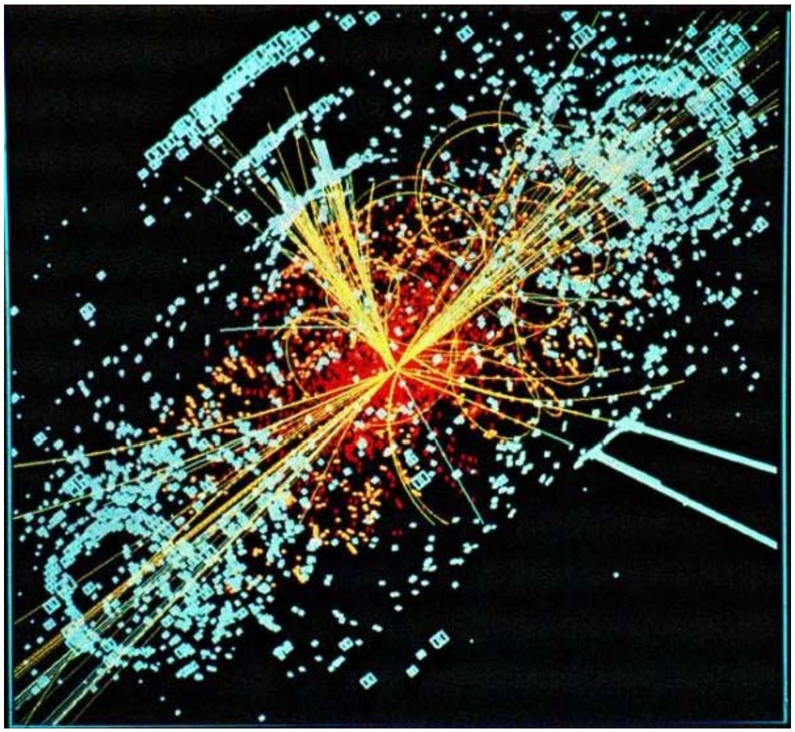
CS 190C
Science Education in Computational Thinking

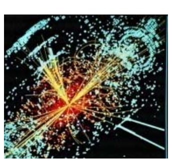


SCIENTIFIC COMPUTING:

- Applications
- High performance computing
- Integration of functions
- Sampling and statistics
- Problem & Algorithm Stability
- Geometric Algebra

Hoffmann, 2008

Sampling and Statistics

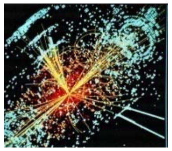


Measurements

- Take a sequence of physical measurements

  $s_1, s_2, \ldots s_n,$

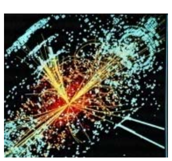
  (temperature, humidity, weight, height, …)
- There is a normal distribution of values around the "true" value
- Compute the "true" value by averaging, but also determine the confidence – the accuracy!



Statistics

- Mean: $\bar{x} = \frac{1}{n}\sum_{k=1}^{n} x_k$
- Variance: $s^2 = \frac{1}{n-1}\sum_{k=1}^{n}(x_k - \bar{x})^2$
- 95% of normally distributed samples lie between -1.96 and $+1.96$, so the 95% confidence interval is

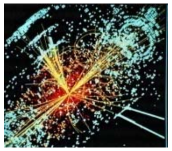
$$\bar{x} \pm 1.96\frac{s}{\sqrt{n}}$$



Example

- Gallup poll of March 13, 2007:
  - Obama leads Clinton 50% to 44%
  - Sampling error ±3%, sampling size 1199 democrats
- Obama 600, Clinton 528, undecided 71
  - Formula for multiple correlation

$$\sqrt{\frac{p(1-p)+q(1-q)+2pq}{n}}$$

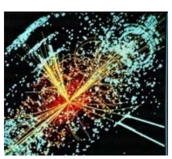


One or Two Passes?

- Computation requires two passes through the data:
  1. Determine the mean
  2. Determine the variance
- Single pass possible from $s^2 = \dfrac{n\sum_k x_k{}^2 - (\sum_k x_k)^2}{n(n-1)}$

