A Quadratically Convergent Algorithm for Structured Low-Rank Approximation

Éric Schost, Pierre-Jean Spaenlehauer

Western University, London, Canada
Max Planck Institute for Mathematics, Bonn, Germany

RAIM
IHP, Nov. 20, 2013

Problem Statement

$p, q, r \in \mathbb{N}$

E a **linear/affine subspace** of $p \times q$ matrices with real entries

For $(M_{i,j})$ a $p \times q$ matrix, $\|M\|_F = \sqrt{\sum_{i,j} M_{i,j}^2}$,

$\langle M_1, M_2 \rangle = \mathrm{trace}(M_1 \cdot M_2^{\intercal})$

Problem Statement

$p, q, r \in \mathbb{N}$

E a **linear/affine subspace** of $p \times q$ matrices with real entries

For $(M_{i,j})$ a $p \times q$ matrix, $\|M\|_F = \sqrt{\sum_{i,j} M_{i,j}^2}$,

$\langle M_1, M_2 \rangle = \mathrm{trace}(M_1 \cdot M_2^{\mathsf{T}})$

Structured Low-Rank Approximation

Given $M \in E$, compute **a matrix** $\hat{M} \in E$ such that

- $\mathrm{Rank}(\hat{M}) \leq r$;
- $\left\| M - \hat{M} \right\|_F$ is **small**.

Problem Statement

$p, q, r \in \mathbb{N}$

E a **linear/affine subspace** of $p \times q$ matrices with real entries

For $(M_{i,j})$ a $p \times q$ matrix, $\|M\|_F = \sqrt{\sum_{i,j} M_{i,j}^2}$,

$\langle M_1, M_2 \rangle = \mathrm{trace}(M_1 \cdot M_2^T)$

Structured Low-Rank Approximation

Given $M \in E$, compute **a matrix** $\hat{M} \in E$ such that

- $\mathrm{Rank}(\hat{M}) \leq r$;
- $\left\| M - \hat{M} \right\|_F$ is **small**.

"Behind every linear data modeling problem there is a (hidden) low-rank approximation problem: the model imposes relations on the data which render a matrix constructed from exact data rank deficient."

Markovsky, 08

Examples and applications

- E =**Sylvester matrices** $\rightsquigarrow$ univariate approximate GCD

$$\begin{bmatrix} a_3 & 0 & b_2 & 0 & 0 \\ a_2 & a_3 & b_1 & b_2 & 0 \\ a_1 & a_2 & b_0 & b_1 & b_2 \\ a_0 & a_1 & 0 & b_0 & b_1 \\ 0 & a_0 & 0 & 0 & b_0 \end{bmatrix}$$

Examples and applications

- $E =$ **Sylvester matrices** $\rightsquigarrow$ univariate approximate GCD
- $E =$ **Hankel matrices** $\rightsquigarrow$ denoising, signal processing

$$\begin{bmatrix} a & b & c & d & e \\ b & c & d & e & f \\ c & d & e & f & g \\ d & e & f & g & h \\ e & f & g & h & i \end{bmatrix}$$

Examples and applications

- $E =$ **Sylvester matrices** $\rightsquigarrow$ univariate approximate GCD
- $E =$ **Hankel matrices** $\rightsquigarrow$ denoising, signal processing
- $E =$ affine **coordinate spaces** $\rightsquigarrow$ matrix completion

$$\begin{bmatrix} 3 & ? & ? & 5 & 5 \\ 1 & 2 & 3 & 2 & ? \\ 10 & 4 & ? & 9 & -4 \\ 6 & ? & 3 & 9 & 10 \\ ? & 5 & -2 & ? & 9 \end{bmatrix}$$

Examples and applications

- $E =$ **Sylvester matrices** $\rightsquigarrow$ univariate approximate GCD
- $E =$ **Hankel matrices** $\rightsquigarrow$ denoising, signal processing
- $E =$ affine **coordinate spaces** $\rightsquigarrow$ matrix completion
- $E =$ **Ruppert matrices** $\rightsquigarrow$ multivariate factorization

$$\begin{bmatrix} 0 & -2 & -a & 0 & -2b & -d \\ -1 & 0 & c & -b & 0 & e \\ a & 2c & 0 & d & 2e & 0 \\ 0 & 0 & 0 & 1 & a & c \\ 0 & 0 & 0 & -b & -d & -e \end{bmatrix}$$

$XY^2 + aXY + bY^2 + cX + dY + e \in \mathbb{C}[X, Y]$ factors $\Leftrightarrow \mathrm{rank} \leq 4$

Main results

$\mathscr{D}_r$: **manifold** of $p \times q$ matrices of rank r
E: **linear/affine subspace** of $p \times q$ matrices

Algorithm NewtonSLRA

`NewtonSLRA`: iterative algorithm with proven local quadratic convergence under mild **transversality assumptions**.

