# structure tensors

Lek-Heng Lim

July 18, 2017

# 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, 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

# motivation

# goal: find fastest algorithms

- fast algorithms are rarely obvious algorithms
- want fast algorithms for bilinear operation $\beta : \mathbb{U} \times \mathbb{V} \to \mathbb{W}$

$$(A, \mathbf{x}) \mapsto A\mathbf{x}, \quad (A, B) \mapsto AB, \quad (A, B) \mapsto AB - BA$$

- embed into appropriate algebra $\mathcal{A}$

$$\begin{array}{ccc} \mathbb{U} \otimes \mathbb{V} & \xrightarrow{\iota} & \mathcal{A} \otimes \mathcal{A} \\ \beta \downarrow & & \downarrow m \\ \mathbb{W} & \xleftarrow[\pi]{} & \mathcal{A} \end{array}$$

- systematic way to discover new algorithms via structure tensors $\mu_\beta$ and $\mu_\mathcal{A}$
- fastest algorithms: rank of structure tensor
- stablest algorithms: nuclear norm of structure tensor

# ubiquitous problems

- linear equations, least squares, eigenvalue problem, etc

$$A\mathbf{x} = \mathbf{b}, \quad \min \|A\mathbf{x} - \mathbf{b}\|, \quad A\mathbf{x} = \lambda\mathbf{x}, \quad \mathbf{x} = \exp(A)\mathbf{b}$$

- backbone of numerical computations
- almost always: $A \in \mathbb{C}^{n \times n}$ has structure
- very often: $A \in \mathbb{C}^{n \times n}$ prohibitively high-dimensional
- impossible to solve without exploiting structure

# structured matrices

- sparse: "any matrix with enough zeros that it pays to take advantage of them" [Wilkinson, 1971]
- classical: circulant, Toeplitz, Hankel

$$T = \begin{bmatrix} t_0 & t_{-1} & & t_{1-n} \\ t_1 & t_0 & \ddots & \\ & \ddots & \ddots & t_{-1} \\ t_{n-1} & & t_1 & t_0 \end{bmatrix}, \quad H = \begin{bmatrix} h_0 & h_1 & \cdots & h_{n-1} \\ h_1 & h_2 & \iddots & h_n \\ \vdots & \iddots & \iddots & \vdots \\ h_{n-1} & h_n & \cdots & h_{2n-2} \end{bmatrix}$$

- many more: banded, triangular, Toeplitz-plus-Hankel, $f$-circulant, symmetric, skew-seymmetric, triangular Toeplitz, symmetric Toeplitz, etc

# multilevel

- 2-level: block-Toeplitz-Toeplitz-blocks (BTTB):

$$T = \begin{bmatrix} T_0 & T_{-1} & & T_{1-n} \\ T_1 & T_0 & \ddots & \\ & \ddots & \ddots & T_{-1} \\ T_{n-1} & & T_1 & T_0 \end{bmatrix} \in \mathbb{C}^{mn \times mn}$$

  where $T_i \in \mathbb{C}^{m \times m}$ are Toeplitz matrices
- 3-level: block-Toeplitz with BTTB blocks
- 4-level: block-BTTB with BTTB blocks
- and so on
- also multilevel versions of:
  - block-circulant-circulant-blocks (BCCB)
  - block-Hankel-Hankel-blocks (BHHB)
  - block-Toeplitz-plus-Hankel-Toeplitz-plus-Hankel-blocks (BTHTHB)

# Krylov subspace methods

- easiest way to exploit structure in $A$
- basic idea: by Cayley–Hamilton,

$$\alpha_0 I + \alpha_1 A + \cdots + \alpha_d A^d = 0$$

  for some $d \leq n$, so

$$A^{-1} = -\frac{\alpha_1}{\alpha_0} I - \frac{\alpha_2}{\alpha_0} A - \cdots - \frac{\alpha_d}{\alpha_0} A^{d-1}$$

  and so $\mathbf{x} = A^{-1}\mathbf{b} \in \operatorname{span}\{\mathbf{b}, A\mathbf{b}, \ldots, A^{d-1}\mathbf{b}\}$
- one advantage: $d$ can be much smaller than $n$, e.g.

