{\varphi : \mathbb{U} \to \mathbb{V} \text{ linear}\}$$
$$\mathbb{U} \otimes \mathbb{V} \otimes \mathbb{W} = \{\beta : \mathbb{U} \times \mathbb{V} \to \mathbb{W} \text{ bilinear}\}$$
$$\cdots$$
$$\mathbb{V}_1 \otimes \mathbb{V}_2 \otimes \cdots \otimes \mathbb{V}_d = \{\tau : \mathbb{V}_1 \times \cdots \times \mathbb{V}_{d-1} \to \mathbb{V}_d \text{ multilinear}\}$$

  elements called tensors of order d or d-tensors

bases and coordinates

- recall: vector spaces have bases
- recall: whenever we choose a basis, we get coordinates
- choose bases

$$\mathbf{u}_1, \dots, \mathbf{u}_m \text{ of } \mathbb{U}, \quad \mathbf{v}_1, \dots, \mathbf{v}_n \text{ of } \mathbb{V}, \quad \mathbf{w}_1, \dots, \mathbf{w}_p \text{ of } \mathbb{W}$$

- order 1: any $\mathbf{u} \in \mathbb{U}$ representable as

$$\mathbf{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix} \in \mathbb{C}^m$$

  where $\mathbf{u} = \sum_{i=1}^m a_i \mathbf{u}_i$
- order 2: ; any linear $\varphi : \mathbb{U} \to \mathbb{V}$ representable as

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \in \mathbb{C}^{m \times n}$$

  where $\varphi(\mathbf{u}_i) = \sum_{j=1}^n a_{ij} \mathbf{v}_j$

bases and coordinates

- order 3: any bilinear $\beta : \mathbb{U} \times \mathbb{V} \to \mathbb{W}$ representable as

$$A = \left[\begin{array}{ccc|ccc|c|ccc} a_{111} & \cdots & a_{1n1} & a_{112} & \cdots & a_{1n2} & & a_{11p} & \cdots & a_{1np} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \ddots & \vdots \\ a_{m11} & \cdots & a_{mn1} & a_{m12} & \cdots & a_{mn2} & & a_{m1p} & \cdots & a_{mnp} \end{array} \right] \in \mathbb{C}^{m \times n \times p}$$

  where $\beta(\mathbf{u}_i, \mathbf{v}_j) = \sum_{k=1}^{p} a_{ijk} \mathbf{w}_k$
- d-tensors representable as d-dimensional hypermatrices
- doesn't this mean that "tensors are multiway arrays"?
- no on multiple levels:
  - "representable" is far from "identical to"
  - hypermatrices are not the same as multiway arrays
  - not true if $\mathbb{U}$, $\mathbb{V}$, $\mathbb{W}$ are not free modules, i.e., no bases

tensor: first appearance of the word

Woldemar Voigt, *Die fundamentalen physikalischen Eigenschaften der Krystalle in elementarer Darstellung*, Verlag Von Veit, Leipzig, 1898.

"An abstract entity represented by an array of components that are functions of coordinates such that, under a transformation of cooordinates, the new components are related to the transformation and to the original components in a definite way."

in modern language

- hypermatrix represents tensor only if it satisfies change-of-basis rule
- choose new bases

$$\mathbf{u}_1', \dots, \mathbf{u}_m' \text{ of } \mathbb{U}, \quad \mathbf{v}_1', \dots, \mathbf{v}_n' \text{ of } \mathbb{V}, \quad \mathbf{w}_1', \dots, \mathbf{w}_p' \text{ of } \mathbb{W}$$

  and let X, Y, Z be corresponding change-of-basis matrices
- if $\mathbf{a} \in \mathbb{C}^m$ represents $\mathbf{u} \in \mathbb{U}$ with respect to old basis and $\mathbf{a}' \in \mathbb{C}^m$ represents it with respect to new basis, then

