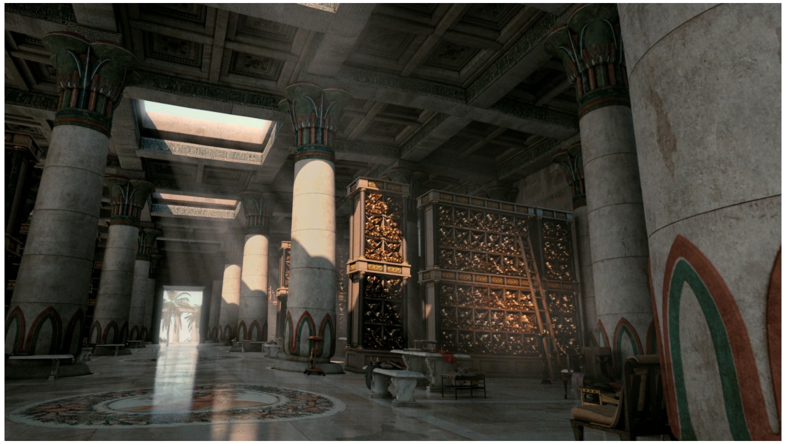
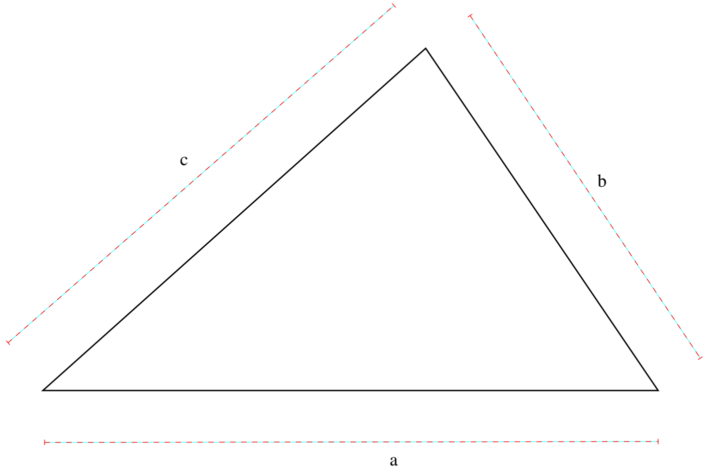
Section 8

Distance Geometry

A gem in Distance Geometry

- *Heron's theorem*
- **Heron lived around year 0**
- Hang out at Alexandria's library





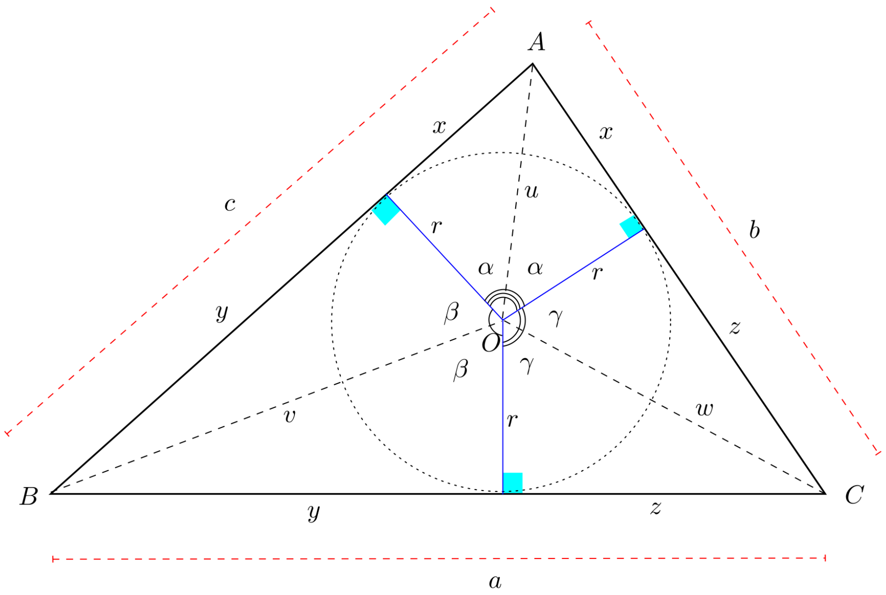
$$A = \sqrt{s(s-a)(s-b)(s-c)}$$

- A = **area of triangle**
- $s = \frac{1}{2}(a+b+c)$

Useful to measure areas of agricultural land

Heron's theorem: *Proof*

A. $2\alpha + 2\beta + 2\gamma = 2\pi \Rightarrow \alpha + \beta + \gamma = \pi$



$$\begin{aligned} r + ix &= ue^{i\alpha} \\ r + iy &= ve^{i\beta} \\ r + iz &= we^{i\gamma} \end{aligned}$$

$\Rightarrow (r+ix)(r+iy)(r+iz) = (uvw)e^{i(\alpha+\beta+\gamma)} =$
$uvw\, e^{i\pi} = -uvw \in \mathbb{R}$

$\Rightarrow \mathrm{Im}((r+ix)(r+iy)(r+iz)) = 0$

$\Rightarrow r^2(x+y+z) = xyz \Rightarrow r = \sqrt{\frac{xyz}{x+y+z}}$

B. $s = \frac{1}{2}(a+b+c) = x+y+z$

$$\begin{aligned} s - a &= x + y + z - y - z = x \\ s - b &= x + y + z - x - z = y \\ s - c &= x + y + z - x - y = z \end{aligned}$$

$$\mathcal{A} = \frac{1}{2}(ra + rb + rc) = r\frac{a+b+c}{2} = rs = \sqrt{s(s-a)(s-b)(s-c)}$$

