Stability, precision and complexity in some numerical problems

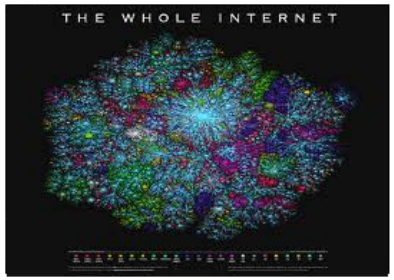
Carlos Beltrán

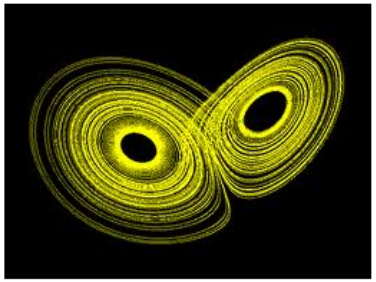
Universidad de Cantabria, Santander, Spain

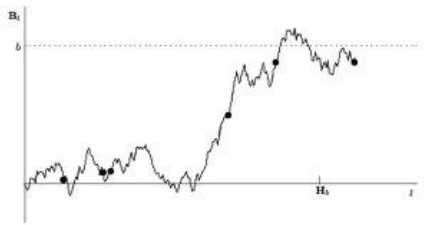
Recent Advances in Real Complexity and Computation

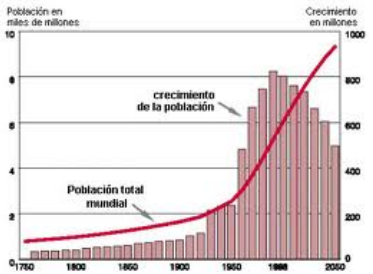
The numerical solution of problems

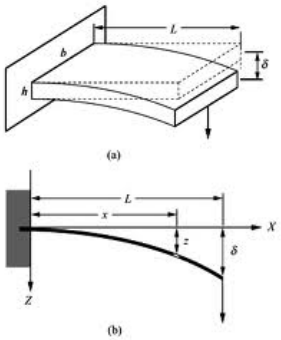
We need to understand

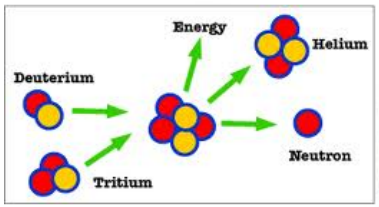


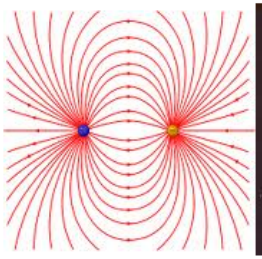






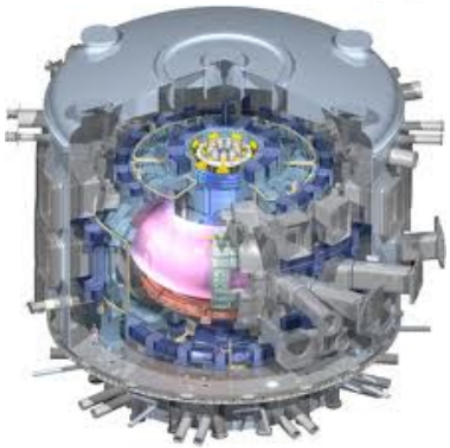


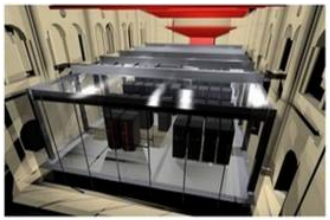




The need for speed

The faster the computer, the more important the speed of algorithms - Ll. Threfeten





The need for precision

The faster the computer, the faster it can screw things up







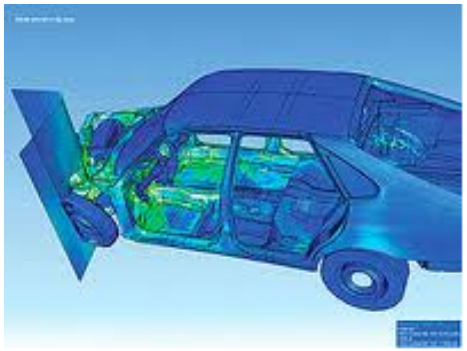

So, people is worried by two things: complexity and precision (stability)