$$d = \text{number of distinct eigenvalues of } A$$

  if $A$ diagonalizable
- another advantage: reduces to forming matrix-vector product $(A, \mathbf{x}) \mapsto A\mathbf{x}$ efficiently

# fastest algorithms

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

$$(a+bi)(c+di) = (ac-bd) + i(bc+ad)$$
$$= (ac-bd) + i[(a+b)(c+d) - ac - bd]$$

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

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

where

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

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

# why minimize multiplications?

- nowadays: latency of `FMUL` $\approx$ latency of `FADD`
- may want other measures of computational cost: e.g. energy consumption, number of gates, code space
- multiplier requires many more gates than adder (e.g. 18-bit: 2200 vs 125) $\to$ more wires/transistors $\to$ more energy
- may not use general purpose CPU: e.g. ASIC, DSP, FPGA, GPU, motion coprocessor, smart chip
- block operations: $A, B, C, D \in \mathbb{R}^{n \times n}$

  $$(A + iB)(C + iD) = (AC - BD) + i[(A + B)(C + D) - AC - BD]$$

  matrix multiplication vastly more expensive than matrix addition

# structure tensors

# structure tensor

- bilinear operator $\beta : \mathbb{U} \times \mathbb{V} \to \mathbb{W}$,

$$\beta(a_1\mathbf{u}_1 + a_2\mathbf{u}_2, \mathbf{v}) = a_1\beta(\mathbf{u}_1, \mathbf{v}) + a_2\beta(\mathbf{u}_2, \mathbf{v}),$$
$$\beta(\mathbf{u}, a_1\mathbf{v}_1 + a_2\mathbf{v}_2) = a_1\beta(\mathbf{u}, \mathbf{v}_1) + a_2\beta(\mathbf{u}, \mathbf{v}_2)$$

- there exists unique 3-tensor $\mu_\beta \in \mathbb{U}^* \otimes \mathbb{V}^* \otimes \mathbb{W}$ such that given any $(\mathbf{u}, \mathbf{v}) \in \mathbb{U} \times \mathbb{V}$ we have

$$\beta(\mathbf{u}, \mathbf{v}) = \mu_\beta(\mathbf{u}, \mathbf{v}, \cdot) \in \mathbb{W}$$

- examples of $\beta : \mathbb{U} \times \mathbb{V} \to \mathbb{W}$,

$$(A, B) \mapsto AB, \quad (A, \mathbf{x}) \mapsto A\mathbf{x}, \quad (A, B) \mapsto AB - BA$$

- call $\mu_\beta$ structure tensor of bilinear map $\beta$

# structure constants

- if we give $\mu_\beta$ coordinates, i.e., choose bases on $\mathbb{U}, \mathbb{V}, \mathbb{W}$, get hypermatrix
$$(\mu_{ijk}) \in \mathbb{C}^{m \times n \times p}$$
where $m = \dim \mathbb{U}$, $n = \dim \mathbb{V}$, $p = \dim \mathbb{W}$,
$$\beta(\mathbf{u}_i, \mathbf{v}_j) = \sum\nolimits_{k=1}^{p} \mu_{ijk} \mathbf{w}_k, \quad i = 1, \dots, m, \ j = 1, \dots, n$$
- $d$-dimensional hypermatrix is $d$-tensor in coordinates
- call $\mu_{ijk}$ structure constants of $\beta$

