Feasibility problems: from alternating projections to matrix completions

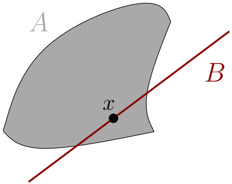
Dmitriy Drusvyatskiy
C&O, Waterloo
Mathematics, Washington

Joint work with
V. Cheung (Waterloo), A.D. Ioffe (Technion), N. Krislock (NIU),
A.S. Lewis (Cornell), and H. Wolkowicz (Waterloo)

June 18, 2014

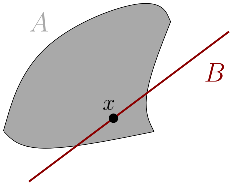
Method of alternating projections

Given sets $A, B \subset \mathbf{R}^n$, find some point $x \in A \cap B$.



Method of alternating projections

Given sets $A, B \subset \mathbf{R}^n$, find some point $x \in A \cap B$.

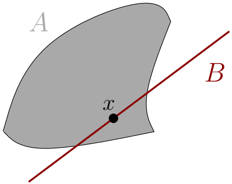


Distance and projection:

$$d_B(x) = \min_{y \in B} |x - y| \quad \text{and} \quad \mathcal{P}_B(x) = \{\text{nearest points of } B \text{ to } x\}.$$

Method of alternating projections

Given sets $A, B \subset \mathbf{R}^n$, find some point $x \in A \cap B$.



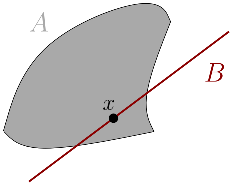
Distance and projection:

$$d_B(x) = \min_{y \in B} |x - y| \quad \text{and} \quad \mathcal{P}_B(x) = \{\text{nearest points of } B \text{ to } x\}.$$

Finding points in $\mathcal{P}_A(x)$ and $\mathcal{P}_B(x)$ is often easy!

Method of alternating projections

Given sets $A, B \subset \mathbf{R}^n$, find some point $x \in A \cap B$.



Distance and projection:

$$d_B(x) = \min_{y \in B} |x - y| \quad \text{and} \quad \mathcal{P}_B(x) = \{\text{nearest points of } B \text{ to } x\}.$$

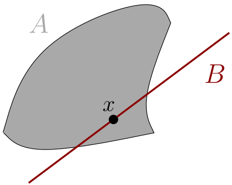
Finding points in $\mathcal{P}_A(x)$ and $\mathcal{P}_B(x)$ is often easy!

Eg: Compressed sensing

$$\{x : \operatorname{rank} x \leq r\} \bigcap \{x : Ax = b\}.$$

Method of alternating projections

Given sets $A, B \subset \mathbf{R}^n$, find some point $x \in A \cap B$.



Distance and projection:

$$d_B(x) = \min_{y \in B} |x - y| \quad \text{and} \quad \mathcal{P}_B(x) = \{\text{nearest points of } B \text{ to } x\}.$$

Finding points in $\mathcal{P}_A(x)$ and $\mathcal{P}_B(x)$ is often easy!

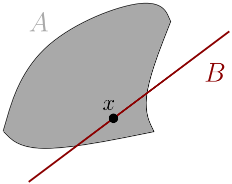
Eg: Compressed sensing

$$\{x : \operatorname{rank} x \leq r\} \bigcap \{x : Ax = b\}.$$

- $\mathcal{P}_{\{x:\, \operatorname{rank} x \leq r\}}(y) \iff$ keep r largest coordinates in absolute value.

Method of alternating projections

Given sets $A, B \subset \mathbf{R}^n$, find some point $x \in A \cap B$.



Distance and projection:

$$d_B(x) = \min_{y \in B} |x - y| \quad \text{and} \quad \mathcal{P}_B(x) = \{\text{nearest points of } B \text{ to } x\}.$$

Finding points in $\mathcal{P}_A(x)$ and $\mathcal{P}_B(x)$ is often easy!

