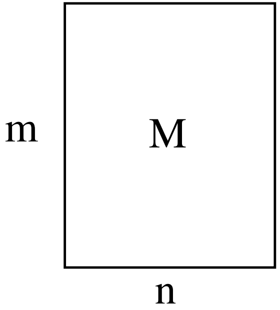
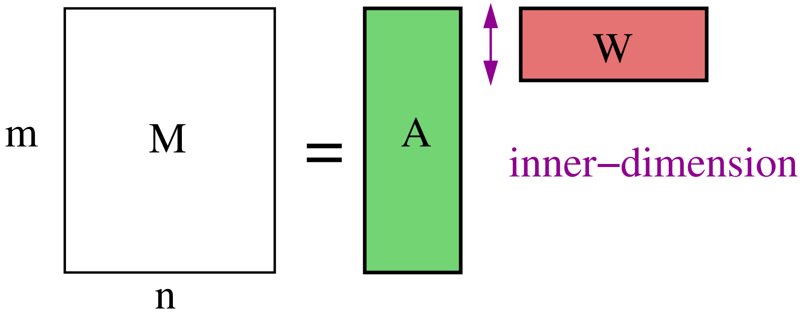
An Almost Optimal Algorithm for Computing Nonnegative Rank

Ankur Moitra

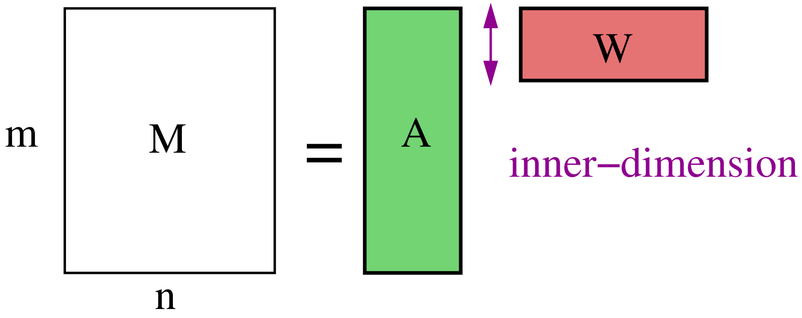
Institute for Advanced Study

January 8, 2013

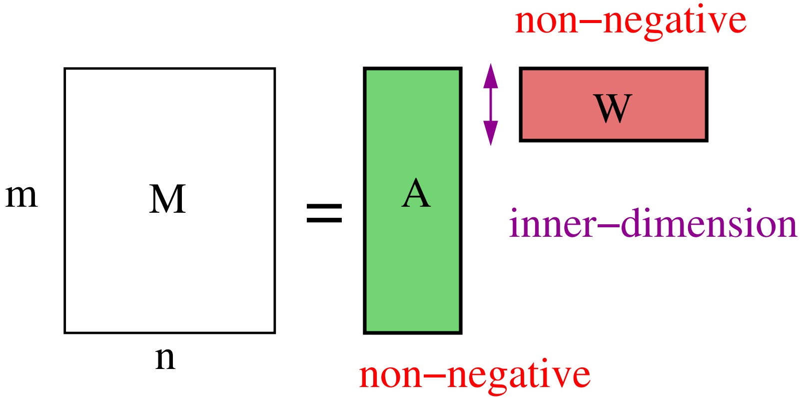

m
M
n

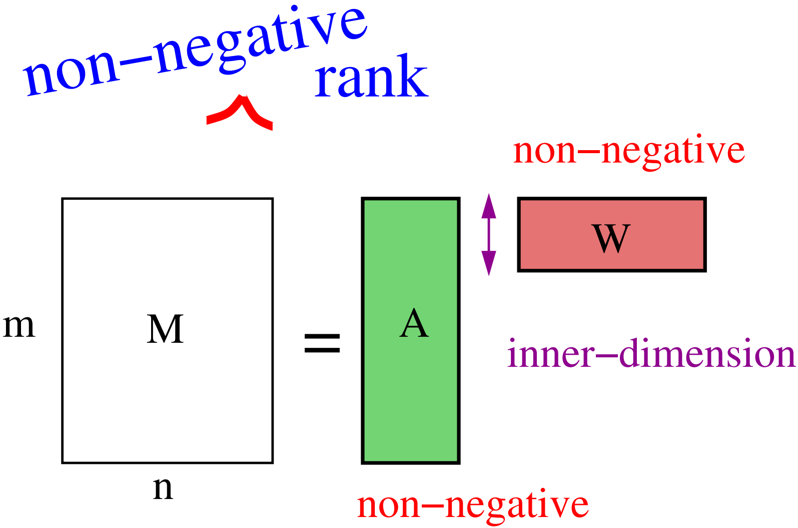

m
M
n
=
A
W
inner–dimension

rank



rank




non–negative
rank
non–negative
m
M
=
A
W
inner–dimension
n
non–negative

Applications

- Statistics and Machine Learning:
  - extract latent relationships in data
  - image segmentation, text classification, information retrieval, collaborative filtering, ...

    [Lee, Seung], [Xu et al], [Hofmann], [Kumar et al], [Kleinberg, Sandler]

Applications

- Statistics and Machine Learning:
  - extract latent relationships in data
  - image segmentation, text classification, information retrieval, collaborative filtering, ...

    [Lee, Seung], [Xu et al], [Hofmann], [Kumar et al], [Kleinberg, Sandler]
- Combinatorics:
  - extended formulation, log-rank conjecture

    [Yannakakis], [Lovász, Saks]

Applications

- Statistics and Machine Learning:
  - extract latent relationships in data
  - image segmentation, text classification, information retrieval, collaborative filtering, ...

    [Lee, Seung], [Xu et al], [Hofmann], [Kumar et al], [Kleinberg, Sandler]
- Combinatorics:
  - extended formulation, log-rank conjecture

    [Yannakakis], [Lovász, Saks]
- Physical Modeling:
  - interaction of components is additive
  - visual recognition, environmetrics

The Complexity of Nonnegative Rank

The Complexity of Nonnegative Rank

[Vavasis]: It is *NP*-hard to compute the nonnegative rank.

The Complexity of Nonnegative Rank