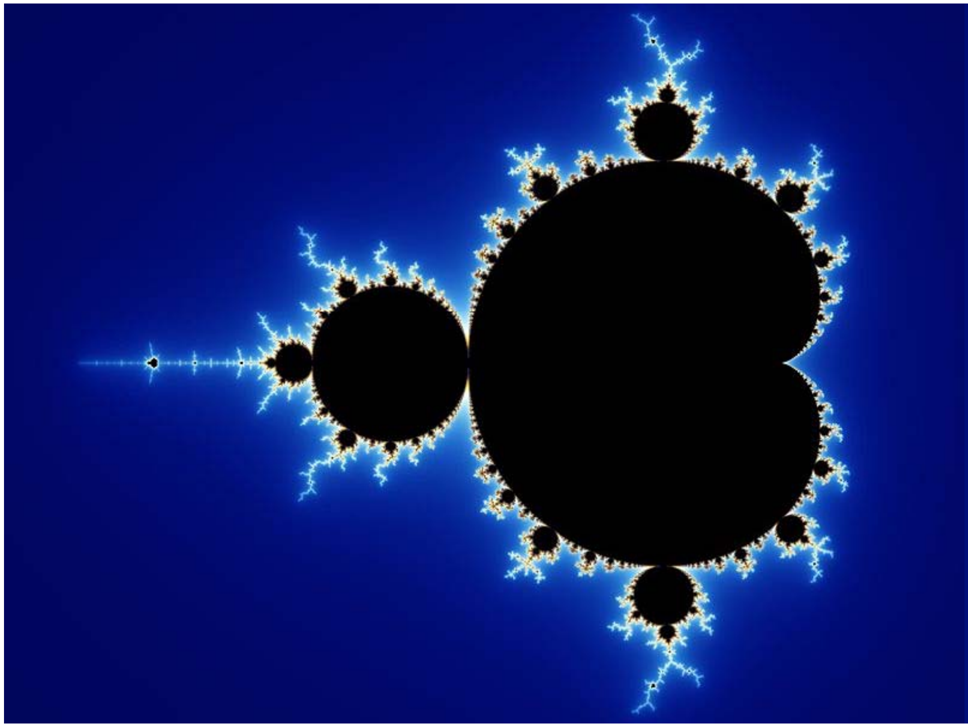
- …but this is unstable numerically

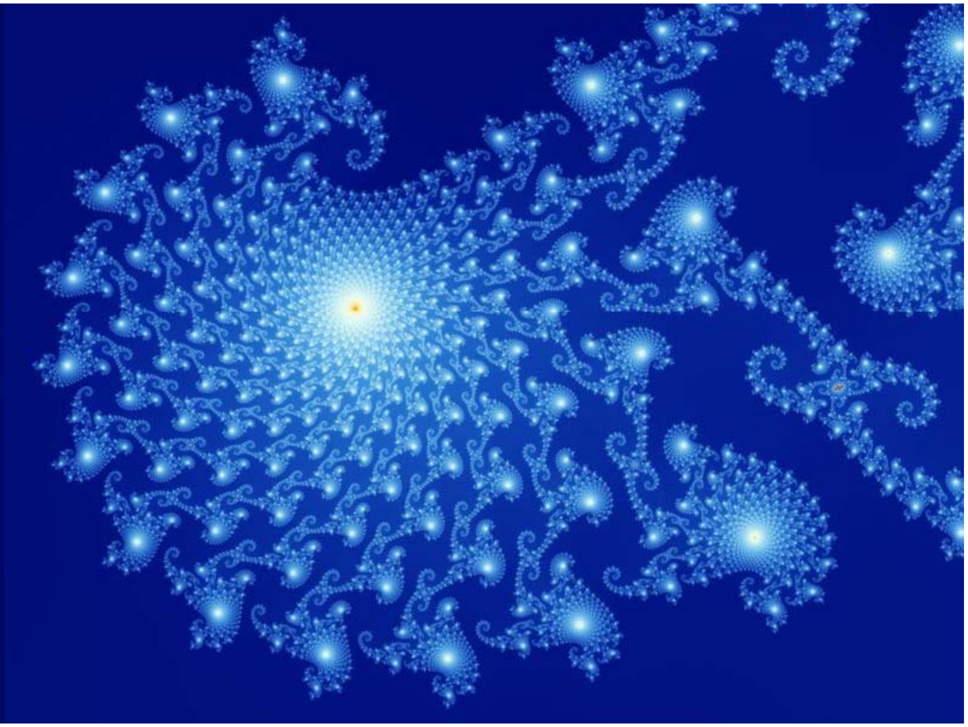
pgm$_1$



Stability

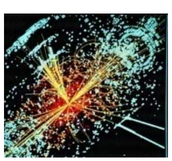
- Is it the algorithm?
- Is it the problem?
  - Iterate $f(z) = z^2+C$ for complex numbers
  - If $|f^k(0)|>2$, then C is not in the Mandelbrot set





