Distributed Geometric Buildup Method for Protein Structure Determination

Zhenli Sheng(盛镇醴)

joint work with Ya-xiang Yuan

Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences

Oct 21, 2012

ORSC2012-Shenyang

Outline

1. Problem Introduction
2. Related Works
3. Distributed Geometric Buildup Method
4. Conclusions and Future work

Distance Geometry(DG) Problem

Find the coordinate vectors $x_1, x_2, \ldots, x_n$ that satisfy several given distances between them.

- data given:
  - exact distances (error-free)
  - inexact distances (with noises)
  - distance bounds
- applications:
  - graph realization
  - protein structure determination (3D)
  - sensor network localization (2D)
  - ...

Related works

- Matrix Decomposition Method (Blumenthal 1953, Torgerson 1958)
- The Embedding Algorithm (Crippen, Havel 1988)
- Global Smoothing Algorithm (Moré, Wu 1997)
- Geometric Buildup Method (Dong, Wu 2002)
- SDP Relaxiation Method (Ye, et al., 2006)
- ...

Matrix Decomposition Method

DG problem with full set of exact distances

Given a full set of distances, $d_{i,j} = \|x_i - x_j\|, \quad i, j = 1, 2, \ldots, n.$

- Set $x_n = (0, 0, 0)^{\mathrm{T}}$, we have

$$\begin{aligned} d_{i,j}^2 &= \|x_i - x_j\|^2 \\ &= \|x_i\|^2 - 2x_i^{\mathrm{T}} x_j + \|x_j\|^2 \\ &= d_{i,n}^2 - 2x_i^{\mathrm{T}} x_j + d_{j,n}^2 \qquad i, j = 1, 2, \ldots, n-1 \end{aligned} \tag{1}$$

- Define $X = (x_1, x_2, \ldots, x_n)^{\mathrm{T}}$ and $D = \{(d_{i,n}^2 - d_{i,j}^2 + d_{j,n}^2)/2 : i, j = 1, 2, \ldots, n-1\}$, (1) $\Rightarrow XX^{\mathrm{T}} = D$.

- Let $D = U\Sigma U^{\mathrm{T}}$, and $V = U(:, 1:3), \Lambda = \Sigma(1:3, 1:3)$. Then $X = V\Lambda^{1/2}$ solves the problem. [Eckart and Young 1936]

Geometric Buildup Method

1. Find four atoms to form a base
   - determine their coordinates to remove the possible translation and rotation/reflection
2. Determine atoms one by one
   - at least four distances from the undetermined atom to determined atoms are known

Zachary Voller, Zhijun Wu(2012), Distance Geometry Methods for Protein Structure Determination.

Determine one unknown atom

Given four determined atoms x_1, x_2, x_3 and x_4, which $x_i = (x_{i1}, x_{i2}, x_{i3})^{\mathrm{T}}$ are known, and four exact distances.

- $d_{i,j}^2 = \|x_i\|^2 - 2x_i^{\mathrm{T}} x_j + \|x_j\|^2 \qquad i = 1, 2, 3, 4.$
- $\Rightarrow Ax_j = b$,

  where $A = 2\begin{pmatrix} x_{11} - x_{21} & x_{12} - x_{22} & x_{13} - x_{23} \\ x_{21} - x_{31} & x_{22} - x_{32} & x_{23} - x_{33} \\ x_{31} - x_{41} & x_{32} - x_{42} & x_{33} - x_{43} \end{pmatrix}$

  and $b = \begin{pmatrix} (\|x_1\|^2 - \|x_2\|^2) - (d_{1,j}^2 - d_{2,j}^2) \\ (\|x_2\|^2 - \|x_3\|^2) - (d_{2,j}^2 - d_{3,j}^2) \\ (\|x_3\|^2 - \|x_4\|^2) - (d_{3,j}^2 - d_{4,j}^2) \end{pmatrix}.$
- Inexact distances?

Linear and Nonlinear Least-squares Approximation

DG problem with inexact distances

Suppose l distances between the unknown atom to the determined atoms are known.