[Vavasis]: It is *NP*-hard to compute the nonnegative rank.

[Cohen and Rothblum]: The nonnegative rank can be computed in time $(nm)^{O(nr+mr)}$.

The Complexity of Nonnegative Rank

[Vavasis]: It is NP-hard to compute the nonnegative rank.

[Cohen and Rothblum]: The nonnegative rank can be computed in time $(nm)^{O(nr+mr)}$.

[Arora, Ge, Kannan and Moitra]: The nonnegative rank can be computed in time $(nm)^{f(r)}$ where $f(r) = O(2^r)$ and any algorithm that runs in time $(nm)^{o(r)}$ would yield a sub exponential time algorithm for 3-SAT.

The Complexity of Nonnegative Rank

[Vavasis]: It is *NP*-hard to compute the nonnegative rank.

[Cohen and Rothblum]: The nonnegative rank can be computed in time $(nm)^{O(nr+mr)}$.

[Arora, Ge, Kannan and Moitra]: The nonnegative rank can be computed in time $(nm)^{f(r)}$ where $f(r) = O(2^r)$ and any algorithm that runs in time $(nm)^{o(r)}$ would yield a sub exponential time algorithm for 3-SAT.

Theorem

The nonnegative rank can be computed in time $(nm)^{O(r^2)}$*.*

The Complexity of Nonnegative Rank

[Vavasis]: It is *NP*-hard to compute the nonnegative rank.

[Cohen and Rothblum]: The nonnegative rank can be computed in time $(nm)^{O(nr+mr)}$.

[Arora, Ge, Kannan and Moitra]: The nonnegative rank can be computed in time $(nm)^{f(r)}$ where $f(r) = O(2^r)$ and any algorithm that runs in time $(nm)^{o(r)}$ would yield a sub exponential time algorithm for 3-SAT.

Theorem

The nonnegative rank can be computed in time $(nm)^{O(r^2)}$.

...these algorithms are about an algebraic question, about how to best encode nonnegative rank as a systems of polynomial inequalities

Is NMF Computable?

Is NMF Computable?

[Cohen, Rothblum]: Yes

Is NMF Computable?