# example: physics

- $\mathfrak{g}$ Lie algebra with basis $\{\mathbf{e}_i : i = 1, \dots, n\}$

$$[\mathbf{e}_i, \mathbf{e}_j] = \sum_{k=1}^{n} c_{ijk} \mathbf{e}_k$$

- $(c_{ijk}) \in \mathbb{C}^{n \times n \times n}$ structure constants measure self-interaction
- structure tensor of $\mathfrak{g}$ is

$$\mu_{\mathfrak{g}} = \sum_{i,j,k=1}^{n} c_{ijk} \mathbf{e}_i^* \otimes \mathbf{e}_j^* \otimes \mathbf{e}_k \in \mathfrak{g}^* \otimes \mathfrak{g}^* \otimes \mathfrak{g}$$

- take $\mathfrak{g} = \mathfrak{so}_3$, real $3 \times 3$ skew symmetric matrices and

$$\mathbf{e}_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}, \ \mathbf{e}_2 = \begin{bmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \ \mathbf{e}_3 = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

- structure tensor of $\mathfrak{so}_3$ is

$$\mu_{\mathfrak{so}_3} = \sum_{i,j,k=1}^{3} \varepsilon_{ijk} \mathbf{e}_i^* \otimes \mathbf{e}_j^* \otimes \mathbf{e}_k,$$

where $\varepsilon_{ijk} = \frac{(i-j)(j-k)(k-i)}{2}$ is Levi-Civita symbol

# example: numerical computations

- for $A = (a_{ij}) \in \mathbb{C}^{m \times n}$, $B = (b_{jk}) \in \mathbb{C}^{n \times p}$,

$$AB = \sum_{i,j,k=1}^{m,n,p} a_{ik} b_{kj} E_{ij} = \sum_{i,j,k=1}^{m,n,p} E_{ik}^*(A) E_{kj}^*(B) E_{ij}$$

  where $E_{ij} = \mathbf{e}_i \mathbf{e}_j^\mathsf{T} \in \mathbb{C}^{m \times n}$ and $E_{ij}^* : \mathbb{C}^{m \times n} \to \mathbb{C}$, $A \mapsto a_{ij}$
- let

$$\mu_{m,n,p} = \sum_{i,j,k=1}^{m,n,p} E_{ik}^* \otimes E_{kj}^* \otimes E_{ij}$$

  write $\mu_n = \mu_{n,n,n}$
- structure tensor of matrix-matrix product

$$\mu_{m,n,p} \in (\mathbb{C}^{m \times n})^* \otimes (\mathbb{C}^{n \times p})^* \otimes \mathbb{C}^{m \times p} \cong \mathbb{C}^{mn \times np \times pm}$$

- later: rank gives minimal number of multiplications required to multiply two matrices [Strassen, 1973]

# example: computer science

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

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

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

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

# example: algebraic geometry

- quantum potential of quantum cohomology

$$\Phi(x,y,z) = \frac{1}{2}(xy^2 + x^2 z) + \sum_{d=1}^{\infty} N(d) \frac{z^{3d-1}}{(3d-1)!} e^{dy}$$

  $N(d)$ is number of rational curves of degree $d$ on the plane passing through $3d-1$ points in general position
- $\Phi(x,y,z) = \frac{1}{2}(xy^2 + x^2 z) + \phi(y,z)$, then $\phi$ satisfies

$$\phi_{zzz} = \phi_{yyz}^2 - \phi_{yyy}\phi_{yzz}$$

- can be transformed into Painlevé-six
- equivalent to third order derivative of $\Phi$ being structure tensor of an associative algebra [Kontsevich–Manin, 1994]

# bilinear complexity = tensor rank

- $A \in \mathbb{C}^{m\times n\times p}$, $\mathbf{u}\otimes\mathbf{v}\otimes\mathbf{w} := (u_i v_j w_k) \in \mathbb{C}^{m\times n\times p}$

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

- 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

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

- see [Bürgisser–Clausen–Shokrollahi, 1997]

# rank, decomposition, nuclear norm

- tensor rank

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

  gives least number of multiplications needed to compute $\beta$

- tensor decomposition

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

  gives an explicit algorithm for computing $\beta$

- tensor nuclear norm [Friedland–LHL, 2016]