- linear least-squares
  - use only the l distances
  - $\min \|b - Ax_j\|$
- nonlinear least-squares
  - use all the distances among the $l+1$ atoms
  - solve a matrix decomposition problem
  - move the same points in two different reference system coincide

Atilla Sit, Zhijun Wu and Ya-xiang Yuan(2009), A geometric buildup algorithm for the solution of the distance geometry problem using least-squares approximation.

Some other problems

- Not enough bases
  - less than four distances can be found
- Bad condition number
  - which means the bases are almost in the same plane!

$\Uparrow$

computational issue

- $\rightsquigarrow$ Move to the last

Motivation

Itr	RmsdErr	Itr	RmsdErr
300	1.02e-012	3000	1.19e-004
600	1.81e-010	3300	3.48e-004
900	2.06e-007	3600	1.95e-003
1200	6.69e-007	3900	1.97e-003
1500	3.36e-006	4200	2.16e-003
1800	4.68e-006	4500	2.23e-003
2100	7.87e-006	4800	2.79e-003
2400	2.06e-005	5100	3.08e-003
2700	6.99e-005	5400	4.28e-003

- 1MQQ, 5681 atoms, cutoff=6Å, 0.75%, exact distances, Buildup method
- Rounding error accumulation!

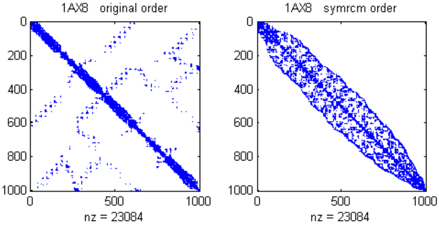
Idea of distributed method

The idea is quite simple. It is a "Divide and Conquer" method.

1. Divide the whole protein into small patches with some overlaps
2. Apply Geometric Buildup method at each patch
3. Make use of the overlap to stitch them together

How to divide

- symrcm: minimize the bandwidth



Pratik Biswas, Kim-Chuan Toh and Yinyu Ye(2007), A Distributed SDP Approach for Large-scale Noisy Anchor-free Graph Realization with Application to Molecular Conformation.

Divide(Cont'd): Laplacian Matrix

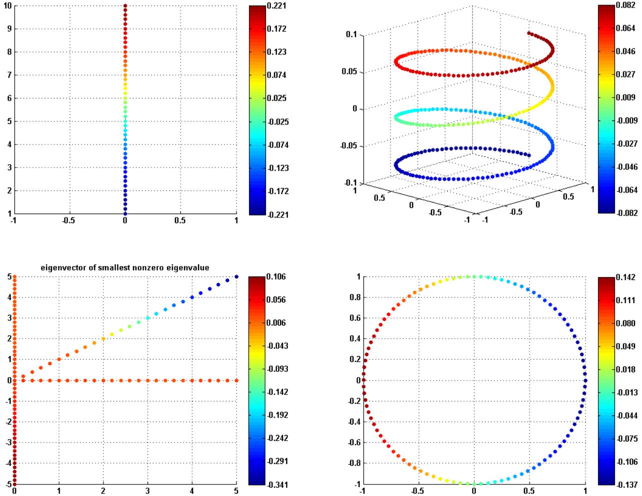
Given a graph (V, E), define its Laplacian matrix by L, whose entries $l_{i,j}$ are given by

- $$l_{i,j} = \begin{cases} deg(v_i) & \text{if } i = j, & \rightarrow -sum(L(i,:)) \\ -1 & \text{if } (i,j) \in E, & \rightarrow -exp(-d_{i,j}^2/2) \\ 0 & \text{otherwise.} & \end{cases}$$

  $L = D - A$, where D is degree matrix, and A is its adjacency matrix.

- Properties:
  - L is always positive-semidefinite.
  - 0 is always its eigenvalue and its corresponding eigenvector is $(1, 1, \dots, 1)^{\mathrm{T}}$.
  - The number of times 0 appears as an eigenvalue in the Laplacian is the number of connected components in the graph.

