Few Complexity Results on Sparse Solution Minimization and Low-Rank Matrix Completion

Yinyu Ye, Fields 2009

Department of Management Science and Engineering and
Institute of Computational and Mathematical Engineering
Stanford University

October 23, 2009

Outline

- L_p $(0 \leq p < 1)$ Minimization and Applications

$$\|\mathbf{x}\|_p = \left(\sum_{1 \leq j \leq n} |x_j|^p \right)^{1/p}.$$

- Low-Rank Semidefinite Programming and Applications

$$\begin{array}{lll} (SDP) & \text{minimize} & A_0 \bullet X \\ & \text{subject to} & A_i \bullet X = b_i \qquad i = 1, \ldots, m, \\ & & X \succeq \mathbf{0}. \end{array}$$

- Universal Rigidity and Graph Realization

L_p Minimization

Consider the problem:

$$\begin{array}{ll} \text{Minimize} & p(\mathbf{x}) = \sum_{1 \le j \le n} x_j^p \\ \text{Subject to} & A\mathbf{x} = \mathbf{b}, \\ & \mathbf{x} \ge \mathbf{0}, \end{array} \tag{1}$$

and

$$\begin{array}{ll} \text{Minimize} & \|\mathbf{x}\|_p^p = \sum_{1 \le j \le n} |x_j|^p \\ \text{Subject to} & A\mathbf{x} = \mathbf{b}; \end{array} \tag{2}$$

where data $A \in R^{m \times n}, \mathbf{b} \in R^m$, and parameter $0 < p < 1$.

Application and Motivation

The original goal is to minimize $\|\mathbf{x}\|_0$, the size of the support set of $\mathbf{x}$, for

- Sparse image reconstruction
- Sparse signal recovering
- Compressed sensing

which is known to be an NP-Hard problem.

Approximation of $\|\mathbf{x}\|_0$

- $\|\mathbf{x}\|_1$ has been used to approximate $\|\mathbf{x}\|_0$, and the relaxation can be exact under certain strong conditions (Donoho 2004, Candès and Tao 2005, etc). The relaxation model is actually a linear program.

- Theoretical and empirical computational results indicate that $\|\mathbf{x}\|_p$ approximation, say $p = .5$, have better performances under weaker conditions, and it is solvable equally efficiently in practice (Fan and Li 2001, Chartrand 2009, Chen et al. 2009, etc).

Image Reconstruction Bounds I

$$\min_{\mathbf{x}} \quad f(\mathbf{x}) = \|A\mathbf{x} - \mathbf{b}\|_2^2 + \lambda \|\mathbf{x}\|_p^p,$$

Theorem

(Chen et al. 2009) Let β *be a positive constant such that for a local minimizer* $\mathbf{x}^*$*:* $\|A^T(A\mathbf{x}^* - \mathbf{b})\| < \beta$*, and let* $L = \left(\frac{\lambda p}{2\beta}\right)^{\frac{1}{1-p}}$*.*
Then, the local minimizer $\mathbf{x}^*$ *possesses the property*

$$x_j^* \in (-L, L) \quad \Rightarrow \quad x_j^* = 0, \;\; j \in \mathcal{N}.$$

Image Reconstruction Bounds II

Theorem

(Chen et al. 2009) Let $L_j = \left(\frac{\lambda p(1-p)}{2\|\mathbf{a}_j\|^2}\right)^{\frac{1}{2-p}}, j \in \mathcal{N}$. *Then for any local minimizer* $\mathbf{x}^*$ *the following statements hold.*

- *(1)* $x_j^* \in (-L_j, L_j) \quad \Rightarrow \quad x_j^* = 0, \;\; j \in \mathcal{N}$.
- *(2) The columns of the sub-matrix* $A_B \in R^{m \times |B|}$ *of* A *are linearly independent, where* $B =$support$(\mathbf{x}^*)$.
- *(3) Let* $\|\mathbf{a}_j\| = 1$ *for all* $i \in \mathcal{N}$ *and* $\mathbf{x}^*$ *be any local minizer satisfying* $f(\mathbf{x}^*) \leq f(\mathbf{0})$. *Then, the number of nonzero entries in* $\mathbf{x}^*$ *is bounded by*

$$|B| \leq \|\mathbf{b}\|^2 \cdot \left(\frac{1}{\lambda}\right)^{\frac{2}{2-p}} \left(\frac{2}{p(1-p)}\right)^{\frac{p}{2-p}}.$$

