# Distance geometry for computing conformations

Ioannis Z. Emiris

## Outline

02. Distance geometry
07. Distance matrices
18. Incomplete data
28. Noisy data: triangular inequality
36. Noisy data: tetrangular inequality
42. Matrix perturbation

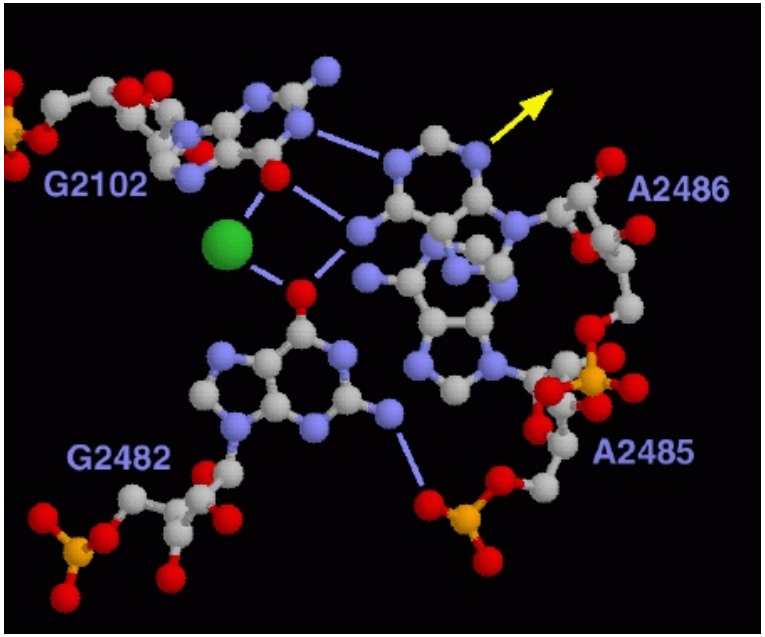
# Distance geometry

# 3D protein conformation

Goal: match pharmacophores, screen databases, identify function. . .

Principle: At a first level of approximation, bond distances and planar angles require too much energy to be altered in a protein
$\Rightarrow$ a 3D conformation depends upon the dihedral angles about the bonds.

Assume point masses, ignore chemical properties (e.g. hydrophobicity), angle limitations (postprocess).



# Structure from Distances

- Treat arbitrary number (e.g. 5000 atoms) of DOFs with:
  - Nuclear Magnetic Resonance (NMR) approximates distances [K.Wüthrich, ETHZ, Chemistry Nobel'02] "for his development of NMR spectroscopy for determining the 3-dimensional structure of biological macromolecules in solution"
  - X-ray crystallography gives accurate distances (error $\leq$ 1Å) but in crystal state, takes $\sim$ 1 research year.

Distance geometry Software effective for molecules of $>$ 100 atoms:

- DYANA [Guntert,Mumenthaler,Wüthrich'97], EMBED [Crippen,Havel'88].
- DISGEO: #507 http://qcpe.chem.indiana.edu [Havel,Wüthrich'98].
- DGSOL: open [Moré,Wu].
- ABBIE [Hendrickson], DGEOM, XPLOR module, . . .
  - http://www.netsci.org/resources/software/modeling/conf

- Main phases of Distance geometry:
  - Triangle/tetrangle inequalities used for bound smoothing
  - minimize error function within distance intervals.

## Nuclear Magnetic Resonance (NMR)

- Specific isotopes have spin ($\pm 1/2$) e.g.: H, $C^{13}$
- Each isotope absorbs and radiates back energy from electromagnetic (EM) pulse of specific "resonance" frequency

Steps of NMR:

1. Constant magnetic field applied to isotopes, spins aligned (polarized).
2. EM pulse applied, specific nuclei stimulated, and radiate energy
3. Stimulated nuclei form pairs whose distance depends on EM frequency

Experiments of NMR spectroscopy:

- Nuclear Overhauser Effect SpectroscopY (NOESY)
- Other 2D (distance) experiments: Correlation SpectroscopY (COSY), J- or Exchange (EXSY) Spectroscopies.
- Not 2D: Residual Dipolar Coupling (RDC) yields angles.

## Nuclear Enhanced Overhauser Spectroscopy (NOESY)

- Different frequencies $f \neq f' \Rightarrow$ different distances $d \neq d'$.
- Triplet $(i, j, k)$ of nuclei, stimulated: $\text{dist}(i, j) = d$, $\text{dist}(j, k) = d'$.

**Problem 1:** Match measurements to specific nuclei pairs.
Semi-automatic, human intervention still important.
**Problem 2:** Given some distances, embed nuclei in $\mathbb{R}^3$

- Only upper / lower bounds of distances are known.
- Only nuclei pairs with distance $\leq 5 - 6$Å.

Global optimization has high time complexity.

[Havel,Guntert,Wüthrich] [E-Nikitopoulos]

## Residual Dipolar Coupling (RDC)

- Consider the vector between each atom and its H.
- Measures angle of vector with a constant but unknown vector.

