Network Localization via Schatten Quasi-Norm Minimization

Anthony Man-Cho So
Department of Systems Engineering & Engineering Management
The Chinese University of Hong Kong

(Joint Work with Senshan Ji, Kam-Fung Sze, Zirui Zhou, Yinyu Ye)

Workshop on Semidefinite Programming & Graph Algorithms
Institute for Computational & Experimental Research in Mathematics (ICERM), Brown University

13 February 2014
Sensor Network Localization

- **Given:**
  - set of sensors $V_s$ and anchors $V_a$
  - set of sensor-sensor edges $E_{ss}$
  - set of sensor-anchor edges $E_{sa}$
  - edge weights $\{d_{ij} \geq 0: (i, j) \in E_{ss}\}$
  - and $\{\bar{d}_{ij} \geq 0: (i, j) \in E_{sa}\}$
  - an integer $d$

- **Goal:**
  - place the vertices of $G$ in $R^d$ so that their coordinates satisfy the anchor and distance constraints
The problem is computationally intractable (Saxe 1979, Aspnes et al. 2004) for \( d \geq 1 \).

This should be contrasted with the complexity of determining whether a generic instance has a unique realization in \( R^d \).

Many heuristics have been proposed:

- global optimization
- \( d \)-lateration
- ad-hoc approaches
- ...
Background

  - Good computational and theoretical properties.
Localization as Rank-Constrained SDP

- The problem can be formulated as follows:
  \[ \| x_i - x_j \|^2 = d_{ij}^2 \quad (i, j) \in E_{ss} \]
  \[ \| a_k - x_j \|^2 = \bar{d}_{kj}^2 \quad (i, j) \in E_{sa} \]

  \( x_i \in \mathbb{R}^d \)
  \( \{a_k\} \) are the positions of “anchors”.

- This turns out to be equivalent to a rank-constrained semidefinite program (SDP).
Localization as Rank-Constrained SDP

- **Step 1:** Variable substitution \( Y_{ij} = x_i^T x_j \)

\[
\|x_i - x_j\|^2 = x_i^T x_i - 2x_i^T x_j + x_j^T x_j \\
Y_{ii} \quad Y_{ij} \quad Y_{jj}
\]

\[
\|a_k - x_j\|^2 = a_k^T a_k - 2a_k^T x_j + x_j^T x_j \\
Y_{jj}
\]

- **Step 2:** Rank connection

\[
Y = X^T X \iff Z = \begin{bmatrix} I & X \\ X^T & Y \end{bmatrix} \succeq 0, \text{rank}(Z) = d
\]
Localization as Rank-Constrained SDP

- Putting things together, the localization problem becomes

find \( Z \) such that

\[
Y_{ii} + Y_{jj} - 2Y_{ij} = d_{ij}^2 \quad (i, j) \in E_{ss}, \\
a_k^T a_k - 2a_k^T x_j + Y_{jj} = d_{kj}^2 \quad (k, j) \in E_{sa},
\]

\[
Z = \begin{bmatrix} I_X & X \\ X^T & Y \end{bmatrix} \succeq 0,
\]

\[
\text{rank}(Z) = d.
\]

(Biswas and Ye 2004)
Putting things together, the localization problem becomes

find $Z$ such that

$$Y_{ii} + Y_{jj} - 2Y_{ij} = d_{ij}^2$$  \hspace{1cm} (i, j) \in E_{ss},
$$a_k^T a_k - 2a_k^T x_j + Y_{jj} = d_{kj}^2$$  \hspace{1cm} (k, j) \in E_{sa},

$$Z = \begin{bmatrix} I & X \\ X^T & Y \end{bmatrix} \succeq 0,$$

rank($Z$) = $d$.