Image Reconstruction Bounds III

These two theorems establish relations between model parameters $p,\ \lambda$ and the desired degree of sparsity of the solution. In particular, it gives a guidance on how to choose the combination of λ and p.

The L_1 minimization does not have such a threshold control on the final solution.

Then, is L_p minimization easier to solve than L_0 minimization?

The Hardness I

Theorem

(Jiang and Y 2009) For a given real number v, it is NP-hard to decide if the minimal objective value of problem (1):

$$\begin{array}{ll} \textit{Minimize} & p(\mathbf{x}) = \sum_{1 \le j \le n} x_j^p \\ \textit{Subject to} & A\mathbf{x} = \mathbf{b}, \\ & \mathbf{x} \ge \mathbf{0}, \end{array}$$

is less than or equal to v.

Proof

An instance of the partition problem can be described as follows: given a set S of integers or rational numbers $\{a_1, a_2, \ldots, a_n\}$, is there a way to partition S into two disjoint subsets S_1 and S_2 such that the sum of the numbers in S_1 equals the sum of the numbers in S_2?

Let vector $\mathbf{a} = (a_1, a_2, \ldots, a_n) \in R^n$. Then, we consider the following reduced minimization problem in form (1):

$$\begin{array}{ll} \text{Minimize} & P(\mathbf{x}, \mathbf{y}) = \sum_{1 \le j \le n} (x_j^p + y_j^p) \\ \text{Subject to} & \mathbf{a}^T(\mathbf{x} - \mathbf{y}) = 0, \\ & x_j + y_j = 1, \ \forall j, \\ & \mathbf{x}, \mathbf{y} \ge \mathbf{0}. \end{array}$$

Proof continued

From the strict concavity of the objective function,

$$x_j^p + y_j^p \geq x_j + y_j = 1, \ \forall j,$$

and they are equal if and only if $(x_j = 1, y_j = 0)$ or $(x_j = 0, y_j = 1)$. Thus, $P(\mathbf{x}, \mathbf{y}) \geq n$ for any (continuous) feasible solution; and if there is a feasible solution pair $(\mathbf{x}, \mathbf{y})$ such that $P(\mathbf{x}, \mathbf{y}) \leq n$, it must be true $x_j^p + y_j^p = 1 = x_j + y_j$ for all j so that $(\mathbf{x}, \mathbf{y})$ is a binary solution, $(x_j = 1, y_j = 0)$ or $(x_j = 0, y_j = 1)$, which generates an equitable partition of the entries of $\mathbf{a}$.

On the other hand, if the entries of $\mathbf{a}$ has an equitable partition, then the reduced problem must have a binary solution pair $(\mathbf{x}, \mathbf{y})$ such that $P(\mathbf{x}, \mathbf{y}) = n$. Therefore, it is NP-hard to decide if there is a feasible solution $(\mathbf{x}, \mathbf{y})$ such that its objective value $P(\mathbf{x}, \mathbf{y}) \leq n$.

The Hardness II

For the same partition problem, consider the following reduced minimization problem in form (2):

$$\begin{array}{ll} \text{Minimize} & \sum_{1 \le j \le n} (|x_j|^p + |y_j|^p) \\ \text{Subject to} & \mathbf{a}^T(\mathbf{x} - \mathbf{y}) = 0, \\ & x_j + y_j = 1, \ \forall j. \end{array}$$

Note that this problem has no non-negativity constraints on variables $(\mathbf{x}, \mathbf{y})$. However, for any feasible solution $(\mathbf{x}, \mathbf{y})$ of the problem, we still have

$$|x_j|^p + |y_j|^p \ge x_j + y_j = 1, \ \forall j.$$

This is because when $x_j + y_j = 1$, the minimal value of $|x_j|^p + |y_j|^p$ is 1, and it equals 1 if and only if $(x_j = 1, y_j = 0)$ or $(x_j = 0, y_j = 1)$.