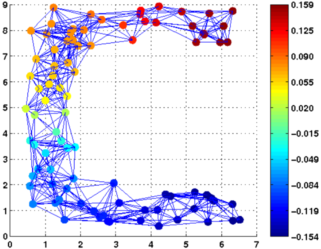
Eigenvector of smallest nonzero eigenvalue



a Conjecture

Conjecture

Given a graph (V, E), *its Laplacian matrix is defined as before, then the eigenvector of the smallest nonzero eigenvalue, which can be viewed as a function of the vertexes, monotonically decrease or increase along the main trend/direction of the graph.*



How to stitch

Given two 3D point sets $\{p_i\}$ and $\{q_i\}$, $i = 1, 2, \ldots, k$.

- $$\begin{aligned} \min_{R,T} \quad & \sum_{i=1}^{k} \|p_i - (Rq_i + T)\|_2^2 \\ \text{s.t.} \quad & R^{\mathrm{T}} R = I. \end{aligned} \tag{2}$$
- T: make their geometric center coincide
- R:

  $$\begin{aligned} \min_{R} \quad & \|P - RQ\|_F^2 \\ \text{s.t.} \quad & R^{\mathrm{T}} R = I. \end{aligned} \tag{3}$$

  Let $C = PQ^{\mathrm{T}}$, and $C = U\Sigma V^{\mathrm{T}}$, then $R = VU^{\mathrm{T}}$ solves (3). [Matrix Computation, Golub]
- Remark: a fundamental problem in Machine Intelligence and Optical Science.

Algorithm framework

```
Distributed Geometric Buildup Method for Protein Structure Determination
```

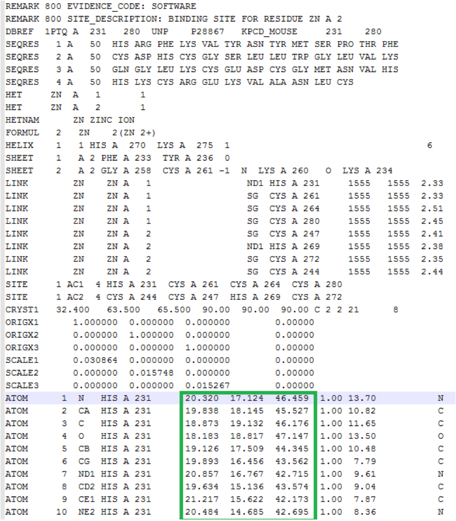
1. Initialize, set parameters: PatchNum, MaxItr
2. Find four atoms that are not in the same plane, determine their coordinates with the distances among them.
3. Construct the Laplacian matrix, sort all the atoms according to eigenvector corresponding to its minimal nonzero eigenvalue, divide the whole protein into several small patches.
4. Solve problem on each patch with Buildup method.
5. Stitch all the patches together.

Numerical experiments

- Download structure data from Protein Data Bank(PDB), obtain the original coordinates X.
- Use *disk graph model* to construct distance matrix, usually set cutoff as 5Å or 6Å.
- Solve the problem with our algorithm to get Computed coordinates Y, then compare it with X, using the criteria defined as below,