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

  quantifies optimal numerical stability of computing $\beta$

# example: Gauss's method

- $\beta : \mathbb{C} \times \mathbb{C} \to \mathbb{C}$, $(z, w) \mapsto zw$ is $\mathbb{R}$-bilinear map
- $\mu_\beta \in (\mathbb{R}^2)^* \otimes (\mathbb{R}^2)^* \otimes \mathbb{R}^2$, as a hypermatrix [Knuth, 1998]

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

- $\mathbf{e}_1 = (1,0)$, $\mathbf{e}_2 = (0,1) \in \mathbb{R}^2$, $\mathbf{e}_1^*$, $\mathbf{e}_2^*$ dual basis in $(\mathbb{R}^2)^*$
- usual multiplication

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

- Gauss multiplication

$$\begin{aligned}\mu_\beta = {} & (\mathbf{e}_1^* + \mathbf{e}_2^*) \otimes (\mathbf{e}_1^* + \mathbf{e}_2^*) \otimes \mathbf{e}_2 \\ & + \mathbf{e}_1^* \otimes \mathbf{e}_1^* \otimes (\mathbf{e}_1 - \mathbf{e}_2) - \mathbf{e}_2^* \otimes \mathbf{e}_2^* \otimes (\mathbf{e}_1 + \mathbf{e}_2)\end{aligned}$$

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

# stability of Gauss's method

- nuclear norm

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

- attained by usual multiplication

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

- but not Gauss multiplication

$$\begin{aligned}\mu_\beta = (\mathbf{e}_1^* + \mathbf{e}_2^*) \otimes (\mathbf{e}_1^* + \mathbf{e}_2^*) \otimes \mathbf{e}_2 \\ + \mathbf{e}_1^* \otimes \mathbf{e}_1^* \otimes (\mathbf{e}_1 - \mathbf{e}_2) - \mathbf{e}_2^* \otimes \mathbf{e}_2^* \otimes (\mathbf{e}_1 + \mathbf{e}_2)\end{aligned}$$

  coefficients (upon normalizing) sums to $2(1 + \sqrt{2})$
- Gauss's algorithm less stable than the usual algorithm
- optimal bilinear complexity and stability:

$$\mu_\beta = \frac{4}{3}\left(\left[\frac{\sqrt{3}}{2}e_1 + \frac{1}{2}e_2\right]^{\otimes 3} + \left[-\frac{\sqrt{3}}{2}e_1 + \frac{1}{2}e_2\right]^{\otimes 3} + (-e_2)^{\otimes 3}\right)$$

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

# sparse, banded, triangular

- matrices with sparsity pattern $\Omega$ is

$$\mathbb{C}^{m\times n}_{\Omega} := \{A \in \mathbb{C}^{m\times n} : a_{ij} = 0 \text{ for all } (i,j) \notin \Omega\}$$

- special case: banded matrices with upper bandwidth $k$ and lower bandwidth $l$

$$\Omega = \{(i,j) \in \{1,\dots,n\} \times \{1,\dots,n\} : k < j - i < l\}$$

  - diagonal if $(k,l) = 0$
  - lower bidiagonal if $(k,l) = (0,1)$
  - upper bidiagonal if $(k,l) = (1,0)$
  - tridiagonal if $(k,l) = (1,1)$
  - pentadiagonal if $(k,l) = (2,2)$
  - lower triangular if $(k,l) = (0,n-1)$
  - upper triangular if $(k,l) = (n-1,0)$

- fastest sparse matrix-vector multiply?

$$\beta_{\Omega} : \mathbb{C}^{m\times n}_{\Omega} \times \mathbb{C}^{n} \to \mathbb{C}^{n}, \quad (A, \mathbf{x}) \mapsto A\mathbf{x}$$

# Toeplitz, Hankel, circulant

Toeplitz $\operatorname{Toep}_n(\mathbb{C}) = \{(t_{ij}) \in \mathbb{C}^{n\times n} : t_{ij} = t_{i-j}\}$