Main results

$\mathscr{D}_r$: **manifold** of $p \times q$ matrices of rank r
E: **linear/affine subspace** of $p \times q$ matrices

Algorithm NewtonSLRA

NewtonSLRA: iterative algorithm with proven local quadratic convergence under mild **transversality assumptions**.

More precisely: for any smooth point $\zeta \in \mathscr{D}_r \cap E$ where $\mathscr{D}_r$ and E **intersect transversely**, there exists a small neighborhood $U \supset \zeta$ such that for any input matrix $M_0 \in U$,

- the sequence of iterates $M_1, M_2, \ldots$ **converges quadratically** towards $M_\infty \in \mathscr{D}_r \cap E$, *i.e.* $\|M_i - M_\infty\| \leq (1/2)^{2^i - 1} \|M_0 - M_\infty\|$

Main results

$\mathscr{D}_r$: **manifold** of $p \times q$ matrices of rank r
E: **linear/affine subspace** of $p \times q$ matrices

Algorithm NewtonSLRA

NewtonSLRA: iterative algorithm with proven local quadratic convergence under mild **transversality assumptions**.

More precisely: for any smooth point $\zeta \in \mathscr{D}_r \cap E$ where $\mathscr{D}_r$ and E **intersect transversely**, there exists a small neighborhood $U \supset \zeta$ such that for any input matrix $M_0 \in U$,

- the sequence of iterates $M_1, M_2, \ldots$ **converges quadratically** towards $M_\infty \in \mathscr{D}_r \cap E$, *i.e.* $\|M_i - M_\infty\| \leq (1/2)^{2^i - 1} \|M_0 - M_\infty\|$
- Let $\hat{M}$ be the **nearest solution**; then $\left\|M_\infty - \hat{M}\right\| = O(\mathrm{dist}(M_0, \mathscr{D}_r \cap E)^2)$.

Singular Value Decomposition

Eckart-Young theorem

Let $M = U \cdot S \cdot V^\intercal$ be the **Singular Value Decomposition** of M, where $S = \mathrm{Diag}(\sigma_1, \ldots, \sigma_q)$ with $\sigma_1 \geq \cdots \geq \sigma_q$.

Set $\widehat{S} = \mathrm{Diag}(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0)$.

Then $\widehat{M} = U \cdot \widehat{S} \cdot V^\intercal$ is the rank r matrix which **minimizes the Frobenius distance to M**.

Singular Value Decomposition

Eckart-Young theorem

Let $M = U \cdot S \cdot V^\intercal$ be the **Singular Value Decomposition** of M, where $S = \mathrm{Diag}(\sigma_1, \ldots, \sigma_q)$ with $\sigma_1 \geq \cdots \geq \sigma_q$.

Set $\widehat{S} = \mathrm{Diag}(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0)$.

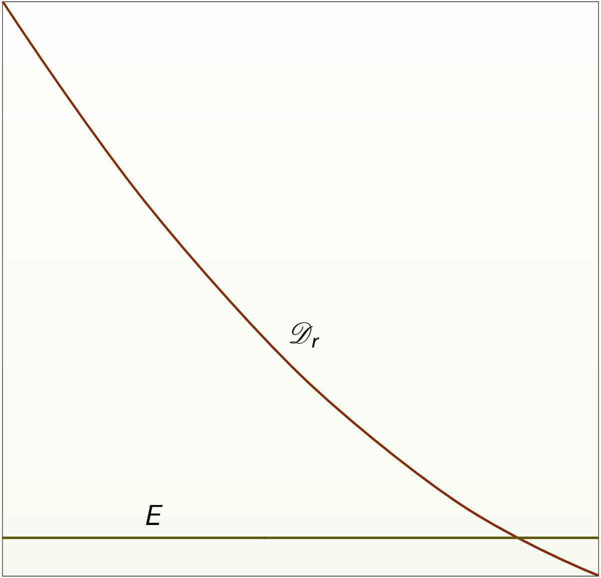
Then $\widehat{M} = U \cdot \widehat{S} \cdot V^\intercal$ is the rank r matrix which **minimizes the Frobenius distance to M**.

Cadzow's algorithm (*Cadzow, 88, Lewis/Malick 08*):

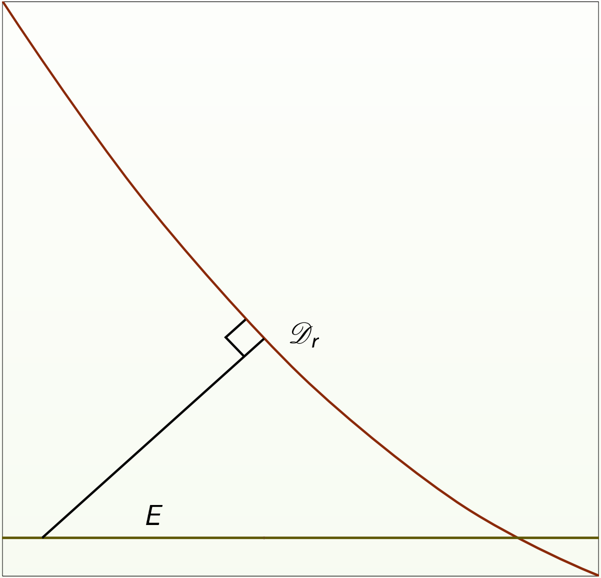
- **project** on $\mathscr{D}_r$ (the **manifold** of matrices of rank r) with **SVD**;
- **project** back on E.