(Biswas and Ye 2004)
Putting things together, the localization problem becomes

find \( Z \) such that

\[
Y_{ii} + Y_{jj} - 2Y_{ij} = d_{ij}^2 \quad (i, j) \in E_{ss},
\]

\[
a_k^T a_k - 2a_k^T x_j + Y_{jj} = d_{kj}^2 \quad (k, j) \in E_{sa},
\]

\[
Z = \begin{bmatrix} I & X \\ X^T & Y \end{bmatrix} \succeq 0,
\]

\[
\text{rank}(Z) = d. \quad \text{difficult constraint}
\]

(Biswas and Ye 2004)
Getting Around the Rank Constraint

• Existing work essentially ignores the rank constraint, resulting in the SDP feasibility problem:

find \( Z \) such that

\[
\begin{align*}
Y_{ii} + Y_{jj} - 2Y_{ij} &= d_{ij}^2 & (i, j) \in E_{SS}, \\
a_k^T a_k - 2a_k^T x_j + Y_{jj} &= d_{kj}^2 & (k, j) \in E_{SA}, \\
Z &= \begin{bmatrix} I & X \\ X^T & Y \end{bmatrix} \succeq 0.
\end{align*}
\]

• Fact: \( \text{rank}(Z) = l \iff \text{localization in } \mathbb{R}^l \)
A fundamental question is, when is the relaxation exact, i.e., when is \( \text{rank}(Z) = d \) ?

(S. and Ye 2005) The relaxation is exact iff the input satisfies the following uniqueness property:

**Unique \( d \)-Realizability**: The input has a unique realization in \( \mathbb{R}^d \), and does not have any non-trivial realization in \( \mathbb{R}^h \) for \( h > d \).

Essentially, the input has to be **universally rigid**.
Universal Rigidity: An Illustration

Universally Rigid

Not universally rigid
Limitations of the SDP Approach

- Consider inputs that are globally rigid in $R^d$, i.e., those with unique (up to congruence) realization in $R^d$. 
Global Rigidity: An Illustration

Universally Rigid

Globally rigid in $\mathbb{R}^2$ but not universally rigid
Limitations of the SDP Approach

- Consider inputs that are globally rigid in $\mathbb{R}^d$, i.e., those with unique (up to congruence) realization in $\mathbb{R}^d$.
- Theorem (Aspnes et al. 2004): For fixed $d$, localizing globally rigid instances in $\mathbb{R}^d$ is intractable.
- Consequence: The Biswas-Ye SDP will necessarily fail on some of the globally rigid instances.
- Question: Can more be done in polynomial time?
Salvaging the Rank Constraint

- In the previous formulation, we drop the rank constraint entirely.
- To recover some of its effects, a natural idea is to use a suitably chosen regularizer $f$, i.e.,

$$
\min f(Z) \quad \text{such that} \\
\begin{aligned}
Y_{ii} + Y_{jj} - 2Y_{ij} &= d_{ij}^2 & (i, j) \in E_{ss}, \\
\alpha_k^T \alpha_k - 2\alpha_k^T x_j + Y_{jj} &= d_{kj}^2 & (k, j) \in E_{sa}, \\
Z &= \begin{bmatrix} I & X \\ X^T & Y \end{bmatrix} \succeq 0.
\end{aligned}
$$
Using Regularizations

- Ideally, we want $f(Z) = \text{rank}(Z)$, but the resulting problem is just as hard as the original.
  - find surrogates of the rank function
- A convex choice: $f(Z) = \text{tr}(Z)$
  - popular in recent work on low-rank matrix recovery due to its tractability and theoretical guarantees
  - performs not so well empirically for the localization problem (!)
Schatten $p$-Quasi-Norm

- Let’s go a bit non-convex (but still continuous) and use
  \[
  f_p(Z) = tr(Z^p) = \sum_{i \geq 1} (\lambda_i(Z))^p, \quad p \in (0,1)
  \]
  This is the so-called Schatten $p$-quasi-norm of $Z$.

- Fact: As $p \searrow 0$, $f_p(Z) \to \text{rank}(Z)$.

- Fact: On the set of psd matrices, $f_p$ is concave.
  - but we are minimizing…

- Fact: Minimizing $f_p$ over system of linear matrix inequalities is NP-hard.
Computability

- Nevertheless…
  
  Theorem (Ji, Sze, Zhou, S., Ye 2013)
  For any fixed $p \in (0,1)$ and $\varepsilon > 0$, an $\varepsilon$-first-order critical point can be found in polynomial time.

