CLP Support-Size and Rank Reduction and Applications

Yinyu Ye
Department of Management Science and Engineering
Stanford University
Stanford, CA 94305, U.S.A.

http://www.stanford.edu/˜yyye

Chapters 3.1-2, 6.4-5

LP Optimality Conditions and Solution Support

$$\left\{ (\mathbf{x},\mathbf{y},\mathbf{s}) \in (\mathcal{R}^n_+, \mathcal{R}^m, \mathcal{R}^n_+) : \begin{array}{rcl} \mathbf{c}^T\mathbf{x} - \mathbf{b}^T\mathbf{y} & = & \mathbf{0} \\ A\mathbf{x} & = & \mathbf{b} \\ -A^T\mathbf{y} - \mathbf{s} & = & -\mathbf{c} \end{array} \right\};$$

or

$$\begin{array}{rcl} \mathbf{x}.\cdot\mathbf{s} & = & \mathbf{0} \\ A\mathbf{x} & = & \mathbf{b} \\ -A^T\mathbf{y} - \mathbf{s} & = & -\mathbf{c}. \end{array}$$

Let $\mathbf{x}^*$ and $\mathbf{s}^*$ be optimal solutions with zero duality gap. Then

$$|\text{supp}(\mathbf{x}^*)| + |\text{supp}(\mathbf{s}^*)| \leq n.$$

There are $\mathbf{x}^*$ and $\mathbf{s}^*$ such that the support sizes of $\mathbf{x}^*$ and $\mathbf{s}^*$ are maximal, respectively.

There are $\mathbf{x}^*$ and $\mathbf{s}^*$ such that the support size of $\mathbf{x}^*$ and $\mathbf{s}^*$ are minimal, respectively.

If there is $\mathbf{s}^*$ such that $|\text{supp}(\mathbf{s}^*)| \geq n - d$, then the support size for $\mathbf{x}^*$ is at most d.

Uniqueness Theorem for LP

Given an optimal solution $\mathbf{x}^*$, how to certify the uniqueness of $\mathbf{x}^*$?

Theorem 1 *An LP optimal solution $\mathbf{x}^*$ is unique if and only if the size of supp$(\mathbf{x}^*)$ is maximal among all optimal solutions and the columns of $A_{\text{supp}(\mathbf{x}^*)}$ are linear independent.*

It is easy to see both conditions are necessary, since otherwise, one can find an optimal solution with a different support size. To see sufficiency, suppose there there is another optimal solution $\mathbf{y}^*$ such that $\mathbf{x}^* - \mathbf{y}^* \neq \mathbf{0}$. We must have $\text{supp}(\mathbf{y}^*) \subset \text{supp}(\mathbf{x}^*)$, since, otherwise, $(0.5\mathbf{x}^* + 0.5\mathbf{y}^*)$ remains optimal and its support size is greater than that of $\mathbf{x}^*$ which is a contradiction. Then we see

$$\mathbf{0} = A\mathbf{x}^* - A\mathbf{y}^* = A(\mathbf{x}^* - \mathbf{y}^*) = A_{\text{supp}(\mathbf{x}^*)}(\mathbf{x}^* - \mathbf{y}^*)_{\text{supp}(\mathbf{x}^*)}$$

which implies that columns of $A_{\text{supp}(\mathbf{x}^*)}$ are linearly dependent.

Corollary 1 *If all optimal solutions of an LP has the same support size, then the optimal solution is unique.*

The Rank Theorem of SDP

$$
\begin{array}{lll}
(SDP) & \text{min} & C \bullet X \\
 & \text{subject to} & A_i \bullet X = b_i, i = 1, 2, ..., m, \; X \succeq 0,
\end{array}
$$

where $C, \; A_i \in \mathcal{S}^n$.