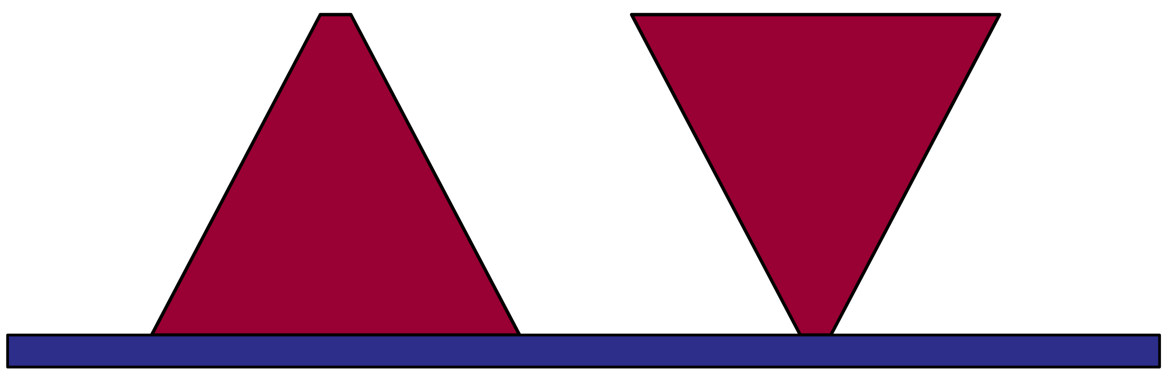
Problem Condition/Stability

- Stability of the problem – a small change of the input leads to a small change in the output
  - a stable problem is *well-conditioned*
  - an unstable problem is *ill-conditioned*
- Stable algorithm?
  - an algorithm that behaves stable on well-conditioned problems
- Likewise, unstable algorithm…



Problem Condition/Stability

- Stability of the problem – a small change of the input leads to a small change in the output
  - a stable problem is *well-conditioned*
  - an unstable problem is *ill-conditioned*
- Example

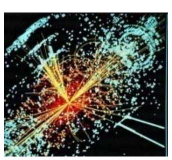


Problem Stability for Linear Algebra

Line equation $ax+by+d = 0$

To evaluate, substitute a point $p = (x, y)$
if the result is (almost) 0, then p is on the line

Because of floating point, we choose a tolerance ε
if abs(result) $< \varepsilon$, then p is on the line



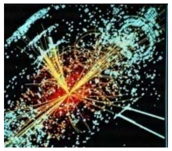
Clicker Question:

Which is more precise to evaluate for $x = 1e^{-150}$

A. $y^2 - x^3 = 0$

B. $1e^{150} y^2 - 1e^{150} x^3 = 0$

C. $1e^{-150} y^2 - 1e^{-150} x^3 = 0$



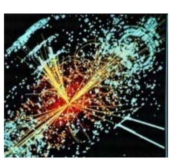
Considerations

- Max exponent for double precision is about $2^{\pm 1024} \approx 10^{\pm 310}$
- So evaluating x^3 at 10^{-150} causes exponent underflow except for C:

A. $y^2 - x^3 = 0$

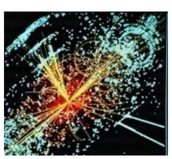
B. $1\mathrm{e}^{150} y^2 - 1\mathrm{e}^{150} x^3 = 0$