[Cohen, Rothblum]: Yes (**DETOUR**)

Semi-algebraic sets: s polynomials, k variables, Boolean function B

$$S = \{x_1, x_2 ... x_k | B(\text{sgn}(f_1), \text{sgn}(f_2), ... \text{sgn}(f_s)) = \text{"true"}\}$$

Semi-algebraic sets: s polynomials, k variables, Boolean function B

$$S = \{x_1, x_2 ... x_k | B(\mathrm{sgn}(f_1), \mathrm{sgn}(f_2), ... \mathrm{sgn}(f_s)) = \text{"true"} \}$$

Question

How many sign patterns arise (as $x_1, x_2, ... x_k$ range over $\mathbb{R}^k$)?

Semi-algebraic sets: s polynomials, k variables, Boolean function B

$$S = \{x_1, x_2 ... x_k | B(\mathrm{sgn}(f_1), \mathrm{sgn}(f_2), ... \mathrm{sgn}(f_s)) = \text{"true"}\}$$

Question

How many sign patterns arise (as $x_1, x_2, ... x_k$ range over $\mathbb{R}^k$)?

Naive bound: 3^s (all of $\{-1, 0, 1\}^s$),

Semi-algebraic sets: s polynomials, k variables, Boolean function B

$$S = \{x_1, x_2 ... x_k | B(\mathrm{sgn}(f_1), \mathrm{sgn}(f_2), ... \mathrm{sgn}(f_s)) = \text{"true"}\}$$

Question

How many sign patterns arise (as $x_1, x_2, ... x_k$ range over $\mathbb{R}^k$)?

Naive bound: 3^s (all of $\{-1, 0, 1\}^s$), [Milnor, Warren]: at most $(ds)^k$, where d is the maximum degree

Semi-algebraic sets: s polynomials, k variables, Boolean function B

$$S = \{x_1, x_2 \ldots x_k | B(\text{sgn}(f_1), \text{sgn}(f_2), \ldots \text{sgn}(f_s)) = \text{"true"} \}$$

Question

How many sign patterns arise (as $x_1, x_2, \ldots x_k$ range over $\mathbb{R}^k$)?

Naive bound: 3^s (all of $\{-1, 0, 1\}^s$), [Milnor, Warren]: at most $(ds)^k$, where d is the maximum degree

In fact, best known algorithms (e.g. [Renegar]) for finding a point in S run in $(ds)^{O(k)}$ time

Is NMF Computable?

[Cohen, Rothblum]: Yes (**DETOUR**)

Is NMF Computable?

[Cohen, Rothblum]: Yes (**DETOUR**)

- Variables: entries in A and W ($nr + mr$ total)

Is NMF Computable?

[Cohen, Rothblum]: Yes (**DETOUR**)

- Variables: entries in A and W ($nr + mr$ total)
- Constraints: $A, W \geq 0$ and $AW = M$ (degree two)

Is NMF Computable?

[Cohen, Rothblum]: Yes (**DETOUR**)

- Variables: entries in A and W ($nr + mr$ total)
- Constraints: $A, W \geq 0$ and $AW = M$ (degree two)

Running time for a solver is exponential in the number of **variables**

Is NMF Computable?

[Cohen, Rothblum]: Yes (**DETOUR**)

- Variables: entries in A and W ($nr + mr$ total)
- Constraints: $A, W \geq 0$ and $AW = M$ (degree two)

Running time for a solver is exponential in the number of **variables**

Question

What is the smallest formulation, measured in the number of variables?

Is NMF Computable?

[Cohen, Rothblum]: Yes (**DETOUR**)

- Variables: entries in A and W ($nr + mr$ total)
- Constraints: $A, W \geq 0$ and $AW = M$ (degree two)

Running time for a solver is exponential in the number of **variables**

Question

What is the smallest formulation, measured in the number of variables? Can we use only $f(r)$ variables?

Is NMF Computable?

