Introduction to Tensors
POEMA Workshop
Florence, 15-17th January, 2020

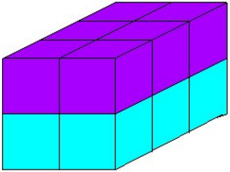
Giorgio Ottaviani

Università di Firenze

January 17, 2020

Tensors

Let V_i be vector spaces over $K = \mathbb{R}$ or $\mathbb{C}$. A tensor is an element $f \in V_1 \otimes \ldots \otimes V_k$, that is a multilinear map $V_1^\vee \times \ldots \times V_k^\vee \to K$
A tensor can be visualized as a multidimensional matrix.

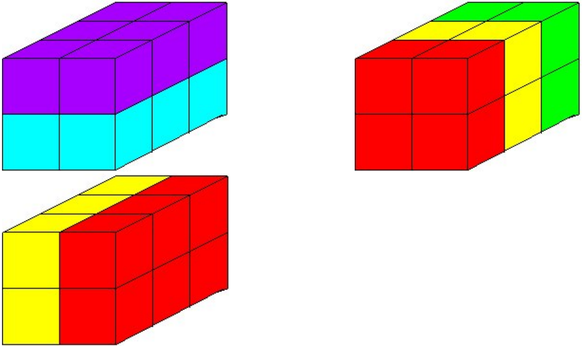


Entries of f are labelled by k indices, as $a_{i_1 \ldots i_k}$
For example, in the case $3 \times 2 \times 2$, with obvious notations, the expression in coordinates of a tensor is

$$a_{000}x_0y_0z_0 + a_{001}x_0y_0z_1 + a_{010}x_0y_1z_0 + a_{011}x_0y_1z_1 +$$
$$a_{100}x_1y_0z_0 + a_{101}x_1y_0z_1 + a_{110}x_1y_1z_0 + a_{111}x_1y_1z_1 +$$
$$a_{200}x_2y_0z_0 + a_{201}x_2y_0z_1 + a_{210}x_2y_1z_0 + a_{211}x_2y_1z_1$$

Slices

Just as matrices can be cutted in rows or in columns, higher dimensional tensors can be cut in slices



The three ways to cut a $3 \times 2 \times 2$ matrix into parallel slices

For a tensor of format $n_1 \times \ldots \times n_k$, there are n_1 slices of format $n_2 \times \ldots \times n_k$.

Multidimensional Gauss elimination

We can operate adding linear combinations of a slice to another slice, just in the case of rows and columns.
This amounts to multiply A of format $n_1 \times \ldots \times n_k$ for $G_1 \in GL(n_1)$, then for $G_i \in GL(n_i)$.

The group acting is quite big
$G = GL(n_1) \times \ldots \times GL(n_k)$.

The group acting, basic computation of dimensions.

The group is big, but not so big...

Let $\dim V_i = n_i$
$\dim V_1 \otimes \ldots \otimes V_k = \prod_{i=1}^k n_i \sim n_1^k$
$\dim GL(n_1) \times \ldots \times GL(n_k) = \sum_{i=1}^k n_i^2 \sim n_1^2$

For $k \geq 3$, the dimension of the group is in general much less that the dimension of the space where it acts.
This makes a strong difference between the classical case $k = 2$ (matrices) and the case $k \geq 3$(tensors).

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

Decomposable tensors, of rank one.

We need some "simple" tensors to start with.

Definition

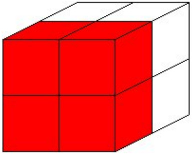
A tensor f is *decomposable* if there exist $x^i \in V_i$ for $i = 1, \ldots, k$ such that $a_{i_1 \ldots i_k} = x^1_{i_1} x^2_{i_2} \ldots x^k_{i_k}$. In equivalent way, $f = x^1 \otimes \ldots \otimes x^k$.

For a (nonzero) usual matrix, decomposable $\Longleftrightarrow$ rank one.
Define the rank of a tensor t as

$$\mathrm{rk}(t) := \min\{r | t = \sum_{i=1}^{r} t_i, t_i \text{ are decomposable}\}$$

For matrices, this coincides with usual rank.