The Hardness III

Thus, it remains NP-hard to decide if there is a feasible solution $(\mathbf{x}, \mathbf{y})$ such that the objective value of the reduced problem is less than or equal to n. This leads to:

Theorem

(Jiang and Y 2009) For a given real number v, it is NP-hard to decide if the minimal objective value of problem (2):

$$\begin{array}{ll} \textit{Minimize} & \|\mathbf{x}\|_p^p = \sum_{1 \le j \le n} |x_j|^p \\ \textit{Subject to} & A\mathbf{x} = \mathbf{b}; \end{array}$$

is less than or equal to v.

The Easiness

We now turn our attention to local minimizers.

In general, finding a local minimizer, or checking if a solution is a local minimizer, remains NP-hard.

Theorem

The set of all basic feasible solutions of (1) is exactly the set of its all local minimizers. The set of all basic solutions of (2) is exactly the set of its all local minimizers.

Interior-Point Algorithms

Naturally, one would start from an interior-point feasible solution such as the analytic center $\mathbf{x}^0$ of the feasible polytope. Without of loss of generality, let $\mathbf{x}^0 = \mathbf{e}$, the vector of all ones.

Consider the Karmarkar potential function

$$\begin{aligned}\phi(\mathbf{x}) &= \rho \log \left(\sum_{j=1}^n x_j^p - \underline{z} \right) - \sum_{j=1}^n \log x_j \\ &= \rho \log(p(\mathbf{x}) - \underline{z}) - p \sum_{j=1}^n \log x_j,\end{aligned}$$

where $\underline{z}$ is a lower bound on the global minimal objective value of (1) and parameter $\rho > n$. For simplicity, we set $\underline{z} = 0$ in the rest of discussion.

Potential Function for L_p Minimization

$$\phi(\mathbf{x}) = \rho \log\left(\sum_{j=1}^{n} x_j^p\right) - \sum_{j=1}^{n} \log x_j = \rho \log(p(\mathbf{x})) - \sum_{j=1}^{n} \log x_j.$$

$$\frac{\sum_{j=1}^{n} x_j^p}{n} \geq \left(\prod_{j=1}^{n} x_j^p\right)^{1/n}$$

so that

$$\frac{n}{p}\log(p(\mathbf{x})) - \sum_{j=1}^{n} \log x_j \geq \frac{n\log n}{p}.$$

Thus, if $\phi(\mathbf{x}) \leq (\rho - n/p)\log(\epsilon) + \frac{n\log n}{p}$, we must have $p(\mathbf{x}) \leq \epsilon$, which implies that $\mathbf{x}$ must be an ϵ-global minimizer.

Complexity of the Potential Reduction Algorithm

Theorem

If we choose $\rho \geq \frac{2n}{\epsilon}$, then the potential redulction algorithm will return an ϵ-stationary point of (1) in no more than $O(\frac{n}{\epsilon}\log\frac{1}{\epsilon})$ iterations for any given $\epsilon < p$.

Semidefinite Programming

Consider the Semidefinite Programming problem:

$$
\begin{array}{llll}
(SDP) & \text{minimize} & A_0 \bullet X & \\
 & \text{subject to} & A_i \bullet X = b_i & i = 1, \ldots, m, \\
 & & X \succeq \mathbf{0} &
\end{array}
$$

where $A_0, A_1, \ldots, A_m$ are given $n \times n$ symmetric matrices and $b_1, \ldots, b_m$ are given scalars, and

$$A \bullet X = \sum_{i,j} a_{ij} x_{ij} = \text{trace}(A^T X).$$

The Dual of SDP

The dual problem to (SDP) can be written as:

$$\begin{array}{lll}(SDD) & \text{maximize} & \mathbf{b}^T\mathbf{y} \\ & \text{subject to} & \sum_i^m y_i A_i + S = A_0,\ S \succeq \mathbf{0},\end{array}$$

where $\mathbf{y} = (y_1; ...; y_m) \in \mathcal{R}^m$.