[Donald et al] [E-Pantos'07]

# Distance matrices

## Distance matrices

**Definition (and structure)**:
A distance matrix $M$ is square, $M_{ii} = 0$, $M_{ij} = M_{ji} \geq 0$.

**Definition.** A distance matrix $M$ is Euclidean and embeddable in $\mathbb{R}^d$ iff

$$\exists \text{ points } p_i \in \mathbb{R}^d \,:\, M_{ij} = \frac{1}{2} \text{ dist}(p_i - p_j)^2.$$

Embeddable matrices in $\mathbb{R}^3$ correspond to 3D conformations.

## Embeddability

**Thm** [Schoenberg'35,Blumenthal'53] Take border (Cayley-Menger) matrix

$$B = \begin{bmatrix} 0 & 1 \cdots 1 \\ 1 & \\ \vdots & M \\ 1 & \end{bmatrix}.$$

Then, distance matrix $M$ embeds in $\mathbb{R}^d$ iff

- $\text{rank}(B) \leq d+2$ [Cayley 1841], and
- $(-1)^k D(i_1, \ldots, i_k) \geq 0$, $\forall k \geq 2$ [Menger 1928],

where $D(\ldots)$ is any minor of $B$ indexed by rows/columns $0, i_1, \ldots, i_k$.

**Cor.** A distance matrix $M$ expresses a 3D conformation iff

- $\text{rank}(B) = 5 \Rightarrow$ all $6 \times 6$ minors $= 0 \Rightarrow$ det Border(5 points) $= 0$.
- satisfies the triangle and tetrangular inequalities.

## Small minors

Given points $i_1, \ldots, i_k$, the corresponding border matrix contains these rows/columns and the artificial (zeroth) row/column of 1's.

- det Border(points $i, j$) $= \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & d_{ij}^2/2 \\ 1 & d_{ij}^2/2 & 0 \end{bmatrix} \geq 0 \Leftrightarrow d_{ij}^2 \geq 0.$

- Triangular inequality: det Border(points 1,2,3) $=$
$(d_{12} + d_{13} + d_{23})(d_{12} + d_{13} - d_{23})(d_{12} + d_{23} - d_{13})(d_{13} + d_{23} - d_{12}) \geq 0$
Recall classical inequalities: $d_{ij} + d_{jk} \geq d_{ik} \geq |d_{ij} - d_{jk}|$.

- Tetrangular inequality: det Border(4 points) $\geq 0$.

## Embeddability, again

**Cor.** Let border (Cayley-Menger) matrix $B$ contain distance matrix $M$; the latter embeds in $\mathbb{R}^d$ and not in $\mathbb{R}^{d-1}$ iff

- $\text{rank}(B) = d + 2$ [Cayley], and
- $(-1)^k D(i_1, \ldots, i_k) > 0$, $k = 2, \ldots, d + 1$ [Menger],

where $D(\ldots)$ is any $(k + 1) \times (k + 1)$ border minor of $B$ indexed by rows/columns $0, i_1, \ldots, i_k$.

Distance matrices express many geometric properties: orthogonality, convexity etc [Michelucci].

## Cyclohexane's distance matrix

$$
\begin{array}{c} \\ p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \\ p_6 \end{array}
\begin{array}{c}
\begin{array}{cccccc} p_1 & p_2 & p_3 & p_4 & p_5 & p_6 \end{array} \\
\left[\begin{array}{cccccc}
0 & 1 & 1 & 1 & 1 & 1 \\
1 & 0 & u & c & x_{14} & c \\
1 & u & 0 & u & c & x_{25} \\
1 & c & u & 0 & u & c \\
1 & x_{14} & c & u & 0 & u \\
1 & c & x_{25} & c & u & 0 \\
1 & u & c & x_{36} & c & u
\end{array}\right.
\end{array}
\begin{array}{c} \\ 1 \\ u \\ c \\ x_{14} \\ c \\ u \end{array}
\begin{array}{c} \\ u \\ c \\ x_{25} \\ c \\ u \\ 0 \end{array}
\left.\begin{array}{c} \\ u \\ c \\ x_{36} \\ c \\ u \\ 0 \end{array}\right]
$$

Known: $u \simeq 1.526$ (adjacent), $\phi \simeq 110.4^{\circ} \Rightarrow c \simeq 2.285$ (triangle).

Rank condition ($= 5$) equivalent to the vanishing of all $6 \times 6$ minors.
This yields a $3 \times 3$ system of quadratic polynomials in the $x_{14}, x_{25}, x_{36}$.

If all $c, u$ same, then 2 isolated conformations, one 1-dim set.
If the $c, u$ perturbed, then $\leq 16$ solutions $\in \mathbb{R}$ [E-Mourrain].