$$\mathbf{a}' = X^{-1}\mathbf{a}$$

- if $A \in \mathbb{C}^{m \times n}$ represents $\varphi : \mathbb{U} \to \mathbb{V}$ with respect to old basis and $A' \in \mathbb{C}^{m \times n}$ represents it with respect to new basis, then

$$A' = XAY^{-1}$$

- if $A \in \mathbb{C}^{m \times n \times p}$ represents $\beta : \mathbb{U} \times \mathbb{V} \to \mathbb{W}$ with respect to old basis and $A' \in \mathbb{C}^{m \times n \times p}$ represents it with respect to new basis, then

$$A' = (X, Y, Z^{-1}) \cdot A$$

main point

for a quantity to be defined on tensors, it must coordinate-independent, i.e., does not depend on a choice of bases

- invariance or equivariance under $\mathrm{GL}(\mathbb{V}_1) \times \cdots \times \mathrm{GL}(\mathbb{V}_d)$ on $\mathbb{V}_1 \otimes \cdots \otimes \mathbb{V}_d$: e.g. $A \in \mathbb{C}^{m \times n}$,

$$\operatorname{rank}(XAY^{-1}) = \operatorname{rank}(A) \quad \text{for all } (X, Y) \in \mathrm{GL}_m(\mathbb{C}) \times \mathrm{GL}_n(\mathbb{C})$$

- invariance or equivariance under action of $\mathrm{GL}(\mathbb{V})$ on $\mathbb{V}^{\otimes d}$: e.g. $A, B \in \mathbb{C}^{n \times n}$,

$$XABX^{-1} = (XAX^{-1})(XBX^{-1}) \quad \text{for all } X \in \mathrm{GL}_n(\mathbb{C})$$

- or invariant under more restrictive changes of bases

$$\det(XAX^{-1}) = \det(A) \quad \text{for all } X \in \mathrm{SL}_n(\mathbb{C})$$
$$\|XAY^*\|_* = \|A\|_* \quad \text{for all } (X, Y) \in \mathrm{U}_m(\mathbb{C}) \times \mathrm{U}_n(\mathbb{C})$$

example

- many recent proposals for "multiplying higher-order tensors" — do they make sense?
- start with the simplest case

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \times \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^n a_{1j} b_{j1} & \sum_{j=1}^n a_{1j} b_{j2} & \cdots & \sum_{j=1}^n a_{1j} b_{jn} \\ \sum_{j=1}^n a_{2j} b_{j1} & \sum_{j=1}^n a_{nj} b_{j2} & \cdots & \sum_{j=1}^n a_{2j} b_{jn} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{j=1}^n a_{nj} b_{j1} & \sum_{j=1}^n a_{nj} b_{j2} & \cdots & \sum_{j=1}^n a_{nj} b_{jn} \end{bmatrix}$$

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \circ \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{bmatrix} = \begin{bmatrix} a_{11} b_{11} & a_{12} b_{12} & \cdots & a_{1n} b_{1n} \\ a_{21} b_{21} & a_{22} b_{22} & \cdots & a_{2n} b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} b_{n1} & a_{n2} b_{n2} & \cdots & a_{nn} b_{nn} \end{bmatrix}$$

- usual product $\times$ defines a product of 2-tensors
- Hadamard product $\circ$ only defines a product of matrices
- nothing to do with ring structure: $(\mathbb{C}^{n \times n}, +, \times)$, $(\mathbb{C}^{n \times n}, +, \circ)$ both rings

tensors in computational math

which bases should we choose?

- order 1: given $\mathbf{u} \in \mathbb{U}$, choose basis so that $\mathbf{u}$ is represented by

$$\lambda \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

  get one-dimensional problem depending on $\lambda \in \mathbb{C}$
- order 2: given $\varphi : \mathbb{U} \to \mathbb{V}$, choose bases so that φ is represented by

$$\begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & \mathbf{0} \end{bmatrix}$$

  get r-dimensional problem depending on $\boldsymbol{\sigma} \in \mathbb{C}^r$ where $r = \operatorname{rank}(\varphi)$
- e.g., best bases could be given by left and and right singular vectors

which bases to choose?