- Achieved by a potential reduction algorithm.
Schatten $p$-Regularized SDP

- In fact, our result applies to the following more general problem:

$$
\Gamma^* = \min \pi(Z) = \text{tr}(CZ) + \mu \cdot \text{tr}(Z^p)
$$

such that

$$
Z \in F = \begin{cases} 
\text{tr}(A_i Z) = b_i & i = 1, \ldots, m, \\
Z \succeq 0.
\end{cases}
$$

- extension of Ge et al. 2011

- This can be used for various sparse vector and low-rank matrix recovery problems.
Some Definitions

Let $Z = U \Lambda U^T \in F$ and $\epsilon \geq 0$ be given. We say

- $Z$ is $\epsilon$-optimal if $\Gamma^* \leq \pi(Z) \leq \Gamma^* + \epsilon$.
- $Z$ is an $\epsilon$-first-order critical point if there exists $y$ such that

$$U^T C U + \mu p \Lambda^{p-1} - \sum_{i=1}^{m} y_i U^T A_i U \geq 0,$$

$$0 \leq \frac{\text{tr}(CZ + \mu p Z^p - \sum_{i=1}^{m} y_i A_i Z)}{\pi(Z)} \leq \epsilon.$$
Some Definitions

- Let $Z = U \Lambda U^T \in F$ and $\varepsilon \geq 0$ be given. We say
  - $Z$ is $\varepsilon$-optimal if $\Gamma^* \leq \pi(Z) \leq \Gamma^* + \varepsilon$.
  - $Z$ is an $\varepsilon$-first-order critical point if there exists $y$ such that

$$
U^T C U + \mu \Lambda^{p-1} - \sum_{i=1}^{m} y_i U^T A_i U \geq 0,
$$

and

$$
0 \leq \frac{\text{tr}(CZ + \mu \Lambda^p - \sum_{i=1}^{m} y_i A_i Z) - \pi(Z)}{\pi(Z)} \leq \varepsilon.
$$
Some Definitions

- Let $Z = U\Lambda U^T \in F$ and $\varepsilon \geq 0$ be given. We say
  - $Z$ is $\varepsilon$-optimal if $\Gamma^* \leq \pi(Z) \leq \Gamma^* + \varepsilon$.
  - $Z$ is an $\varepsilon$-first-order critical point if there exists $y$ such that

$$U^T CU + \mu p \Lambda^{p-1} - \sum_{i=1}^{m} y_i U^T A_i U \geq 0,$$

$$0 \leq \frac{\text{tr}(CZ + \mu p Z^p - \sum_{i=1}^{m} y_i A_i Z)}{\pi(Z)} \leq \varepsilon.$$
Some Definitions

- Let \( Z = U \Lambda U^T \in F \) and \( \varepsilon \geq 0 \) be given. We say
  - \( Z \) is \( \varepsilon \)-optimal if \( \Gamma^* \leq \pi(Z) \leq \Gamma^* + \varepsilon \).
  - \( Z \) is an \( \varepsilon \)-first-order critical point if there exists \( y \) such that

\[
U^T C U + \mu p \Lambda^{p-1} - \sum_{i=1}^{m} y_i U^T A_i U \geq 0,
\]

\[
0 \leq \frac{\text{tr}(CZ + \mu p Z^p - \sum_{i=1}^{m} y_i A_i Z)}{\pi(Z)} \leq \varepsilon.
\]

Question: What are the implications for localization?
Theoretical Implications

- A first-order point (i.e., $\varepsilon = 0$) is still feasible, so we can still extract from it a localization (possibly lying in a higher dimension than $d$).
- If original SDP relaxation recovers a rank-$d$ solution, then so does Schatten $p$-minimization.
  - non-convex optimization does not “mess things up”
  - a direct consequence of S. and Ye 2005
Schatten $p$-Regularized SDP

- Back to our problem:

\[
\Gamma^* = \min \pi(Z) = \text{tr}(CZ) + \mu \cdot \text{tr}(Z^p)
\]

such that

\[
Z \in F = \begin{cases} 
\text{tr}(A_i Z) = b_i & i = 1, \ldots, m, \\
Z \succeq 0.
\end{cases}
\]
Proof Sketch of the Theorem

- Use a potential function to keep track of our progress:
  \[ \phi(Z) = \rho \log(\pi(Z)) - \log \det Z \]

- Let \( Z \in F \) be strictly feasible. We update the iterate via \( Z^+ = Z + D_Z \), where \( \text{tr}(A_i D_Z) = 0 \).