## Point coordinates from distances

**Input**: distances $d_{ij}$ among $n+1$ points $(i,j=0,\ldots,n)$ in $\mathbb{R}^3$.
**Output**: point coordinates (up to rigid transforms) in $\mathbb{R}^3$.

Consider points $p_i \in \mathbb{R}^d$ as (column) vectors, $i=0,\ldots,n$, where

$$d_{ij}^2 = |p_i - p_j|^2,\ d_{i0}^2 = |p_i|^2, \quad \text{setting } p_0 = 0.$$

$$\text{Now } d_{ij}^2 = |p_i|^2 - 2p_i^T p_j + |p_j|^2 \ \Leftrightarrow\ p_i^T p_j = \frac{d_{i0}^2 - d_{ij}^2 + d_{j0}^2}{2} =: G_{ij}$$

defines (Gram) $n \times n$ matrix $G$ of inner products:

$$G = \begin{pmatrix} p_1^T p_1 & p_1^T p_2 & \cdots & p_1^T p_n \\ p_2^T p_1 & p_2^T p_2 & \cdots & p_2^T p_n \\ \vdots & \vdots & \ddots & \vdots \\ p_n^T p_1 & p_n^T p_2 & \cdots & p_n^T p_n \end{pmatrix}$$

## Coordinates from inner products

Gram matrix expresses coordinate vectors:

$$[G_{ij}] = [p_i^T p_j] = P^T \cdot P, \text{ where } P = [p_1, \ldots, p_n] \text{ is } 3 \times n.$$

$G$ is real symmetric, so eigen-values/vectors $\in \mathbb{R}$ and $U = V$:
The Singular Value Decomposition (SVD) yields

$$G = V\Sigma V^T, \quad V^T V = I, \; \Sigma = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix}, \; \sigma_i = |\text{eigenval}_i|.$$

Since $\text{rank}(d_{ij}) = 3 \Rightarrow \text{rank}(G) = 3$, therefore can have:

- singular vector matrix $V$ be $n \times 3$, and
- singular value matrix $\Sigma$ positive, $3 \times 3$ diagonal.

Then, $G = V\sqrt{\Sigma}\sqrt{\Sigma}V^T = P^T P \Rightarrow P := \sqrt{\Sigma}V^T$.

## Embeddability Theorem

**Theorem.** $\{p_1, \ldots, p_n\}$ embed in $\mathbb{R}^3$ (and not $\mathbb{R}^2$) iff
Gram matrix $G = P^T P$ defined by coordinates $P$ has rk$G = 3$.

[$\Leftarrow$] SVD: symmetric real $G = VSV^T$ (previous slide),
where $V^T V = VV^T = I$, $S =$diag$[\sigma_1, \ldots, \sigma_n], \sigma_i =$ |eigenvalue$_i$| $\geq 0$.
rk$G = 3 \Rightarrow S =$ diag$[\sigma_1, \sigma_2, \sigma_3, 0, \cdots]$, $\sqrt{S} = [\sqrt{\sigma_i}]$, $G = V\sqrt{S} \cdot \sqrt{S}V^T$.
$P := V\sqrt{S}$ is $n \times 3$: defines $n$ points $p_i \in \mathbb{R}^3$.

- SVD $= O(n^3)$ and well implemented, e.g. Matlab.

[$\Rightarrow$] $P = [p_1, \ldots, p_n]$ is $3 \times n$.
rk$(AB) =$ min{rk$A$,rk$B$} $\Rightarrow$ rk$G =$rk$P$.
Embed $\Rightarrow \exists p_a, p_b, p_c$ linearly independent $\Rightarrow$ rk$P = 3$.

- $\forall y \in \mathbb{R}^n$, $y^T P^T P y = (Py)^T (Py) \geq 0 \Rightarrow G = P^T P \succeq 0$.

## Positive (semi)definite matrices

Def. An $n \times n$ real matrix $M$ is positive (semi)definite if

$$x^T M x > 0, \quad (x^T M x \geq 0), \quad \forall\, x \neq 0.$$

Denoted $M \succ 0$, $M \succeq 0$.

Examples: $\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$ $\begin{bmatrix} 3 & -1 & -2 \\ -2 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix}$

Lemma [Sylvester]: For all $i \times i$ upper-left minor $M_i$:
$M \succ 0 \Leftrightarrow \det M_i > 0$, $M \succeq 0 \Rightarrow \det M_i \geq 0$.

## Symmetric P(S)D matrices

Hypothesis: Real symmetric $M = M^T$ (or complex Hermitian $M = M^*$)

- $M \succeq 0 \Leftrightarrow$ all eigenvalues $\geq 0$, $M \succ 0 \Leftrightarrow$ all eigenvalues $> 0$.
- $M \succ 0 \Leftrightarrow \exists\, B : \det B \neq 0, M = BB^T$.
- $M \succ 0 \Leftrightarrow \exists$ low-triangular $B : M = B^2$, then $B = \sqrt{M}$.

Unique $B : B \succ 0, B = B^T$.