Weierstrass Theorem about Tensor Decomposition in $n \times n \times 2$ case



Theorem (Weierstrass)

A general tensor t of format $n \times n \times 2$ has a unique tensor decomposition as a sum of n decomposable tensors

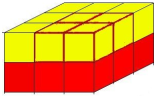
There is a algorithm to actually decompose such tensors. We see how it works in a $3 \times 3 \times 2$ example.

Tensor decomposition in a $3 \times 3 \times 2$ example.

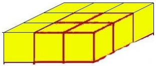
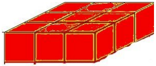
We consider the following “random” real tensor

$$
\begin{aligned}
f = & 6x_0y_0z_0 && +2x_1y_0z_0 + 6x_2y_0z_0 \\
& - 2014x_0y_1z_0 && +121x_1y_1z_0 - 11x_2y_1z_0 \\
& + 48x_0y_2z_0 && -13x_1y_2z_0 - 40x_2y_2z_0 \\
& - 31x_0y_0z_1 && +93x_1y_0z_1 + 97x_2y_0z_1 \\
& + 63x_0y_1z_1 && +41x_1y_1z_1 - 94x_2y_1z_1 \\
& - 3x_0y_2z_1 && +47x_1y_2z_1 + 4x_2y_2z_1
\end{aligned}
$$

We divide into two 3×3 slices, like in

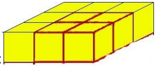
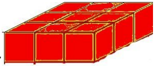


$\Longrightarrow$

z_0  $+z_1$ 

Two slices

Sum the yellow slice plus t times the red slice.

$f_0 + tf_1 =$  $+t$ 

$$f_0 + tf_1 = \begin{pmatrix} -31t+6 & 63t-2014 & -3t+48 \\ 93t+2 & 41t+121 & 47t-13 \\ 97t+6 & -94t-11 & 4t-40 \end{pmatrix}$$

Singular combination of slices

We compute the determinant, which is a cubic polynomial in t

$$\det(f_0 + tf_1) = 159896t^3 - 8746190t^2 - 5991900t - 69830$$

with roots $t_0 = -.0118594$, $t_1 = -.664996$, $t_2 = 55.3761$.

This computation gives a "guess" about the three summands for z_i, (note the sign change!)

$$f = A_0(.0118594z_0 + z_1) + A_1(.664996z_0 + z_1) + A_2(-55.3761z_0 + z_1)$$

where A_i are 3×3 matrices of rank one, that we have to find. Indeed, we get

$$f_0 + tf_1 = A_0(.0118594 + t) + A_1(.664996 + t) + A_2(-55.3761 + t)$$

and for the three roots $t = t_i$ one summand vanishes, it remains a matrix of rank 2, *with only two colors*, hence with zero determinant.

Finding the three matrices from kernels.

In order to find A_i, let $a_0 = \begin{pmatrix} -.0589718 & -.964899 & .255916 \end{pmatrix}$, left kernel of $f_0 + t_0 f_1$
let $b_0 = \begin{pmatrix} -.992905 & -.00596967 & -.118765 \end{pmatrix}$, transpose of right kernel of $f_0 + t_0 f_1$.
In the same way, denote
$a_1 =$ left kernel of $f_0 + t_1 f_1$, $a_2 =$ left kernel of $f_0 + t_2 f_1$
$b_1 =$ transpose of right kernel of $f_0 + t_1 f_1$, $b_2 =$ transpose of right kernel of $f_0 + t_2 f_1$,

$$aa = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} -.0589718 & -.964899 & .255916 \\ -.014181 & -.702203 & .711835 \\ .959077 & .0239747 & .282128 \end{pmatrix}$$

$$bb = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} -.992905 & -.00596967 & -.118765 \\ .582076 & -.0122361 & -.813043 \\ .316392 & .294791 & -.901662 \end{pmatrix}$$

Inversion and summands of tensor decomposition

Now we invert the two matrices

$$aa^{-1} = \begin{pmatrix} .450492 & -.582772 & 1.06175 \\ -1.43768 & .548689 & -.0802873 \\ -1.40925 & 1.93447 & -.0580488 \end{pmatrix}$$

$$bb^{-1} = \begin{pmatrix} -.923877 & .148851 & -.0125305 \\ -.986098 & -3.43755 & 3.22958 \\ -.646584 & -1.07165 & -.0575754 \end{pmatrix}$$

The first summand A_0 is given by a scalar c_0 multiplied by

$(.450492x_0 - 1.43768x_1 - 1.40925x_2)(-.923877y_0 - .986098y_1 - .646584y_2)$

the same for the other colors.