Hankel $\operatorname{Hank}_n(\mathbb{C}) = \{(h_{ij}) \in \mathbb{C}^{n\times n} : h_{ij} = h_{i+j}\}$

Circulant $\operatorname{Circ}_n(\mathbb{C}) = \{(c_{ij}) \in \mathbb{C}^{n\times n} : c_{ij} = c_{i-j \bmod n}\}$

- all vector spaces but $\operatorname{Circ}_n(\mathbb{C})$ is an algebra

$$\dim \operatorname{Toep}_n(\mathbb{C}) = \dim \operatorname{Hank}_n(\mathbb{C}) = 2n-1, \quad \dim \operatorname{Circ}_n(\mathbb{C}) = n$$

- structured matrix-vector multiplication:

$$\beta_t : \operatorname{Toep}_n(\mathbb{C}) \times \mathbb{C}^n \to \mathbb{C}^n, \quad (T, \mathbf{x}) \mapsto T\mathbf{x}$$
$$\beta_h : \operatorname{Hank}_n(\mathbb{C}) \times \mathbb{C}^n \to \mathbb{C}^n, \quad (H, \mathbf{x}) \mapsto H\mathbf{x}$$
$$\beta_c : \operatorname{Circ}_n(\mathbb{C}) \times \mathbb{C}^n \to \mathbb{C}^n, \quad (C, \mathbf{x}) \mapsto C\mathbf{x}$$

- what are the fastest algorithms?

# other structures

- symmetric and skew-symmetric matrices: $\mathsf{S}^2(\mathbb{C}^n)$, $\Lambda^2(\mathbb{C}^n)$
- $f$-circulant

$$\begin{bmatrix} x_1 & x_2 & \dots & x_{n-1} & x_n \\ fx_n & x_1 & \dots & x_{n-2} & x_{n-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ fx_3 & fx_4 & \dots & x_1 & x_2 \\ fx_2 & fx_3 & \dots & fx_n & x_1 \end{bmatrix} \in \mathbb{C}^{n\times n}$$

- Toeplitz-plus-Hankel: $\mathsf{Toep}_n(\mathbb{C}) + \mathsf{Hank}_n(\mathbb{C})$
- multilevel structures

  2-level block Toeplitz with Toeplitz blocks (BTTB)
  3-level block Toeplitz with BTTB blocks
  4-level block BTTB with BTTB blocks
  $k$-level and so on

- mixed: e.g. block BCCB with Toeplitz-plus-Hankel blocks

# fastest algorithms

- want the tensor ranks of

$$\mu_t \in \mathsf{Toep}_n(\mathbb{C})^* \otimes (\mathbb{C}^n)^* \otimes \mathbb{C}^n, \quad \mu_h \in \mathsf{Hank}_n(\mathbb{C})^* \otimes (\mathbb{C}^n)^* \otimes \mathbb{C}^n,$$

  and other structured matrices
- without structure: $\operatorname{rank}(\mu_{m,n}) = mn$ [Ye–LHL, 2016]

$$\beta_{m,n} : \mathbb{C}^{m\times n} \times \mathbb{C}^n \to \mathbb{C}^m, \quad (A, \mathbf{x}) \mapsto A\mathbf{x}$$

- ditto for sparse matrices [Ye–LHL, 2016]

$$\operatorname{rank}(\mu_\Omega) = \#\Omega$$

# generalizing Cohn–Umans

# representation theory

- $G$ finite group, $\mathbb{C}[G]$ group algebra
- $S, T, U \subseteq G$ of sizes $m, n, p$ with triple product property

$$stu = s't'u' \;\Rightarrow\; s = s', t = t', u = u'$$

for all $s, s' \in S,\ t, t' \in T,\ u, u' \in U$ [Cohn–Umans, 2003]