## Complexity

- Given complete set of exact distances: Embed $\in$ P.

- Given incomplete set of exact distances: Embed-$\mathbb{R}$ $\in$ NP-hard.

Reduction of set-partition [Saxe'79].

- Embed-$\mathbb{R}^k \in$ NP-hard, for $k \geq 2$, even if weights $\in \{1, 2\}$ [Saxe'79].

- (approximate) Embed-$\mathbb{R}^2 \in$ NP-hard for planar graphs: weights $= 1$

[Demaine,RoteEtal]

- Given distances $\pm\epsilon$, approximate-Embed $\in$ NP-hard [Moré,Wu'96].

# Incomplete data

## Matrix completion

Prob. Given incomplete $M$, complete it to PSD (or distance) matrix.
This is a feasibility question on a PSD program.

Def. PSD program (feasibility): Given matrices $A_1, \ldots, A_m$, $b_1, \ldots, b_m \in \mathbb{R}$, $\exists$ ? matrix $X$ s.t.:

$$X \succeq 0,\ b_k = A_k \cdot X := \sum_{i,j} A_{i,j}^{(k)} X_{i,j},\ k = 1, \ldots, m.$$

Note $A_k \cdot X = \text{trace}(A_k^T X)$.
Full PSD program includes also linear objective function on the $X_{i,j}$.

- If $M$ contains determined line/column, completion reduces to that of submatrix.

## Complexity of PSD

- Solving PSD programs with arbitrary precision $\in P_{\mathbb{R}}$:

interior-point [Karmarkar], ellipsoid method [Khachiyan].

- Unknown whether PSD-program $\in \mathrm{NP}_{bit}$.
- Unknown whether PSD-feasibility $\in \mathrm{P}_{bit}$.

- LP $\in \mathrm{P}_{bit}$ (weak polytime) interior-point [Karmarkar], ellipsoid [Khachiyan]
- Open whether LP in $\mathrm{P}_{\mathbb{R}}$ (strong polytime).

- Typically, symmetric $M = M^T$, diagonal entries known.
- Focus on polynomial instances of PSD completion [Laurent].

## Further into partial PSD matrices

**Def.** $M$ is partial-PSD if $\forall$ principal specified submatrix is PSD.

**Lem.** If incomplete matrix $M$ has a P(S)D / distance matrix completion,
$\Rightarrow$ $M$ is partial-P(S)D / partial-distance matrix.
Necessary (not sufficient) conditions for completing $M$.

**Counter-example:** Partial-PSD but $\nexists$ PSD-completion:

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

Note: cycle of length 4.

## Chordal Graphs: poly-1

Defn. A graph is chordal if its largest cycle is of length 3.

Defn. Given symmetric matrix $M$, construct graph $G$ with nodes corresponding to rows/columns and edges to every specified matrix entry.

Thm [Johnson,Wolkowitz,etal'84,Bakonyi,Johnson'95].
Every partial-PSD (or partial-distance) matrix $M$ with graph $G$ has a PSD (or distance-matrix) completion iff $G$ is chordal.

Proof: [$\Leftarrow$] algorithm in $\mathsf{P}_{bit}$ [Laurent'98].
Moreover: $\forall$ principal submatrices embed in $\mathbb{R}^k \Rightarrow$ entire matrix in $\mathbb{R}^k$.

## Chordal filling: poly-2

Thm [Laurent]. Given $M, G$, if min #edges added to make $G$ chordal is $O(1)$, then PSD-completion $\in \mathsf{P}_{bit}$.

Proof. PSD program with #constraints $= O(1)$.

- Generally, minimizing #edges to make $G$ chordal is NP-hard.

## No $K_4$, class $G_{p,q}$: poly-3, -4

Def Graph $G'$ is similar to $G$ if obtained by adding vertices to edges of $G$

Thm [Laurent]. Given $M$ and $G$, where $G$ has no subgraph similar to clique $K_4$. Then, $M$ can be completed to a PSD matrix in $\mathsf{P}_{\mathbb{R}}$.

Def. Let be $p, q \geq 1$. Class $G_{p,q}$ contains graphs $G = (V, E)$ s.t.:

1. $\exists\, V_1, V_2 \subset V$, $V_1 \cap V_2 = \emptyset$, $\min(|V_1|, |V_2|) = p$,
2. $F := \{ij | i \in V_1, j \in V_2\} \Rightarrow F \cap E = \emptyset$,
3. Graph $H := (V, E \cup F)$ is chordal,
4. $H$ has $q$ max-cliques that are not cliques of $G$.

Thm [Laurent]. For integers $p, q \geq 1$, matrices with graph $G \in G_{p,q}$ can be completed to a PSD matrix in weak Polytime.