Eg: Compressed sensing

$$\{x : \operatorname{rank} x \le r\} \bigcap \{x : Ax = b\}.$$

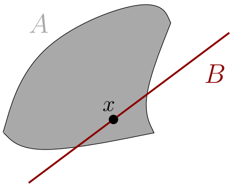
- $\mathcal{P}_{\{x:\, \operatorname{rank} x \le r\}}(y) \iff$ keep r largest coordinates in absolute value.

Eg: Low-order control

$$\{X \succeq 0 : \operatorname{rank} X \le r\} \bigcap \{X : \mathcal{A}(X) = b\}$$

Method of alternating projections

Given sets $A, B \subset \mathbf{R}^n$, find some point $x \in A \cap B$.



Distance and projection:

$$d_B(x) = \min_{y \in B} |x - y| \quad \text{and} \quad \mathcal{P}_B(x) = \{\text{nearest points of } B \text{ to } x\}.$$

Finding points in $\mathcal{P}_A(x)$ and $\mathcal{P}_B(x)$ is often easy!

Eg: Compressed sensing

$$\{x : \operatorname{rank} x \le r\} \bigcap \{x : Ax = b\}.$$

- $\mathcal{P}_{\{x:\, \operatorname{rank} x \le r\}}(y) \iff$ keep r largest coordinates in absolute value.

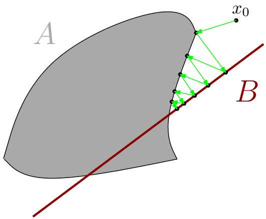
Eg: Low-order control

$$\{X \succeq 0 : \operatorname{rank} X \le r\} \bigcap \{X : \mathcal{A}(X) = b\}$$

- $\mathcal{P}_{\{X \succeq 0:\, \operatorname{rank} X \le r\}}(Y) \iff$ diagonalize, keep r largest positive eigenvalues.

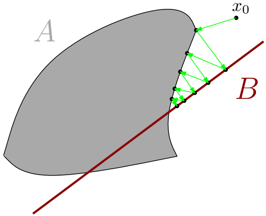
Method of alternating projections

$$\begin{array}{|lcl|}\hline x_{k+1} & \in & \mathcal{P}_B(x_k) \\ x_{k+2} & \in & \mathcal{P}_A(x_{k+1}) \\ \hline\end{array}$$



Method of alternating projections

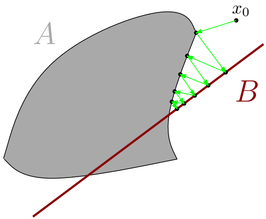
$$\boxed{\begin{array}{lcl} x_{k+1} & \in & \mathcal{P}_B(x_k) \\ x_{k+2} & \in & \mathcal{P}_A(x_{k+1}) \end{array}}$$



The "angle" between A and B drives the convergence!

Method of alternating projections

$$\boxed{\begin{array}{lcl} x_{k+1} & \in & \mathcal{P}_B(x_k) \\ x_{k+2} & \in & \mathcal{P}_A(x_{k+1}) \end{array}}$$

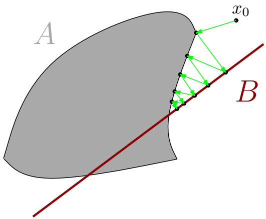


The "angle" between A and B drives the convergence!

von Neumann '33, Aronszajn '50 for subspaces
Bregman '65, Gubin-Polyak-Raik '67 for convex sets
Lewis-Malick '07, Bauschke et al '11 for nonconvex sets
Noll-Rondepierre '13 for nonconvex semi-algebraic sets

Method of alternating projections

$$\begin{array}{lcl} x_{k+1} & \in & \mathcal{P}_B(x_k) \\ x_{k+2} & \in & \mathcal{P}_A(x_{k+1}) \end{array}$$



The "angle" between A and B drives the convergence!