Converges **linearly** towards a point in $\mathscr{D}_r \cap E$.
Does not converge to the **nearest solution**.

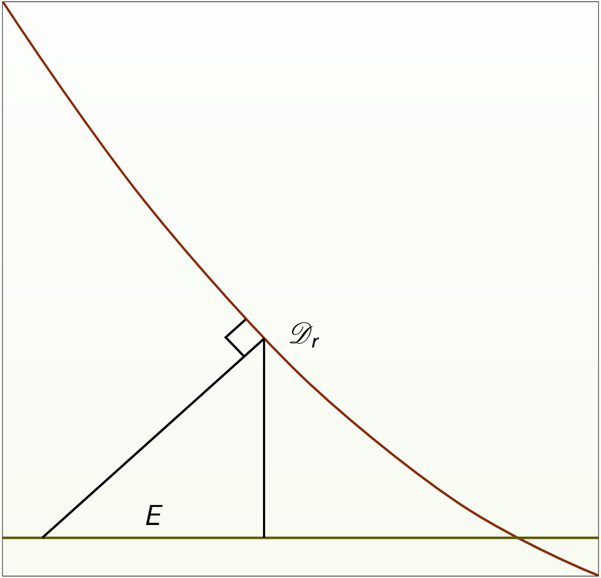
Cadzow's algorithm



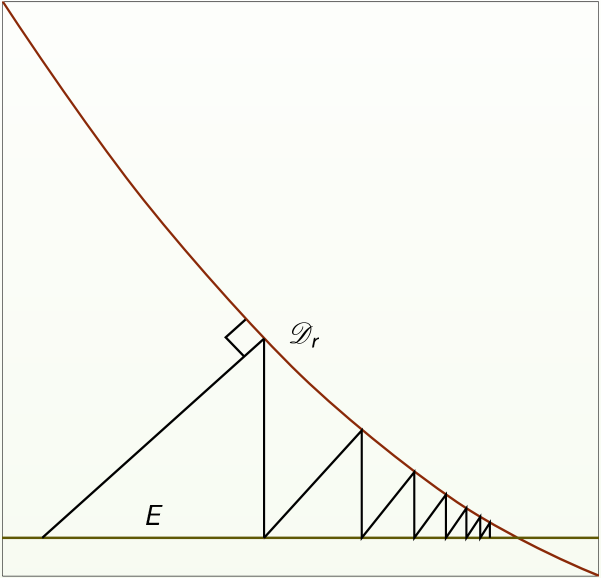
Cadzow's algorithm



Cadzow's algorithm



Cadzow's algorithm



Newton's method

Classical **Newton's method** for $f : \mathbb{R}^n \to \mathbb{R}^n$

$$N_f(x) = Df(x)^{-1}(f(x)).$$

Quadratic convergence when Df is locally invertible.

Newton's method

Classical **Newton's method** for $f : \mathbb{R}^n \to \mathbb{R}^n$

$$N_f(x) = Df(x)^{-1}(f(x)).$$

Quadratic convergence when Df is locally invertible.

Newton's method for **underdetermined systems**:
$f : \mathbb{R}^m \to \mathbb{R}^n$, $N_f(x) = Df(x)^{\dagger}(f(x))$.
$Df^{\dagger}$: **Moore-Penrose pseudo-inverse**.

Newton's method

Classical **Newton's method** for $f : \mathbb{R}^n \to \mathbb{R}^n$

$$N_f(x) = Df(x)^{-1}(f(x)).$$

Quadratic convergence when Df is locally invertible.

Newton's method for **underdetermined systems**:
$f : \mathbb{R}^m \to \mathbb{R}^n$, $N_f(x) = Df(x)^\dagger(f(x))$.
$Df^\dagger$: **Moore-Penrose pseudo-inverse**.

Quadratic convergence towards a point x_∞ such that $f(x_\infty) = 0$ when Df is locally surjective.

Newton's method

Classical **Newton's method** for $f : \mathbb{R}^n \to \mathbb{R}^n$

$$N_f(x) = Df(x)^{-1}(f(x)).$$

Quadratic convergence when Df is locally invertible.

Newton's method for **underdetermined systems**:
$f : \mathbb{R}^m \to \mathbb{R}^n$, $N_f(x) = Df(x)^{\dagger}(f(x))$.
$Df^{\dagger}$: **Moore-Penrose pseudo-inverse**.