Heron's gifted disciple

- This proof by **Miles Edwards** as a high school student in 2007 `lhsblogs.typepad.com/files/a-proof-of-heron-formula-miles-edwards.pdf`
  (tried to contact him, never got an answer)
- Beats all other proofs for compactness and elegance
  ...Other people think so too!
  `jwilson.coe.uga.edu/emt725/Heron/HeronComplex.html`
- He was ranked 16th in the Putnam Competition 2010 `newsinfo.iu.edu/news/page/normal/13885.html`
- Want to see what kind of exercises he was able to solve? `kskedlaya.org/putnam-archive/2010.pdf`
- An example:

  Given that A, B, C are noncollinear points in the plane with integer coordinates such that the distances AB, AC and BC are integers, what is the smallest possible value of AB?

Another gem in DG

- [I. Schoenberg, *Remarks to Maurice Fréchet's article "Sur la définition axiomatique d'une classe d'espaces distanciés vectoriellement applicable sur l'espace de Hilbert"*, Ann. Math., 1935]
- **Question:** Given $n \times n$ symmetric matrix D, what are necessary and sufficient conditions s.t. D is a EDM[1] corresponding to n points $x_1, \ldots, x_n \in \mathbb{R}^K$ with K minimum?
- **Main theorem**:
  Thm.
  $D = (d_{ij})$ is an EDM iff $\frac{1}{2}(d_{1i}^2 + d_{1j}^2 - d_{ij}^2 \mid 2 \leq i, j \leq n)$ is PSD of rank K
- Gave rise to one of the most important results in data science: **Classic Multidimensional Scaling**

[1] *Euclidean Distance Matrix*

Gram in function of EDM

- $x = (x_1, \ldots, x_n) \subseteq \mathbb{R}^K$, written as $n \times K$ matrix
- matrix $G = xx^\top = (x_i \cdot x_j)$ is the *Gram matrix* of x
- **Schoenberg's theorem:** ***relation between EDMs and Gram matrices***

$$G = -\frac{1}{2} J D^2 J \qquad (\S)$$

- $D^2 = (d_{ij}^2)$, $J = I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^\top$

Multidimensional scaling (MDS)

- **Often get approximate EDMs $\tilde{D}$ from raw data**
  (dissimilarities, discrepancies, differences)
- $\tilde{G} = -\frac{1}{2} J \tilde{D}^2 J$ **is an approximate Gram matrix**
- Approximate Gram $\Rightarrow$ spectral decomposition $P \tilde{\Lambda} P^\top$ has $\tilde{\Lambda} \not\geq 0$
- **Let Λ closest PSD diagonal matrix to $\tilde{\Lambda}$:**
  zero the negative components of $\tilde{\Lambda}$
- $x = P\sqrt{\Lambda}$ **is an "approximate realization" of $\tilde{D}$**

Classic MDS: Main result

1. **Prove** $G = -\frac{1}{2} J \tilde{D}^2 J$
2. **Prove matrix is Gram iff it is PSD**

Classic MDS: Proof 1/3

- Assume zero centroid WLOG (can translate x as needed)
- Expand: $d_{ij}^2 = \|x_i - x_j\|_2^2 = (x_i - x_j)(x_i - x_j) = x_i x_i + x_j x_j - 2 x_i x_j \quad (*)$
- Aim at "inverting" $(*)$ to express $x_i x_j$ in function of d_{ij}^2
- Sum $(*)$ over i: $\sum_i d_{ij}^2 = \sum_i x_i x_i + n x_j x_j - 2 x_j \sum_i x_i$ (0 by zero centroid)
- Similarly for j and divide by n, get:

$$\frac{1}{n}\sum_{i\le n} d_{ij}^2 = \frac{1}{n}\sum_{i\le n} x_i x_i + x_j x_j \quad (\dagger)$$

$$\frac{1}{n}\sum_{j\le n} d_{ij}^2 = x_i x_i + \frac{1}{n}\sum_{j\le n} x_j x_j \quad (\ddagger)$$

- Sum $(\dagger)$ over j, get:

$$\frac{1}{n}\sum_{i,j} d_{ij}^2 = n\frac{1}{n}\sum_i x_i x_i + \sum_j x_j x_j = 2\sum_i x_i x_i$$

- Divide by n, get:

$$\frac{1}{n^2}\sum_{i,j} d_{ij}^2 = \frac{2}{n}\sum_i x_i x_i \quad (**)$$

Classic MDS: Proof 2/3

- Rearrange (∗), (†), (‡) as follows:

$$
\begin{aligned}
2x_i x_j &= x_i x_i + x_j x_j - d_{ij}^2 && (5)\\
x_i x_i &= \frac{1}{n}\sum_j d_{ij}^2 - \frac{1}{n}\sum_j x_j x_j && (6)\\
x_j x_j &= \frac{1}{n}\sum_i d_{ij}^2 - \frac{1}{n}\sum_i x_i x_i && (7)
\end{aligned}
$$

- Replace LHS of Eq. (6)-(7) in Eq. (5), get

$$2x_i x_j = \frac{1}{n}\sum_k d_{ik}^2 + \frac{1}{n} d_{kj}^2 - d_{ij}^2 - \frac{2}{n}\sum_k x_k x_k$$