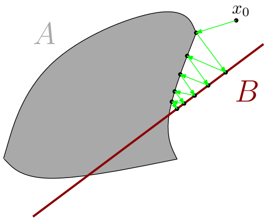
von Neumann '33, Aronszajn '50 for subspaces
Bregman '65, Gubin-Polyak-Raik '67 for convex sets
Lewis-Malick '07, Bauschke et al '11 for nonconvex sets
Noll-Rondepierre '13 for nonconvex semi-algebraic sets

All arguments are based on two considerations:

1. **(set property)** A or B is "convex-like"
2. **(intersection property)** $A \cap B$ is "rigid"

Method of alternating projections

$$\begin{array}{|lcl|}\hline x_{k+1} & \in & \mathcal{P}_B(x_k) \\ x_{k+2} & \in & \mathcal{P}_A(x_{k+1}) \\ \hline\end{array}$$



The "angle" between A and B drives the convergence!

von Neumann '33, Aronszajn '50 for subspaces
Bregman '65, Gubin-Polyak-Raik '67 for convex sets
Lewis-Malick '07, Bauschke et al '11 for nonconvex sets
Noll-Rondepierre '13 for nonconvex semi-algebraic sets

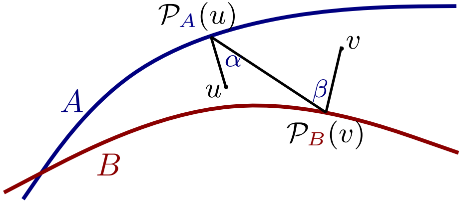
All arguments are based on two considerations:

1. **(set property)** A or B is "convex-like"
2. **(intersection property)** $A \cap B$ is "rigid"

We eliminate 1 and keep 2 loose.

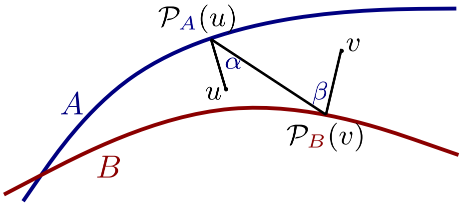
Most general result known

Intrinsic Transversality:



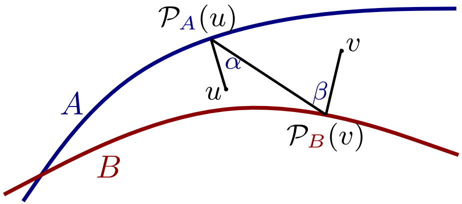
Most general result known

Intrinsic Transversality: $\boxed{\max\{\alpha, \beta\} \geq \kappa > 0}$



Most general result known

Intrinsic Transversality: $\boxed{\max\{\alpha, \beta\} \geq \kappa > 0}$



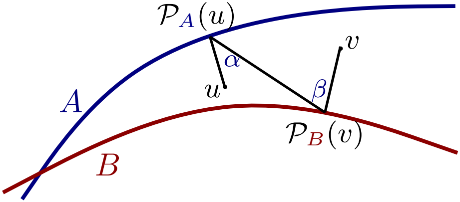
Convergence of alternating projections (D-Ioffe-Lewis '13):

A and B intrinsically transverse at $\bar{x}$

$\implies$ local **R**-linear convergence.

Most general result known

Intrinsic Transversality: $\boxed{\max\{\alpha, \beta\} \geq \kappa > 0}$



Convergence of alternating projections (D-Ioffe-Lewis '13):

A and B intrinsically transverse at $\bar{x}$

$\implies$ local **R**-linear convergence.

Remark: If A, B convex, this is essentially the Slater condition:

$$\operatorname{ri} A \cap \operatorname{ri} B \neq \emptyset.$$

Generic transversality

Is intrinsic transversality typical?

Generic transversality

Is intrinsic transversality typical?