$$RMSD(X,Y) = \min_{Q,T} \|X - YQ - T\|_F / \sqrt{n}$$

PDB file



```
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE ZN A 2
DBREF  1PTQ A  231   280  UNP    P28867   KPCD_MOUSE     231    280
SEQRES   1 A   50  HIS ARG PHE LYS VAL TYR ASN TYR MET SER PRO THR PHE
SEQRES   2 A   50  CYS ASP HIS CYS GLY SER LEU LEU TRP GLY LEU VAL LYS
SEQRES   3 A   50  GLN GLY LEU LYS CYS GLU ASP CYS GLY MET ASN VAL HIS
SEQRES   4 A   50  HIS LYS CYS ARG GLU LYS VAL ALA ASN LEU CYS
HET     ZN  A   1       1
HET     ZN  A   2       1
HETNAM      ZN ZINC ION
FORMUL   2   ZN    2(ZN 2+)
HELIX    1   1 HIS A  270  LYS A  275  1                                   6
SHEET    1   A 2 PHE A 233  TYR A 236  0
SHEET    2   A 2 GLY A 258  CYS A 261 -1  N  LYS A 260   O  LYS A 234
LINK        ZN    ZN A   1                 ND1 HIS A 231     1555   1555  2.33
LINK        ZN    ZN A   1                 SG  CYS A 261     1555   1555  2.33
LINK        ZN    ZN A   1                 SG  CYS A 264     1555   1555  2.51
LINK        ZN    ZN A   1                 SG  CYS A 280     1555   1555  2.45
LINK        ZN    ZN A   2                 SG  CYS A 247     1555   1555  2.41
LINK        ZN    ZN A   2                 ND1 HIS A 269     1555   1555  2.38
LINK        ZN    ZN A   2                 SG  CYS A 272     1555   1555  2.35
LINK        ZN    ZN A   2                 SG  CYS A 244     1555   1555  2.44
SITE     1 AC1  4 HIS A 231  CYS A 261  CYS A 264  CYS A 280
SITE     1 AC2  4 CYS A 244  CYS A 247  HIS A 269  CYS A 272
CRYST1   32.400   63.500   65.500  90.00  90.00  90.00 C 2 2 21      8
ORIGX1      1.000000  0.000000  0.000000        0.00000
ORIGX2      0.000000  1.000000  0.000000        0.00000
ORIGX3      0.000000  0.000000  1.000000        0.00000
SCALE1      0.030864  0.000000  0.000000        0.00000
SCALE2      0.000000  0.015748  0.000000        0.00000
SCALE3      0.000000  0.000000  0.015267        0.00000
ATOM      1  N   HIS A 231      20.320  17.124  46.459  1.00 13.70           N
ATOM      2  CA  HIS A 231      19.838  18.145  45.527  1.00 10.82           C
ATOM      3  C   HIS A 231      18.873  19.132  46.176  1.00 11.65           C
ATOM      4  O   HIS A 231      18.183  18.817  47.147  1.00 13.50           O
ATOM      5  CB  HIS A 231      19.126  17.509  44.345  1.00 10.48           C
ATOM      6  CG  HIS A 231      19.893  16.456  43.562  1.00  7.79           C
ATOM      7  ND1 HIS A 231      20.857  16.767  42.715  1.00  9.61           N
ATOM      8  CD2 HIS A 231      19.634  15.136  43.574  1.00  9.04           C
ATOM      9  CE1 HIS A 231      21.217  15.622  42.173  1.00  7.87           C
ATOM     10  NE2 HIS A 231      20.484  14.685  42.695  1.00  8.36           N
```

1PTQ.pdb

Data information

exact distances					
PdbID	Num	cutoff	degree	cutoff	degree
1PTQ	402	5	5.46%	6	8.79%
1HOE	558	5	4.05%	6	6.55%
1LFB	641	5	3.40%	6	5.57%
1PHT	811	5	3.35%	6	5.37%
1POA	914	5	2.51%	6	4.07%
1AX8	1003	5	2.30%	6	3.74%
1F39	1534	5	1.47%	6	2.43%
1RGS	2015	5	1.12%	6	1.87%
1KDH	2846	5	0.83%	6	1.36%
1BPM	3671	5	0.66%	6	1.12%
1RHJ	3740	5	0.65%	6	1.10%
1HQQ	3944	5	0.60%	6	1.00%
1TOA	4292	5	0.56%	6	0.94%
1MQQ	5681	5	0.44%	6	0.75%

We test these 14 proteins which was used in Prof. Ye' paper as mentioned before.

Notice that the atom number of these proteins varies from hundreds to more that five thousand.