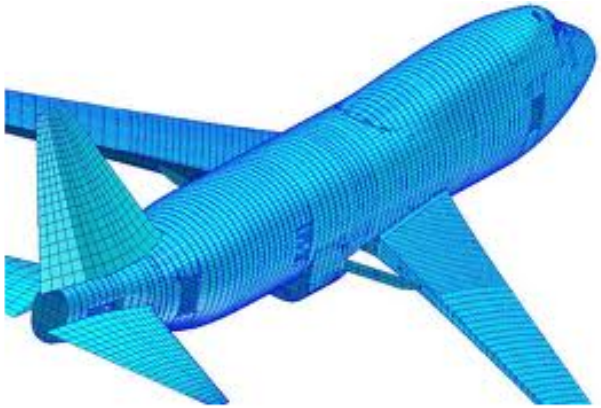
And these are studied in all important (numerical) problems around:

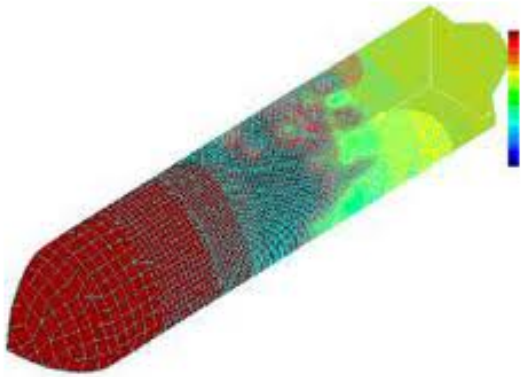
1. Linear algebra routines (solving $Ax = b$, finding kernels, LSQ, matrix decompositions...)
2. Solving systems of Ordinary Differential Equations (ODEs), Differential Algebraic Equations (DAEs) and Partial Differential Equations (PDEs).
3. Solving $f(x) = 0$ where $f : \mathbb{K}^n \to \mathbb{K}^n$.
4. And in general, *any* problem which is to be solved by a computer.

We will shortly deal with the first and the last of these two problems

Virtually every non–trivial numerical computation involves LA as a subroutine







Virtually every non–trivial numerical computation involves LA as a subroutine

$Ax = b$ where A is $n \times n$, $\det(A) \neq 0$

Input: $A \in \mathcal{M}_n(\mathbb{C})$, $det(A) \neq 0$. $b \in \mathbb{C}^n$. **Output:** $x \in \mathbb{C}^n$ such that $Ax = b$. **Stability of the solution:** Relation between x and x' where:

$$Ax = b \qquad A'x' = b, \qquad A \approx A'.$$

Virtually every non–trivial numerical computation involves LA as a subroutine

$Ax = b$ where A is $n \times n$, $\det(A) \neq 0$

Input: $A \in \mathcal{M}_n(\mathbb{C})$, $det(A) \neq 0$. $b \in \mathbb{C}^n$. **Output:** $x \in \mathbb{C}^n$ such that $Ax = b$. **Stability of the solution:** Relation between x and x' where:

$$Ax = b \qquad A'x' = b, \qquad A \approx A'.$$

Condition number: $\kappa(A) := \|A\|_2 \|A^{-1}\|_2$ [Turing].

$$\frac{\|x - x'\|}{\|x'\|} \leq \kappa(A) \frac{\|A - A'\|_2}{\|A\|_2}.$$

Alternatively: $\kappa_D(A) := \|A\|_F \|A^{-1}\|_2$ [Smale, Demmel].

Virtually every non–trivial numerical computation involves LA as a subroutine

$Ax = b$ where A is $n \times n$, $\det(A) \neq 0$

Input: $A \in \mathcal{M}_n(\mathbb{C})$, $det(A) \neq 0$. $b \in \mathbb{C}^n$. **Output:** $x \in \mathbb{C}^n$ such that $Ax = b$. **Stability of the solution:** Relation between x and x' where:

$$Ax = b \qquad A'x' = b, \qquad A \approx A'.$$

Condition number: $\kappa(A) := \|A\|_2 \|A^{-1}\|_2$ [Turing].

$$\frac{\|x - x'\|}{\|x'\|} \leq \kappa(A) \frac{\|A - A'\|_2}{\|A\|_2}.$$

Alternatively: $\kappa_D(A) := \|A\|_F \|A^{-1}\|_2$ [Smale, Demmel].
Computing the condition number is more difficult than computing x.

Virtually every non–trivial numerical computation involves LA as a subroutine

$Ax = b$ where A is $n \times n$, $\det(A) \neq 0$

Input: $A \in \mathcal{M}_n(\mathbb{C})$, $det(A) \neq 0$. $b \in \mathbb{C}^n$. **Output:** $x \in \mathbb{C}^n$ such that $Ax = b$. **Stability of the solution:** Relation between x and x' where:

$$Ax = b \qquad A'x' = b, \qquad A \approx A'.$$

Condition number: $\kappa(A) := \|A\|_2 \|A^{-1}\|_2$ [Turing].

$$\frac{\|x - x'\|}{\|x'\|} \leq \kappa(A) \frac{\|A - A'\|_2}{\|A\|_2}.$$

Alternatively: $\kappa_D(A) := \|A\|_F \|A^{-1}\|_2$ [Smale, Demmel].
Computing the condition number is more difficult than computing x. **What can we hope? Will $\kappa(A)$ be "in general" a small quantity?**

A geometric property of κ_D

[Eckardt, Young, Smicht, Mirski, and even Banach (?)]'s theorem

We note that $\kappa_D(A)$ depends only on the projective class of A.