Quadratic convergence towards a point x_∞ such that $f(x_\infty) = 0$ when Df is locally surjective.
If x_0 is the starting point of the iteration, let

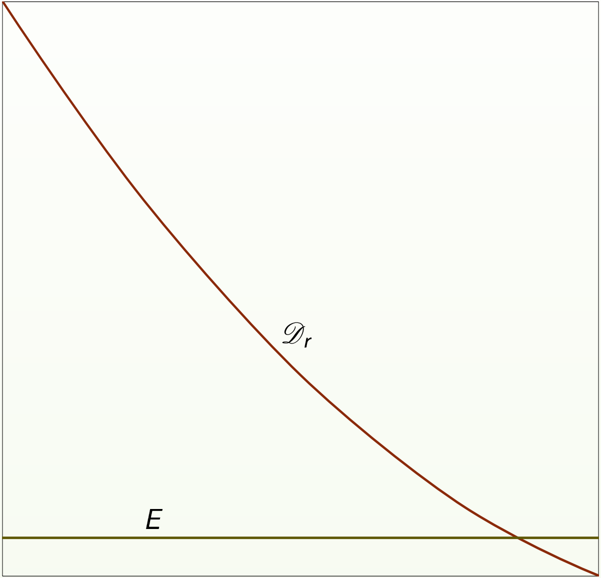
$$\hat{x} = \operatorname{argmin}_{f(y)=0} \|y - x_0\|.$$

Does not converge to the nearest solution $\hat{x}$, but

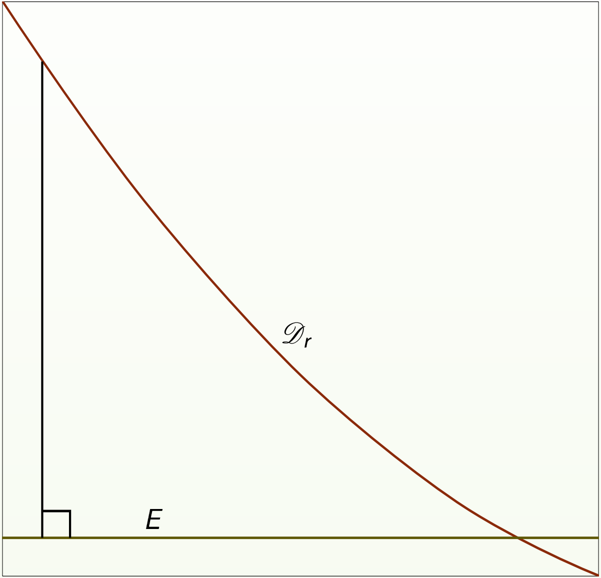
$$\|x_\infty - \hat{x}\| = O\left(\|x_0 - \hat{x}\|^2\right).$$

Ben-Israel 66, Allgower/Georg 90, Beyn 93, Shub/Smale 96, Dedieu/Shub 00, Dedieu/Kim 02, Dedieu 06