Sard '36, Whitney '65:

If A and B are semi-algebraic, then

$$A + a \text{ and } B \text{ are transverse for a.e. } a$$

Generic transversality

Is intrinsic transversality typical?

Sard '36, Whitney '65:
If A and B are semi-algebraic, then

$$A + a \text{ and } B \text{ are transverse for a.e. } a$$

On the other hand:

Transversality often fails in practice (!!)

Generic transversality

Is intrinsic transversality typical?

Sard '36, Whitney '65:
If A and B are semi-algebraic, then
$A + a$ and B are transverse for a.e. a

On the other hand:

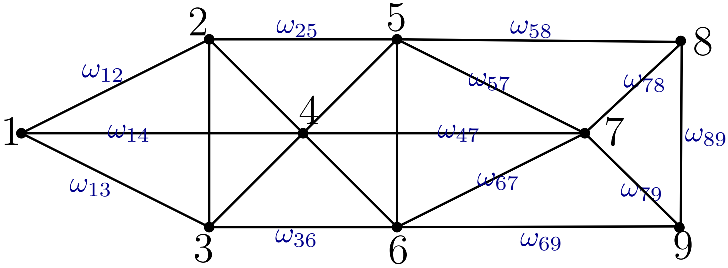
Transversality often fails in practice (!!)

Eg: Structured data

$$\begin{bmatrix} 1 & 1 & ? \\ 1 & 1 & 1 \\ ? & 1 & 1 \end{bmatrix} \succeq 0$$

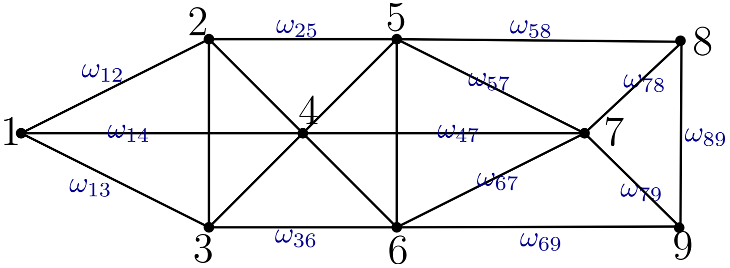
Example: Isometric graph embeddings

Problem: given a weighed graph $G = (V, E, \omega)$



Example: Isometric graph embeddings

Problem: given a weighed graph $G = (V, E, \omega)$

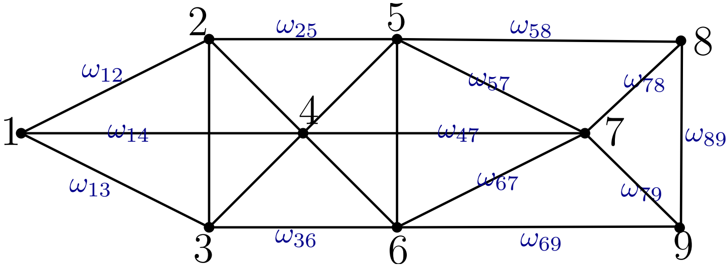


find a **realization**:

$$x_1, \ldots, x_n \in \mathbf{R}^r \quad \text{with} \quad \omega_{ij} = |x_i - x_j|^2.$$

Example: Isometric graph embeddings

Problem: given a weighed graph $G = (V, E, \omega)$



find a **realization**:

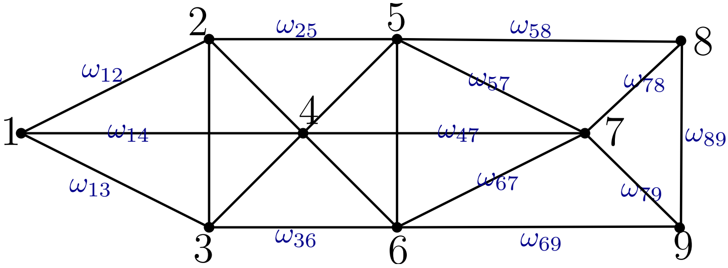
$$x_1, \ldots, x_n \in \mathbf{R}^r \quad \text{with} \quad \omega_{ij} = |x_i - x_j|^2.$$