- for $A = (a_{ij}), B = (b_{jk}) \in \mathbb{C}^{n \times n}$, set

$$\widehat{A} = \sum\nolimits_{i,j=1}^{n} a_{ij} s_i t_j^{-1},\ \widehat{B} = \sum\nolimits_{j,k=1}^{n} b_{jk} t_j u_k^{-1} \in \mathbb{C}[G]$$

- $AB$ can be read off from entries of $\widehat{A}\widehat{B} \in \mathbb{C}[G]$
- use non-abelian FFT [Wedderburn, 1908] to compute $\widehat{A}\widehat{B}$

$$\mathbb{C}[G] \cong \bigoplus\nolimits_{i=1}^{k} \mathbb{V}_i \otimes \mathbb{V}_i^* \cong \bigoplus\nolimits_{i=1}^{k} \mathbb{C}^{d_i \times d_i}$$

$\mathbb{V}_1, \ldots, \mathbb{V}_k$ irreducible representations of $G$

# what we did

- do this for more general bilinear operations

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

  with $\mathbb{U}, \mathbb{V}, \mathbb{W}$ in place of $\mathbb{C}^{n \times n}$
- do this for more general algebraic object $\mathcal{A}$ in place of $\mathbb{C}[G]$
- generalize triple product property

$$\begin{array}{ccc} \mathbb{U} \otimes \mathbb{V} & \xrightarrow{\iota} & \mathcal{A} \otimes \mathcal{A} \\ \beta \downarrow & & \downarrow m \\ \mathbb{W} & \xleftarrow[\pi]{} & \mathcal{A} \end{array}$$

- relate ranks of multiplication tensors $\mu_\beta$ and $\mu_\mathcal{A}$
- apply these to answer our earlier questions

# generalizing Cohn–Umans

- $\beta : \mathbb{U} \times \mathbb{V} \to \mathbb{W}$ bilinear map
- $\mathcal{A}$ algebra with multiplication $m : \mathcal{A} \times \mathcal{A} \to \mathcal{A}$
- $\iota : \mathbb{U} \times \mathbb{V} \to \mathcal{A} \times \mathcal{A}$ embedding of vector spaces
- $\pi : \mathcal{A} \to \mathbb{W}$ projection of vector spaces
- if following diagram commutes [Ye–LHL, 2016]

$$\begin{array}{ccc} \mathbb{U} \otimes \mathbb{V} & \xrightarrow{\ \iota\ } & \mathcal{A} \otimes \mathcal{A} \\ {\scriptstyle \beta}\big\downarrow & & \big\downarrow{\scriptstyle m} \\ \mathbb{W} & \xleftarrow[\ \pi\ ]{} & \mathcal{A} \end{array}$$

  then we may determine $\beta(\mathbf{u}, \mathbf{v})$ by computing within $\mathcal{A}$
- if $\mathbb{U} = \mathbb{V} = \mathbb{W} = \mathcal{B}$ algebra, then $\operatorname{rank}(\mu_{\mathcal{A}}) = \operatorname{rank}(\mu_{\mathcal{B}})$

# example: Cohn–Umans

- apply this to

$$\begin{array}{ccc} \mathbb{C}^{n\times n} \otimes \mathbb{C}^{n\times n} & \xrightarrow{\ \iota\ } & \mathbb{C}[G] \otimes \mathbb{C}[G] \\ \beta \big\downarrow & & \big\downarrow m \\ \mathbb{C}^{n\times n} & \xleftarrow[\pi]{} & \mathbb{C}[G] \end{array}$$

- define $\iota : \mathbb{C}^{n\times n} \otimes \mathbb{C}^{n\times n} \to \mathbb{C}[G] \otimes \mathbb{C}[G]$ by

$$\iota(A, B) = \left( \sum\nolimits_{i,j=1}^{n} a_{ij} s_i t_j^{-1}, \sum\nolimits_{j,k=1}^{n} b_{jk} t_j u_k^{-1} \right) = (\widehat{A}, \widehat{B})$$