Let X^* and S^* be a solution pair with zero duality gap. Then

$$\text{rank}(X^*) + \text{rank}(S^*) \leq n.$$

Thus, if there is S^* such that $\text{rank}(S^*) \geq n - d$, then the max rank of X^* is bounded above by d.

Known Results

- The SDP interior-point algorithm finds an ϵ-approximate solution where solution time is linear in $\log(1/\epsilon)$ and polynomial in m and n.
- Barvinok 95 (earlier results ?)showed that if the problem is solvable, then there exists a solution X^* whose rank r satisfies $r(r+1) \leq 2m$. (A constructive proof can be based on Carathéodory's theorem.)
- And the rank bound is essentially tight.
- A such optimal solution can be found in polynomial time; Pataki (1999), and Alfakih/Wolkowicz (1999).

Fixed-Low-Rank SDP Solution?

- We are interested in finding a fixed low–rank (say d) solution to the above system.
- However, there are some issues:
  - Such a solution may not exist!
  - Even if it does, one may not be able to find it efficiently.
- So we consider an approximation of the problem.

Approximate Low-Rank SDP Solution

For simplicity, consider finding X satisfies

$$A_i \bullet X = b_i \qquad i = 1, \ldots, m, \quad X \succeq \mathbf{0}$$

where $A_1, \ldots, A_m$ are positive semidefinite matrices and scalars $(b_1, \ldots, b_m) \geq \mathbf{0}$.

We consider finding an $\hat{X} \succeq 0$ of rank at most d that satisfies every SDP constraint approximately and uniformly:

$$\beta(m, n, d) \cdot b_i \leq A_i \bullet \hat{X} \leq \alpha(m, n, d) \cdot b_i \quad \forall\ i = 1, \ldots, m.$$

Here, $\alpha(\cdot) \geq 1$ and $\beta(\cdot) \in (0, 1]$ are called the distortion factors. Clearly, the closer are both to 1, the better the solution quality.

Approximate Low-Rank Theorem (So, Y and Zhang 07)

Let $r = \max\{\text{rank}(A_i)\}$. Then, for any $d \geq 1$, there exists an $\hat{X} \succeq \mathbf{0}$ with $\text{rank}(\hat{X}) \leq d$ such that

$$\alpha(m,n,d) = \begin{cases} 1 + \dfrac{12\ln(4mr)}{d} & \text{for } 1 \leq d \leq 12\ln(4mr) \\ 1 + \sqrt{\dfrac{12\ln(4mr)}{d}} & \text{for } d > 12\ln(4mr) \end{cases}$$

$$\beta(m,n,d) = \begin{cases} \dfrac{1}{e(2m)^{2/d}} & \text{for } 1 \leq d \leq 4\ln(2m) \\ \max\left\{\dfrac{1}{e(2m)^{2/d}},\ 1 - \sqrt{\dfrac{4\ln(2m)}{d}}\right\} & \text{for } d > 4\ln(2m) \end{cases}$$

Moreover, there exists an efficient randomized algorithm for finding such an $\hat{X}$.

Some Remarks

- There is always a low-rank approximate SDP solution with bounded distortion factors.
- As the allowable rank increases, the distortion become smaller and smaller. In particular, when $d = O(\ln(m))$, the distortion factors are both equal a constant close to 1.
- The lower distortion factor is independent of n and the rank of A_is.
- The factors are sharp; but they can be improved if we only consider one–sided inequalities.
- This result contains as special cases several well-known results in the literature.