If possible, then

$$\text{embdim}\, G = \text{ minimal } r.$$

Example: Isometric graph embeddings

Problem: given a weighed graph $G = (V, E, \omega)$



find a **realization**:

$$x_1, \ldots, x_n \in \mathbf{R}^r \quad \text{with} \quad \omega_{ij} = |x_i - x_j|^2.$$

If possible, then

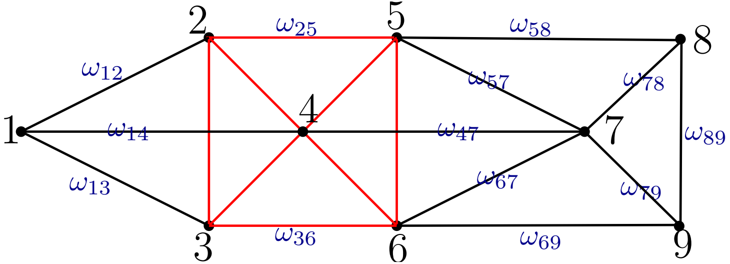
$$\text{embdim}\, G = \text{ minimal } r.$$

Eg: Sensor network localization and molecular conformation

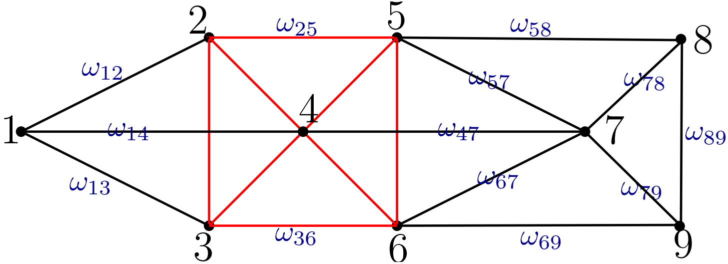
Graph embeddings

Natural substructures: cliques.

Natural substructures: cliques.



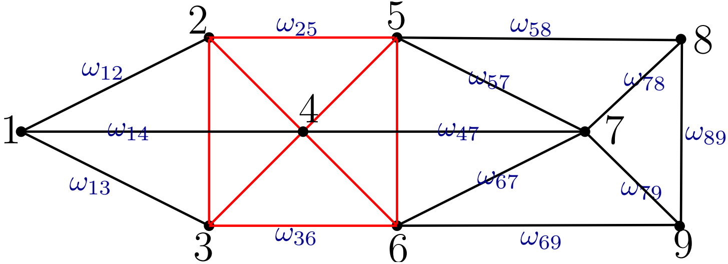
Natural substructures: cliques.



Idea: "Collapse" cliques

- Not clear how to do this

Natural substructures: cliques.



Idea: "Collapse" cliques

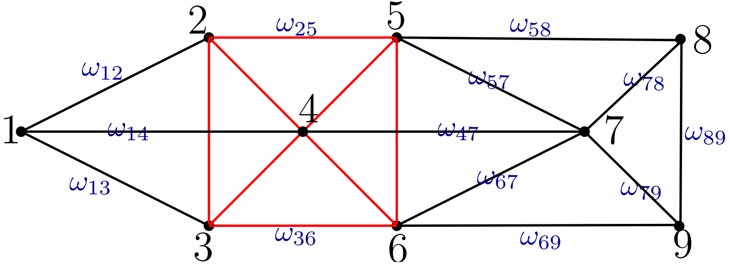
- Not clear how to do this

Different approach: model as a semi-definite program

$$\{X \succeq 0 : \mathcal{A}(X) = \omega\}.$$

Graph embeddings

Natural substructures: cliques.