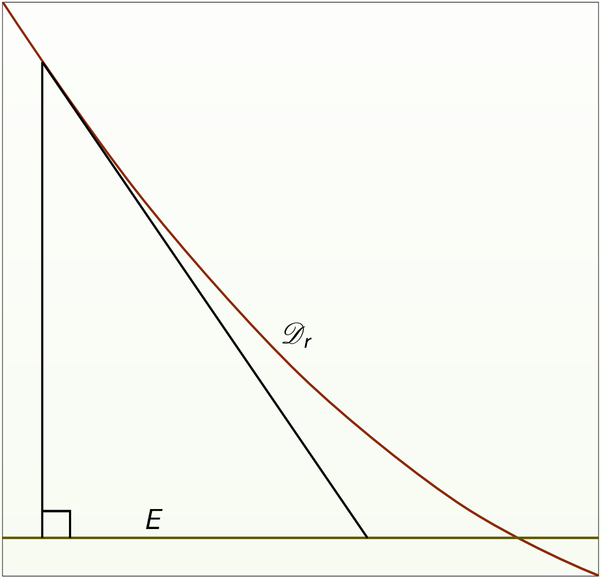
Newton's method



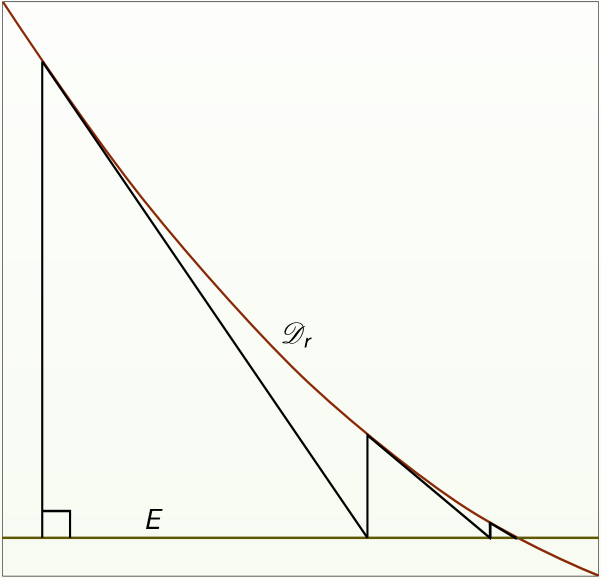
Newton's method



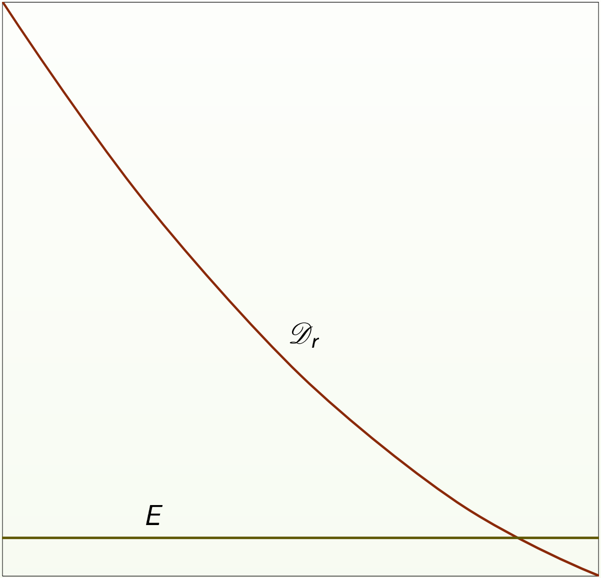
Newton's method



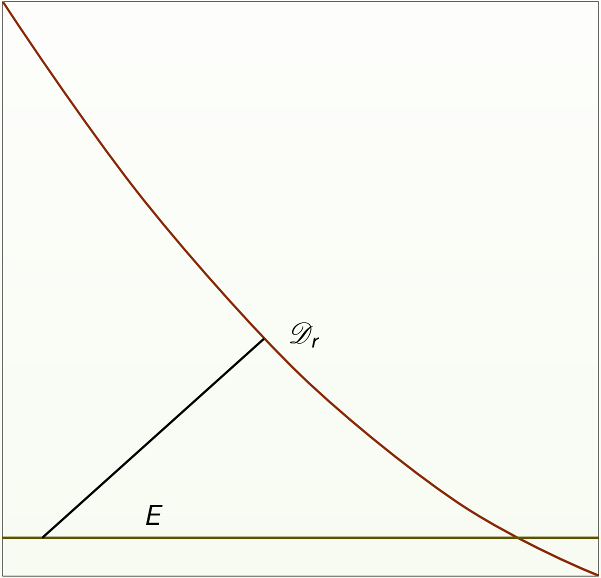
Newton's method

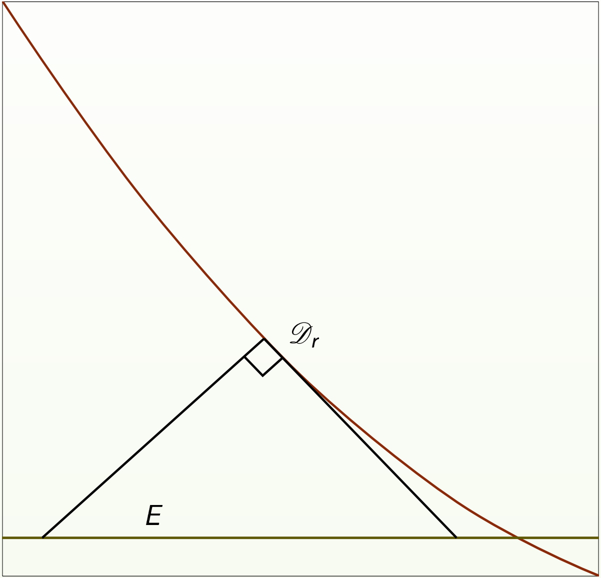


NewtonSLRA

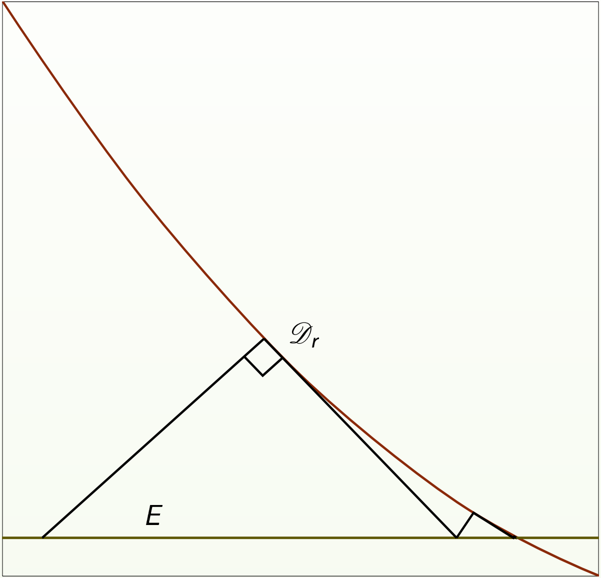


NewtonSLRA




$\mathscr{D}_r$
E

NewtonSLRA



Normal space of determinantal varieties

$\overline{\mathscr{D}_r}$: **algebraic variety** of matrices of rank **at most** r.
$\rightsquigarrow$ well-studied in algebraic geometry/commutative algebra
Bruns, Conca, Eisenbud, Herzog, Lascoux, Room, Sturmfels,...

Normal space of determinantal varieties

$\overline{\mathscr{D}_r}$: **algebraic variety** of matrices of rank **at most** r.
$\rightsquigarrow$ well-studied in algebraic geometry/commutative algebra
Bruns, Conca, Eisenbud, Herzog, Lascoux, Room, Sturmfels,...

Classical theorem

Let M be $p \times q$ matrix of rank r.
Then the **normal space** to $\mathscr{D}_r$ at M is

$$\mathrm{Ker}(M^\intercal) \otimes \mathrm{Ker}(M).$$

Bases of the kernels of M and $M^\intercal$ can be read off from the Singular Value Decomposition of M.

NewtonSLRA

procedure NewtonSLRA($M \in E$, $(E_1, \dots, E_d)$ an orthonormal basis of E, $r \in \mathbb{N}$)

  $(U, S, V) \leftarrow \mathrm{SVD}(M)$

  $S_r \leftarrow r \times r$ top-left submatrix of S

  $U_r \leftarrow$ first r columns of U

  $V_r \leftarrow$ first r columns of V

  $\widetilde{M} \leftarrow U_r \cdot S_r \cdot V_r^\intercal$

  $\widetilde{u_1}, \dots, \widetilde{u_{p-r}} \leftarrow$ last $p - r$ columns of U