Or simply for the SDP Feasibility problem:

$$
\text{Solve} \quad A_i \bullet X = b_i, i = 1, 2, ..., m, \; X \succeq 0,
$$

Solution Rank for SDP

$$\begin{array}{rcl} C \bullet X - \mathbf{b}^T\mathbf{y} & = & 0 \\ \mathcal{A}X & = & \mathbf{b} \\ -\mathcal{A}^T\mathbf{y} - S & = & -C \\ X, S & \succeq & \mathbf{0}, \end{array}, \quad \text{or} \quad \begin{array}{rcl} XS & = & \mathbf{0} \\ \mathcal{A}X & = & \mathbf{b} \\ -\mathcal{A}^T y - S & = & -C \\ X, S & \succeq & \mathbf{0} \end{array}$$

Let X^* and S^* be optimal solutions with zero duality gap. Then

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

Hint of the Proof: for any symmetric PSD matrix $P \in S^n$ with rank r, there is a factorization $P = V^TV$ where $V \in R^{r\times n}$ and columns of V are nonzero-vectors and orthogonal to each other.

There are X^* and S^* such that the ranks of X^* and S^* are maximal, respectively.

There are X^* and S^* such that the ranks of X^* and S^* are minimal, respectively.

If there is S^* such that $\text{rank}(S^*) \ge n - d$, then the maximal rank of X^* is at most d.

Uniqueness Theorem for SDP

Given an SDP optimal and complementary solution X^*, how to certify the uniqueness of X^*?

Theorem 2 *An SDP optimal and complementary solution X^* is unique if and only if the rank of X^* is maximal among all optimal solutions and $V^* A_i (V^*)^T$, $i = 1, ..., m$, are linearly independent, where $X^* = (V^*)^T V^*$, $V^* \in \mathcal{R}^{r \times n}$, and r is the rank of X^*.*

It is easy to see why the rank of X^* being maximal is necessary.

Note that for any optimal dual slack matrix S^*, we have $S^* \bullet (V^*)^T V^* = 0$ which implies that $S^* (V^*)^T = \mathbf{0}$. Consider any matrix

$$X = (V^*)^T U V^*$$

where $U \in \mathcal{S}^r_+$ and

$$b_i = A_i \bullet (V^*)^T U V^* = V^* A_i (V^*)^T \bullet U, \; i = 1, ..., m.$$

One can see that X remains an optimal SDP solutions for any such $U \in \mathcal{S}^r_+$, since it makes X feasible and remain complementary to any optimal dual slack matrix. If $V^* A_i (V^*)^T$, $i = 1, ..., m$, are not linearly independent, then one can find

$$V^* A_i (V^*)^T \bullet W = 0, \;\; i = 1, ..., m, \; \mathbf{0} \neq W \in \mathcal{S}^r.$$

Now consider

$$X(\alpha) = (V^*)^T (I + \alpha \cdot W) V^*,$$

and then we can choose $\alpha \neq 0$ such that $X(\alpha) \succeq \mathbf{0}$ is another optimal solution.

To see sufficiency, suppose there there is another optimal solution Y^* such that $X^* - Y^* \neq \mathbf{0}$. We must have $Y^* = (V^*)^T U V^*$ for some $I \neq U \in \mathcal{S}^r_+$. Then we see

$$V^* A_i (V^*)^T \bullet (I - U) = 0, \;\; i = 1, ..., m,$$

contradicts that they are linear independent.

Corollary 2 *If all optimal solutions of an SDP has the same rank, then the optimal solution is unique.*

Rank-Reductions for CLP

In most applications, we may not be lucky and need an effort to search a rank-minimal SDP solution for SDP:

$$
\begin{array}{lll}
(SDP) & \text{min} & C \bullet X \\
 & \text{subject to} & A_i \bullet X = b_i, i = 1, 2, ..., m, \;\; X \succeq 0,
\end{array}
$$

where $C, \; A_i \in \mathcal{S}^n$.