Let Σ^{n-1} be the set of all singular matrices of $\mathbb{P}(\mathcal{M}_n(\mathbb{C}))$ (which is an algebraic variety.)

Theorem

$$\kappa_D(A) = \frac{1}{d_{\mathbb{P}(\mathcal{M}_n(\mathbb{C}))}(A, \Sigma^{n-1})}.$$

A geometric property of κ_D

[Eckardt, Young, Smicht, Mirski, and even Banach (?)]'s theorem

We note that $\kappa_D(A)$ depends only on the projective class of A.

Let Σ^{n-1} be the set of all singular matrices of $\mathbb{P}(\mathcal{M}_n(\mathbb{C}))$ (which is an algebraic variety.)

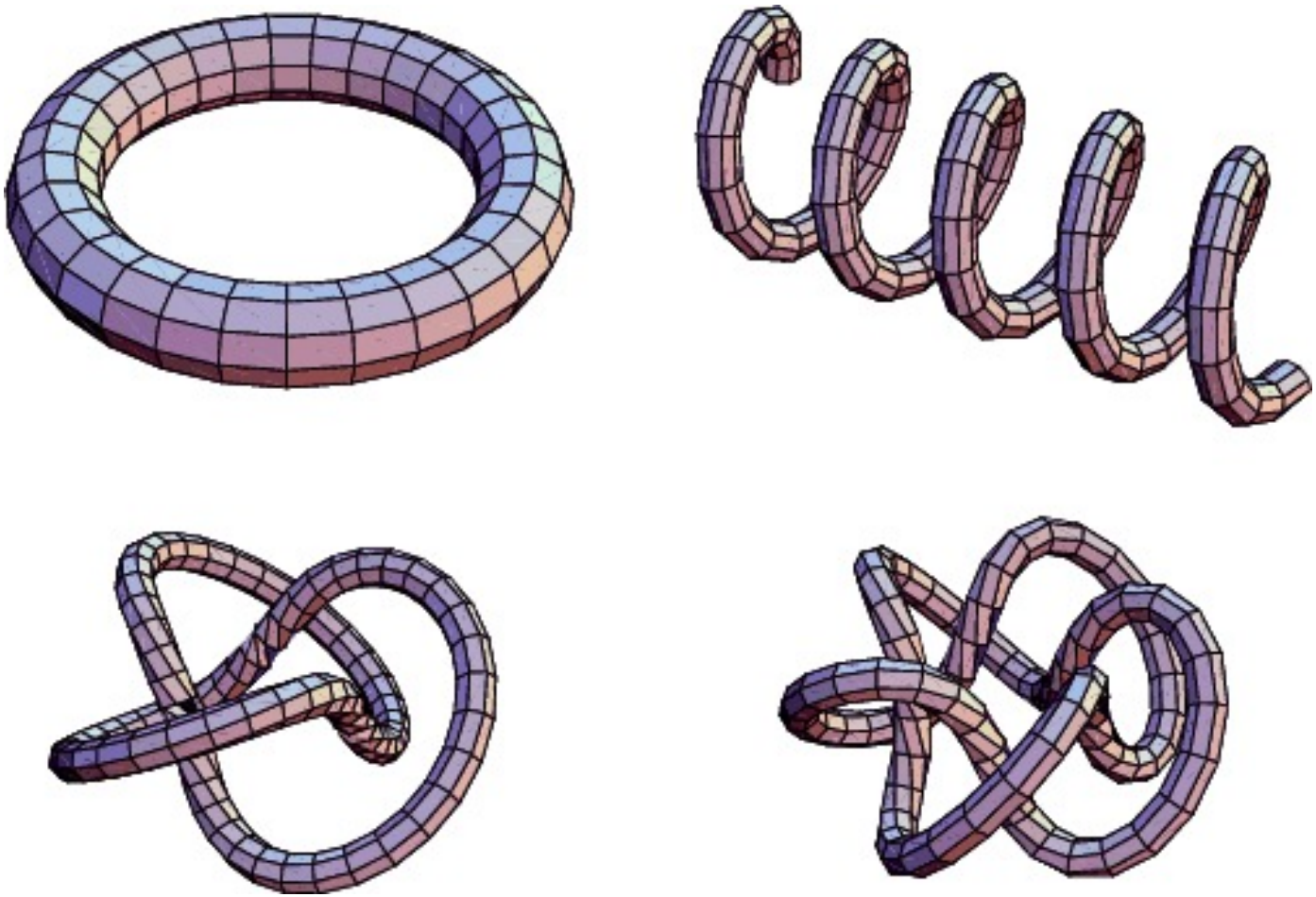
Theorem

$$\kappa_D(A) = \frac{1}{d_{\mathbb{P}(\mathcal{M}_n(\mathbb{C}))}(A, \Sigma^{n-1})}.$$

The probability that the condition number is big... equals the probability of being close to Σ^{n-1}.