- By (∗∗) replace $\frac{2}{n}\sum_i x_i x_i$ with $\frac{1}{n^2}\sum_{i,j} d_{ij}^2$, get

$$2x_i x_j = \frac{1}{n}\sum_k (d_{ik}^2 + d_{kj}^2) - d_{ij}^2 - \frac{1}{n^2}\sum_{h,k} d_{hk}^2 \quad (\S)$$

which expresses $x_i x_j$ in function of D

Classic MDS: Proof 3/3

- *Gram* $\subseteq$ *PSD*
  - x **is an** $n \times K$ **real matrix**
  - $G = xx^\top$ **its Gram matrix**
  - **For each** $y \in \mathbb{R}^n$ **we have**

$$yGy^\top = y(xx^\top)y^\top = (yx)(x^\top y^\top) = (yx)(yx)^\top = \|yx\|_2^2 \geq 0$$

  - $\Rightarrow G \succeq 0$
- *PSD* $\subseteq$ *Gram*
  - **Let** $G \succeq 0$ **be** $n \times n$
  - ***Spectral decomposition:*** $G = P\Lambda P^\top$

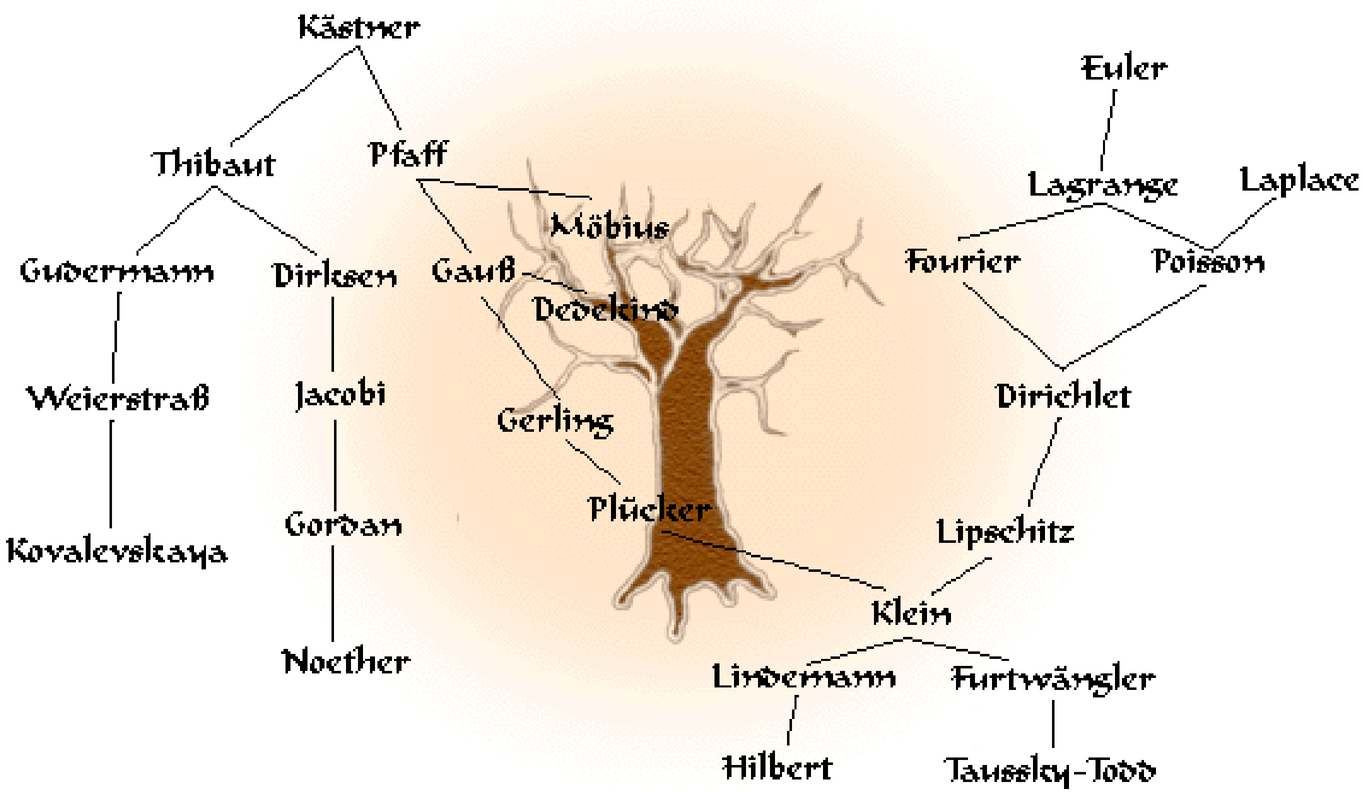
    (P orthogonal, $\Lambda \geq 0$ diagonal)
  - $\Lambda \geq 0 \Rightarrow \sqrt{\Lambda}$ **exists**
  - $G = P\Lambda P^\top = (P\sqrt{\Lambda})(\sqrt{\Lambda}^\top P^\top) = (P\sqrt{\Lambda})(P\sqrt{\Lambda})^\top$
  - **Let** $x = P\sqrt{\Lambda}$**, then** G **is the Gram matrix of** x

Principal Component Analysis (PCA)

- **You want to draw $x = P\sqrt{\Lambda}$ in 2D or 3D**
  but $\mathsf{rank}(\Lambda) = K > 3$
- **Only keep 2 or 3 largest components of Λ**
  zero the rest
- **Get realization in desired space**