Or simply for the SDP feasibility problem:

$$
\text{Solve} \quad A_i \bullet X = b_i, i = 1, 2, ..., m, \;\; X \succeq 0,
$$

A Bound on Support/Rank

Theorem 3 *(Carathéodory's theorem)*

- *If there is a minimizer for (LP), then there is a minimizer of (LP) whose support size r satisfying $r \leq m$.*
- *If there is a minimizer for (SDP), then there is a minimizer of (SDP) whose rank r satisfying $\frac{r(r+1)}{2} \leq m$. Moreover, such a solution can be find in polynomial time.*

How Sharp is the Rank Bound? The rank bound is sharp: consider $n = 4$ and the SDP problem:

$$\begin{aligned} (\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)^T \bullet X &= 1, \ \forall i < j = 1, 2, 3, 4, \\ X &\succeq 0, \end{aligned}$$

Applications: Finding the extreme eigenvalue of a symmetric matrix and the singular value of any matrix are convex optimization!

Application of the Rank Theorem

Consider the spherical minimization

$$\begin{array}{ll}\min & \mathbf{x}^T Q\mathbf{x} + 2\mathbf{c}^T\mathbf{x} \\ \text{s.t.} & \|\mathbf{x}\|^2 = 1.\end{array} \quad \text{or} \quad \begin{array}{ll}\min & \mathbf{x}^T Q\mathbf{x} + 2x_{n+1}\cdot\mathbf{c}^T\mathbf{x} \\ \text{s.t.} & \|\mathbf{x}\|^2 = 1, \\ & x_{n+1}^2 = 1.\end{array}$$

The SDP relaxation is

$$\begin{array}{ll}\min & \begin{pmatrix} Q & \mathbf{c} \\ \mathbf{c}^T & 0 \end{pmatrix} \cdot Z \\ \text{s.t.} & \begin{pmatrix} I & \mathbf{0} \\ \mathbf{0}^T & 0 \end{pmatrix} \cdot Z = 1, \quad \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0}^T & 1 \end{pmatrix} \cdot Z = 1 \\ & Z \succeq \mathbf{0}\end{array}$$

where $Z = (\mathbf{x}; x_{n+1})(\mathbf{x}; x_{n+1})^T \in S^{n+1}$. The relaxation is EXACT since it has a rank-one optimal solution matrix.

Application of the Rank Theorem

Consider the spherical minimization

$$\begin{array}{ll} \min & \mathbf{x}^T Q\mathbf{x} + 2\mathbf{c}^T\mathbf{x} \\ \text{s.t.} & \|\mathbf{x}\|^2 = 1. \end{array} \quad \text{or} \quad \begin{array}{ll} \min & \mathbf{x}^T Q\mathbf{x} + 2x_{n+1}\cdot\mathbf{c}^T\mathbf{x} \\ \text{s.t.} & \|\mathbf{x}\|^2 = 1, \\ & x_{n+1}^2 = 1. \end{array}$$

The SDP relaxation is

$$\begin{array}{ll} \min & \begin{pmatrix} Q & \mathbf{c} \\ \mathbf{c}^T & 0 \end{pmatrix} \cdot Z \\ \text{s.t.} & \begin{pmatrix} I & \mathbf{0} \\ \mathbf{0}^T & 0 \end{pmatrix} \cdot Z = 1, \quad \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0}^T & 1 \end{pmatrix} \cdot Z = 1 \\ & Z \succeq \mathbf{0} \end{array}$$

where $Z = (\mathbf{x}; x_{n+1})(\mathbf{x}; x_{n+1})^T \in S^{n+1}$. The relaxation is EXACT since it has a rank-one optimal solution matrix.

The Null-Space Support-Reduction for LP

1. Start at any feasible solution $\mathbf{x}^0$ and, without loss of generality, assume $\mathbf{x}^0 > \mathbf{0}$, and let $k = 0$ and $A^0 = A$.

2. Find any $A^k\mathbf{d} = \mathbf{0},\ \mathbf{d} \neq \mathbf{0}$, and let $\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha\mathbf{d}$ where α is chosen such as $\mathbf{x}^{k+1} \geq \mathbf{0}$ and at least one of $\mathbf{x}^{k+1}$ equals 0.