## Dense Graphs: poly-5

- Consider a clique missing edges incident to a vertex $v$. Then, in poly-time, $M$ is PSD-completed and its minimum embedding dimension (MED) computed. E.g.:

$$\mathsf{MED}(v - G_1 \equiv G_2) = \max\{\mathsf{MED}(v - G_1), \mathsf{MED}(G_1 \equiv G_2)\}.$$

- Same for star of cliques $K_i$:

$$\mathsf{MED}(\text{star}) = \max_i\{\mathsf{MED}(K_i)\}.$$

[E-Fragoudakis-Markou]

## Graph embedding

- [Bourgain'93]. Every graph with $n$ vertices can be embedded in *some* Euclidean space *(of any dimension)* with distortion $O(\log n)$.

- [Barvinok]. If graph embedded in $\mathbb{R}^d$, for some $d$, then embedded in dimension $d^*$ s.t.:

$$d^* = \lfloor \frac{\sqrt{8E+1}-1}{2} \rfloor, \quad E+1 \leq \binom{d^*+2}{2}.$$

- Conj. $d^* <$ #vertices, or $\leq \max \deg(\text{graph})$.

# Noisy data: triangular inequality

## Minors

- $6 \times 6$ minors vanish $\Leftrightarrow$ det Border(5 points) $= 0$.

- det Border(2 points) $= \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & d_{ij}^2/2 \\ 1 & d_{ij}^2/2 & 0 \end{bmatrix} \geq 0 \Leftrightarrow d_{ij}^2 \geq 0.$

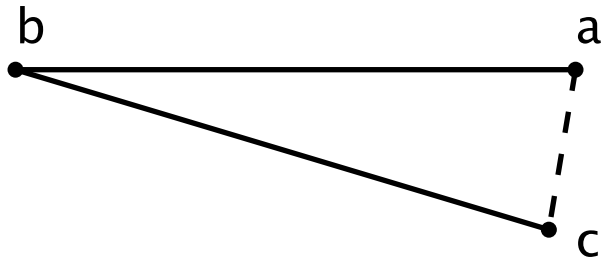
- Triangular inequality: det Border(3 points) $=$
$(d_{12} + d_{13} + d_{23})(d_{12} + d_{13} - d_{23})(d_{12} + d_{23} - d_{13})(d_{13} + d_{23} - d_{12}) \geq 0$

- Tetrangular inequality: det Border(4 points) $\geq 0$.

## Smoothing triangular inequalities

• Triangle inequalities (equality iff coliner): For any 3 points in Euclidean space of any dimension (including $\mathbb{R}^3$) the triangle inequality holds:

$$|d_{ik} - d_{kj}| \leq d_{ij} \leq d_{ik} + d_{kj}.$$



## Bound smoothing

- Improve upper bound by forcing:

$$u_{ij} \leq u_{ik} + u_{kj}.$$

All-min-paths in single pass: $\Omega(VE)$,
$O(V^3)$ [FloydWarshal], $O(V(V+m)\lg V), m$ edges in minpaths [KargerEtal91]

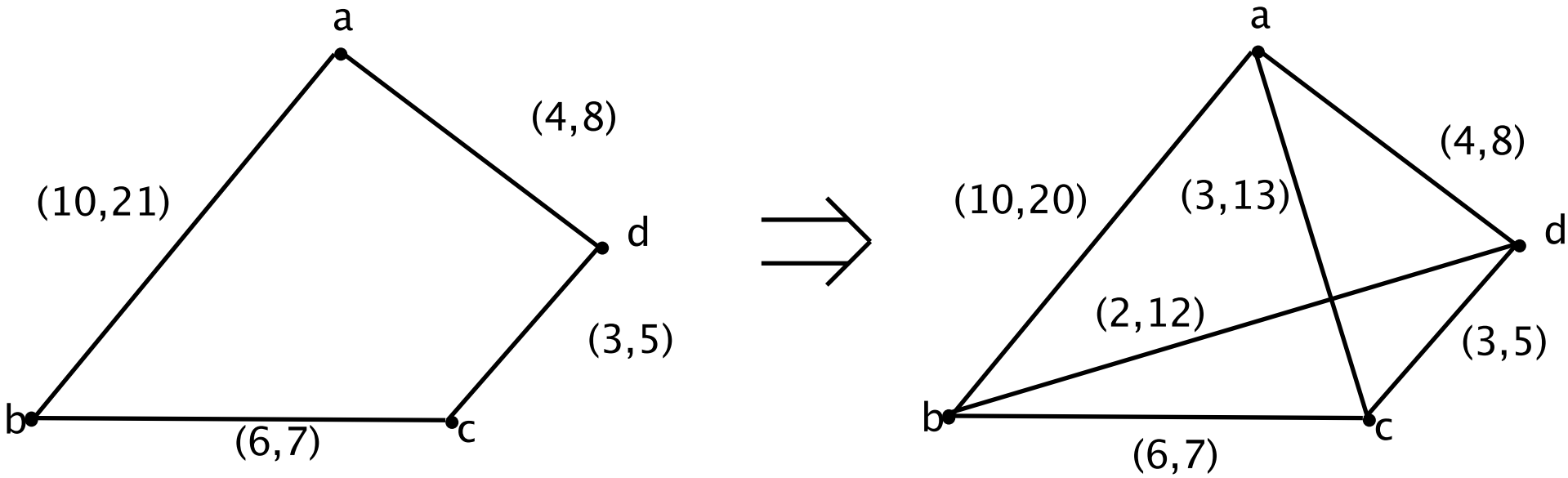
- Improve lower bound by forcing:

$$l_{ij} \geq l_{ik} - u_{kj},\ l_{kj} - u_{ik} \Rightarrow l_{ij} \geq l_{km} - u_{ik} - u_{mj},$$

where indices $i, j, k, m$ not necessarily all distinct.
Independent of upper bounds; also reduces to all-minpaths.

# Example

$$l_{bd} \geq l_{ab} - u_{ad} = 2$$
$$l_{ac} \geq l_{ab} - u_{bc} = 3$$
$$u_{bd} \leq u_{bc} + u_{cd} = 12$$
$$u_{ac} \leq u_{ad} + u_{cd} = 13$$
$$u_{ab} \leq u_{ad} + u_{bd} = 20$$



## Simple algorithm for upper bounds

1. Check all triplets of nodes, applying the inequality for upper bounds. Repeat until no improvement is possible.

2. The runtime $= 2\binom{n}{r} = O(n^3)$. Second pass merely confirms that 1st pass found best possible bounds.

3. Algorithm is essentially Floyd's, $O(n^3)$ is best for dense graphs.

## A better algorithm

1. For every graph node $i$ consider all $j, k$ such $(i, j), (i, k), (j, k) \in E$.
2. This correctly checks all triplets with 3 edges. Runtime is

$$n \sum_i deg(i) = O(nm)$$

3. For edges not in $E$, apply the procedure "Fillin" below.
4. Connected graph, $m = O(n)$: best $\binom{n}{2}$ bounds in $= O(n^2)$

"Fillin" procedure:

3.1. For every $i$ constuct BFS-tree by picking "closest" neighbors first depending on refined upper bounds.

3.2. $G$ connected $\Rightarrow$ tree is spanning.

3.3. $\forall j$ s.t. $(i, j) \notin E$, set as $\widehat{u}_{ij}$ the distance of $i, j$ on tree.

3.3. The running time is linear per $i$, hence $O(n^2)$ in total.

New (?) algorithm [Palios'05].

## Exotic data structures

- Improve Floyd using Dijkstra's algorithm for single-source min paths. Using binary heaps Dijkstra $= O((m+n)\log n)$; with Fibonacci heaps it is $O(m + n\log n)$, but with high hidden constant.

- [Karger,Koller,Philips'91] single-source minpaths $= O((n+\mu)\log n)$ using binary heaps; $O(\mu + n\log n)$ using Fibonacci heaps, where $\mu \leq m$ is number of edges actually participating in some minpath.

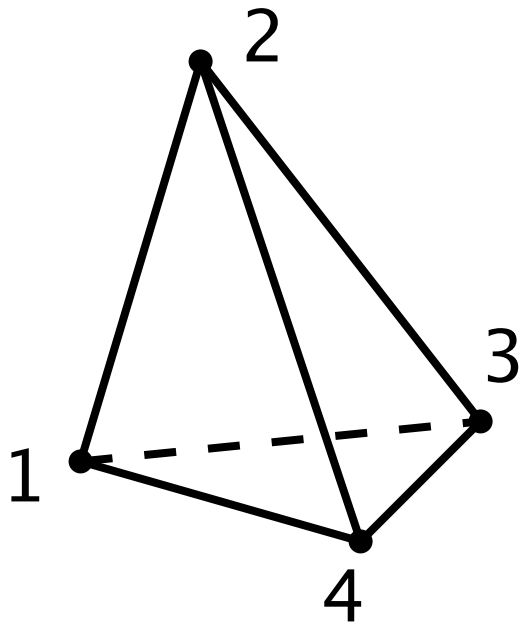
- If $m = \Theta(n) \Rightarrow$ runtime for $n$ executions becomes $O(n^2 \log n)$ with either kind of heaps.