- triple product property ensures commutativity
- $\pi : \mathbb{C}[G] \to \mathbb{C}^{n\times n}$ reads entries of $AB$ from entries of $\widehat{A}\widehat{B}$

# example: fast integer multiplications

- apply this to

$$\begin{array}{ccc} \mathbb{Z} \otimes_{\mathbb{Z}} \mathbb{Z} & \xrightarrow{j_p} & \mathbb{Z}[x] \otimes_{\mathbb{Z}} \mathbb{Z}[x] \\ \downarrow \beta & & \downarrow \beta' \\ \mathbb{Z} & \xleftarrow[\mathrm{ev}_p]{} & \mathbb{Z}[x] \end{array}$$

- for $n \in \mathbb{Z}$

$$f_n(x) := \sum_{i=0}^{d} a_i x^i \in \mathbb{Z}[x]$$

  where $n = \sum_{i=0}^{d} a_i p^i$ is $p$-adic expansion
- embedding $j_p$ is

$$j_p(m \otimes n) = f_m(x) \otimes f_n(x)$$

- evaluation map $\mathrm{ev}_p$ sends $f(x) \in \mathbb{Z}[x]$ to $f(p) \in \mathbb{Z}$
- divide-and-conquer, interpolation, discrete Fourier transform, fast Fourier transform for polynomials gives Karatsuba, Toom–Cook, Schönhage–Strassen, Fürer for integers

# example: circulant matrices

- apply this to

$$\begin{array}{ccc} \mathrm{Circ}_n(\mathbb{C}) \otimes \mathbb{C}^n & \xrightarrow{\ \iota\ } & \mathbb{C}[\mathsf{C}_n] \otimes \mathbb{C}[\mathsf{C}_n] \\ \Big\downarrow{\scriptstyle \beta_c} & & \Big\downarrow{\scriptstyle m} \\ \mathbb{C}^n & \xleftarrow[\ \pi\ ]{} & \mathbb{C}[\mathsf{C}_n] \end{array}$$

  where $\mathsf{C}_n = \{1, \omega, \dots, \omega^{n-1}\}$ and $\omega = e^{2\pi i/n}$
- in this case $\iota$ and $\pi$ determined by isomorphism

$$\begin{bmatrix} c_0 & c_1 & \dots & c_{n-2} & c_{n-1} \\ c_{n-1} & c_0 & \dots & c_{n-3} & c_{n-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ c_2 & c_3 & \dots & c_0 & c_1 \\ c_1 & c_2 & \dots & c_{n-1} & c_0 \end{bmatrix} \mapsto \sum_{k=0}^{n-1} c_k \omega^k$$

- may show $\mathrm{rank}(\beta_c) = \overline{\mathrm{rank}}(\beta_c) = n$ [Ye–LHL, 2016]

# Toeplitz and Hankel?

- note that $ST \notin \mathsf{Toep}_n(\mathbb{C})$ even if $S, T \in \mathsf{Toep}_n(\mathbb{C})$
- however any $T_n \in \mathsf{Toep}_n(\mathbb{C})$ can be embedded as

$$\begin{bmatrix} T_n & S_n \\ S_n & T_n \end{bmatrix} \in \mathsf{Circ}_{2n}(\mathbb{C})$$

- extends to Hankel: $H \in \mathsf{Hank}_n(\mathbb{C})$ iff $JH$ or $HJ \in \mathsf{Toep}_n(\mathbb{C})$

$$J = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 1 & 0 \\ \vdots & \vdots & \iddots & \vdots & \vdots \\ 0 & 1 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \end{bmatrix}$$

- use these to define $\iota$ and $\pi$ with $\mathcal{A} = \mathbb{C}[\mathsf{C}_{2n}]$
- may show

$$\operatorname{rank}(\beta_t) = \overline{\operatorname{rank}}(\beta_t) = 2n - 1, \quad \operatorname{rank}(\beta_h) = \overline{\operatorname{rank}}(\beta_h) = 2n - 1$$