3. Eliminate the the variable(s) in $\mathbf{x}^{k+1}$ and column(s) in A^k corresponding to $x_j^{k+1} = 0$, and let the new narrower matrix be A^{k+1}.

4. Set $k = k + 1$ and return to step 2.

This process is called rounding, or purification, procedure in linear programming.

I: The Null-Space Reduction for SDP

Let X^* be an optimal SDP solution. If the rank, r, of X^* satisfies the inequality of the theorem, then we need do nothing. Thus, we assume $r(r+1)/2 > m$, and let

$$V^T V = X^*, \quad V \in R^{r \times n}.$$

Then consider

$$\begin{array}{ll} \text{Minimize} & VCV^T \bullet U \\ \\ \text{Subject to} & VA_iV^T \bullet U = b_i, \ i = 1, ..., m \\ \\ & U \succeq 0. \end{array} \tag{1}$$

Note that VCV^T, VA_iV^Ts and U are $r \times r$ symmetric matrices and, in particular,

$$VCV^T \bullet I = C \bullet V^T V = C \bullet X^* = z^*.$$

Moreover, for any feasible solution of (1) one can construct a feasible matrix solution for (**??**) using

$$X(U) = V^T U V \quad \text{and} \quad C \bullet X(U) = VCV^T \bullet U. \tag{2}$$

Thus, the minimal value of (1) is also z^*, and $U = I$ is a minimizer of (1).

Now we show that any feasible solution U to (1) is a minimizer for (1); thereby $X(U)$ of (2) is a minimizer for (**??**). Consider the dual of (1)

$$\begin{array}{rll} z^* := & \text{Maximize} & \mathbf{b}^T\mathbf{y} = \sum_{i=1}^m b_i y_i \\ & \text{Subject to} & VCV^T \succeq \sum_{i=1}^m y_i V A_i V^T. \end{array} \tag{3}$$

Let $\mathbf{y}^*$ be a dual maximizer. Since $U = I$ is an interior optimizer for the primal, the strong duality condition holds, i.e.,

$$I \bullet (VCV^T - \sum_{i=1}^m y_i^* V A_i V^T) = 0$$

so that we have

$$VCV^T - \sum_{i=1}^{m} y_i^* VA_iV^T = \mathbf{0}.$$

Then, any feasible solution of (1) satisfies the strong duality condition so that it must be also optimal.

Consider the system of homogeneous linear equations

$$VA_iV^T \bullet W = 0, \; i = 1, ..., m$$

where W is a $r \times r$ symmetric matrices (does not need to be definite). This system has $r(r+1)/2$ real number variables and m equations. Thus, as long as $r(r+1)/2 > m$, we must be able to find a symmetric matrix $W \neq \mathbf{0}$ to satisfy all m equations. Without loss of generality, let W be either indefinite or negative semidefinite (if it is positive semidefinite, we take $-W$ as W), that is, W has at least one negative eigenvalue, and consider

$$U(\alpha) = I + \alpha W.$$

Choosing $\alpha^* = 1/|\bar{\lambda}|$ where $\bar{\lambda}$ is the least eigenvalue of W, we have

$$U(\alpha^*) \succeq \mathbf{0}$$

and it has at least one 0 eigenvalue or $\text{rank}(U(\alpha^*)) < r$, and

$$VA_iV^T \bullet U(\alpha^*) = VA_iV^T \bullet (I + \alpha^* W) = VA_iV^T \bullet I = b_i,\ i = 1, ..., m.$$

That is, $U(\alpha^*)$ is a feasible and so it is an optimal solution for (1). Then,

$$X(U(\alpha^*)) = V^T U(\alpha^*) V$$

is a new minimizer for SDP, and $\text{rank}(X(U(\alpha^*))) < r$.

This process can be repeated till the system of homogeneous linear equations has only all zero solution, which is necessarily given by $r(r+1)/2 \le m$. The total number of such reduction steps is bounded by $n-1$ and each step uses no more than $O(m^2 n)$ arithmetic operations and finds the least eigenvalue of W, which is a polynomial time.