Example 1/3

Mathematical genealogy skeleton



Example 2/3

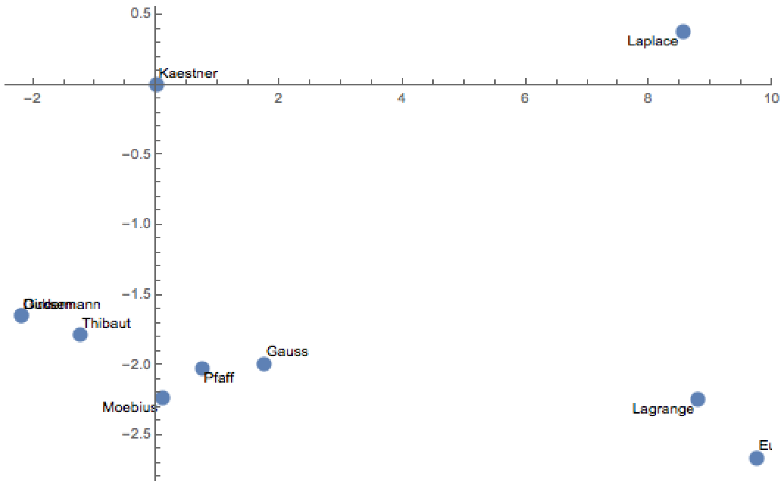
A partial view

	Euler	Thibaut	Pfaff	Lagrange	Laplace	Möbius	Gudermann	Dirksen	Gauss
Kästner	10	1	1	9	8	2	2	2	2
Euler		11	9	1	3	10	12	12	8
Thibaut			2	10	10	3	1	1	3
Pfaff				8	8	1	3	3	1
Lagrange					2	9	11	11	7
Laplace						9	11	11	7
Möbius							4	4	2
Gudermann								2	4
Dirksen									4

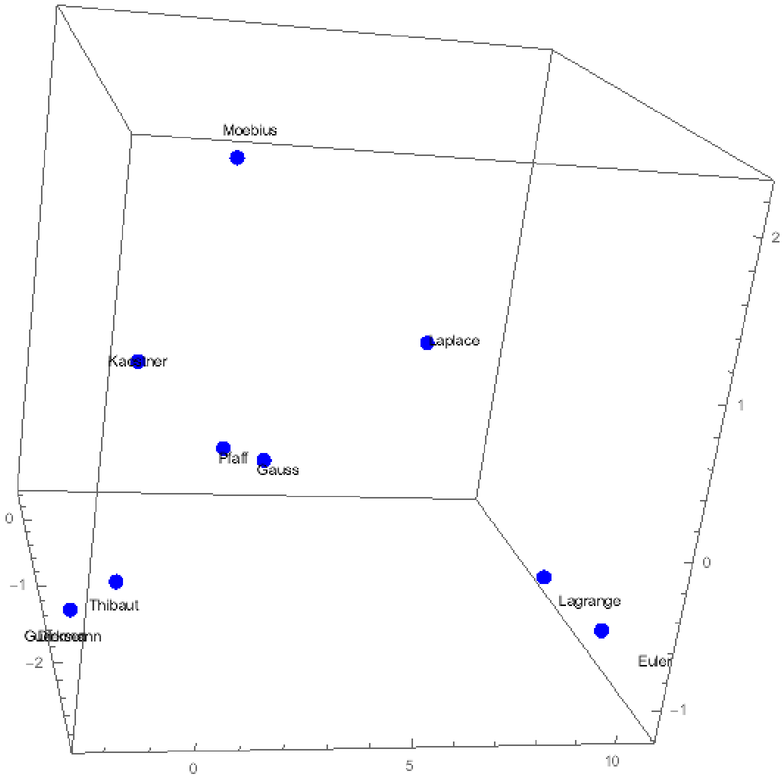
$$D = \begin{pmatrix} 0 & 10 & 1 & 1 & 9 & 8 & 2 & 2 & 2 & 2 \\ 10 & 0 & 11 & 9 & 1 & 3 & 10 & 12 & 12 & 8 \\ 1 & 11 & 0 & 2 & 10 & 10 & 3 & 1 & 1 & 3 \\ 1 & 9 & 2 & 0 & 8 & 8 & 1 & 3 & 3 & 1 \\ 9 & 1 & 10 & 8 & 0 & 2 & 9 & 11 & 11 & 7 \\ 8 & 3 & 10 & 8 & 2 & 0 & 9 & 11 & 11 & 7 \\ 2 & 10 & 3 & 1 & 9 & 9 & 0 & 4 & 4 & 2 \\ 2 & 12 & 1 & 3 & 11 & 11 & 4 & 0 & 2 & 4 \\ 2 & 12 & 1 & 3 & 11 & 11 & 4 & 2 & 0 & 4 \\ 2 & 8 & 3 & 1 & 7 & 7 & 2 & 4 & 4 & 0 \end{pmatrix}$$

Example 3/3

In 2D



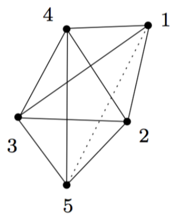
In 3D



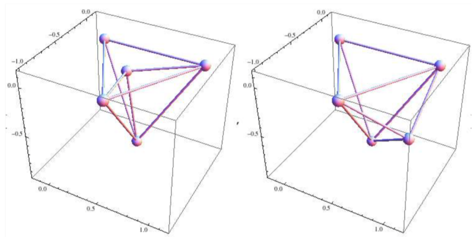
The Distance Geometry Problem (DGP)