C. $1\mathrm{e}^{-150} y^2 - 1\mathrm{e}^{-150} x^3 = 0$

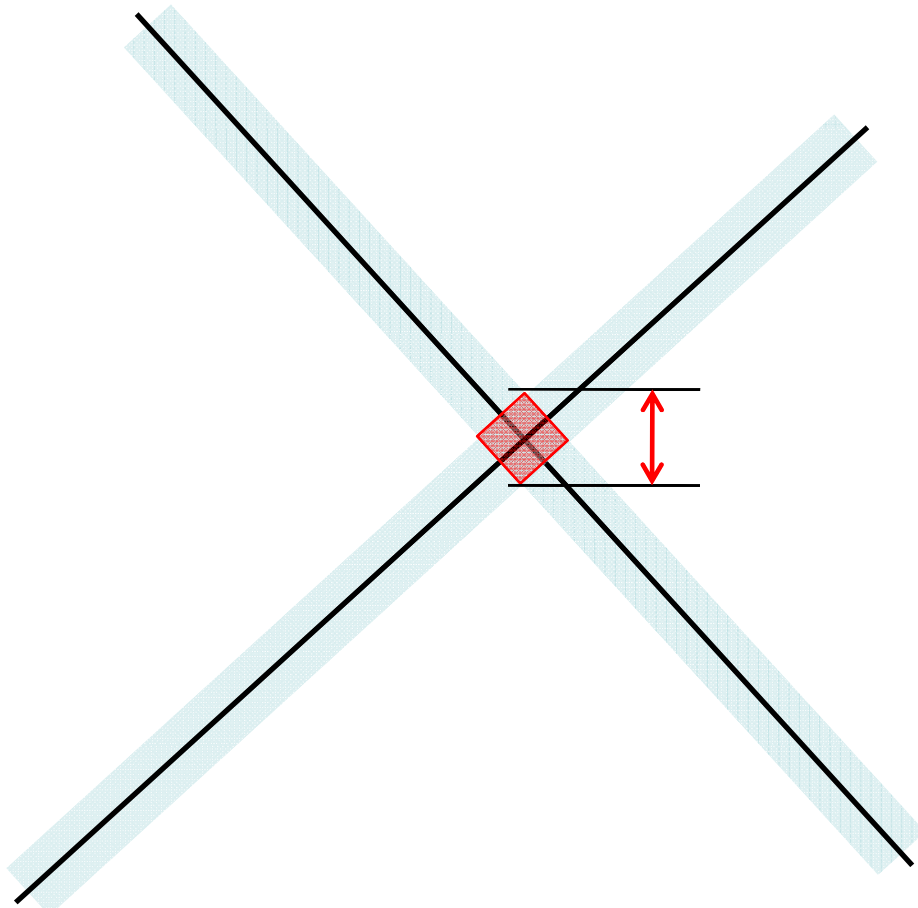


Lines

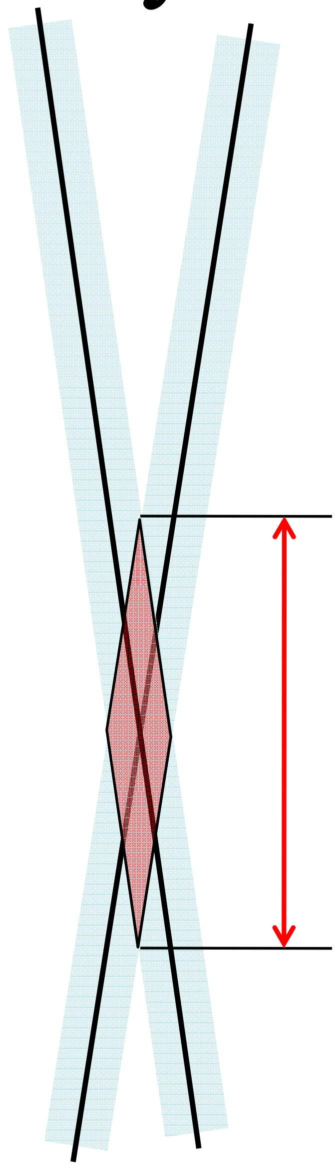
- Implicit equation $ax^2 + by^2 + c = 0$ normalized such that $a^2+b^2 = 1$
- Stability of intersection a geometric issue