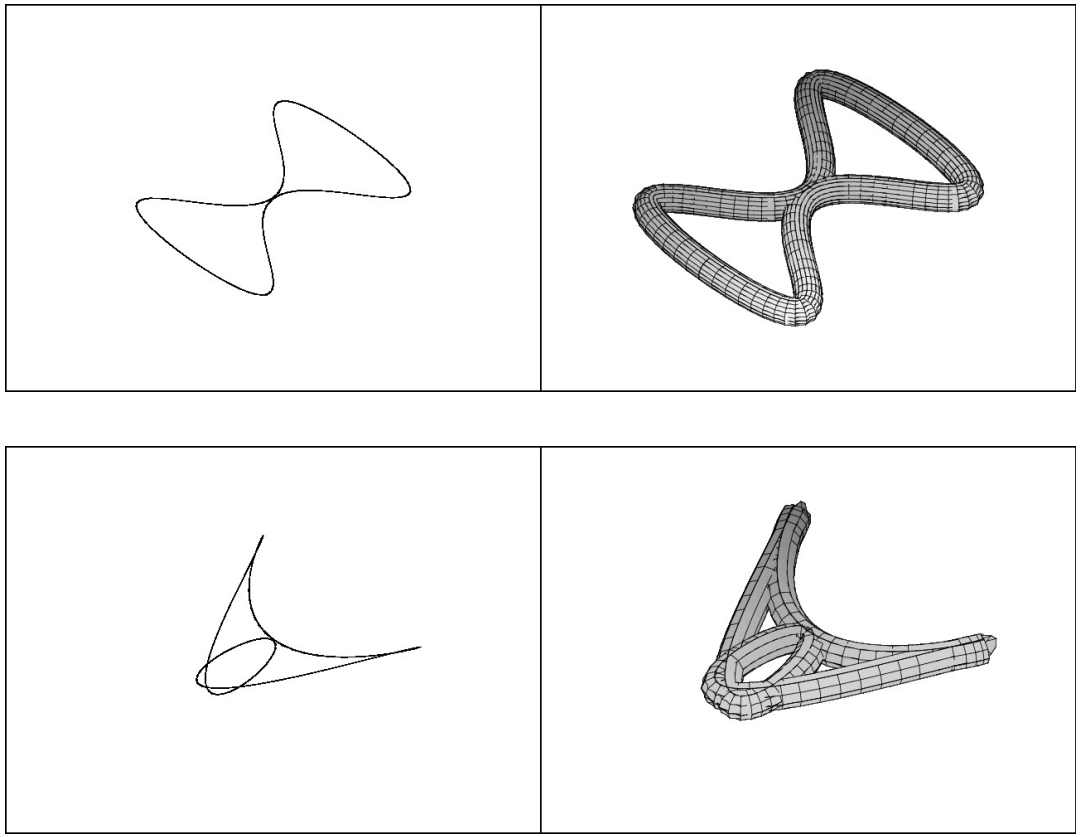
The volume of tubes

The classical result by Weyl (Gray for the projective version) is only valid for smooth varieties and small radius... not our case



The volume of tubes

The classical result by Weyl (Gray for the projective version) is only valid for smooth varieties and small radius... not our case. But there *is* a way to do it.