Computational order

cutoff=6Å, exact distances, Buildup Method, 'linear'									
PDB	Total	Rmsd	CPU	Rmsd	CPU	Rmsd	CPU	Rmsd	CPU
ID	Num	Err	time	Err	time	Err	time	Err	time
		original		greedy		randperm		randperm	
1PTQ	402	8.02e-012	0.40	1.15e-012	0.43	9.35e-013	0.56	6.21e-011	0.52
1HOE	558	2.13e-012	0.57	1.50e-012	0.69	2.15e-010	1.24	1.75e-008	0.75
1LFB	641	1.16e-010	0.68	7.59e-010	0.81	1.98e-008	1.46	5.67e-009	1.03
1PHT	811	1.38e-009	0.97	1.61e-011	1.06	2.24e-009	1.23	4.43e-011	1.77
1POA	914	4.53e-010	1.01	6.17e-010	1.26	1.53e-011	1.90	2.42e-011	2.74
1AX8	1003	3.74e-006	1.22	1.24e-011	1.49	8.74e-011	2.07	3.84e-009	3.11
1F39	1534	2.52e-007	1.90	2.32e-006	3.52	4.09e-003	6.90	7.17e-007	2.88
1RGS	2015	2.24e-002	2.54	1.08e-001	7.65	3.34e-004	7.48	8.68e-004	8.84
1KDH	2846	1.45e-003	3.74	7.15e-004	7.71	2.34e-003	42.33	2.12e-004	18.08
1BPM	3671	6.38e-002	5.70	4.45e-005	9.00	8.90e-005	16.24	7.51e-005	21.17
1RHJ	3740	7.07e+000	6.12	3.47e-008	9.83	6.92e-005	113.34	4.55e-007	67.71
1HQQ	3944	2.03e-003	6.58	4.77e-006	11.69	8.07e-005	22.15	9.26e-004	40.18
1TOA	4292	2.88e+000	6.58	2.35e+001	26.36	1.47e-005	67.02	6.64e+001	65.18
1MQQ	5681	1.13e+001	9.52	4.31e-003	46.41	1.05e+001	21.84	1.80e+000	82.65

Computational order(Cont'd)

cutoff=6Å, exact distances, Buildup Method, 'nonlinear'									
PDB ID	Total Num	Rmsd Err	CPU time	Rmsd Err	CPU time	Rmsd Err	CPU time	Rmsd Err	CPU time
		original		greedy		randperm		randperm	
1PTQ	402	1.34e-013	0.59	2.26e-014	0.98	5.56e-014	0.88	6.89e-014	0.88
1HOE	558	3.21e-013	0.92	7.80e-014	1.13	1.05e-013	1.10	7.34e-014	1.11
1LFB	641	5.53e-014	0.94	1.42e-013	1.18	4.96e-014	1.40	1.90e-013	1.49
1PHT	811	7.29e-013	1.52	6.04e-014	1.65	1.19e-013	1.71	2.20e-013	1.83
1POA	914	1.56e-013	1.41	9.14e-014	1.69	2.60e-013	3.26	8.53e-014	2.16
1AX8	1003	8.16e-013	1.68	6.28e-014	2.71	2.07e-013	4.31	8.01e-014	2.56
1F39	1534	2.52e-013	2.53	7.65e-013	4.69	5.23e-013	6.14	1.35e-012	18.56
1RGS	2015	1.93e-012	3.37	6.66e-013	8.61	3.20e-012	5.39	1.14e-011	17.50
1KDH	2846	2.31e-011	5.09	8.43e-013	10.20	6.36e-013	10.63	1.58e-010	16.19
1BPM	3671	1.99e-011	7.16	1.01e-012	14.44	1.11e-012	15.22	1.84e-012	171.51
1RHJ	3740	2.27e-010	7.65	2.25e-012	12.71	2.10e-012	49.17	1.03e-012	39.95
1HQQ	3944	4.66e-011	8.47	2.73e-012	14.25	8.22e-013	16.85	2.03e-012	93.09
1TOA	4292	1.19e-009	8.74	7.90e-011	29.27	6.83e-011	115.35	2.61e-012	54.58
1MQQ	5681	4.03e-008	12.59	4.21e-011	54.21	6.95e-010	367.92	3.61e-011	94.60

From now on, we use greedy order to implement our algorithm.