  $\widetilde{v_1}, \dots, \widetilde{v_{q-r}} \leftarrow$ last $q - r$ columns of V

  for $i \in \{1, \dots, p - r\}, j \in \{1, \dots, q - r\}$ **do**

    $N_{(i-1)(q-r)+j} \leftarrow \widetilde{u_i} \cdot \widetilde{v_j}^\intercal$

  end for

  $A \leftarrow (\langle N_i, E_j \rangle)_{i,j}$

  $b \leftarrow (\langle N_i, \widetilde{M} - M \rangle)_i$

  return $M + \begin{bmatrix} E_1 & \dots & E_d \end{bmatrix} \cdot A^\dagger \cdot b$

end procedure

Rate of convergence

Quadratic convergence

For any **smooth** point $\zeta \in \mathscr{D}_r \cap E$ where $\mathscr{D}_r$ and E intersect **transversely**, there exists a small neighborhood $U \supset \zeta$ such that for any input matrix $M_0 \in U$,

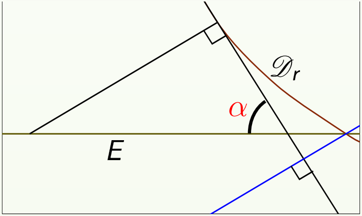
- the sequence of iterates $M_1, M_2, \ldots$ **converges quadratically** towards a matrix $M_\infty \in \mathscr{D}_r \cap E$, *i.e.* $\|M_i - M_\infty\| \leq (1/2)^{2^i-1} \|M_0 - M_\infty\|$
- Let $\hat{M}$ be the **nearest solution**; then $\left\|M_\infty - \hat{M}\right\| = O(\mathrm{dist}(M_0, \mathscr{D}_r \cap E)^2)$.

Rate of convergence

Quadratic convergence

For any **smooth** point $\zeta \in \mathscr{D}_r \cap E$ where $\mathscr{D}_r$ and E intersect **transversely**, there exists a small neighborhood $U \supset \zeta$ such that for any input matrix $M_0 \in U$,

- the sequence of iterates $M_1, M_2, \ldots$ **converges quadratically** towards a matrix $M_\infty \in \mathscr{D}_r \cap E$, *i.e.* $\|M_i - M_\infty\| \leq (1/2)^{2^i-1} \|M_0 - M_\infty\|$
- Let $\hat{M}$ be the **nearest solution**; then $\left\|M_\infty - \hat{M}\right\| = O(\text{dist}(M_0, \mathscr{D}_r \cap E)^2)$.