- best bases depend on the tensor in your problem
- depend on your problem too: may want smallest r so that

$$A = \sum_{i=1}^{r} \sigma_i \mathbf{u} \otimes \mathbf{v}$$

  i.e., rank of φ
- or may want smallest $\sigma_1 + \cdots + \sigma_r$, i.e., nuclear norm of A
- note that rank (invariant under general linear transformation) and nuclear norm (invariant under unitary transformations) defined on 2-tensors
- extends to higher-order tensors

rank, decomposition, nuclear norm

- goal: compute bilinear operation

$$\beta : \mathbb{U} \times \mathbb{V} \to \mathbb{W}$$

- tensor rank [Hitchcock, 1927]

$$\operatorname{rank}(\beta) = \min \left\{ r : \beta = \sum\nolimits_{i=1}^{r} \lambda_i \mathbf{u}_i \otimes \mathbf{v}_i \otimes \mathbf{w}_i \right\}$$

gives least number of multiplications needed to compute β

- tensor decomposition

$$\beta = \sum\nolimits_{i=1}^{r} \lambda_i \mathbf{u}_i \otimes \mathbf{v}_i \otimes \mathbf{w}_i$$

gives an explicit algorithm for computing β

- tensor nuclear norm [LHL–Comon, 2010; Derksen, 2016]

$$\|\beta\|_* = \inf \left\{ \sum\nolimits_{i=1}^{r} |\lambda_i| : \beta = \sum\nolimits_{i=1}^{r} \lambda_i \mathbf{u}_i \otimes \mathbf{v}_i \otimes \mathbf{w}_i,\ r \in \mathbb{N} \right\}$$

quantifies optimal numerical stability of computing β

fast(est) algorithms

- bilinear complexity: counts only multiplication of variables, ignores addition, subtraction, scalar multiplication
- Gauss's method

$$\begin{aligned}(a+bi)(c+di) &= (ac-bd)+i(bc+ad)\\ &= (ac-bd)+i[(a+b)(c+d)-ac-bd]\end{aligned}$$

- usual: 4 $\times$'s and 2 $\pm$'s; Gauss: 3 $\times$'s and 5 $\pm$'s
- Strassen's algorithm

$$\begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix}\begin{bmatrix} b_1 & b_2 \\ b_3 & b_4 \end{bmatrix} = \begin{bmatrix} a_1b_1 + a_2b_2 & \beta+\gamma+(a_1+a_2-a_3-a_4)b_4 \\ \alpha+\gamma+a_4(b_2+b_3-b_1-b_4) & \alpha+\beta+\gamma \end{bmatrix}$$

where

$$\alpha = (a_3-a_1)(b_3-b_4),\ \beta=(a_3+a_4)(b_3-b_1),\ \gamma=a_1b_1+(a_3+a_4-a_1)(b_1+b_4-b_3)$$

- usual: 8 $\times$'s and 8 $\pm$'s; Strassen: 7 $\times$'s and 15 $\pm$'s

complexity of Gauss's method

- $\beta : \mathbb{C} \times \mathbb{C} \to \mathbb{C}$, $(z, w) \mapsto zw$ is $\mathbb{R}$-bilinear map
- $\beta \in \mathbb{R}^2 \otimes \mathbb{R}^2 \otimes \mathbb{R}^2$ is a tensor
- choose basis $\mathbf{e}_1 = (1, 0)$, $\mathbf{e}_2 = (0, 1) \in \mathbb{R}^2$, get hypermatrix

$$\beta = \left[\begin{array}{cc|cc} 1 & 0 & 0 & 1 \\ 0 & -1 & 1 & 0 \end{array}\right] \in \mathbb{R}^{2\times 2\times 2}$$