Linear VS. Nonlinear

cutoff=5Å, exact distances, Buildup Method							
PDB ID	Total Num	RmsdErr	CPU time	Det Num	RmsdErr	CPU time	Det Num
		linear			nonlinear		
1PTQ	402	4.80e-011	0.49	402	2.05e-014	0.59	402
1HOE	558	3.27e-007	0.85	558	6.52e-014	0.93	558
1LFB	641	2.52e-006	0.85	641	3.49e-014	0.96	641
1PHT	811	8.87e-007	1.14	806	2.52e-013	1.45	806
1POA	914	3.52e-004	1.62	914	3.23e-013	1.96	914
1AX8	1003	9.28e-005	1.89	1003	7.15e-014	2.16	1003
1F39	1534	6.74e-005	3.74	1534	8.26e-014	3.86	1534
1RGS	2015	8.40e+001	7.67	2010	4.46e-013	8.22	2010
1KDH	2846	7.43e+005	29.14	2845	1.03e-011	28.05	2846
1BPM	3671	3.22e+005	18.45	3665	2.86e-011	13.81	3668
1RHJ	3740	1.83e+005	28.70	3734	1.56e-012	25.75	3740
1HQQ	3944	2.85e+001	17.31	3938	2.46e-013	20.23	3938
1TOA	4292	1.08e+003	22.30	4280	2.90e-012	26.82	4280
1MQQ	5681	2.81e+000	35.01	5681	7.47e-013	35.73	5681

nonlinear: generally, a little more time, much more accurate!

Linear VS. Nonlinear(Cont'd)

cutoff=6Å, exact distances, Buildup Method							
PDB	Total	RmsdErr	CPU	Det	RmsdErr	CPU	Det
ID	Num		time	Num		time	Num
		linear			nonlinear		
1PTQ	402	1.15e-012	0.43	402	2.26e-014	0.98	402
1HOE	558	1.50e-012	0.69	558	7.80e-014	1.13	558
1LFB	641	7.59e-010	0.81	641	1.42e-013	1.18	641
1PHT	811	1.61e-011	1.06	811	6.04e-014	1.65	811
1POA	914	6.17e-010	1.26	914	9.14e-014	1.69	914
1AX8	1003	1.24e-011	1.49	1003	6.28e-014	2.71	1003
1F39	1534	2.32e-006	3.52	1534	7.65e-013	4.69	1534
1RGS	2015	1.08e-001	7.65	2015	6.66e-013	8.61	2015
1KDH	2846	7.15e-004	7.71	2846	8.43e-013	10.20	2846
1BPM	3671	4.45e-005	9.00	3671	1.01e-012	14.44	3671
1RHJ	3740	3.47e-008	9.83	3740	2.25e-012	12.71	3740
1HQQ	3944	4.77e-006	11.69	3944	2.73e-012	14.25	3944
1TOA	4292	2.35e+001	26.36	4292	7.90e-011	29.27	4292
1MQQ	5681	4.31e-003	46.41	5681	4.21e-011	54.21	5681

nonlinear: generally, a little more time, much more accurate!

distributed: symrcm VS. Laplacian

cutoff=6Å, exact distances							
PDB ID	Total Num	RmsdErr	CPU time	Det Num	RmsdErr	CPU time	Det Num
		symrcm			Laplacian		
1PTQ	402	2.24e-014	0.69	402	3.54e-014	0.62	402
1HOE	558	2.28e-013	0.95	558	8.76e-014	1.03	558
1LFB	641	7.75e-014	1.04	641	1.43e-013	1.08	641
1PHT	811	7.08e-014	1.61	811	1.08e-012	1.59	811
1POA	914	1.39e-013	1.71	914	1.80e-013	1.73	914
1AX8	1003	2.62e-013	2.08	1003	1.04e-013	1.94	1003
1F39	1534	9.30e-014	3.09	1534	1.41e-013	3.09	1534
1RGS	2015	1.47e-012	4.21	2015	1.03e-012	4.61	2015
1KDH	2846	4.30e-013	6.51	2846	1.25e-012	6.09	2846
1BPM	3671	2.35e+002	8.05	3671	3.58e-013	8.14	3671
1RHJ	3740	1.36e-012	8.56	3740	6.69e-011	8.60	3740
1HQQ	3944	1.02e+006	10.21	3944	4.29e-013	8.58	3944
1TOA	4292	8.41e+006	9.24	4292	1.92e-012	9.51	4292
1MQQ	5681	2.19e+001	14.89	5681	3.22e-012	13.33	5681

Laplacian: almost the same time, much more accurate!