Low Rank SDP Applications

The low–rank SDP problem arises in many applications, e.g.:

- metric embedding/dimension reduction (e.g., Johnson and Lindenstrauss 84, Matousek 90, Sun, Xiao and Boyd 06, etc.)
- approximating non-convex (real, complex) quadratic optimization (e.g., Goemans and Williamson 95, Nesterov 98, Y 98, Nemirovskii, Roos and Terlaky 99, Luo, Sidiropoulos, Tseng and Zhang 06, So, Zhang and Y 07, etc.)
- distance matrix completion (e.g., Laurent 97, Alfakih, Khandani and Wolkowicz 99, etc.)
- low-rank matrix completion (e.g., ISMP 2009 ...)
- graph realization/sensor network localization (e.g., Biswas and Y 04, So and Y 04, Biswas, Toh, and Y 06, Jin and Saunders 07, Wang, Zheng, Boyd and Y 08, Kim, Kojima and Waki 08, Pong and Tseng 08, Krislock and Wolkowicz 08, etc.)

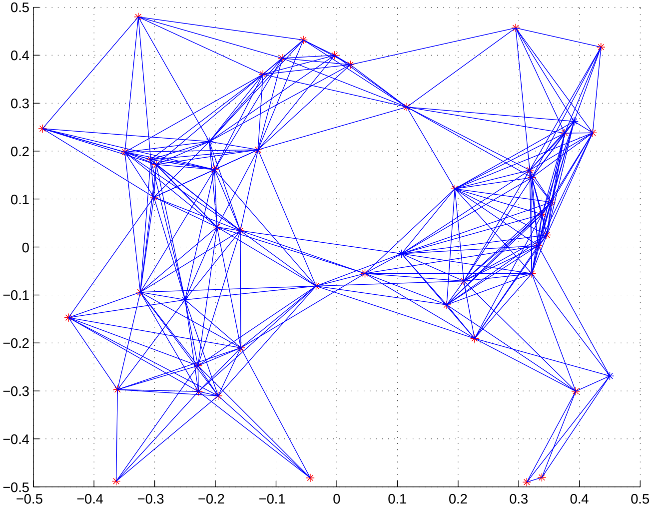
Graph Realization and Sensor Network Localization

Given a graph $G = (V, E)$ and sets of non–negative weights, say $\{d_{ij} : (i,j) \in E\}$, the goal is to compute a realization of G in the Euclidean space $\mathbf{R}^d$ for a given low dimension d, i.e.

- to place the vertices of G in $\mathbf{R}^d$ such that
- the Euclidean distance between a pair of adjacent vertices (i,j) equals to (or bounded by) the prescribed weight $d_{ij} \in E$.

In what follows, we assume that the Euclidean distance measurements are drawn from a graph already in $\mathbf{R}^d$ with general positions.

Unit-Disk Sensor Network: 50-node in 2-D



SDP Formulation

Find a symmetric matrix $Z \in \mathbf{R}^{n \times n}$ such that

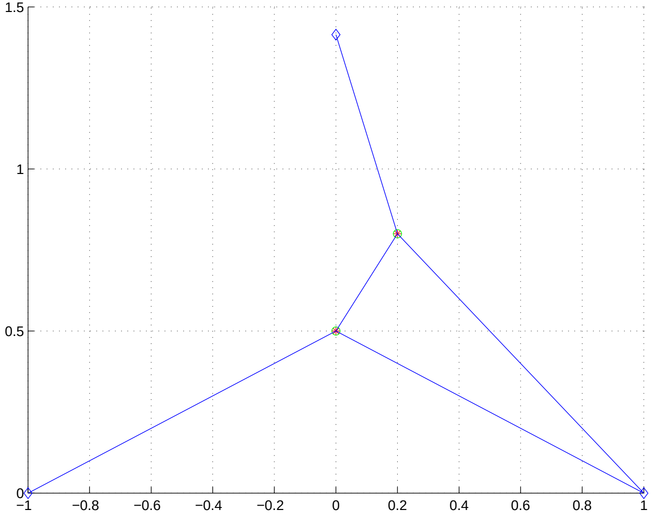
$$(\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)^T \bullet Z = d_{ij}^2, \ \forall\ i,j \in E,\ i < j,$$
$$Z \succeq \mathbf{0},\ \text{rank}(Z) = d.$$

SDP Relaxation:

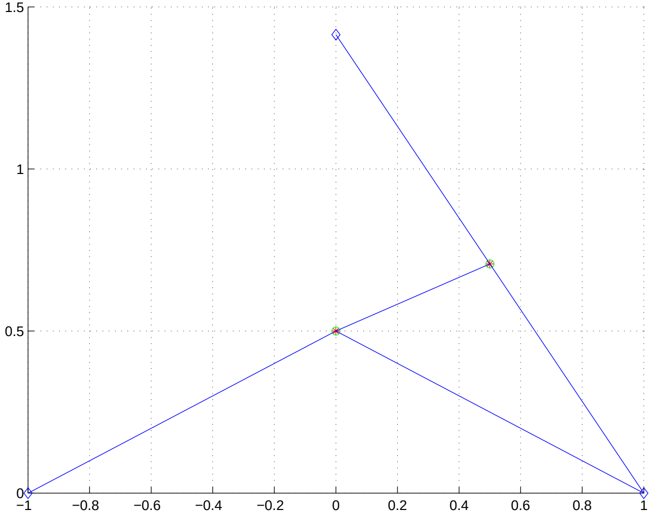
$$\begin{array}{rl} \text{minimize} & I \bullet Z \\ \text{s.t.} & (\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)^T \bullet Z = d_{ij}^2, \ \forall\ i,j \in E,\ i < j, \\ & Z \succeq \mathbf{0}. \end{array}$$

We say that the graph is uniquely d-realizable if the SDP relaxation returns a unique rank-d solution, and the graph is strongly d-realizable if, in addition, its dual has a rank-d slack matrix.

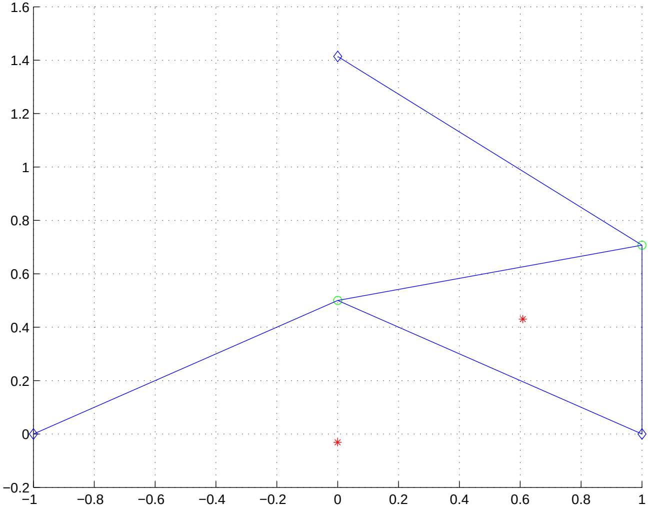
Two sensor-Three anchors: Strongly Realizable



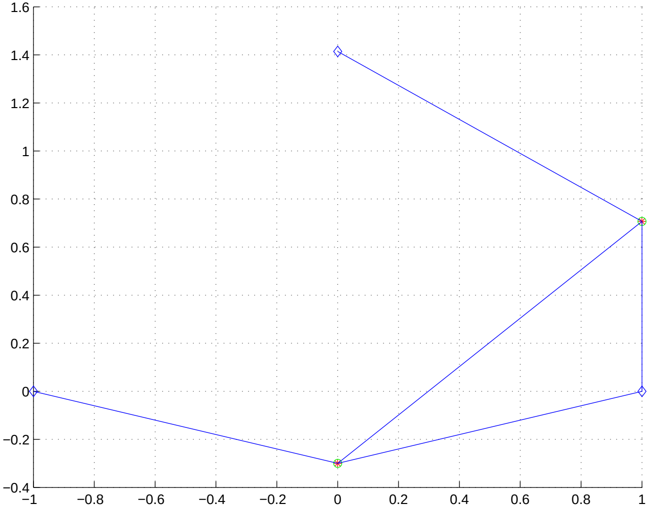
Two sensor-Three anchors: Realizable but not Strongly



Two sensor-Three anchors: Not Realizable



Two sensor-Three anchors: Strongly Realizable



Generically Unique Realizability

- The d-realizability depends on graph E combinatorics as well as distance measurements d_{ij}.
- Is there a sparse graph that is generically d-realizable, that is, independent of distance measurements?
- The degree of freedom of rank-d symmetric and PSD matrix is $d \cdot n$, could we use only $O(d \cdot n)$ distance measurements to realize a graph?