- usual multiplication

$$\beta = (\mathbf{e}_1 \otimes \mathbf{e}_1 - \mathbf{e}_2 \otimes \mathbf{e}_2) \otimes \mathbf{e}_1 + (\mathbf{e}_1 \otimes \mathbf{e}_2 + \mathbf{e}_2 \otimes \mathbf{e}_1) \otimes \mathbf{e}_2$$

- Gauss multiplication

$$\beta = (\mathbf{e}_1 + \mathbf{e}_2) \otimes (\mathbf{e}_1 + \mathbf{e}_2) \otimes \mathbf{e}_2 \\ + \mathbf{e}_1 \otimes \mathbf{e}_1 \otimes (\mathbf{e}_1 - \mathbf{e}_2) - \mathbf{e}_2 \otimes \mathbf{e}_2 \otimes (\mathbf{e}_1 + \mathbf{e}_2)$$

- $\operatorname{rank}(\beta) = 3 = \overline{\operatorname{rank}}(\beta)$ [De Silva–LHL, 2008]

stability of Gauss's method

- nuclear norm

$$\|\beta\|_* = 4$$

- attained by usual multiplication

$$\beta = (\mathbf{e}_1 \otimes \mathbf{e}_1 - \mathbf{e}_2 \otimes \mathbf{e}_2) \otimes \mathbf{e}_1 + (\mathbf{e}_1 \otimes \mathbf{e}_2 + \mathbf{e}_2 \otimes \mathbf{e}_1) \otimes \mathbf{e}_2$$

- but not Gauss multiplication

$$\beta = (\mathbf{e}_1 + \mathbf{e}_2) \otimes (\mathbf{e}_1 + \mathbf{e}_2) \otimes \mathbf{e}_2 + \mathbf{e}_1 \otimes \mathbf{e}_1 \otimes (\mathbf{e}_1 - \mathbf{e}_2) - \mathbf{e}_2 \otimes \mathbf{e}_2 \otimes (\mathbf{e}_1 + \mathbf{e}_2)$$

  coefficients (upon normalizing) sums to $2(1 + \sqrt{2})$

- Gauss's algorithm less stable than the usual algorithm
- optimal bilinear complexity and stability:

$$\beta = \frac{4}{3}\left(\left[\frac{\sqrt{3}}{2}\mathbf{e}_1 + \frac{1}{2}\mathbf{e}_2\right]^{\otimes 3} + \left[-\frac{\sqrt{3}}{2}\mathbf{e}_1 + \frac{1}{2}\mathbf{e}_2\right]^{\otimes 3} + (-\mathbf{e}_2)^{\otimes 3}\right)$$

  attains both $\operatorname{rank}(\beta)$ and $\|\beta\|_*$ [Friedland–LHL, 2016]

matrix multiplication tensor

$$\begin{bmatrix} c_1 & c_2 \\ c_3 & c_4 \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} \begin{bmatrix} b_1 & b_2 \\ b_3 & b_4 \end{bmatrix}$$

- write
$$c_k = \sum_{i=1}^{n^2} \sum_{j=1}^{n^2} \mu_{ijk} a_i b_j$$
- suppose
$$\mu_{ijk} = \sum_{p=1}^{r} u_{ip} v_{jp} w_{kp}$$
  then
$$c_k = \sum_{p=1}^{r} w_{kp} \Bigl(\sum_{i=1}^{n^2} u_{ip} a_i\Bigr) \Bigl(\sum_{j=1}^{n^2} v_{jp} b_j\Bigr)$$
- more generally, hypermatrix $\mu_{m,n,p} = (\mu_{ijk}) \in \mathbb{C}^{mn \times np \times mp}$ represents matrix multiplication tensor
$$\mu_{m,n,p} : \mathbb{C}^{m \times n} \times \mathbb{C}^{n \times p} \to \mathbb{C}^{m \times p}$$