Idea: "Collapse" cliques

- Not clear how to do this

Different approach: model as a semi-definite program

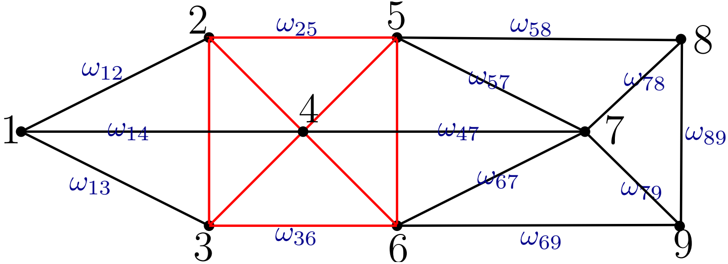
$$\{X \succeq 0 : \mathcal{A}(X) = \omega\}.$$

Krislock-Wolkowicz '10: For "any" cliques $\chi_1, \ldots, \chi_m$ in G

$$\{X \succeq 0 : \mathcal{A}(X) = \omega\} \quad \subseteq \quad \mathcal{S}^n_+ \cap \left(Y_1^\perp \cap \ldots \cap Y_m^\perp\right)$$

Graph embeddings

Natural substructures: cliques.



Idea: "Collapse" cliques

- Not clear how to do this

Different approach: model as a semi-definite program

$$\{X \succeq 0 : \mathcal{A}(X) = \omega\}.$$

Krislock-Wolkowicz '10: For "any" cliques $\chi_1, \ldots, \chi_m$ in G

$$\{X \succeq 0 : \mathcal{A}(X) = \omega\} \quad \subseteq \quad \mathcal{S}^n_+ \cap \left(Y_1^\perp \cap \ldots \cap Y_m^\perp \right)$$

- Slater inherently fails.
- Collapse occurs in the SDP.

Shortcoming of KW algorithm

Realistic problems have noise and a known embedding dimension.

Shortcoming of KW algorithm

Realistic problems have noise and a known embedding dimension.

Least-squares formulation:

$$\begin{aligned} \min \quad & \|\mathcal{A}(X) - \omega\|^2 \\ \text{s.t.} \quad & \operatorname{rank} X = r \\ & X \succeq 0 \end{aligned}$$

This is a large-scale, noisy, nonconvex problem.

Shortcoming of KW algorithm

Realistic problems have noise and a known embedding dimension.

Least-squares formulation:

$$\begin{aligned} \min \quad & \|\mathcal{A}(X) - \omega\|^2 \\ \text{s.t.} \quad & \operatorname{rank} X = r \\ & X \succeq 0 \end{aligned}$$

This is a large-scale, noisy, nonconvex problem.

- Local methods output irrelevant local minimizers.
- “Ignore rank & solve, project, boost” (Biswas, Ye ’04).
- Too big for SDP.

Randomized rounding

Our problem:

$$\begin{aligned} \min\ & \|\mathcal{A}(X) - \omega\|^2 \\ \text{s.t.}\quad & \operatorname{rank} X = r \\ & X \succeq 0 \end{aligned}$$

Randomized rounding

Our problem:

$$\begin{aligned} \min \ & \|\mathcal{A}(X) - \omega\|^2 \\ \text{s.t.} \quad & \operatorname{rank} X = r \\ & X \succeq 0 \end{aligned}$$

Key idea:

$$\mathcal{S}_+^n \cap \Big(Y_1^\perp \cap \ldots \cap Y_m^\perp \Big) \quad = \quad \mathcal{S}_+^n \cap \Big(Y_1 + \ldots + Y_m \Big)^\perp$$

Randomized rounding

Our problem:

$$\begin{aligned} \min\ & \|\mathcal{A}(X) - \omega\|^2 \\ \text{s.t.}\quad & \operatorname{rank} X = r \\ & X \succeq 0 \end{aligned}$$