II. The Principle-Component or Eigenvalue Reduction

Let $\bar{X}$ be an SDP solution with rank r and

$$\bar{X} = \sum_{i=1}^{r} \lambda_i \mathbf{v}_i \mathbf{v}_i^T$$

where

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n.$$

Then, let

$$\hat{X} = \sum_{i=1}^{d} \lambda_i \mathbf{v}_i \mathbf{v}_i^T$$

III. Continuous Randomized Reduction

Let $\bar{X}$ be an SDP solution with rank r and

$$\bar{X} = VV^T$$

where $V \in R^{n \times r}$ is any factorization matrix of $\bar{X}$

Then, let random matrix

$$R = \sum_{i=1}^{d} \xi_i \xi_i^T, \quad \xi_i \in N(\mathbf{0}, \frac{1}{d}I); \quad \text{or} \quad \xi_i \in \text{Binary}(\mathbf{0}, \frac{1}{d}I)$$

that is, each entry either 1 or -1 in the latter case. Then assign

$$\hat{X} = VRV^T.$$

Note that $(V\xi_i)(V\xi_i)^T \in N(\mathbf{0}, \frac{1}{d}\bar{X})$ and

$$E[\hat{X}] = VE[R]V^T = VV^T = \bar{X}.$$

Approximate Low-Rank SDP Theorem

For simplicity, consider the SDP feasibility problem

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

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

$$x_1 + x_2 + x_3 = 1,$$
$$\begin{pmatrix} x_1 & x_2 \\ x_2 & x_3 \end{pmatrix} \succeq \mathbf{0}.$$

We try to find an approximate $\hat{X} \succeq 0$ of rank at most d:

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

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

The Main Theorem

Theorem 4 *Let* $r = \max\{\textit{rank}(A_i)\}$ *and* $\bar{X} = VV^T$ *be a feasible solution. Then, for any* $d \geq 1$, *the randomly generated*

$$\hat{X} = V[\sum_{i=1}^{d} \xi_i \xi_i^T]V^T, \quad \xi_i \in N(\mathbf{0}, \frac{1}{d}I)$$

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

and

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

Some Remarks and Open Questions

- There is always a low-rank, or sparse, approximate SDP solution with respect to a bounded relative residual distortion. As the allowable rank increases, the distortion bounds become smaller and smaller.
- The lower distortion factor is independent of n and the rank of A_is.
- The factors can be improved if we only consider one–sided inequalities.
- This result contains as special cases several well-known results in the literature.
- Can the distortion upp bound be improved such that it's independent of rank of A_i?
- Is there deterministic rank-reduction procedure? Choose the largest d eigenvalue component of X?
- General symmetric A_i?
- In practical applications, we see much smaller distortion, why?

IV. $\{-1, 1\}$ Randomized Reduction

Let X be an SDP solution with rank r and

$$X = VV^T.$$

Then, let random vector

$$\mathbf{u} \in N(\mathbf{0}, I) \quad \text{and} \quad \hat{\mathbf{x}} = \text{Sign}(V\mathbf{u})$$

where

$$\text{Sign}(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ -1 & \text{otherwise.} \end{cases}$$

Note that $V\mathbf{u} \in N(\mathbf{0}, X)$. It was proved by Sheppard (1900):

$$\mathsf{E}[\hat{x}_i \hat{x}_j] = \frac{2}{\pi} \arcsin(\bar{X}_{ij}), \quad i, j = 1, 2, \ldots, n.$$

This is the basis for proving the Max-Cut approximation algoroithm.

V. Objective-Guided Reduction

Construct a suitable objective for the SDP solution set

$$
\begin{array}{lll}
\text{Minimize} & R \bullet X & \\
\text{Subject to} & A_i \bullet X & = b_i,\ i = 1, \ldots, m, \\
 & C \bullet X & \leq \alpha \cdot z^*, \\
 & X \succeq \mathbf{0}, &
\end{array}
$$

where z^* is the minimal objective value of the SDP relaxation, and α is a tolerance factor.