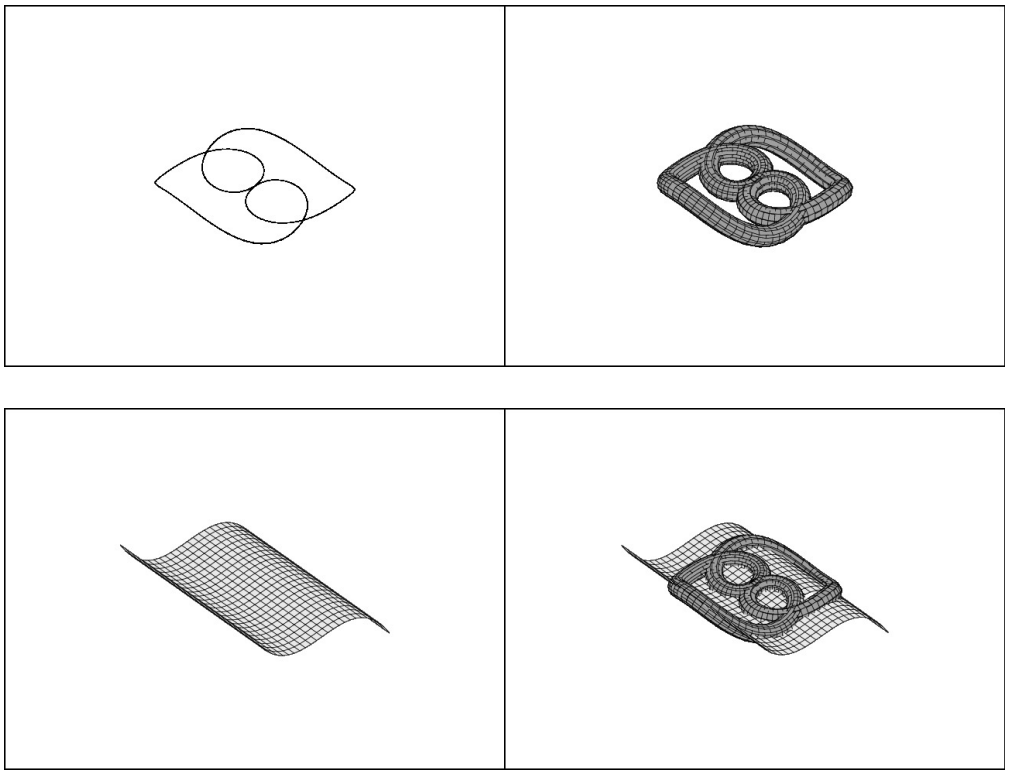


This approach produced the first theoretical statistical study of the condition number. Smale, Renegar, Demmel and others.

Computing kernels

This was probably first observed by Kahan. Let A of size $m \times n$ and $rank(A) = s$

The condition number of A for the problem of computing the subspace such that $Ax = 0$ is the inverse of the distance to the set of one–rank–less matrices.



[B.,Pardo]: a first technique for bounding studying the case $\mathbb{C}$.

The estate of the art in the probabilistic estimation of the condition number of matrices

B., Chen, Dongarra, Edelman, Pardo, Shub, Sutton...

A of size $m \times n$ and rank r. Let $t \geq n + m - 2r + 1$. Then,

$$\mathrm{P}\left[\frac{\kappa(A)}{\frac{n+m-r}{n+m-2r+1}} > t\right] \approx \frac{1}{(2\pi)^{\beta/2}} \left(\frac{C}{t}\right)^{\beta(n+m-2r+1)}$$

Moreover,

$$\mathrm{E}[\log(\kappa(A))] \leq \log \frac{n+m-r}{n+m-2r+1} + 2.6.$$

Here, $\beta = 1$ (resp. $\beta = 2$) if the matrices are real (resp. complex).

How much precision should we use for solving "random" problems $Ax = b$ where A is an $m \times n$ matrix of rank r?

B., Chen, Dongarra, Edelman, Pardo, Shub, Sutton...

$$\mathrm{E}[\log(\kappa(A)) : A \in \Sigma] \leq \log \frac{n+m-r}{n+m-2r+1} + 2.6.$$

How much precision should we use for solving "random" problems $Ax = b$ where A is an $m \times n$ matrix of rank r?

B., Chen, Dongarra, Edelman, Pardo, Shub, Sutton...

$$\mathrm{E}[\log(\kappa(A)) : A \in \Sigma] \leq \log \frac{n+m-r}{n+m-2r+1} + 2.6.$$

For example, for $10^6 \times 10^6 + 1$ matrices of rank $10^6 - 200$ that quantity is 6 so you should be using at least 8 decimal digits of precision, at least in the output. **This is doable in our double precision IEEE standard.**

How much precision should we use for solving "random" problems $Ax = b$ where A is an $m \times n$ matrix of rank r?

B., Chen, Dongarra, Edelman, Pardo, Shub, Sutton...

$$\mathrm{E}[\log(\kappa(A)) : A \in \Sigma] \leq \log \frac{n+m-r}{n+m-2r+1} + 2.6.$$

For example, for $10^6 \times 10^6 + 1$ matrices of rank $10^6 - 200$ that quantity is 6 so you should be using at least 8 decimal digits of precision, at least in the output. **This is doable in our double precision IEEE standard.**
But for $10^{14} \times 10^{14}$ matrices of rank $10^{14} - 1$, that quantity is around 17. Too much precision for our machines, right?

How much precision should we use for solving "random" problems $Ax = b$ where A is an $m \times n$ matrix of rank r?