[Cohen, Rothblum]: Yes (**DETOUR**)

- Variables: entries in A and W ($nr + mr$ total)
- Constraints: $A, W \geq 0$ and $AW = M$ (degree two)

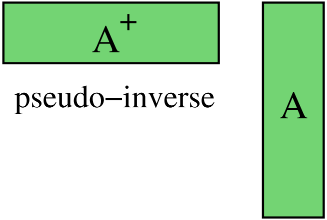
Running time for a solver is exponential in the number of **variables**

Question

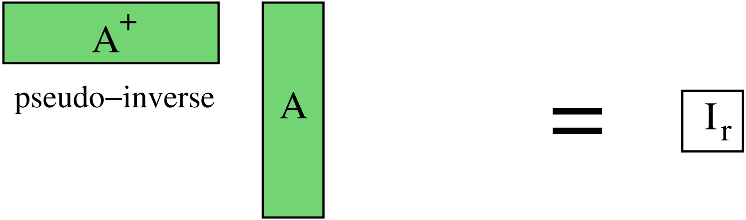
What is the smallest formulation, measured in the number of variables? Can we use only $f(r)$ variables? $O(r^2)$ variables?

Easy Case: A has Full Column Rank (AGKM)

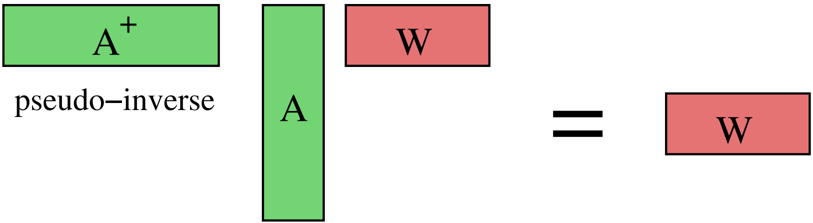
Easy Case: A has Full Column Rank (AGKM)



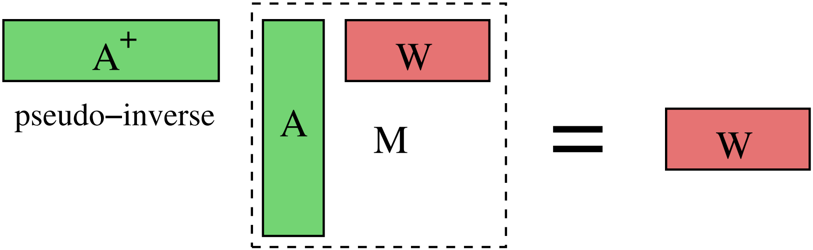
Easy Case: A has Full Column Rank (AGKM)



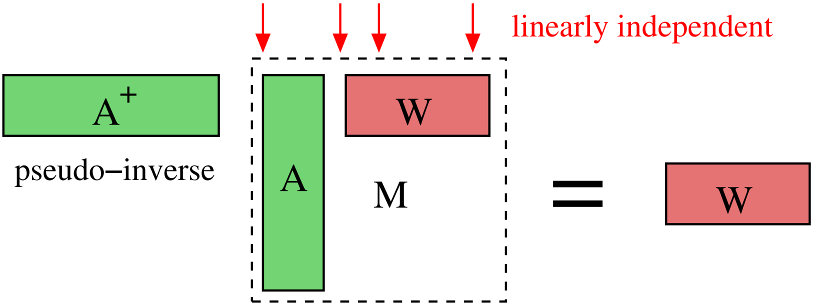
Easy Case: A has Full Column Rank (AGKM)



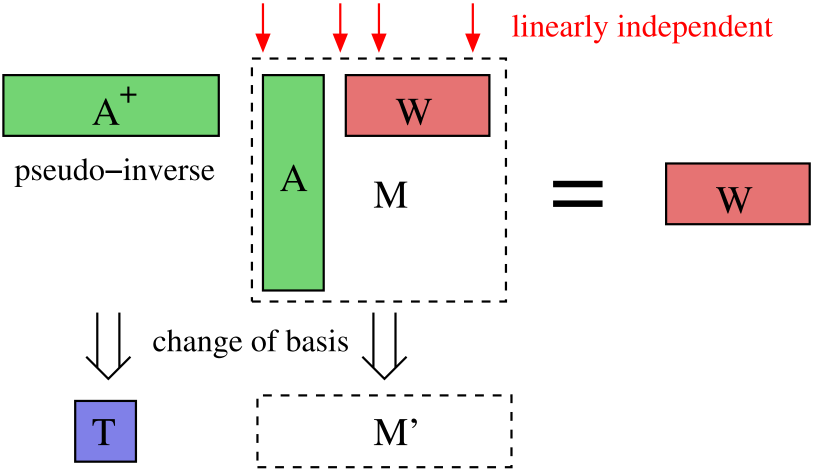
Easy Case: A has Full Column Rank (AGKM)