Trilateration Graphs

A trilaterative ordering in dimension d for a graph G is an ordering of the vertices $1, \cdots, d+1, d+2, \cdots, n$ such that K_{d+1}, the complete graph of the first $d+1$ vertices, is in G, and every vertex $j > d+1$ has $d+1$ edges connected to its preceding vertices on the sequence.

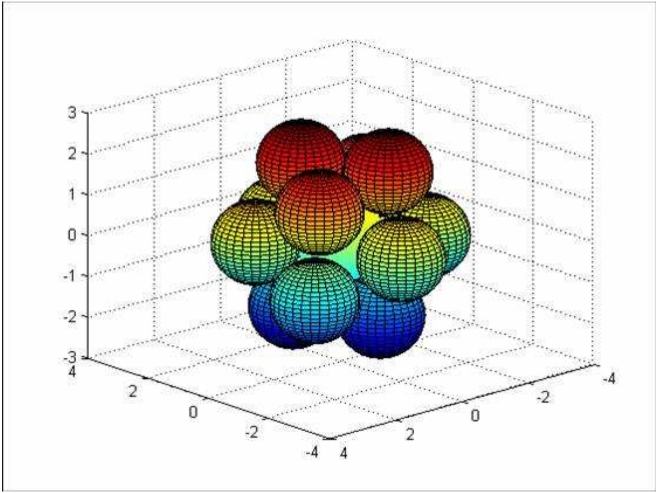
Graphs for which a trilaterative ordering exists in dimension d are called trilateration graphs in dimension d (or d-trilateration graphs). A spanning d-trilateration graph is a d-trilateration and contains every vertex of the graph.

Theorem

(Zhu, So and Y 2009) The spanning trilateration graph in dimension d is generically d-realizable. Moreover, it is a near optimal (with only $(d+1) \cdot n$ edges) in terms of information-theoretical complexity and generically d-realizable graph.

The Kissing Problem

- Given a unit center sphere, the maximum number of unit spheres, in d dimensions, can touch or kiss the center sphere?
- General Solutions does not exist.
- Delsarte Method uses linear programming to provide an upper bound on the number of spheres.
- K(1)=2, K(2)=6, K(3)= 12, K(8) = 240, K(24) = 196650.
- K(4) = 24: proved using Delsarte Method by Oleg Musin only 3 years ago.
- For other dimensions, lower bounds have been provided by constructing a lattice structure. There also exists a bound using the Riemann zeta function, but is non-constructive.



3
2
1
0
-1
-2
-3
4
2
0
-2
-4
4
2
0
-2
-4

The Kissing Problem as Realization

Given n unit-balls, find the lowest-rank solution matrix to

$$\begin{aligned} (\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)^T \bullet Z &\geq 1, \ \forall i < j \leq n, \\ \mathbf{e}_i \mathbf{e}_i^T \bullet Z &= 1, \ \forall i, \\ Z &\succeq \mathbf{0}. \end{aligned}$$

From the Approximate Low-Rank Theorem,

Corollary

One can have n-balls kissed in dimension-$O(\log(n))$ space where the distance error is below a fixed ϵ.

Current Work: PSD Matrix Completion with other Measurements

Find a symmetric matrix $Z \in \mathbf{R}^{n \times n}$ such that

$$\begin{array}{l} \mathbf{e}_i \mathbf{e}_j^T \bullet Z = a_{ij}, \ \forall \ i, j \in E, \ i < j, \\ Z \succeq \mathbf{0}, \ \text{rank}(Z) = d. \end{array}$$

- Covariance matrix
- Motion structure reconstruction
- Recommendation system