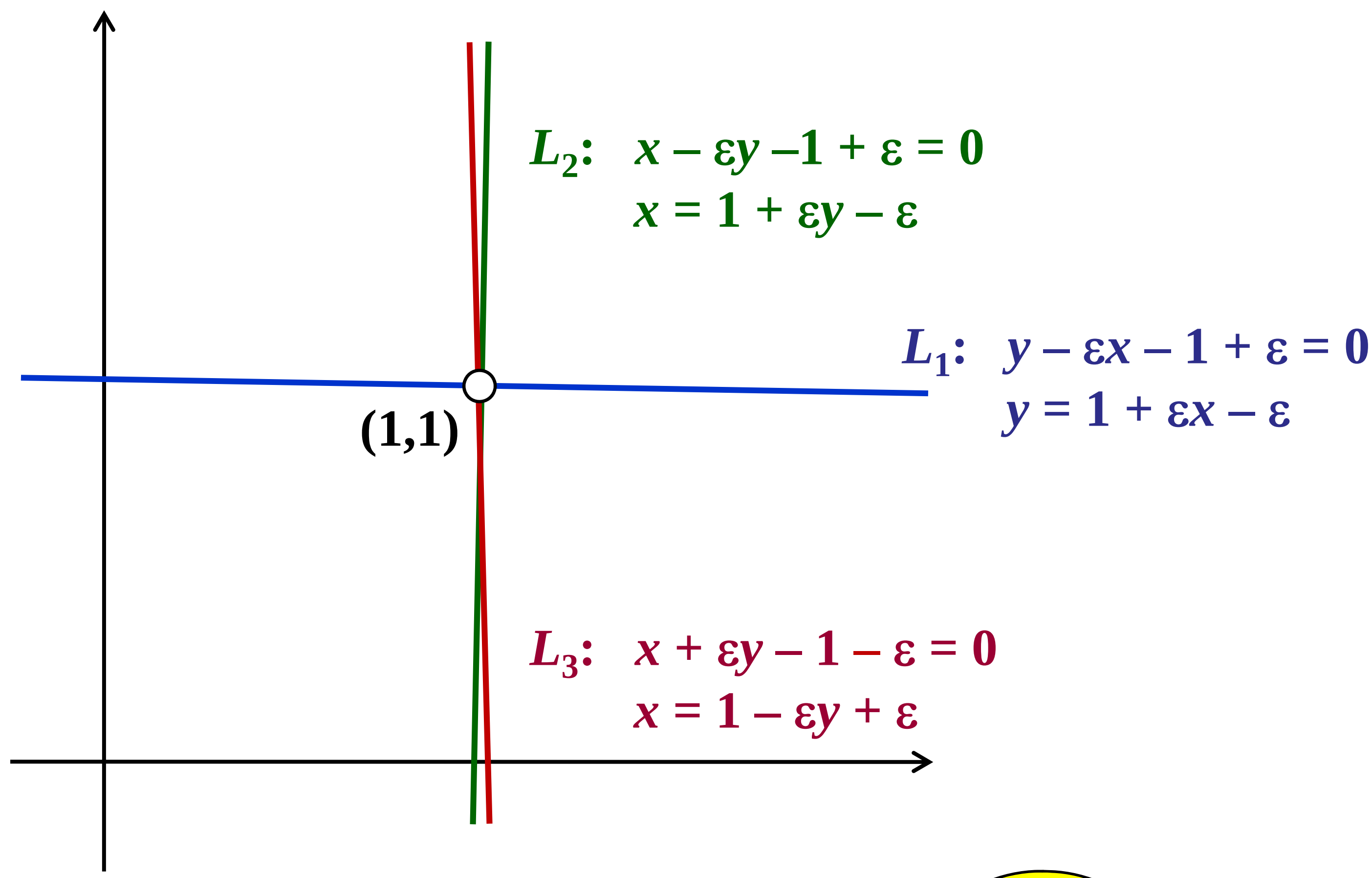
Very Stable



Not Very Stable

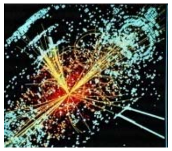


Example

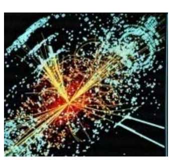


pgm$_3$

$\varepsilon = 1.0\,\mathrm{e}{-8}$



Geometric (Linear) Algebra



Lines and Planes

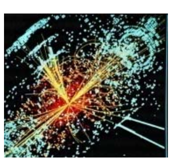
- $ax + by + d = 0$
- $ax + by + d = 0$
  $a'x + b'y + d' = 0$

$$\begin{pmatrix} a & b \\ a' & b' \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = -\begin{pmatrix} d \\ d' \end{pmatrix}$$

- Matrix notation:

$$Au = -d$$

$$u = -A^{-1} d$$



Lines and Planes

- $ax + by + cz + d = 0$

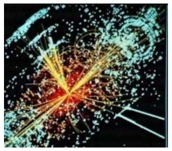
- $ax + by + cz + d = 0$
  $a'x + b'y + c'z + d' = 0$
  $a''x + b''y + c''z + d'' = 0$

$$\begin{pmatrix} a & b & c \\ a' & b' & c' \\ a'' & b'' & c'' \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = - \begin{pmatrix} d \\ d' \\ d'' \end{pmatrix}$$

- Matrix notation

  $Au + d = 0$

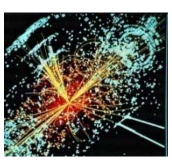
  $u = -A^{-1} d$



Computational Linear Algebra

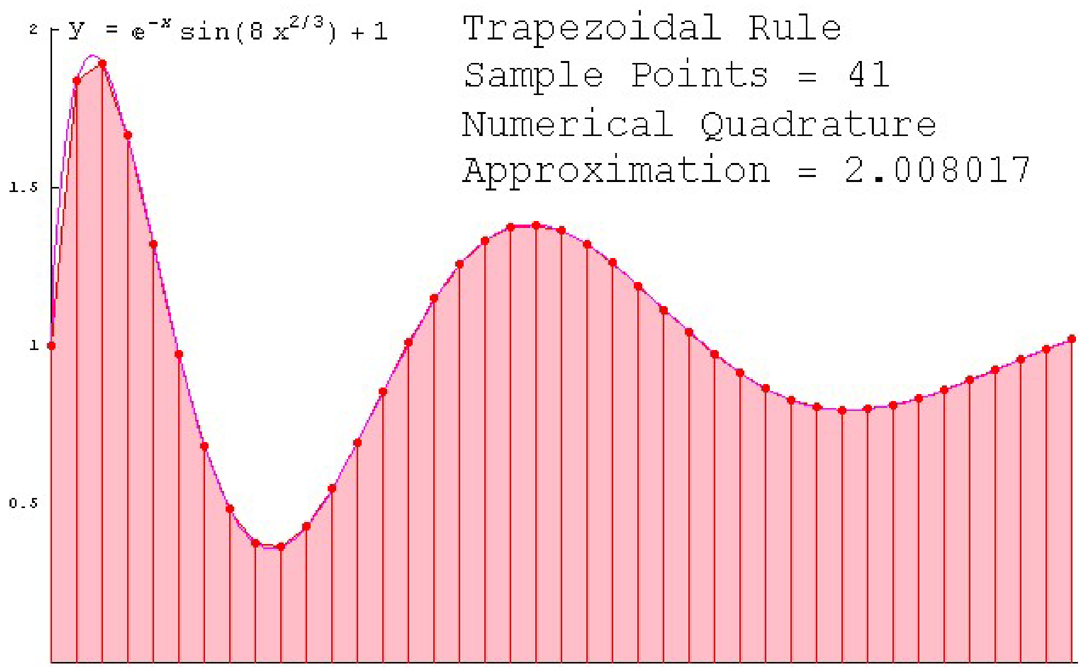
- Solves large systems of linear equations
- Develops algorithms that are ideally both stable and efficient
- Achieves high efficiency for systems with specific structure:
  - sparse systems
  - banded systems
  - block structured systems
  - symmetric matrices