The selection of matrix R is problem dependent. Examples include the L_1 norm function, the tensegrity graph approach, etc.

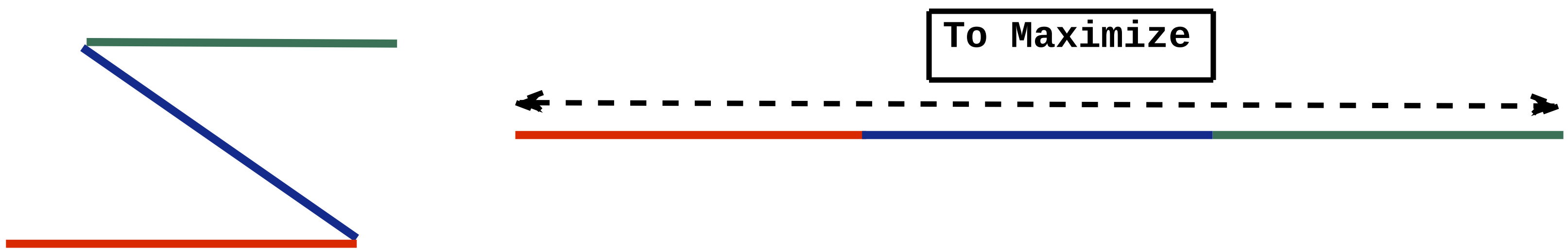
Example: The Kissing Problem (matlab demo).

Tensegrity (Tensional-Integrity) Objective for SNL: a Chain Graph

Anchor-free SNL: let $\mathbf{e}_i$ be the unit vector (one for the ith entry and zeros for the else)

$$\begin{aligned} (\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)^T \bullet X &= d_{ij}^2, \ \forall\, (i,j) \in E,\ i<j, \\ X &\succeq \mathbf{0}. \end{aligned}$$

For certain graphs, to select a subset edges to maximize and/or a subset of edges to minimize is guaranteed to finding the lowest rank SDP solution – Tensegrity Method.



The Chain Graph Example

Consider:

$$
\begin{array}{ll}
\text{max} & \mathbf{e}_3\mathbf{e}_3 \bullet X \\
\text{s.t.} & \mathbf{e}_1\mathbf{e}_1^T \bullet X = 1, \\
& (\mathbf{e}_1 - \mathbf{e}_2)(\mathbf{e}_1 - \mathbf{e}_2)^T \bullet X = 1, \\
& (\mathbf{e}_2 - \mathbf{e}_3)(\mathbf{e}_2 - \mathbf{e}_3)^T \bullet X = 1, \\
& X \succeq \mathbf{0} \in \mathcal{S}^3,
\end{array}
$$

where its maximal solution $X^* = (1;\ 2;\ 3)^T(1;\ 2;\ 3)$. The dual is

$$
\begin{array}{ll}
\text{min} & y_1 + y_2 + y_3 \\
\text{s.t.} & y_1\mathbf{e}_1\mathbf{e}_1^T + y_2(\mathbf{e}_1 - \mathbf{e}_2)(\mathbf{e}_1 - \mathbf{e}_2)^T + y_3(\mathbf{e}_2 - \mathbf{e}_3)(\mathbf{e}_2 - \mathbf{e}_3)^T - S = \mathbf{e}_3\mathbf{e}_3, \\
& S \succeq \mathbf{0} \in \mathcal{S}^3,
\end{array}
$$

The dual has a rank-two solution with $(y_1 = 3,\ y_2 = 3,\ y_3 = 3)$.