B., Chen, Dongarra, Edelman, Pardo, Shub, Sutton...

$$\mathrm{E}[\log(\kappa(A)) : A \in \Sigma] \leq \log \frac{n+m-r}{n+m-2r+1} + 2.6.$$

For example, for $10^6 \times 10^6 + 1$ matrices of rank $10^6 - 200$ that quantity is 6 so you should be using at least 8 decimal digits of precision, at least in the output. **This is doable in our double precision IEEE standard.**

But for $10^{14} \times 10^{14}$ matrices of rank $10^{14} - 1$, that quantity is around 17. Too much precision for our machines, right?

And for $10^{20} \times 10^{25}$ matrices of rank 10^{19}?

How much precision should we use for solving "random" problems $Ax = b$ where A is an $m \times n$ matrix of rank r?

B., Chen, Dongarra, Edelman, Pardo, Shub, Sutton...

$$\mathrm{E}[\log(\kappa(A)) : A \in \Sigma] \leq \log \frac{n+m-r}{n+m-2r+1} + 2.6.$$

For example, for $10^6 \times 10^6 + 1$ matrices of rank $10^6 - 200$ that quantity is 6 so you should be using at least 8 decimal digits of precision, at least in the output. **This is doable in our double precision IEEE standard.**

But for $10^{14} \times 10^{14}$ matrices of rank $10^{14} - 1$, that quantity is around 17. Too much precision for our machines, right?

And for $10^{20} \times 10^{25}$ matrices of rank 10^{19}?The value of this quantity is less than 3. So, that is doable!

A question, or maybe a conjecture

This is strongly suggested by the facts above

Prove or disprove: The set Σ of rank r matrices in the set of $m \times n$ matrices is a minimal variety in the vector space of matrices. That is, for any open set U, the set $U \cap \Sigma$ has minimal Hausdorff measure among all rectifiable (for the same Hausdorff dimension) sets with the same boundary that $U \cap \Sigma$.

A question, or maybe a conjecture

This is strongly suggested by the facts above

Prove or disprove: The set Σ of rank r matrices in the set of $m \times n$ matrices is a minimal variety in the vector space of matrices. That is, for any open set U, the set $U \cap \Sigma$ has minimal Hausdorff measure among all rectifiable (for the same Hausdorff dimension) sets with the same boundary that $U \cap \Sigma$.
In the complex case, this is consequence of a classical fact: every complex algebraic variety has that minimizing property.

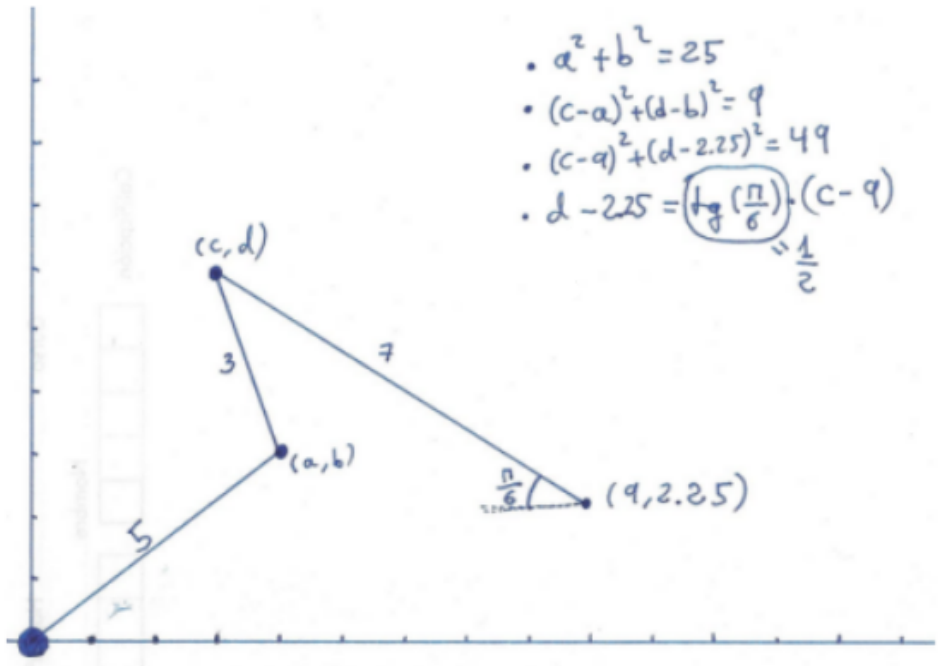
Further

That was just a very particular thing!