Sketch of proof:

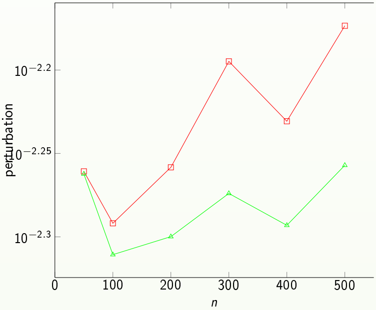
- lower bound for α;
- Taylor approximation of $\Pi_{\mathscr{D}_r}$;
- manage corrective terms when $\dim(\mathscr{D}_r \cap E) > 0$.

Other features of NewtonSLRA

- Combines the **generality of alternating projections** and the **quadratic convergence of Newton's method**.
- Computationally most intensive step: computing the SVD (polynomial in p, q at fixed precision).
- Algorithm for SLRA with **proven quadratic rate of convergence**.

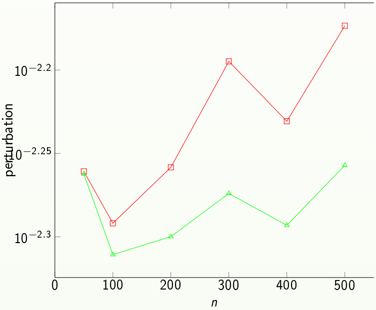
Comparison with GPGCD, *Terui, ISSAC'09*.

	sizes of iteration steps	
iteration	NewtonSLRA	GPGCD
1	0.9e-1	0.9e-1
2	0.5e-3	0.5e-3
3	0.6e-8	0.2e-5
4	0.1e-17	0.8e-8
5	0.1e-36	0.4e-10



Comparison with GPGCD, *Terui, ISSAC'09*.

	sizes of iteration steps	
iteration	NewtonSLRA	GPGCD
1	0.9e-1	0.9e-1
2	0.5e-3	0.5e-3
3	0.6e-8	0.2e-5
4	0.1e-17	0.8e-8
5	0.1e-36	0.4e-10



Fast convergence towards $\mathscr{D}_r \cap E$

$\rightsquigarrow$ starting point for a certified **Gauss-Newton iteration**

Yakoubsohn/Masmoudi/Chèze/Auroux 06

Conclusions and perspectives

Linear sections of determinantal varieties

rich **structure** with a lot of facets
(numeric/symbolic, finite fields/characteristic 0, real solutions)
which appears in many applications.

- Low-rank matrix completion, Hankel matrices.
- **Algebraic** properties of special linear subspaces
  $\rightsquigarrow$ Euclidean distance degree, *Ottaviani/S./Sturmfels '13*.
- **Certification** of `NewtonSLRA` *a la Dedieu*: α,γ theorems?

Conclusions and perspectives

Linear sections of determinantal varieties

rich **structure** with a lot of facets
(numeric/symbolic, finite fields/characteristic 0, real solutions)
which appears in many applications.

- Low-rank matrix completion, Hankel matrices.
- **Algebraic** properties of special linear subspaces
  $\rightsquigarrow$ Euclidean distance degree, *Ottaviani/S./Sturmfels '13*.
- **Certification** of `NewtonSLRA` *a la Dedieu*: α,γ theorems?

Thank you!

Application to approximate GCD

Approximate GCD

Let $m, n, d \in \mathbb{N}$, $f, g \in \mathbb{R}[x]$ with $\deg(f) = m, \deg(g) = n$.
Find $f^*, g^* \in \mathbb{R}[x]$, $\deg(f^*) = m$, $\deg(g^*) = n$ such that

$$\deg(\mathrm{GCD}(f^*, g^*)) \geq d$$

and (f^*, g^*) are close to (f, g).

- needs a **distance** on the pairs (f, g):

$$\|(\sum_{i=0}^{m} f_i x^i, \sum_{j=0}^{n} g_j x^j)\|^2 = \sum_{i=0}^{m} f_i^2 + \sum_{j=0}^{n} g_j^2.$$

- What does **"close"** mean
  $\rightsquigarrow$ quasi-GCD, *Schönage 85*
  $\rightsquigarrow$ ε-GCD, *Emiris/Galligo/Lombardi 97, Zeng/Dayton 04, Bini/Boito 06-09*
  $\rightsquigarrow$ nearest pair for a given norm, *Karmarkar/Lakshman 98, Kaltofen/Zhi/Yang 05-08, Terui 09*

Matrix Completion

Unknown matrix of **rank** r:

$$\begin{bmatrix} ? & ? & ? & ? \\ ? & ? & ? & ? \\ ? & ? & ? & ? \\ ? & ? & ? & ? \end{bmatrix}$$

Uncover m **entries** at random.

Matrix Completion

Unknown matrix of **rank** r:

$$\begin{bmatrix} ? & ? & ? & ? \\ ? & ? & ? & ? \\ ? & ? & ? & ? \\ ? & ? & ? & ? \end{bmatrix}$$

Uncover m **entries** at random.

Matrix Completion

Unknown matrix of **rank** r:

$$\begin{bmatrix} ? & 4 & ? & ? \\ ? & ? & ? & ? \\ ? & ? & ? & ? \\ ? & ? & ? & ? \end{bmatrix}$$

Uncover m **entries** at random.