# symmetric matrices

- $\mathsf{S}^2(\mathbb{C}^n) := \{(a_{ij}) \in \mathbb{C}^{n\times n} : a_{ij} = a_{ji}\}$, want

$$\beta_s : \mathsf{S}^2(\mathbb{C}^n) \times \mathbb{C}^n \to \mathbb{C}^n, \quad (A, \mathbf{x}) \mapsto A\mathbf{x}$$

- express as sum of Hankel matrices

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix}, \quad \begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix} = \begin{bmatrix} a & b & c \\ b & c & e \\ c & e & f \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & d-c & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$\begin{bmatrix} a & b & c & d \\ b & e & f & g \\ c & f & h & i \\ d & g & i & j \end{bmatrix} = \begin{bmatrix} a & b & c & d \\ b & c & d & g \\ c & d & g & i \\ d & g & i & j \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & e-c & f-d & 0 \\ 0 & f-d & e-c & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & h-g-e+c & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

- apply result for Hankel matrices to get [Ye–LHL, 2016]

$$\operatorname{rank}(\beta_s) = \overline{\operatorname{rank}}(\beta_s) = \frac{n(n+1)}{2}$$

# multilevels

- use Kronecker product $\circledast$

$$\mathsf{Toep}_m(\mathbb{C}) \circledast \mathsf{Toep}_n(\mathbb{C}) = \mathrm{BTTB}_{m,n}(\mathbb{C})$$
$$\mathsf{Toep}_m(\mathbb{C}) \circledast \mathsf{Toep}_n(\mathbb{C}) \circledast \mathsf{Toep}_p(\mathbb{C}) = \mathsf{Toep}_m(\mathbb{C}) \circledast \mathrm{BTTB}_{n,p}(\mathbb{C})$$

- $\mathbb{U} \subseteq \mathbb{C}^{m\times m}$ and $\mathbb{V} \subseteq \mathbb{C}^{n\times n}$ linear subspaces

$$\beta_{\mathbb{U}} : \mathbb{U}\times\mathbb{C}^m \to \mathbb{C}^m, \quad \beta_{\mathbb{V}} : \mathbb{V}\times\mathbb{C}^n \to \mathbb{C}^n, \quad \beta_{\mathbb{U}\circledast\mathbb{V}} : (\mathbb{U}\circledast\mathbb{V})\times\mathbb{C}^{mn} \to \mathbb{C}^{mn}$$

  matrix-vector products with structure tensors $\mu_{\mathbb{U}}, \mu_{\mathbb{V}}, \mu_{\mathbb{U}\circledast\mathbb{V}}$

- if $\operatorname{rank}(\mu_{\mathbb{U}}) = \dim \mathbb{U}$, $\operatorname{rank}(\mu_{\mathbb{V}}) = \dim \mathbb{V}$, then [Ye–LHL, 2016]

$$\operatorname{rank}(\mu_{\mathbb{U}\circledast\mathbb{V}}) = \operatorname{rank}(\mu_{\mathbb{U}}) \operatorname{rank}(\mu_{\mathbb{V}})$$

- e.g. structure tensor of $\beta_{\mathrm{BTTB}} : \mathrm{BTTB}_{m,n}(\mathbb{C}) \times \mathbb{C}^{mn} \to \mathbb{C}^{mn}$ has rank $= (2m-1)(2n-1)$
- extends to arbitrary number of levels

# no time for these

- other subspaces of matrices:
    - skew-symmetric
    - Toeplitz-plus-Hankel
    - *f*-circulant
- other operations
    - matrix-matrix product
    - simultaneous matrix product
    - commutator
- other algebras:
    - coordinate rings of schemes
    - cohomology rings of manifolds
    - polynomial identity rings

# references


- S. Friedland and L.-H. Lim, "Nuclear norm of higher-order tensors," *Math. Comp.*, (2016), `ams:journals/mcom/earlyview/mcom3239`
- 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), `doi:10.1007/s10208-016-9332-x`