1. Smooth analysis of the condition number (Spielman & Chen) has been expanded in different directions with similar techniques (Burgisser & Cucker and others).
2. Average results for other problems as linear optimization, polynomial system solving, eigenvalue solving etc. also exist (Smale, Renegar, Shub, B., Pardo, Armentano and others).
3. Other probability distributions have been analysed with some success (Tao & Vu and others).
4. Many open questions. For example, average of the condition number when the entries are uniformly distributed in $[-1, 1]$ or in $\{-1, 1\}$.
5. Do you want a paper in Annals of Math.? Prove a version for sparse matrices.

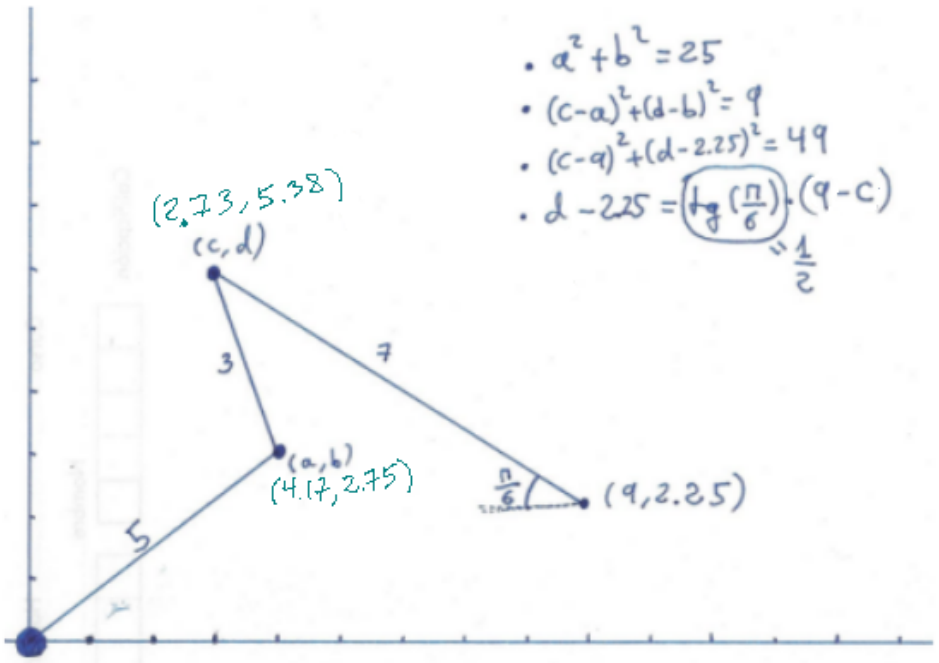
Now let us look at the problem of solving $f : \mathbb{K}^n \to \mathbb{K}^n$

Robots in car factories



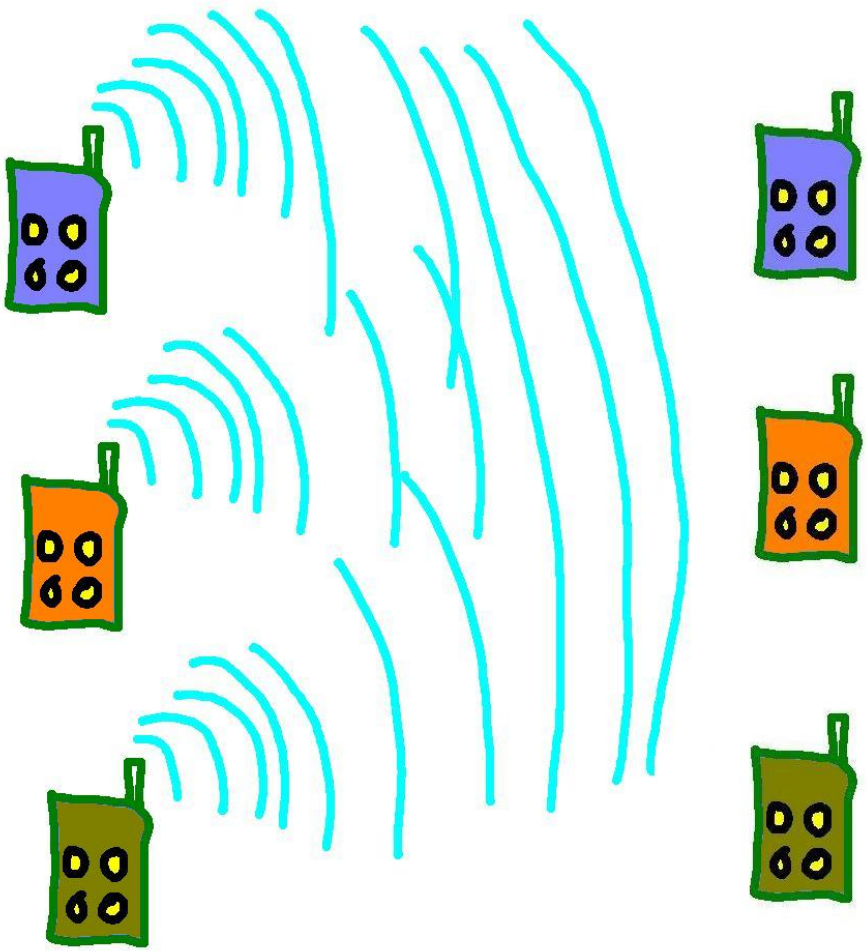
Now let us look at the problem of solving $f : \mathbb{K}^n \to \mathbb{K}^n$