Decomposition as sum of three terms

By solving a linear system, we get the scalars c_i

$$(.450492x_0 - 1.43768x_1 - 1.40925x_2)(-.923877y_0 - .986098y_1 - .646584y_2)(.809777z_0 + 68.2814z_1) +$$

$$(-.582772x_0 + .548689x_1 + 1.93447x_2)(.148851y_0 - 3.43755y_1 - 1.07165y_2)(18.6866z_0 + 28.1003z_1) +$$

$$(1.06175x_0 - .0802873x_1 - .0580488x_2)(-.0125305y_0 + 3.22958y_1 - .0575754y_2)(-598.154z_0 + 10.8017z_1)$$

and the sum is

$$\begin{aligned}
6x_0y_0z_0 + 2x_1y_0z_0 \quad & +6x_2y_0z_0 \\
-2014x_0y_1z_0 + 121x_1y_1z_0 \quad & -11x_2y_1z_0 \\
+48x_0y_2z_0 - 13x_1y_2z_0 \quad & -40x_2y_2z_0 \\
-31x_0y_0z_1 + 93x_1y_0z_1 \quad & +97x_2y_0z_1 \\
+63x_0y_1z_1 + 41x_1y_1z_1 \quad & -94x_2y_1z_1 \\
-3x_0y_2z_1 + 47x_1y_2z_1 \quad & +4x_2y_2z_1
\end{aligned}$$

The rank of the tensor f is 3, because we have 3 summands, and no less.

Uniqueness of the decomposition

The decomposition we have found is *unique*, up to reordering the summands.
This is a strong difference with the case of matrices, where any decomposition with at least two summands is *never unique*.

For tensors f of rank ≤ 2,the characteristic polynomial vanishes identically.
We understand this phenomenon geometrically, in a while.

General symmetric tensors of rank s fill an open subset of an irreducible projective variety, which is called the s-th secant variety to the Veronese variety $v_d(\mathbb{P}^n)$ and it is denoted by $\sigma_s v_d(\mathbb{P}^n)$.

The rational normal curve $v_d(\mathbb{P}^1)$ is called also the moment curve since it has the parametrization

$$(1, t, t^2, \ldots, t^d)$$

Six differences between matrix rank and tensor rank

1) Tensor Rank may be different from the dimension of image

2) Tensor Rank may be larger than the dimensions of the factors

Six differences between matrix rank and tensor rank

3) Tensor Rank may increase or a decrease in a limit (for matrices may only decrease)

4) Maximum Tensor rank may be larger than Generic Tensor Rank

Six differences between matrix rank and tensor rank

5) Tensor Rank may depend on the field

6) Tensor Rank is NP-hard to be computed

A seventh difference which marks a point for Tensors !

7) Tensor Decomposition is in general unique, unless the case of matrices.

See Luca Chiantini talk, this afternoon.

In the case $V_1 = \ldots = V_k = V$ we may consider symmetric tensors $f \in S^d V$.

Elements of $S^d V$ can be considered as homogeneous polynomials of degree d in $x_0, \ldots x_n$, basis of V.

So polynomials have rank (as all tensors) and also symmetric rank (next slides).

Symmetric Tensor Decomposition (Waring)

A *Waring decomposition* of $f \in S^d V$ is

$$f = \sum_{i=1}^{r} c_i (l_i)^d \qquad \text{with } l_i \in V$$

with minimal r, which is called the symmetric rank

Example: $7x^3 - 30x^2y + 42xy^2 - 19y^3 = (-x+2y)^3 + (2x-3y)^3$
$\mathrm{rk}\left(7x^3 - 30x^2y + 42xy^2 - 19y^3\right) = 2$

Shitov counterexample

It is obvious that for a symmetric tensor f we have

$$\mathrm{rank}(f) \leq \mathrm{symmetricrank}(f)$$

In 2017 Shitov shocked the tensor community by exhibiting a polynomial f of degree 3 in 800 variables, say a symmetric tensor of format $800 \times 800 \times 800$ such that

$$\mathrm{rank}(f) \leq 903 < 904 \leq \mathrm{symmetricrank}(f)$$

The tensors where rank and symmetric rank differ are not yet understood. Do they form a set of measure zero ?