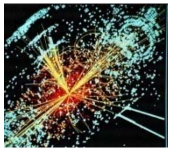
  …



Why?

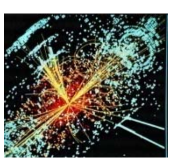
- Analysis of real structures and artifacts is often done by systems of differential equations
- Those systems can be solved by various strategies of "linearization"





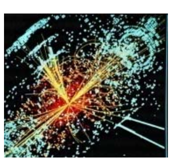
Lines and Planes

- Solving linear equations is fundamental and involves inverting a matrix
- Simple methods:
  - Determinants (dim 3 or less)
  - Gauss elimination
  - Various decomposition methods



2D and Determinant Theory

- Two distinct lines intersect in one point
- Two distinct points are connected by one line
- *Exceptions:*
  - Lines are parallel: “point is at infinity”
- A method to compute those with a redundant coordinate system $(x, y, w) \leftrightarrow (x/w, y/w)$
  - If $w = 0$ it means “at infinity”



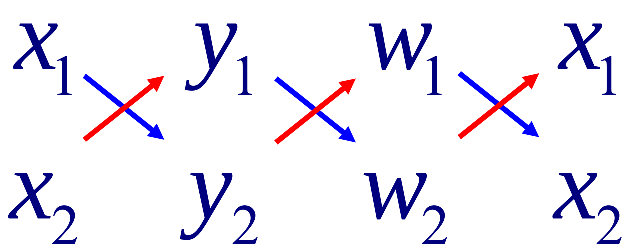
Uniformity

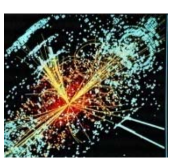
- One line passes through two points. For example:

$$P_1: \quad ax_1 + by_1 + cw_1 = 0$$

$$P_2: \quad ax_2 + by_2 + cw_2 = 0$$

$$(y_1 w_2 - y_2 w_1, w_1 x_2 - w_2 x_1, x_1 y_2 - x_2 y_1)$$



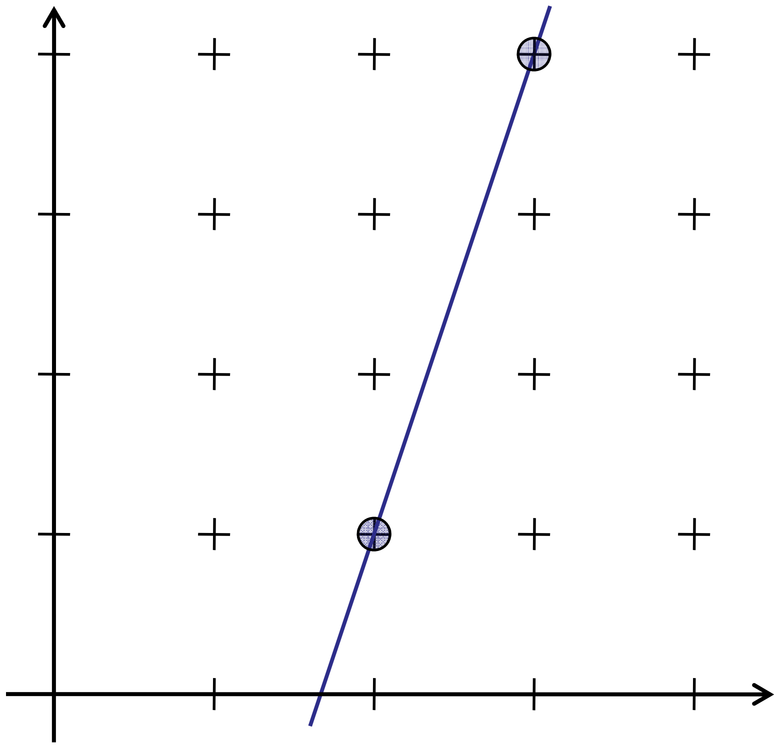


Example

- Line connecting (3,4) and (2,1):