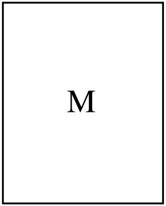
Easy Case: *A* has Full Column Rank (AGKM)



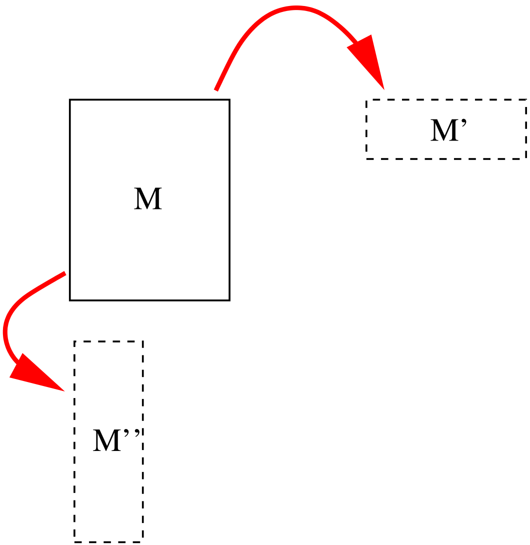
Easy Case: A has Full Column Rank (AGKM)



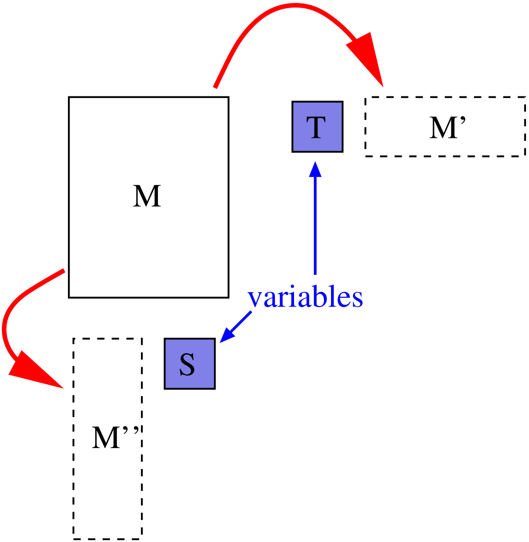
Putting it Together: $2r^2$ Variables



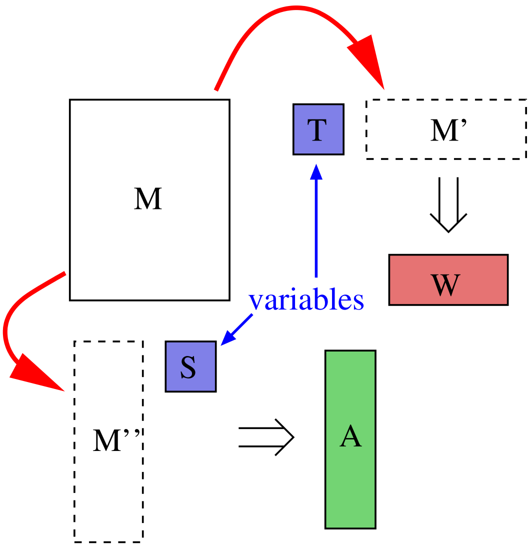
Putting it Together: $2r^2$ Variables



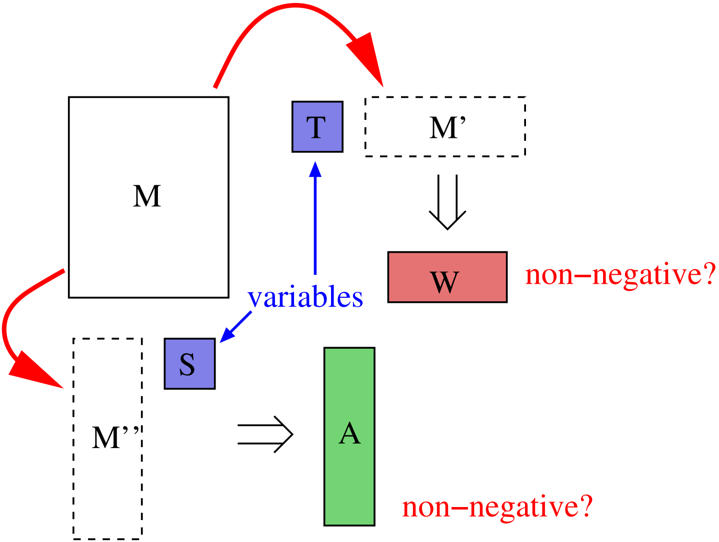
Putting it Together: $2r^2$ Variables



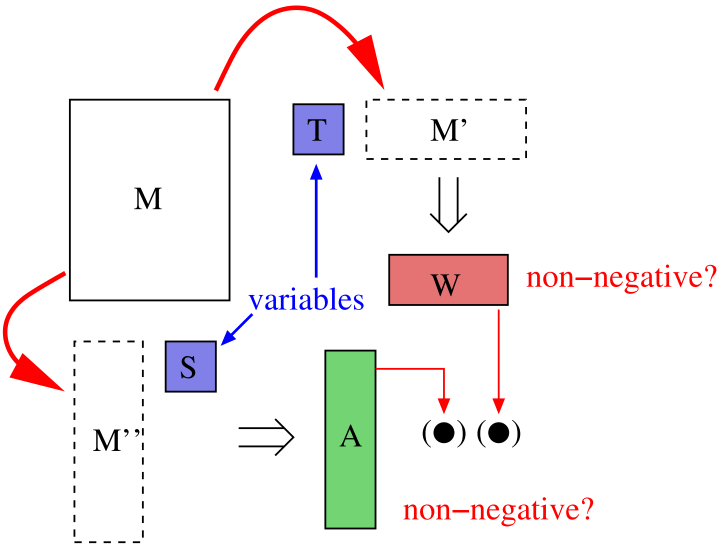