poema
Polynomial Optimization, Efficiency through Moments and Algebra

Symmetric case: the Alexander-Hirschowitz Theorem

Theorem (Campbell, [1891] Terracini, [1916] Alexander-Hirschowitz [1995])

The general $f \in S^d\mathbb{C}^{n+1}$ $(d \geq 3)$ *has rank*

$$\lceil \frac{\binom{n+d}{d}}{n+1} \rceil$$

which is called the generic rank, with the only exceptions

- $S^4\mathbb{C}^{n+1}$, $2 \leq n \leq 4$, *where the generic rank is* $\binom{n+2}{2}$
- $S^3\mathbb{C}^5$, *where the generic rank is* 8, *sporadic case*

Apolarity and Waring decomposition, I

For any $l = \alpha x_0 + \beta x_1 \in \mathbb{C}^2$ we denote $l^{\perp} = -\beta\partial_0 + \alpha\partial_1 \in \mathbb{C}^{2^{\vee}}$.
Note that

$$l^{\perp}(l^d) = 0 \tag{1}$$

so that $l^{\perp}$ is well defined (without referring to coordinates) up to scalar multiples. Let e be an integer. Any $f \in S^d\mathbb{C}^2$ defines
$C_f^e\colon S^e(\mathbb{C}^{2^{\vee}}) \to S^{d-e}\mathbb{C}^2$
Elements in $S^e(\mathbb{C}^{2^{\vee}})$ can be decomposed as $(l_1^{\perp} \circ \ldots \circ l_e^{\perp})$ for some $l_i \in \mathbb{C}^2$.

Apolarity and Waring decomposition, II

Proposition

Let l_i be distinct for $i = 1, \ldots, e$. There are $c_i \in K$ such that $f = \sum_{i=1}^{e} c_i (l_i)^d$ if and only if $(l_1^{\perp} \circ \ldots \circ l_e^{\perp}) f = 0$

Proof: The implication $\Longrightarrow$ is immediate from (1). It can be summarized by the inclusion
$< (l_1)^d, \ldots, (l_e)^d > \subseteq \ker(l_1^{\perp} \circ \ldots \circ l_e^{\perp})$. The other inclusion follows by dimensional reasons, because both spaces have dimension e. $\square$

The previous Proposition is the core of the Prony-Sylvester algorithm, because the differential operators killing f allow to define the decomposition of f, as we see in the next slide.

Prony-Sylvester algorithm for Waring decomposition in two variables

Prony-Sylvester algorithm for general f Compute the decomposition of a general $f \in S^d U$

- Pick a generator g of $\ker C_f^a$ with $a = \lfloor \frac{d+1}{2} \rfloor$.
- Decompose g as product of linear factors, $g = (l_1^\perp \circ \ldots \circ l_r^\perp)$
- Solve the system $f = \sum_{i=1}^r c_i (l_i)^d$ in the unknowns c_i.

Remark When d is odd the kernel is one-dimensional and the decomposition is unique. When d is even the kernel is two-dimensional and there are infinitely many decompositions.

The catalecticant matrices for two variables

If $f(x,y) = a_0x^4 + 4a_1x^3y + 6a_2x^2y^2 + 4a_3xy^3 + a_4y^4$ then

$$C_f^1 = \begin{bmatrix} a_0 & a_1 & a_2 & a_3 \\ a_1 & a_2 & a_3 & a_4 \end{bmatrix}$$

and

$$C_f^2 = \begin{bmatrix} a_0 & a_1 & a_2 \\ a_1 & a_2 & a_3 \\ a_2 & a_3 & a_4 \end{bmatrix}$$

The catalecticant algorithm at work

The catalecticant matrix associated to
$f = 7x^3 - 30x^2 + 42x - 19 = 0$ is

$$A_f = \begin{bmatrix} 7 & -10 & 14 \\ -10 & 14 & -19 \end{bmatrix}$$

$\ker A_f$ is spanned by $\begin{bmatrix} 6 \\ 7 \\ 2 \end{bmatrix}$ which corresponds to
$6\partial_x^2 + 7\partial_x\partial_y + 2\partial_y^2 = (2\partial_x + \partial_y)(3\partial_x + 2\partial_y)$
Hence the decomposition

$$7x^3 - 30x^2y + 42xy^2 - 19y^3 = c_1(-x + 2y)^3 + c_2(2x - 3y)^3$$

Solving the linear system, we get
$c_1 = c_2 = 1$

The singular n-ples (Optimization Problem)