Matrix Completion

Unknown matrix of **rank** r:

$$\begin{bmatrix} ? & 4 & ? & ? \\ ? & ? & ? & ? \\ 1 & ? & ? & ? \\ ? & ? & ? & ? \end{bmatrix}$$

Uncover m **entries** at random.

Matrix Completion

Unknown matrix of **rank** r:

$$\begin{bmatrix} ? & 4 & ? & ? \\ ? & ? & 7 & ? \\ 1 & ? & ? & ? \\ ? & ? & ? & ? \end{bmatrix}$$

Uncover m **entries** at random.

Matrix Completion

Unknown matrix of **rank** r:

$$\begin{bmatrix} ? & 4 & ? & ? \\ ? & ? & 7 & ? \\ 1 & ? & 9 & ? \\ ? & ? & ? & ? \end{bmatrix}$$

Uncover m **entries** at random.

Matrix Completion

Unknown matrix of **rank** r:

$$\begin{bmatrix} ? & 4 & ? & ? \\ ? & ? & 7 & ? \\ 1 & ? & 9 & ? \\ ? & ? & ? & 7 \end{bmatrix}$$

Uncover m **entries** at random.

Matrix Completion

Unknown matrix of **rank** r:

$$\begin{bmatrix} ? & 4 & ? & ? \\ ? & ? & 7 & ? \\ 1 & ? & 9 & ? \\ ? & ? & ? & 7 \end{bmatrix}$$

Uncover m **entries** at random.
How many entries do we need? How to reconstruct the matrix?

Matrix Completion

Unknown matrix of **rank** r:

$$\begin{bmatrix} ? & 4 & ? & ? \\ ? & ? & 7 & ? \\ 1 & ? & 9 & ? \\ ? & ? & ? & 7 \end{bmatrix}$$

Uncover m **entries** at random.
How many entries do we need? How to reconstruct the matrix?

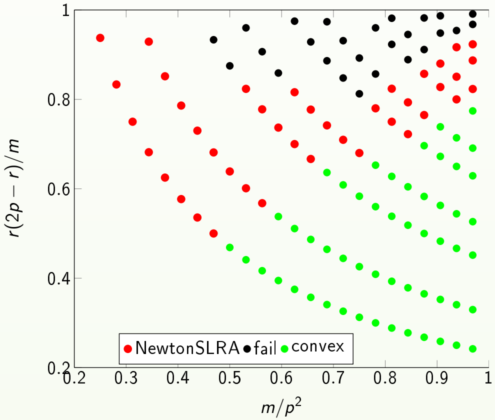
- Alternating minimization, *Jain, Netrapalli, Sanghavi, 12*
- Riemannian optimization, *Absil/Amodei/Meyer 12, Vandereycken 12*
- Convex relaxation, *Candes, Tao, Plan, Recht, 09-13*

Experimental results

Overdetermined SLRA problems

Transversality assumption do not hold $\rightsquigarrow$ no quadratic convergence.

Square matrix of size $p = 40$



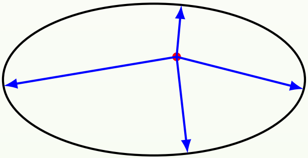
An algebraic approach to SLRA

The Euclidean distance degree
Draisma/Horobet/Ottaviani/Sturmfels/Thomas 13

$V \in \mathbb{C}^n$ an algebraic variety, $\mathbf{u} \in \mathbb{C}^n$ a generic point. The **EDdegree** of V is the number of **complex critical points** of the function

$$\lambda_1(x_1 - u_1)^2 + \cdots + \lambda_n(x_n - u_n)^2$$

on the smooth locus of V.



$\mathrm{EDdegree}(\text{ellipse}) = 4.$

Nearest solution of SLRA:

critical point of the distance function on a **linear section of a determinantal variety** $\mathscr{D}_r \cap E$.

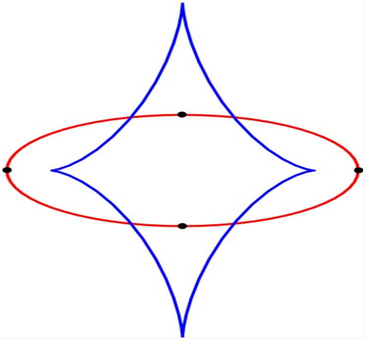
EDdegree of linear sections of determinantal varieties

Proposition (Draisma/Horobet/Ottaviani/Sturmfels/Thomas)

Under **transversality assumptions** with a special quadric and for generic weights, the **EDdegree** of a projective variety is the sum of the degrees of its **polar classes**.

How many **real solutions**? Important information for numerical algorithms.

$\rightsquigarrow$ **ED discriminant**



What happens if the variety is a **generic/special linear section of a determinantal variety**?