Given $K \in \mathbb{N}$ ***and*** $G = (V, E, d)$ ***with*** $d : E \to \mathbb{R}_+$***,***
find $x : V \to \mathbb{R}^K$ ***s.t.***

$$\forall \{i, j\} \in E \quad \|x_i - x_j\|_2^2 = d_{ij}^2$$

Given a weighted graph 

, draw it so edges are drawn as

segments with lengths = weights 

Some applications

- **clock synchronization ($K = 1$)**
- **sensor network localization ($K = 2$)**
- **molecular structure from distance data ($K = 3$)**
- **autonomous underwater vehicles ($K = 3$)**
- **distance matrix completion (whatever K)**

Clock synchronization

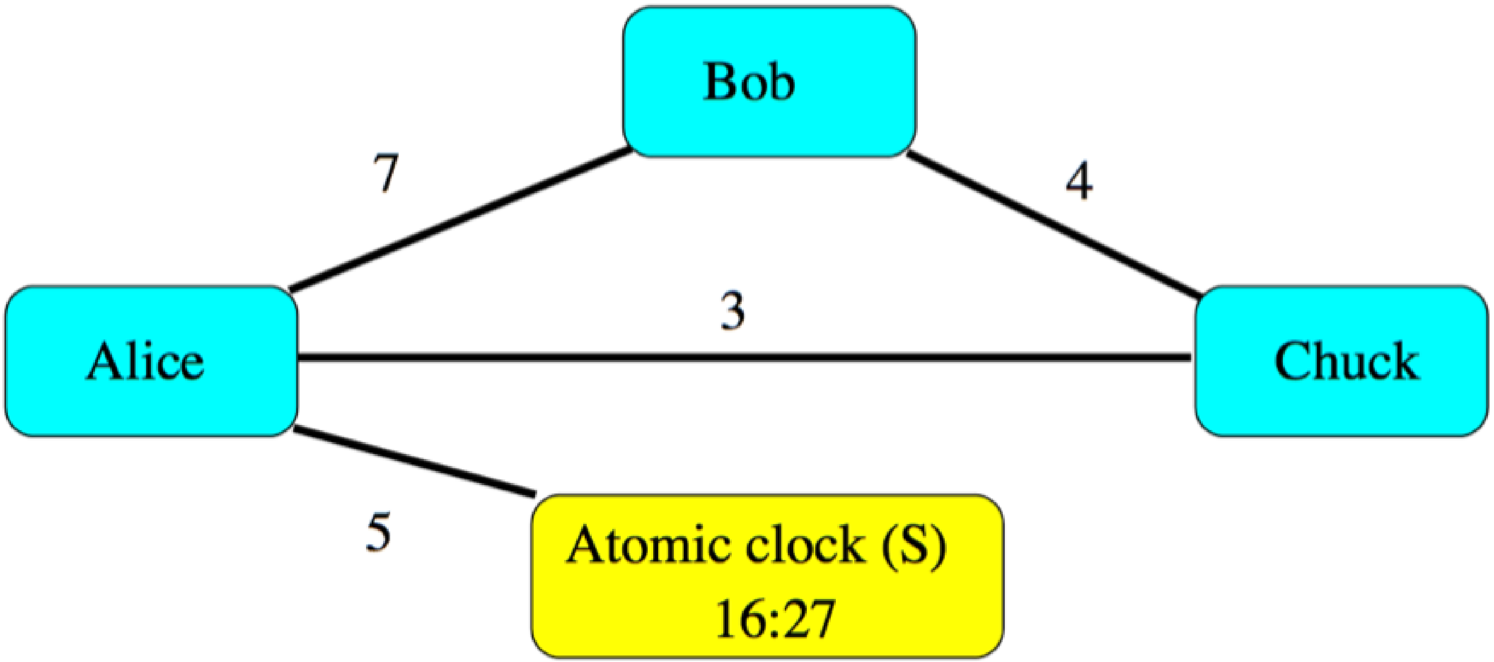
From [Singer, *Appl. Comput. Harmon. Anal.* 2011]

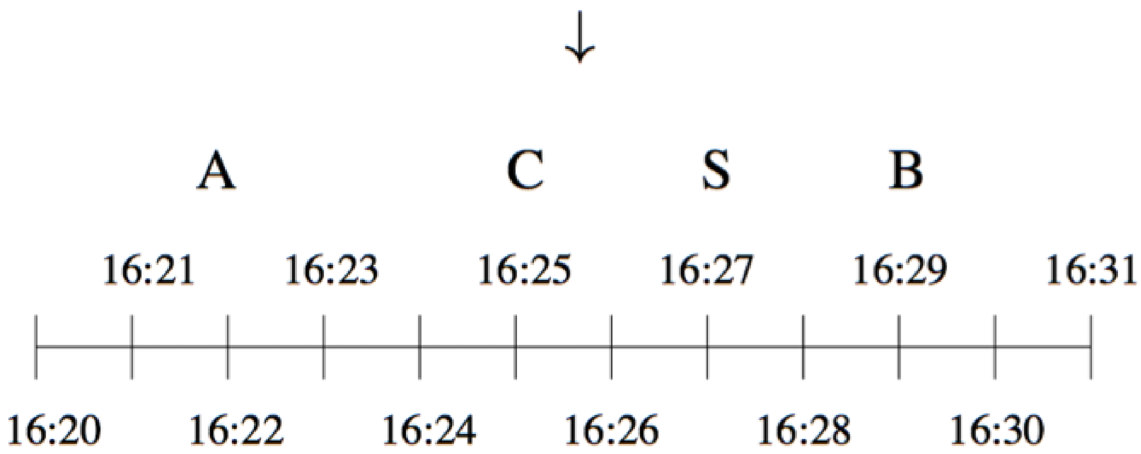
Determine a set of unknown timestamps from a partial measurements of their time differences