Any tensor $A \in \mathbb{R}^{m_1} \otimes \ldots \otimes \mathbb{R}^{m_d}$ (m_i can be different) defines by contraction a function f_A over the product $S = S^{m_1-1} \times \ldots \times S^{m_d-1}$ of the corresponding spheres.

$$\begin{array}{rccc} f_A\colon & S & \to & \mathbb{R} \\ & x & \mapsto & A \times x \end{array}$$

In alternative, f_A could be defined over the affine cone of decomposable tensors.

Theorem ([Lim, Qi], Variational principle)

The critical points of f_A corresponds to d-ples $(x_1, \ldots, x_d) \in S$ such that

$$A(x_1, \ldots, \hat{x}_i, \ldots, x_d) = \lambda x_i \qquad \forall i = 1, \ldots, d$$

References and notations

• L.-H. Lim, *Singular values and eigenvalues of tensors: a variational approach*, Proc. IEEE, (CAMSAP '05), (2005).
• L. Qi, *Eigenvalues of a real supersymmetric tensor*, J. Symbolic Comput. (2005)

The critical points of the theorem are called singular d-ples, or E-eigenvectors in the symmetric case ([Hu-Qi]).
• S. Hu, L, Qi, *The E-eigenvectors of tensors*, Linear and Multilinear Algebra, (2013)

Passage to complex numbers.

The euclidean quadratic form $< x, y >= \sum_{i=1}^{m} x_i y_i$, where $x, y \in \mathbb{R}^m$ is often extended to the complex case as the Hermitian form $< x, y >= \sum_{i=1}^{m} x_i \overline{y}_i$, where $x, y \in \mathbb{C}^m$.

To use algebro-geometric techniques it is more convenient to extend it as $< x, y >= \sum_{i=1}^{m} x_i y_i$, where $x, y \in \mathbb{C}^m$.

This is enough to identify each vector space $\mathbb{C}^m$ with its dual. Moreover the complex orthogonal group $O(m_i, \mathbb{C})$ is acting over any $\mathbb{C}^{m_i}$. Anyway this "trick" will have some annoying consequences, about transversality, we will see later....

Orthogonal group acts, preserving singular d-ples

The group $O(\mathbf{m}, \mathbb{C}) := O(m_1, \mathbb{C}) \times \ldots \times O(m_d, \mathbb{C})$ acts over $\mathbb{C}^{m_1} \otimes \ldots \otimes \mathbb{C}^{m_d}$ in natural way.

Lemma

Let $A \in \mathbb{C}^{m_1} \otimes \ldots \otimes \mathbb{C}^{m_d}$ be a tensor.

$$\left.\begin{array}{r} (x_1, \ldots, x_d) \\ \textit{is a singular } d\textit{-ple of } A \end{array}\right\} \Longleftrightarrow \left\{\begin{array}{l} g \cdot (x_1, \ldots, x_d) \\ \textit{is a singular } d\textit{-ple of } \ g \cdot A \\ \forall g \in O(\mathbf{m}, \mathbb{C}). \end{array}\right.$$

Poema
Polynomial Optimization, Efficiency through Moments and Algebra

Two-dimensional case

For $d = 2$, tensors correspond to usual matrices A, and singular 2-ples are the classical singular pairs of vectors (x, y) such that

$$Ax = \lambda y \qquad A^{\top} y = \lambda x$$

If A is general matrix of format $m \times n$, it has $\min(m, n)$ singular pairs of vectors.

Question How many singular d-ples has a general tensor A ?

The number of singular d-ples

Theorem ([Friedland-O])

The number of singular d-ples of a general tensor A over $\mathbb{C}$, of format $m_1 \times \ldots \times m_d$, is the coefficient of $\prod_{i=1}^{d} t_i^{m_i-1}$ in the polynomial

$$\prod_{i=1}^{d} \frac{\hat{t}_i^{\,m_i} - t_i^{m_i}}{\hat{t}_i - t_i}$$

where $\hat{t}_i = \sum_{j \neq i} t_j$.
We denote this coefficient by $c(m_1, \ldots, m_d)$.