Putting it Together: $2r^2$ Variables



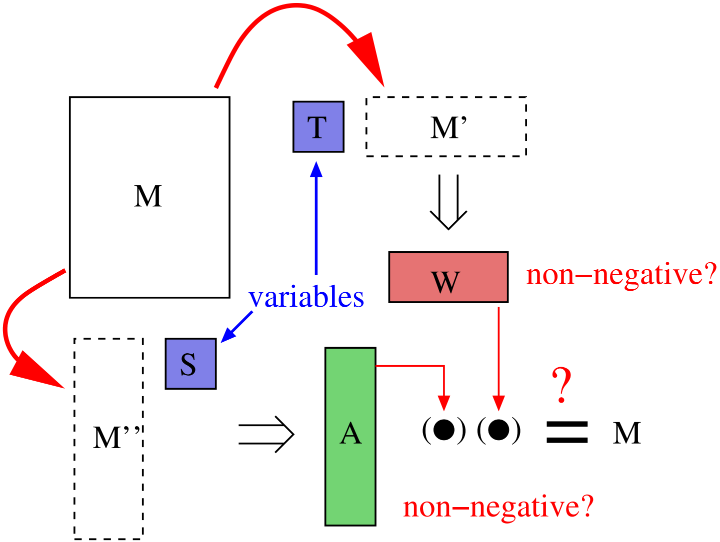
Putting it Together: $2r^2$ Variables



Putting it Together: $2r^2$ Variables



Putting it Together: $2r^2$ Variables



Most interesting case in e.g. extended formulations:

$$\#\text{facets}\{conv(A)\} \gg \#\text{vertices}\{conv(A)\}$$

which can only happen if (say) $rank(A) = r/2$

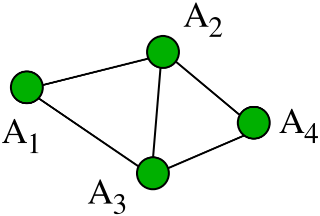
Most interesting case in e.g. extended formulations:

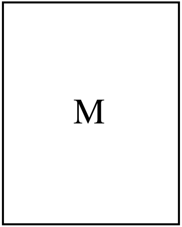
$$\#\text{facets}\{conv(A)\} \gg \#\text{vertices}\{conv(A)\}$$

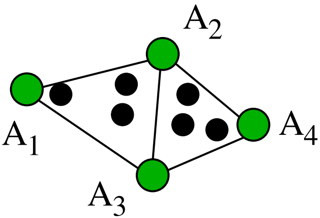
which can only happen if (say) $rank(A) = r/2$

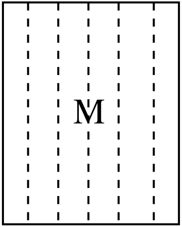
Question

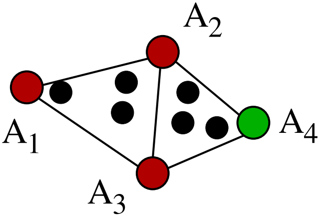
Can we still find the rows of W from **many** *linear transformations of rows of M?*

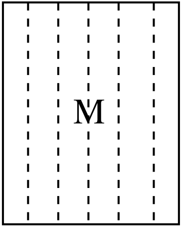

A2
A1
A4
A3


M

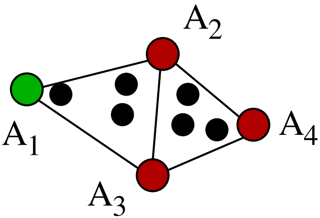

A_2
A_1
A_4
A_3

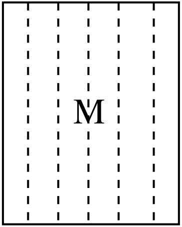

M





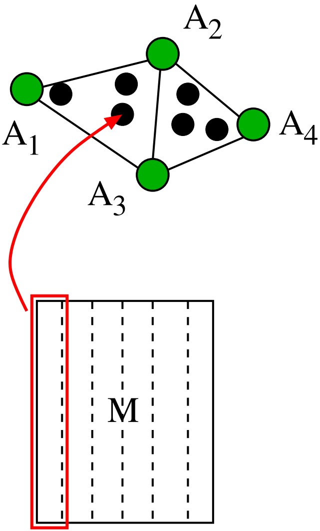
$$T_1 = (A_1 A_2 A_3)^+$$





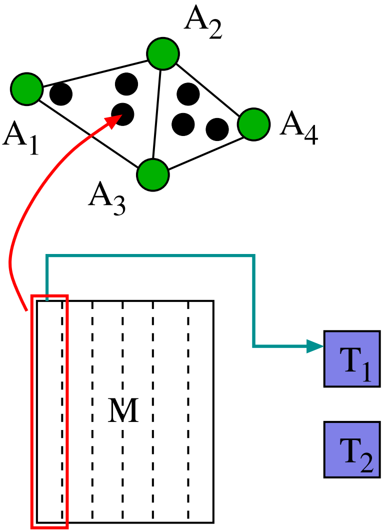
$$T_1 = (A_1 A_2 A_3)^+$$

$$T_2 = (A_2 A_3 A_4)^+$$



$$T_1 = (A_1 A_2 A_3)^+$$

$$T_2 = (A_2 A_3 A_4)^+$$



$$T_1 = \left(A_1 A_2 A_3\right)^+$$

$$T_2 = \left(A_2 A_3 A_4\right)^+$$

Most interesting case in e.g. extended formulations:

$$\#\text{facets}\{conv(A)\} \gg \#\text{vertices}\{conv(A)\}$$

which can only happen if (say) $rank(A) = r/2$

Question

Can we still find the rows of W from **many** *linear transformations of rows of M?*

Most interesting case in e.g. extended formulations:

$$\#\text{facets}\{conv(A)\} \gg \#\text{vertices}\{conv(A)\}$$

which can only happen if (say) $rank(A) = r/2$

Question

Can we still find the rows of W from **many** *linear transformations of rows of M?*

This could require exponentially many (2^r) linear transformations (e.g. covering the cross polytope by simplices)

Most interesting case in e.g. extended formulations:

$$\#\text{facets}\{conv(A)\} \gg \#\text{vertices}\{conv(A)\}$$

which can only happen if (say) $rank(A) = r/2$

Question

Can we still find the rows of W from **many** *linear transformations of rows of M?*

This could require exponentially many (2^r) linear transformations (e.g. covering the cross polytope by simplices)

Key

These linear transformations can be defined using a common set of r^2 variables!