$$\begin{vmatrix} 3 & 4 & 1 \\ 2 & 1 & 1 \end{vmatrix}$$

- So (4–1, 2–3, 3–8), i.e., $3x - y - 5 = 0$



Uniformity

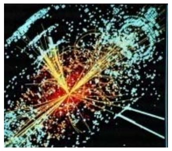
- Two lines intersect in one point. Examples:

$$L_1: \quad a_1 x + b_1 y + c_1 w = 0$$

$$L_2: \quad a_2 x + b_2 y + c_2 w = 0$$

$$(b_1 c_2 - b_2 c_1, c_1 a_2 - c_2 a_1, a_1 b_2 - a_2 b_1)$$

$$\begin{array}{cccc} a_1 & b_1 & c_1 & a_1 \\ a_2 & b_2 & c_2 & a_2 \end{array}$$

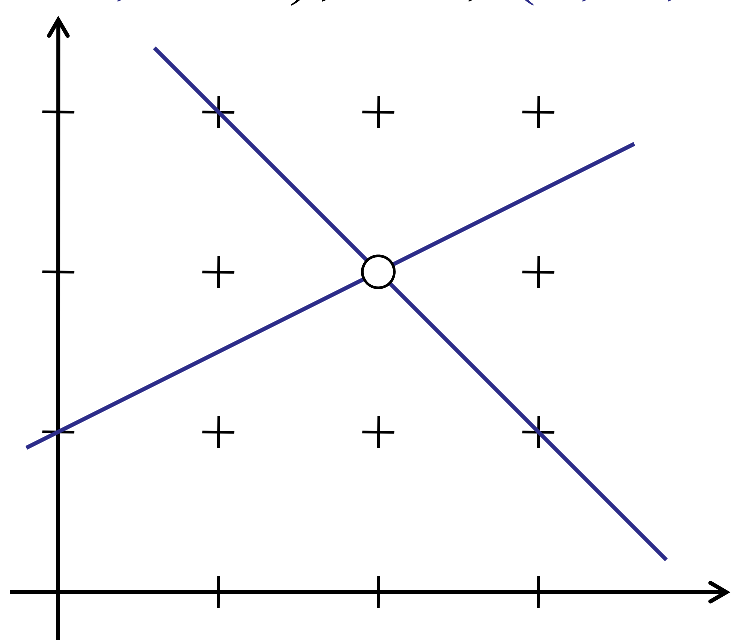


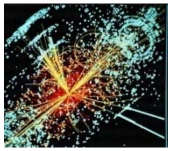
Example

- Intersection of lines
  $L_1: x - 2y + 2 = 0$
  $L_2: x + y - 4 = 0$

$$\begin{vmatrix} 1 & -2 & 2 \\ 1 & 1 & -4 \end{vmatrix}$$

- So (8–2, 2+4, 1+2), i.e., (6, 6, 3) = (2,2)



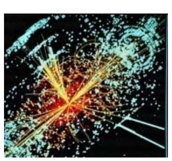


From Lines to Curves

- Consider line L: $ax+by+c=0$ as the linear polynomial $L=(ax+by+c)$
- The polynomial vanishes on the line L, or, if a point is on the line, the polynomial must evaluate to zero
- Take four lines and "blend" them for some s:

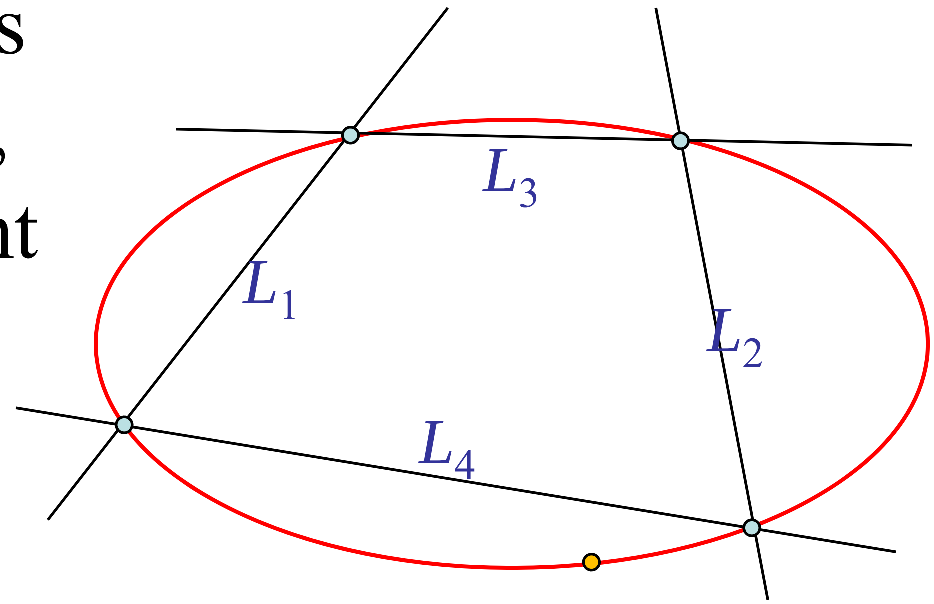
  $$C:\ (1-s)\,L_1L_2 + s\,L_3\,L_4 = 0$$