- S. Friedland, G. Ottaviani, *The number of singular vector tuples and uniqueness of best rank one approximation of tensors*, Journal Found. Comp. Math. 2014.

The real case

Theorem ([Friedland-O])

The number of singular d-ples of a general tensor A over $\mathbb{R}$ of format $m_1 \times \ldots \times m_d$ is $\leq c(m_1, \ldots, m_d)$. Moreover there are no singular d-ples corresponding to zero singular value and all singular d-ples are simple.

Note that $c(m_1, m_2) = \min(m_1, m_2)$. For $d \geq 3$ the numbers $c(m_1, \ldots, m_d)$ are quite large. For example $c(\underbrace{2, \ldots, 2}_{d}) = d!$.

The proof needs Chern classes on Segre variety.

Let $X = \mathbb{P}(\mathbb{C}^{m_1}) \times \ldots \times \mathbb{P}(\mathbb{C}^{m_d})$ be the projective variety of rank one tensors, called the Segre variety. Let Q_i be the pullback on X of the quotient bundle from the i-th factor.

The singular d-ples of a tensor A are zero locus of a section, corresponding to A, of the bundle on X

$[Q_1 \otimes \mathcal{O}(0,1,1,\ldots,1)] \oplus [Q_2 \otimes \mathcal{O}(1,0,1,\ldots,1)] \oplus \ldots$

where $\mathcal{O}(a_1,\ldots,a_d)$ is the line bundle obtained by tensoring the pullback of $\mathcal{O}(a_i)$ from i-th factor.

The formula comes from computing the top Chern class of this bundle. The multiplicity of a d-ple can be defined as the multiplicity in the zero locus of the section corresponding to A.

Polynomial Optimization, Efficiency through Moments and Algebra

List of the number $c(m_1, m_2, m_3)$ of singular 3-ples

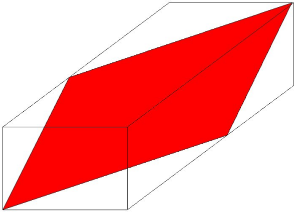
m_1, m_2, m_3	$c(m_1, m_2, m_3)$	
$2, 2, 2$	6	
$2, 2, m_3$	8	$m_3 \geq 3$
$2, 3, 3$	15	
$2, 3, m_3$	18	$m_3 \geq 4$
$2, m, m$	$m(2m-1)$	
$2, m_2, m_3$	$2m_2^2$	$m_3 \geq m_2 + 1$
$3, 3, 3$	37	
$3, 3, 4$	55	
$3, 3, m_3$	61	$m_3 \geq 5$
$3, 4, 4$	104	
$3, 4, 5$	138	
$3, 4, m_3$	148	$m_3 \geq 6$
$3, 5, 5$	225	
$3, 5, 6$	280	
$3, 5, m_3$	295	$m_3 \geq 7$
$3, m_2, m_3$	$\frac{8}{3}m_2^3 - 2m_2^2 + \frac{7}{3}m_2$	$m_3 \geq m_2 + 2$

The boundary format and the diagonal.

The format (of a tensor) $m_1 \times \ldots \times m_d$ with $m_1 \leq \ldots \leq m_d$ is called boundary format if

$$m_d - 1 = \sum_{i=1}^{d-1} (m_i - 1)$$

In the boundary format case it is well defined a unique diagonal given by elements $a_{i_1 \ldots i_d}$ satisfying $i_d = \sum_{j=1}^{d-1} i_j$



(indices start from zero)

In $d = 2$ case, boundary format means square.

Stabilization beyond boundary format.

The number of critical points *stabilizes* for formats larger than boundary format, that is

Corollary ([Friedland-O])

For $m_d - 1 \geq \sum_{i=1}^{d-1}(m_i - 1)$*, the number of critical points, from a general tensor to the variety of rank one tensors, does not depend on* m_d*.*

Open question Is there a direct proof of Corollary, without using the formula ?

Thanks

Poema
Polynomial Optimization, Efficiency
through Moments and Algebra

Thanks !!

European Commission
Horizon 2020
European Union funding
for Research & Innovation

Poema
Polynomial Optimization, Efficiency
through Moments and Algebra