Goal

Encode nonnegative rank as a semi-algebraic set with $2r^2$ variables

Goal

Encode nonnegative rank as a semi-algebraic set with $2r^2$ *variables*

(i.e. the set is non-empty iff $rank^+(M) \leq r$)

Goal

Encode nonnegative rank as a semi-algebraic set with $2r^2$ variables

(i.e. the set is non-empty iff $rank^+(M) \leq r$)

- We give a new **normal form** for nonnegative matrix factorization (that uses exponentially many $r \times r$ unknown linear transformations)

Goal

Encode nonnegative rank as a semi-algebraic set with $2r^2$ variables

(i.e. the set is non-empty iff $rank^+(M) \leq r$)

- We give a new **normal form** for nonnegative matrix factorization (that uses exponentially many $r \times r$ unknown linear transformations)
- We use **Cramer's Rule** to write all of these linear transformations using a common set of r^2 variables

Goal

Encode nonnegative rank as a semi-algebraic set with $2r^2$ variables

(i.e. the set is non-empty iff $rank^+(M) \leq r$)

- We give a new **normal form** for nonnegative matrix factorization (that uses exponentially many $r \times r$ unknown linear transformations)
- We use **Cramer's Rule** to write all of these linear transformations using a common set of r^2 variables
- These transformations recover the factorization A, W, and we can check that it is a valid nonnegative matrix factorization

Goal

Encode nonnegative rank as a semi-algebraic set with $2r^2$ variables

(i.e. the set is non-empty iff $rank^+(M) \leq r$)

- We give a new **normal form** for nonnegative matrix factorization (that uses exponentially many $r \times r$ unknown linear transformations)
- We use **Cramer's Rule** to write all of these linear transformations using a common set of r^2 variables
- These transformations recover the factorization A, W, and we can check that it is a valid nonnegative matrix factorization

Theorem

The nonnegative rank can be computed in time $(nm)^{O(r^2)}$.

Concluding Remarks

This algorithm is based on answering a purely algebraic question: How many variables do we need in a semi-algebraic set to encode nonnegative rank?

Concluding Remarks

This algorithm is based on answering a purely algebraic question: How many variables do we need in a semi-algebraic set to encode nonnegative rank?

Question

Are there other examples of a better understanding of the expressive power of semi-algebraic sets can lead to a new algorithm?

Concluding Remarks

This algorithm is based on answering a purely algebraic question: How many variables do we need in a semi-algebraic set to encode nonnegative rank?

Question

Are there other examples of a better understanding of the expressive power of semi-algebraic sets can lead to a new algorithm?

Similar: recent work [Anandkumar et al] on topic modeling is based on understanding how many moments are needed to find the parameters.

Concluding Remarks

This algorithm is based on answering a purely algebraic question: How many variables do we need in a semi-algebraic set to encode nonnegative rank?

Question

Are there other examples of a better understanding of the expressive power of semi-algebraic sets can lead to a new algorithm?

Similar: recent work [Anandkumar et al] on topic modeling is based on understanding how many moments are needed to find the parameters.

Observation

The number of variables plays an analogous role to VC-dimension

Concluding Remarks

This algorithm is based on answering a purely algebraic question: How many variables do we need in a semi-algebraic set to encode nonnegative rank?

Question

Are there other examples of a better understanding of the expressive power of semi-algebraic sets can lead to a new algorithm?

Similar: recent work [Anandkumar et al] on topic modeling is based on understanding how many moments are needed to find the parameters.

Observation

The number of variables plays an analogous role to VC-dimension

Is there an elementary proof of the Milnor-Warren bound?

Any Questions?

Thanks!