# Noisy data: tetrangular inequality

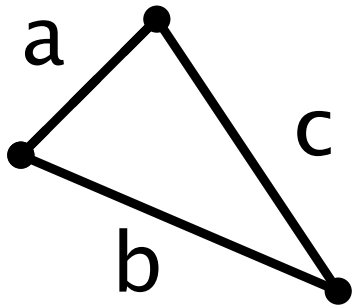
## Four points

Four atoms (1, 2, 3, 4), Cayley-Menger determinant:

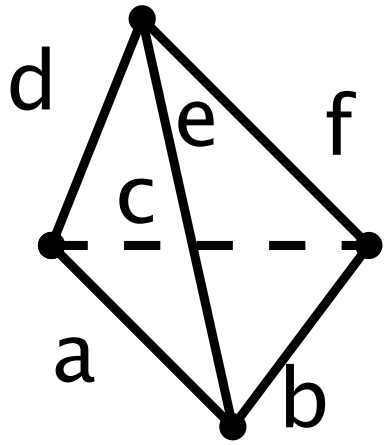
$$\begin{vmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & d_{12}^2 & d_{13}^2 & d_{14}^2 \\ 1 & d_{21}^2 & 0 & d_{23}^2 & d_{24}^2 \\ 1 & d_{31}^2 & d_{32}^2 & 0 & d_{34}^2 \\ 1 & d_{41}^2 & d_{42}^2 & d_{43}^2 & 0 \end{vmatrix} > 0$$



## Heron's formula ($'H\rho\omega\nu$)




T




T

- Triangular: $CM(a,b,c) = 16(\text{Area of T})^2 =$

$$= (a+b+c)(-a+b+c)(a-b+c)(a+b-c).$$

- Tetrangular: $CM(a,b,c,d,e,f) = 288(\text{Volume of T})^2.$

## Independence

- To tighten bounds for a point pair, we have to solve 64 inequalities.
- If given bounds satisfy triangle inequality, only 7 are non-redundant.
- Consider the (3,4) distance: For upper limit $u_{34}$:

$$CM(l_{12}, u_{13}, u_{14}, u_{23}, u_{24}, u_{34}) \geq 0,$$
$$CM(u_{12}, l_{13}, l_{14}, u_{23}, u_{24}, u_{34}) > 0,$$
$$CM(u_{12}, u_{13}, u_{14}, l_{23}, l_{24}, u_{34}) > 0,$$

For the lower limit $l_{34}$ we have:

$$CM(u_{12}, u_{13}, l_{14}, l_{23}, u_{24}, l_{34}) > 0,$$
$$CM(u_{12}, l_{13}, u_{14}, u_{23}, l_{24}, l_{34}) > 0,$$
$$CM(l_{12}, l_{13}, u_{14}, l_{23}, u_{24}, l_{34}) > 0,$$
$$CM(l_{12}, u_{13}, l_{14}, u_{23}, l_{24}, l_{34}) > 0.$$

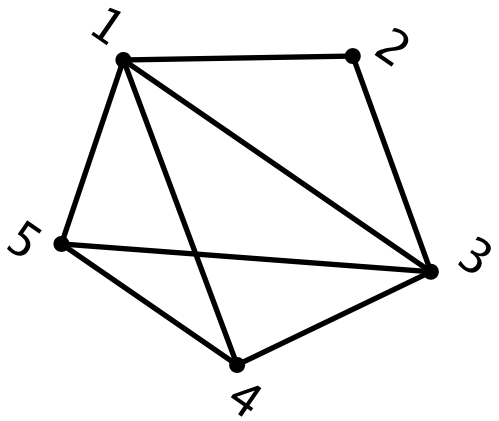
## Bound smoothing

Algorithm:

0. Fix a Tolerance value.
1. Initialize using the triangle inequality.
2. Check all $\binom{n}{4}$ quadruples of nodes, applying 7 inequalities.
3. Repeat until max change in any bound is $<$ Tolerance.

- Order of quadruples/inequalities does not affect output.
- BUT: may get in cycle and progress very slowly to final result.
- Tetrangle inequalities much tighter than triangular, but slow.

## Example of cycling




- 5 edges: $u = l = 1$,
- 3 shown diagonals: $l = 1.617$,
- true diagonal $= 1.618 = 2\cos 36^o$

- Triangle-Bound-Smoothing yields upper bound $= 2$ for each of 5 diagonals. New $u_{24}$ by using quadruple (2,3,4,5), which is used in (2,4,5,1) to update $u_{25}$.

- After 30 passes, tolerance $= 10^{-14}$, we have

$u_{24} = 1.6207323507579925$, $u_{25} = 1.6207323507579441$.