- $K = 1$
- V: timestamps
- $\{u, v\} \in E$ if known time difference between u, v
- d: values of the time differences

Used in time synchronization of distributed networks

Clock synchronization





Sensor network localization

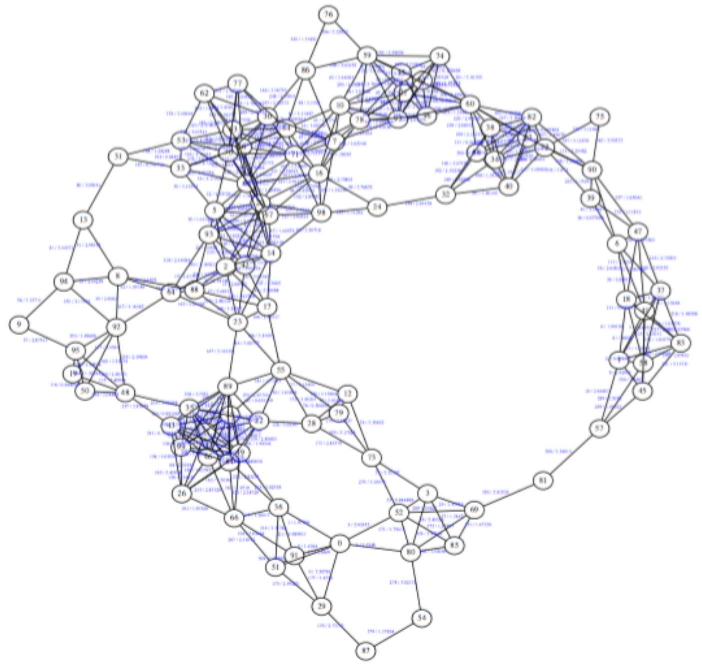
From [Yemini, *Proc. CDSN*, 1978]

The positioning problem arises when it is necessary to locate a set of geographically distributed objects using measurements of the distances between some object pairs

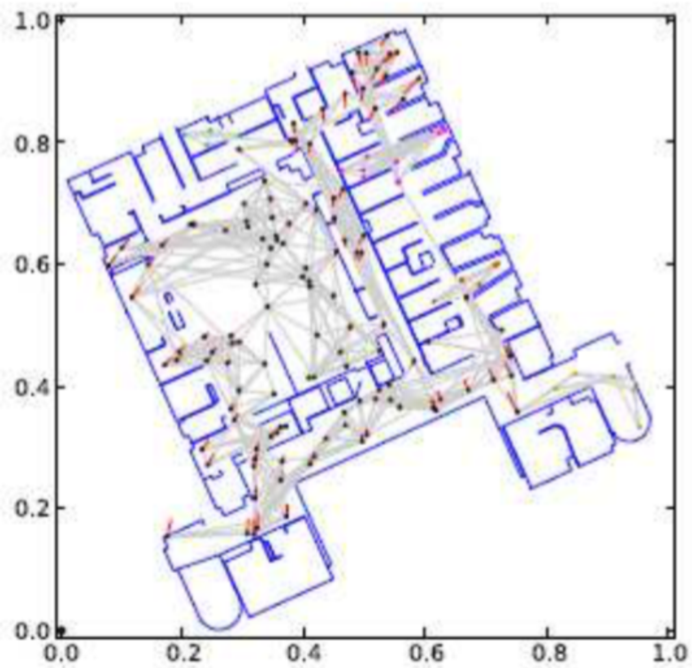
- $K = 2$
- V: (mobile) sensors
- $\{u, v\} \in E$ iff distance between u, v is measured
- d: distance values

Used whenever GPS not viable (e.g. underwater)
$d_{uv} \propto$ battery consumption in P2P communication betw. u, v

Sensor network localization

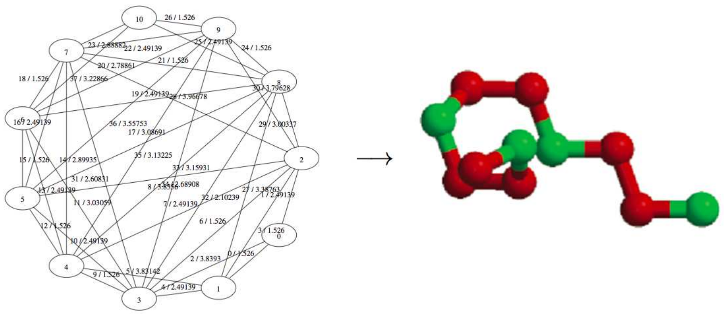


→



Molecular structure from distance data

From [Liberti et al., *SIAM Rev.*, 2014]



- $K = 3$
- V: atoms
- $\{u, v\} \in E$ iff distance between u, v is known
- d: distance values