Key idea:

$$\mathcal{S}^n_+ \cap \left(Y_1^\perp \cap \ldots \cap Y_m^\perp \right) \quad = \quad \mathcal{S}^n_+ \cap \left(Y_1 + \ldots + Y_m \right)^\perp$$

Algorithmic framework (Cheung-D-Krislock-Wolkowicz'14):

1. Fix a set of cliques χ^i
2. Form "approximate exposing matrices" Y_i from χ_i
3. Form the aggregate

$$Y = Y_1 + \ldots + Y_m.$$

4. Round down Y to a nearest rank $n - r$ matrix $\mathcal{N}$
5. Solve Least Squares on $\mathcal{S}^n_+ \cap \mathcal{N}^\perp \cong \mathcal{S}^r_+$

Randomized rounding

Our problem:

$$\begin{aligned} \min\ & \|\mathcal{A}(X) - \omega\|^2 \\ \text{s.t.}\quad & \operatorname{rank} X = r \\ & X \succeq 0 \end{aligned}$$

Key idea:

$$\mathcal{S}^n_+ \cap \left(Y_1^\perp \cap \ldots \cap Y_m^\perp \right) \quad = \quad \mathcal{S}^n_+ \cap \left(Y_1 + \ldots + Y_m \right)^\perp$$

Algorithmic framework (Cheung-D-Krislock-Wolkowicz'14):

1. Fix a set of cliques χ^i
2. Form "approximate exposing matrices" Y_i from χ_i
3. Form the aggregate

$$Y = Y_1 + \ldots + Y_m.$$

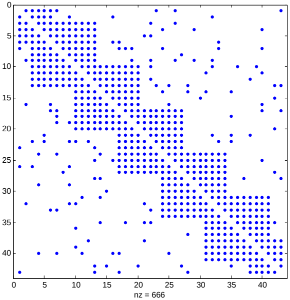
4. Round down Y to a nearest rank $n - r$ matrix $\mathcal{N}$
5. Solve Least Squares on $\mathcal{S}^n_+ \cap \mathcal{N}^\perp \cong \mathcal{S}^r_+$

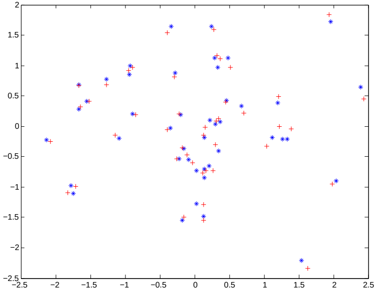
Reasonable conditions $\Longrightarrow$

$$\text{output error} \leq \kappa(\text{input noise}).$$

Randomized rounding

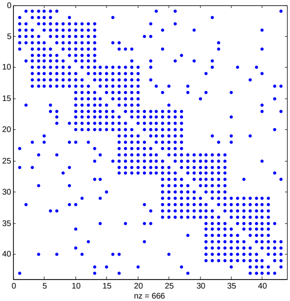
5% noise $(n = 45, r = 2)$:

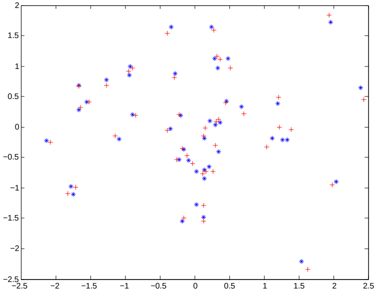




Randomized rounding

5% noise $(n = 45, r = 2)$:





Eigenvalues of $Y_1 + \ldots + Y_m$ (multiples of 10^{-6}):

$$0 \quad 1 \quad 200 \quad 300 \quad 350 \quad \ldots$$

Conclusion & Open questions

- Alternating projections: simple, powerful, versatile (when it works).
- Transversality can fail for structured problems.
- This can be an advantage.
  - Illustration: noisy, low-rank Euclidean distance completions.

Thank you.