- Key: Understand how the potential values change.

- Idea: In each iteration, either
  - the potential value decreases by a sufficient amount, in which case we continue; or
  - an approximate first-order critical point is found.
Proof Sketch of the Theorem

- Let $D = Z^{1/2} DZ Z^{1/2}$. We have
  
  $Z \mapsto \log(\pi(Z))$ is concave, $\forall \text{tr}(Z^p) = pZ^{p-1}$ on $S^n_{++}$,
  
  $\log \det Z - \log \det Z^+ \leq -\text{tr}(D) + \frac{\beta^2}{2(1 - \beta)} \quad \forall \|D\|_F^2 \leq \beta < 1$

  Then, one can show

  $\phi(Z^+) - \phi(Z) \leq \frac{\rho}{\pi(Z)} \text{tr}((\tilde{C} + \mu pZ^p)D) - \text{tr}(D) + \frac{\beta^2}{2(1 - \beta)}$

  where $\tilde{C} = Z^{1/2} C Z^{1/2}$.

- Idea: Minimize the RHS w.r.t. $D$. 

Proof Sketch of the Theorem

- **Observation:** The problem

\[
\min \frac{\rho}{\pi(Z)} \text{tr}((\bar{C} + \mu p Z^p)D) - \text{tr}(D)
\]

s. t. \( \text{tr}(Z^{1/2}A_i Z^{1/2} D) = 0 \) for \( i = 1, \ldots, m, \)

\[
\|D\|_F^2 \leq \beta
\]

admits a closed form solution.

- Then, it can be shown that

\[
\phi(Z^+) - \phi(Z) \leq -\beta \cdot \|f(A, C, Z)\|_F + \frac{\beta^2}{2(1 - \beta)}
\]

for some \( f, \) where \( A \) is the linear operator defined by the \( A_i \)'s.
Proof Sketch of the Theorem

- From
  \[
  \phi(Z^+) - \phi(Z) \leq -\beta \cdot \|f(A, C, Z)\|_F + \frac{\beta^2}{2(1 - \beta)}
  \]
  we see that if \(\|f(A, C, Z)\|_F \geq 1\), then \(\beta\) can be chosen so that \(\phi(Z^+) - \phi(Z) \leq -1/4\).

- Otherwise, we can prove that an approximate first-order critical point has been reached, so the algorithm can terminate.
Proof Sketch of the Theorem

- What if sufficient decrease in the potential value is achieved in every iteration?
- Proposition: Suppose that $\Gamma^* \geq 0$ and $\rho > n/p$. If $Z \in F$ is strictly feasible and

$$\phi(Z) \leq \left( \rho - \frac{n}{p} \right) \log \varepsilon + \frac{n}{p} \log(\mu n)$$

then $Z$ is $\varepsilon$-optimal.
- This gives another stopping criterion for the algorithm.
Proof Sketch of the Theorem

- To establish complexity estimates, it remains to bound the initial potential value.
- Proposition: Suppose that a strictly feasible $Z_0$ is available, with $\|Z_0\|_F \leq R$ and $\lambda_{\min}(Z_0) \geq r$. Then,
  \[
  \phi(Z_0) \leq \rho \log(R\|C\|_F + \mu R^p n^{1-p/2}) - n \log r
  \]
- From this, we can establish the desired polynomial complexity result.
Preliminary Computational Results

- 50 sensors, 3 anchors: Average rank (over 100 instances) of solutions
Preliminary Computational Results

- 50 sensors, 3 anchors: Specific GGR instance
Preliminary Computational Results

- 50 sensors, underlying graph GGR: Exact recovery performance
Concluding Remarks/Open Questions

- Original SDP relaxation has nice rigidity-theoretic interpretation
  - dual variable: stress matrix (S. and Ye 2006)
- How about Schatten $p$-minimization?
  - some sort of nonlinear stress matrix, perhaps?
Concluding Remarks/Open Questions

- A disconnect between linear algebra and geometry
  - Why trace regularization typically fails in localization problems, while it works fine in general low-rank matrix recovery?
- How to generate globally rigid graphs uniformly at random?
- Other applications?
Thank You!