- Then C defines a curve of degree 2, a conic.

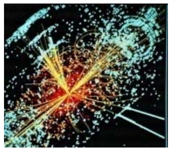


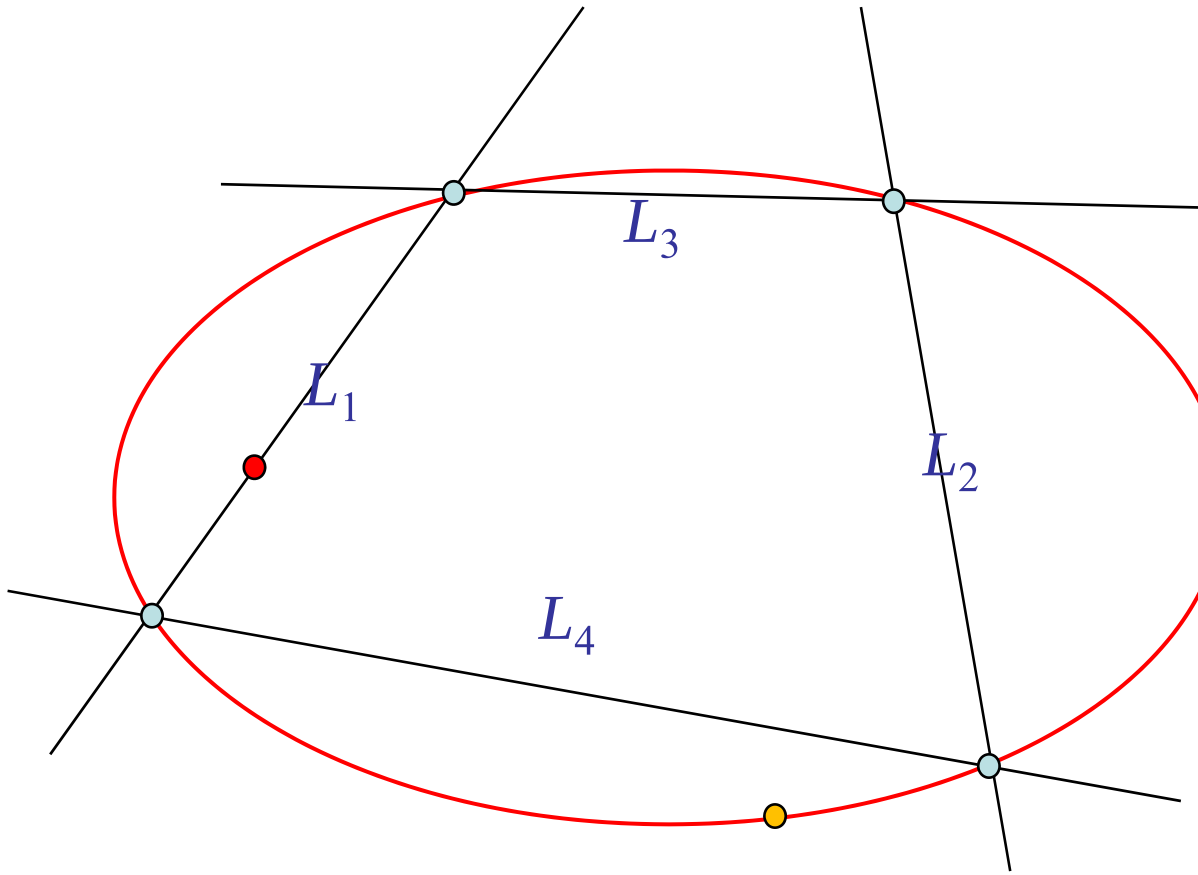
Liming's Conics

- Implicit equation is
  $$(1 - s)\, L_1 L_2 + s\, L_3\, L_4 = 0$$
- Which conic depends on *s*, or equivalently, on an additional point



pgm$_2$





$$(1-s)\,L_1 L_2 + s\,L_3\,L_4 = 0$$

Stable Implementation

- Exploits point/line duality
- Differentiates minimally between points and lines
- Includes structures at infinity

pgm_4