Applications

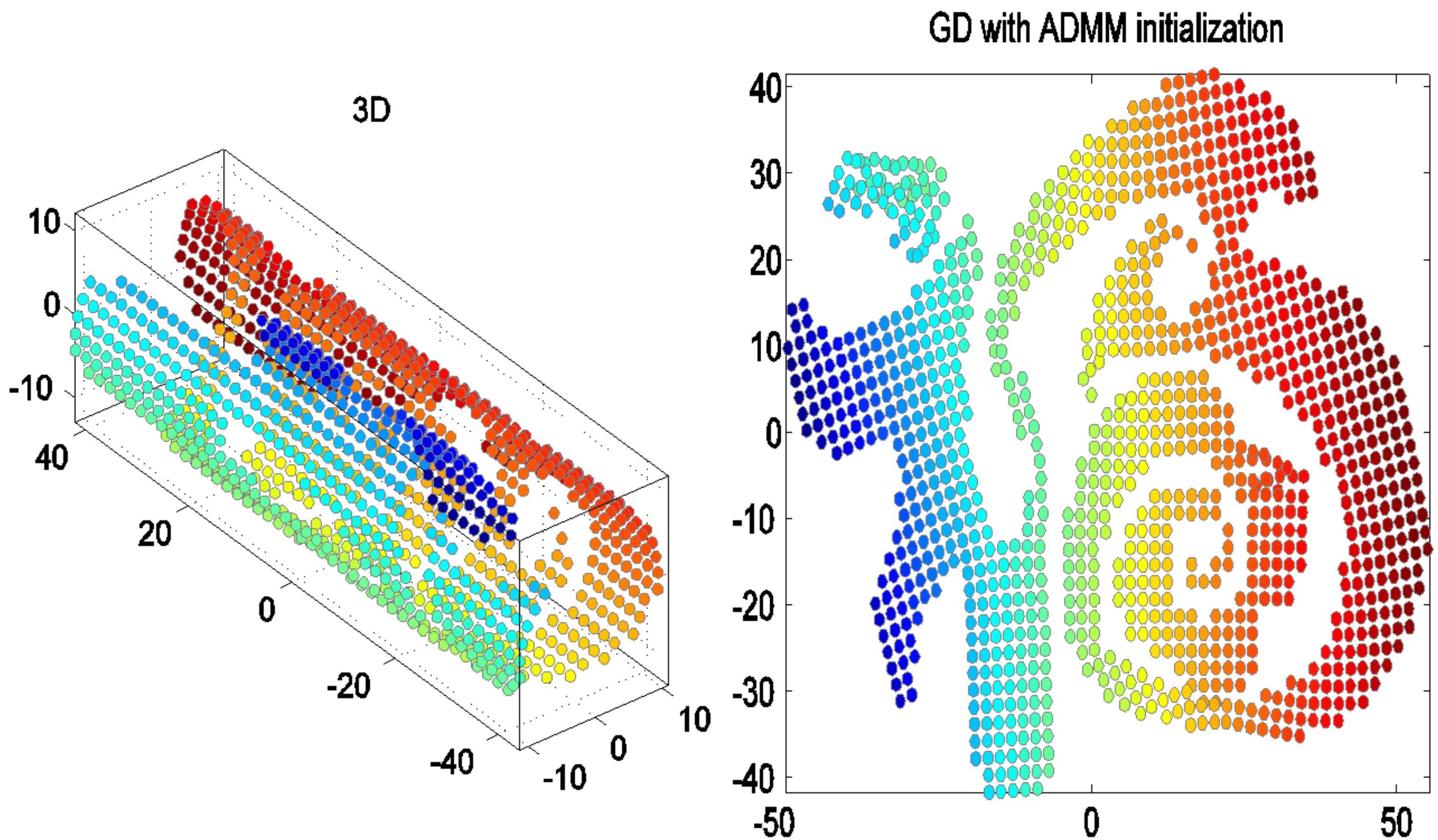


Figure 1: Dimension Reduction – Unfolding Scroll of Happiness

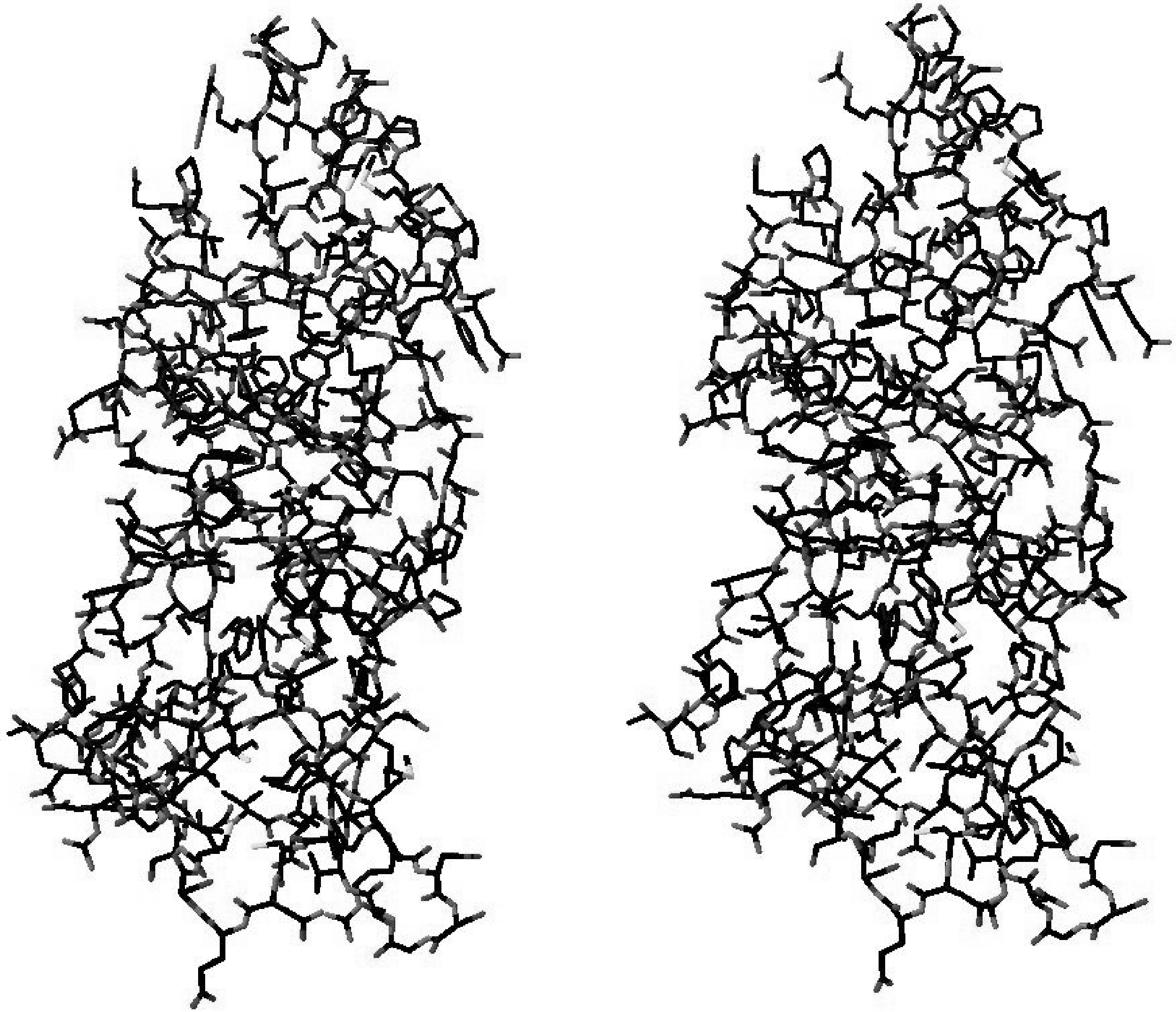


Figure 2: Molecular Conformation – 1F39(1534 atoms) with 85% of distances below $6\mathbf{r}$A and 10% noise on upper and lower bounds

VI. Steepest-Descent Reduction

Use the approximate constructed from the SDP relaxation as the initial solution and apply the gradientdescent mathod in minimizing the onlinear square constraint errors.

For SNL example, it would be

$$\min \quad \sum_{(i,j)\in E} \left(\|\mathbf{x}_i - \mathbf{x}_j\|^2 - d_{ij}^2\right)^2$$