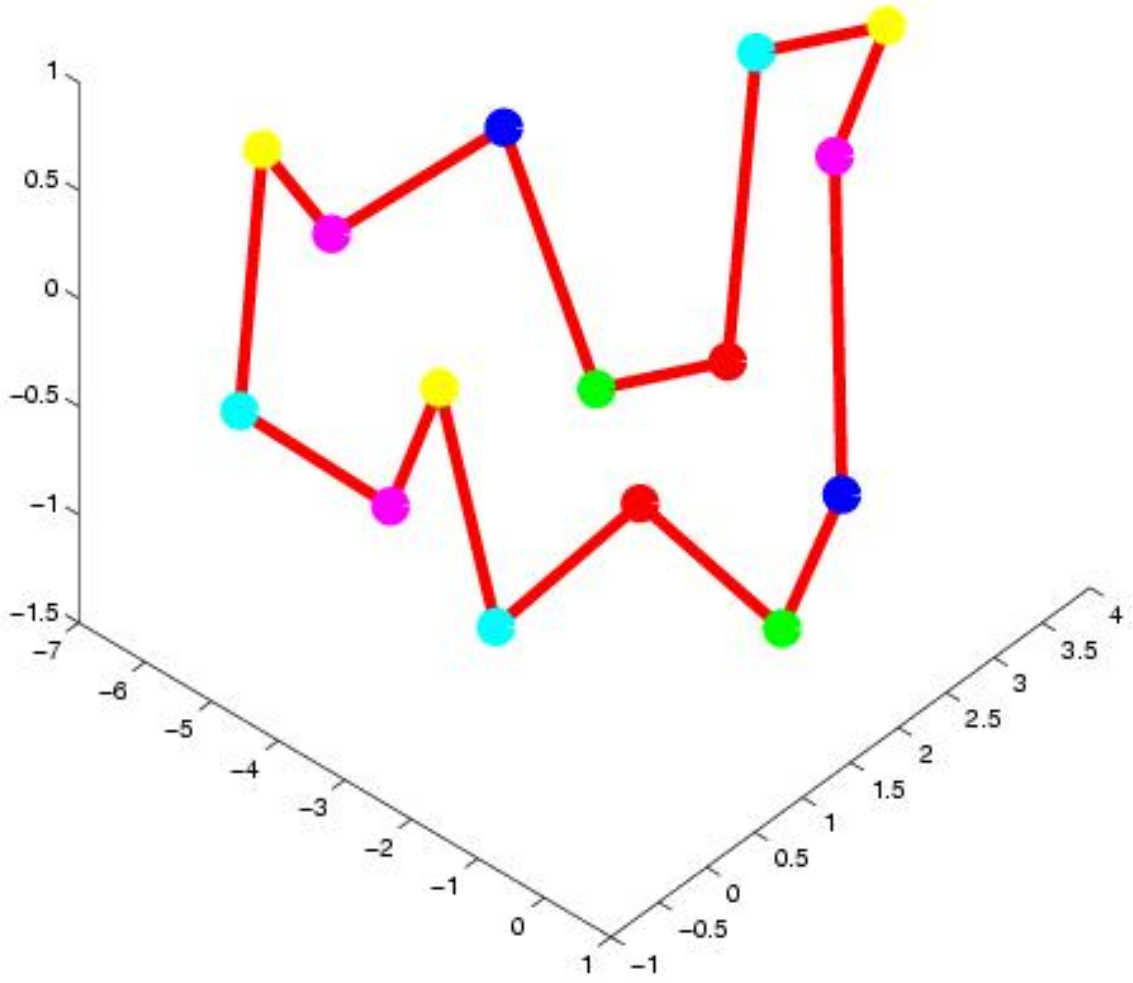
# Matrix perturbations

## Performance on ring molecules

Matlab/Scilab implementation reduces the 6-th eigenvalue; precision = 16 decimal digits, 500-MHz Pentium-III. [Nikitop-E'02].

| #atoms | Init. eigval | Final eigval | Iterations | Time [sec.] | KFlop |
|---|---|---|---|---|---|
| 7 | 2.98e-02 | 6.64e-14 | 3 | 0.01 | 36 |
| 8 | 2.57e-02 | 4.43e-12 | 3 | 0.05 | 49 |
| 9 | 2.10e-02 | 6.29e-11 | 3 | 0.05 | 73 |
| 10 | 2.38e-02 | 2.95e-13 | 3 | 0.11 | 109 |
| 11 | 3.16e-02 | 2.60e-12 | 3 | 0.16 | 165 |
| 12 | 8.13e-02 | 1.20e-07 | 3 | 0.22 | 282 |
| 13 | 8.09e-02 | 8.49e-08 | 3 | 0.30 | 450 |
| 14 | 3.72e-02 | 6.04e-13 | 3 | 0.49 | 606 |
| 15 | 3.53e-02 | 2.02e-14 | 3 | 0.77 | 940 |
| 16 | 3.78e-02 | 1.72e-12 | 3 | 1.15 | 1404 |
| 17 | 3.83e-02 | 1.70e-13 | 3 | 1.54 | 2082 |
| 18 | 3.53e-02 | 3.93e-13 | 3 | 2.14 | 3039 |
| 19 | 3.80e-02 | 4.59e-14 | 3 | 2.91 | 4344 |
| 20 | 4.00e-02 | 7.09e-13 | 3 | 3.79 | 6136 |

# Example: 15-atom molecular backbone



## Distance matrices: done / to do

- Triangle and tetrangle inequalities for bound smoothing.
- Find conformation if started in its neighborhood.
- Bound locally the number of DOFs on the conformation manifold

[E-Nikitopoulos].

- Treat family of known proteins, or just with common features.
- Partial distance information ("on-line" or false data), or probability distribution, to avoid full NMR-spectra assignment.
- Determine the minimum set of required distances.