Buildup VS. Distributed Buildup

cutoff=6Å, exact distances							
PDB ID	Total Num	RmsdErr	CPU time	Det Num	RmsdErr	CPU time	Det Num
		Buildup			Distributed Buildup		
1PTQ	402	2.26e-014	0.98	402	3.54e-014	0.62	402
1HOE	558	7.80e-014	1.13	558	8.76e-014	1.03	558
1LFB	641	1.42e-013	1.18	641	1.43e-013	1.08	641
1PHT	811	6.04e-014	1.65	811	1.08e-012	1.59	811
1POA	914	9.14e-014	1.69	914	1.80e-013	1.73	914
1AX8	1003	6.28e-014	2.71	1003	1.04e-013	1.94	1003
1F39	1534	7.65e-013	4.69	1534	1.41e-013	3.09	1534
1RGS	2015	6.66e-013	8.61	2015	1.03e-012	4.61	2015
1KDH	2846	8.43e-013	10.20	2846	1.25e-012	6.09	2846
1BPM	3671	1.01e-012	14.44	3671	3.58e-013	8.14	3671
1RHJ	3740	2.25e-012	12.71	3740	6.69e-011	8.60	3740
1HQQ	3944	2.73e-012	14.25	3944	4.29e-013	8.58	3944
1TOA	4292	7.90e-011	29.27	4292	1.92e-012	9.51	4292
1MQQ	5681	4.21e-011	54.21	5681	3.22e-012	13.33	5681

Distributed: almost the same accurate, much less time!

Buildup VS. Distributed Buildup(Cont'd)

cutoff=6Å, inexact distances, d=(1+2*(0.5-rand)*noise) *d, noise=1e-4							
PDB ID	Total Num	RmsdErr	CPU time	Det Num	RmsdErr	CPU time	Det Num
		Buildup			Distributed Buildup		
1PTQ	402	6.89e-004	0.64	402	5.97e-004	0.80	402
1HOE	558	7.01e-003	1.32	558	8.04e-003	0.97	558
1LFB	641	2.70e-003	1.47	641	5.68e-003	1.17	641
1PHT	811	1.73e-003	1.69	811	1.64e-001	1.80	811
1POA	914	6.05e-003	1.94	914	1.02e-002	1.76	914
1AX8	1003	6.06e-003	2.11	1003	3.98e-003	2.06	1003
1F39	1534	8.03e-002	4.10	1534	1.87e-002	3.28	1534
1RGS	2015	3.61e-002	8.74	2015	5.29e+003	5.75	2015
1KDH	2846	3.93e-002	8.92	2846	7.85e-002	6.36	2846
1BPM	3671	6.46e-002	9.94	3671	1.63e+000	9.15	3671
1RHJ	3740	2.55e+001	11.96	3740	6.97e+000	9.32	3740
1HQQ	3944	3.83e+010	12.87	3944	2.84e+007	9.08	3944
1TOA	4292	1.25e+008	26.85	4292	5.90e+004	10.02	4292
1MQQ	5681	4.04e+007	48.52	5681	4.73e+006	17.62	5681

Distributed: almost the same accurate, much less time!

Conclusions and Future work

- What we have done:
  - explore deeply the Geometric Buildup method
  - propose a distributed algorithm
  - finish some preliminary numerical experiments, which seems promising
- Future work:
  - theoretical analysis on eigenvector of Laplacian matrix
  - develop techniques to handle large noise

Thank you for your attention!