complexity = tensor rank

- write $\mu_n = \mu_{n,n,n}$, i.e., matrix multiplication tensor for square matrices
- number of multiplications given by $\operatorname{rank}(\mu_n)$
- asymptotic growth
  - usual: $O(n^3)$
  - earliest: $O(n^{\log_2 7})$ [Strassen, 1969]
  - longest: $O(n^{2.375477})$ [Coppersmith–Winograd, 1990]
  - recent: $O(n^{2.3728642})$ [Williams, 2011]
  - latest: $O(n^{2.3728639})$ [Le Gall, 2014]
  - exact: $O(n^{\omega})$ where

$$\omega := \inf\{\alpha : \operatorname{rank}(\mu_n) = O(n^{\alpha})\}$$

self-concordance

- convex $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ self-concordant at $\mathbf{x} \in \Omega$ if

$$\left[\nabla^3 f(\mathbf{x})(\mathbf{h},\mathbf{h},\mathbf{h})\right]^2 \le 4\sigma\left[\nabla^2 f(\mathbf{x})(\mathbf{h},\mathbf{h})\right]^3$$

  for all $\mathbf{h} \in \mathbb{R}^n$ [Nesterov–Nemirovskii, 1994]

$$\nabla^2 f(\mathbf{x})(\mathbf{h},\mathbf{h}) = \sum_{i,j=1}^{n} \frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j} h_i h_j, \quad \nabla^3 f(\mathbf{x})(\mathbf{h},\mathbf{h},\mathbf{h}) = \sum_{i,j,k=1}^{n} \frac{\partial^3 f(\mathbf{x})}{\partial x_i \partial x_j \partial x_k} h_i h_j h_k$$

- convex programming problem may be solved to arbitrary ε-accuracy in polynomial time if it has self-concordant barrier functions (e.g. LP, QP, SOCP, SDP, GP)
- affine invariance of self-concordance implies that it is a property defined on the tensors $\nabla^2 f(\mathbf{x})$ and $\nabla^3 f(\mathbf{x})$

Grothendieck inequality

- $A \in \mathbb{R}^{m \times n}$, there exists $K_G > 0$ such that

$$\max_{\mathbf{x}_1,\dots,\mathbf{x}_m,\mathbf{y}_1,\dots,\mathbf{y}_n \in \mathbb{S}^{m+n-1}} \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}\langle \mathbf{x}_i, \mathbf{y}_j\rangle \le K_G \max_{\varepsilon_1,\dots,\varepsilon_m,\delta_1,\dots,\delta_n \in \{-1,+1\}} \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}\varepsilon_i\delta_j.$$

- remarkable: K_G independent of m and n [Grothendieck, 1953]
- important: unique games conjecture and SDP relaxations of NP-hard problems
- best known bounds: $1.676 \le K_G \le 1.782$
- Grothendieck's constant is injective norm of matrix multiplication tensor [LHL, 2016]

$$\|\mu_{m,n,m+n}\|_{1,2,\infty} := \max_{A,X,Y \neq 0} \frac{\mu_{m,n,m+n}(A, X, Y)}{\|A\|_{\infty,1}\|X\|_{1,2}\|Y\|_{2,\infty}}$$

pointers

- S. Friedland and L.-H. Lim, "Nuclear norm of higher-order tensors," *Math. Comp.*, (2017), to appear.
- L.-H. Lim, "Self-concordance is NP-hard," *J. Global Optim.*, **68** (2017), no. 2, pp. 357–366.
- L.-H. Lim and P. Comon, "Blind multilinear identification," *IEEE Trans. Inform. Theory*, **60** (2014), no. 2, pp. 1260–1280.
- K. Ye and L.-H. Lim, "Algorithms for structured matrix-vector product of optimal bilinear complexity," *Proc. IEEE Inform. Theory Workshop* (ITW), **16** (2016), pp. 310–314.
- K. Ye and L.-H. Lim, "Fast structured matrix computations: tensor rank and Cohn–Umans method," *Found. Comput. Math.*, (2016), to appear.
- K. Ye and L.-H. Lim, "Can one multiply higher order tensors?" *preprint*, (2017).

thank you!