Used whenever X-ray crystallography does not apply (e.g. liquid)
Covalent bond lengths and angles known precisely
Distances $\lessapprox 5.5$ measured approximately by NMR

Complexity

- **DGP$_1$ with $d : E \to \mathbb{Q}_+$ is in NP**
  - if instance YES $\exists$ realization $x \in \mathbb{R}^{n\times 1}$
  - if some component $x_i \notin \mathbb{Q}$ translate x so $x_i \in \mathbb{Q}$
  - consider some other x_j
  - let $\ell =$ (length sh. path$p : i \to j) = \sum\limits_{\{u,v\}\in p} d_{uv} \in \mathbb{Q}$
  - then $x_j = x_i \pm \ell \to x_j \in \mathbb{Q}$
  - $\Rightarrow$ verification of

$$\forall \{i,j\} \in E \quad |x_i - x_j| = d_{ij}$$

    in polytime
- **DGP$_K$ may not be in NP for $K > 1$**

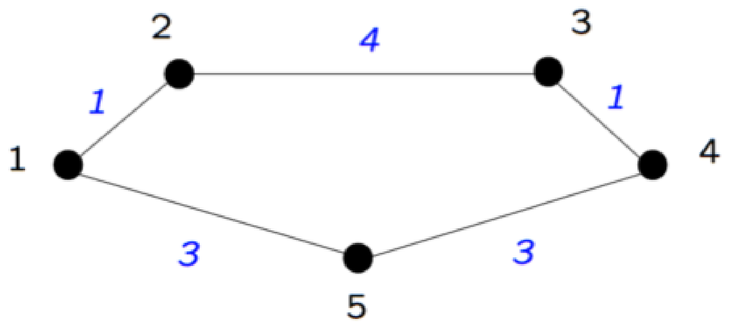
  don't know how to verify $\|x_i - x_j\|_2 = d_{ij}$ for $x \notin \mathbb{Q}^{nK}$

Hardness

Partition *is* NP*-hard*

Given $a = (a_1, \ldots, a_n) \in \mathbb{N}^n$, $\exists\, I \subseteq \{1, \ldots, n\}$ **s.t.** $\sum_{i \in I} a_i = \sum_{i \notin I} a_i$ **?**

- Reduce Partition to DGP_1
- $a \longrightarrow$ **cycle** C
  $V(C) = \{1, \ldots, n\}$, $E(C) = \{\{1, 2\}, \ldots, \{n, 1\}\}$
- **For** $i < n$ **let** $d_{i,i+1} = a_i$
  $d_{n,n+1} = d_{n1} = a_n$
- *E.g. for* $a = (1, 4, 1, 3, 3)$*, get cycle graph:*



[Saxe, 1979]

PARTITION is YES ⇒ DGP$_1$ is YES

- **Given:** $I \subset \{1,\ldots,n\}$ **s.t.** $\sum_{i\in I} a_i = \sum_{i\notin I} a_i$
- **Construct:** realization x of C in $\mathbb{R}$
  1. $x_1 = 0$ `// start`
  2. **induction step:** suppose x_i known
     if $i \in I$
     let $x_{i+1} = x_i + d_{i,i+1}$ `// go right`
     else
     let $x_{i+1} = x_i - d_{i,i+1}$ `// go left`
- **Correctness proof:** by the same induction
  but careful when $i = n$: have to show $x_{n+1} = x_1$

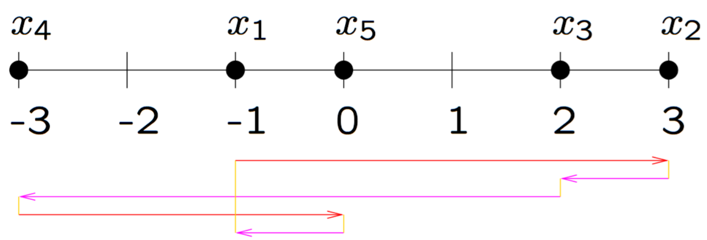
PARTITION is YES $\Rightarrow$ DGP$_1$ is YES

$$\begin{aligned}
(1) = \sum_{i\in I}(x_{i+1} - x_i) &= \sum_{i\in I} d_{i,i+1} = \\
= \sum_{i\in I} a_i &= \sum_{i\notin I} a_i = \\
= \sum_{i\notin I} d_{i,i+1} &= \sum_{i\notin I}(x_i - x_{i+1}) = (2)
\end{aligned}$$

$$\begin{aligned}
(1) = (2) \Rightarrow \sum_{i\in I}(x_{i+1} - x_i) = \sum_{i\notin I}(x_i - x_{i+1}) \Rightarrow \sum_{i\leq n}(x_{i+1} - x_i) &= 0 \\
\Rightarrow (x_{n+1} - x_n) + (x_n - x_{n-1}) + \cdots + (x_3 - x_2) + (x_2 - x_1) &= 0 \\
\Rightarrow x_{n+1} &= x_1
\end{aligned}$$

PARTITION is NO ⇒ DGP_1 is NO

- By contradiction: suppose DGP_1 is YES, x realization of C
- $F = \{\{u,v\} \in E(C) \mid x_u \le x_v\}$,
  $E(C) \smallsetminus F = \{\{u,v\} \in E(C) \mid x_u > x_v\}$
- Trace $x_1, \ldots, x_n$: follow edges in F ($\rightarrow$) and in $E(C) \smallsetminus F$ ($\leftarrow$)



$$
\begin{aligned}
\sum_{\{u,v\}\in F} (x_v - x_u) &= \sum_{\{u,v\}\notin F} (x_u - x_v) \\
\sum_{\{u,v\}\in F} |x_u - x_v| &= \sum_{\{u,v\}\notin F} |x_u - x_v| \\
\sum_{\{u,v\}\in F} d_{uv} &= \sum_{\{u,v\}\notin F} d_{uv}
\end{aligned}
$$

- Let $J = \{i < n \mid \{i, i+1\} \in F\} \cup \{n \mid \{n, 1\} \in F\}$

$$\Rightarrow \quad \sum_{i \in J} a_i = \sum_{i \notin J} a_i$$

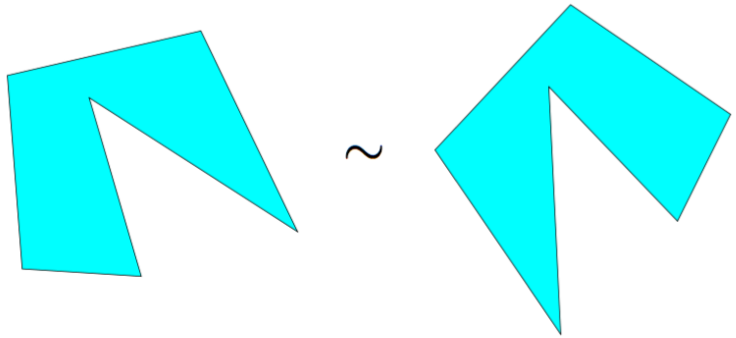
- So J solves Partition instance, contradiction
- ⇒ DGP is NP-hard, DGP_1 is NP-complete

Number of solutions: with congruences

- (G, K): **DGP instance**
- $\tilde{X} \subseteq \mathbb{R}^{Kn}$: **set of solutions**
- ***Congruence***: **composition of translations, rotations, reflections**
- $C =$ **set of congruences in** $\mathbb{R}^K$
- $x \sim y$ **means** $\exists \rho \in C \; (y = \rho x)$**:**
  distances in x **are preserved in** y **through** ρ
- $\Rightarrow$ **if** $|\tilde{X}| > 0$**,** $|\tilde{X}| = 2^{\aleph_0}$

Number of solutions: without congruences

- Congruence is an *equivalence relation* $\sim$ on $\tilde{X}$ (reflexive, symmetric, transitive)



- Partitions $\tilde{X}$ into *equivalence classes*
- $X = \tilde{X}/\sim$: sets of representatives of equivalence classes
- **Focus on $|X|$ rather than $|\tilde{X}|$**

Rigidity, flexibility and $|X|$

- **infeasible** $\Leftrightarrow |X| = 0$
- **rigid graph** $\Leftrightarrow |X| < \aleph_0$
- **globally rigid graph** $\Leftrightarrow |X| = 1$
- **flexible graph** $\Leftrightarrow |X| = 2^{\aleph_0}$
- $|X| = \aleph_0$**: impossible by Milnor's theorem**

Milnor's theorem implies $|X| \neq \aleph_0$

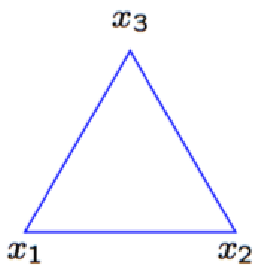
- System S of polynomial equations of degree 2

$$\forall i \leq m \quad p_i(x_1, \ldots, x_{nK}) = 0$$

- Let X be the set of $x \in \mathbb{R}^{nK}$ satisfying S
- **Number of connected components of X is $O(3^{nK})$**
  [Milnor 1964]
- If $|X|$ is countably ∞ then G cannot be flexible
  $\Rightarrow$ *incongruent elts of X are separate connected components*
  $\Rightarrow$ by Milnor's theorem, there's finitely many of them

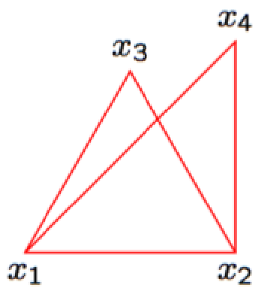
Examples

$V^1 = \{1, 2, 3\}$
$E^1 = \{\{u, v\} \mid u < v\}$
$d^1 = 1$



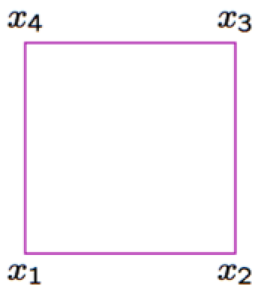
ρ congruence in $\mathbb{R}^2$
$\Rightarrow \rho x$ valid realization
$|X| = 1$

$V^2 = V^1 \cup \{4\}$
$E^2 = E^1 \cup \{\{1, 4\}, \{2, 4\}\}$
$d^2 = 1 \wedge d_{14} = \sqrt{2}$



ρ reflects x_4 wrt $\overline{x_1, x_2}$
$\Rightarrow \rho x$ valid realization
$|X| = 2$ ()

$V^3 = V^2$
$E^3 = \{\{u, u+1\} | u \leq 3\} \cup \{1, 4\}$
$d^1 = 1$



ρ rotates $\overline{x_2 x_3}$, $\overline{x_1 x_4}$ by θ
$\Rightarrow \rho x$ valid realization
$|X|$ is uncountable
(, ...)