Robots in car factories



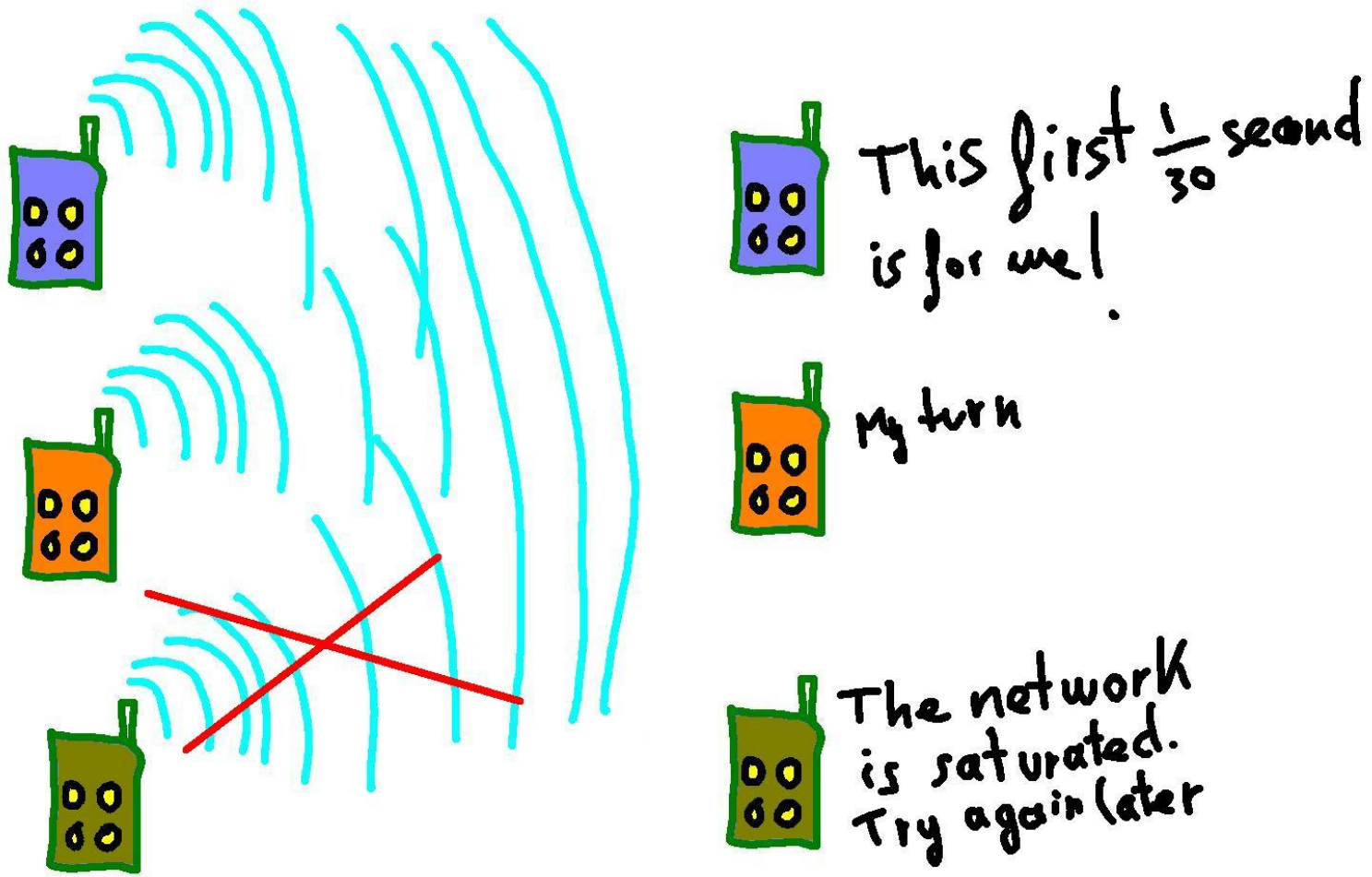
Now let us look at the problem of solving $f : \mathbb{K}^n \to \mathbb{K}^n$

20 people can use their mobiles at the same time in the same room. How is it possible?



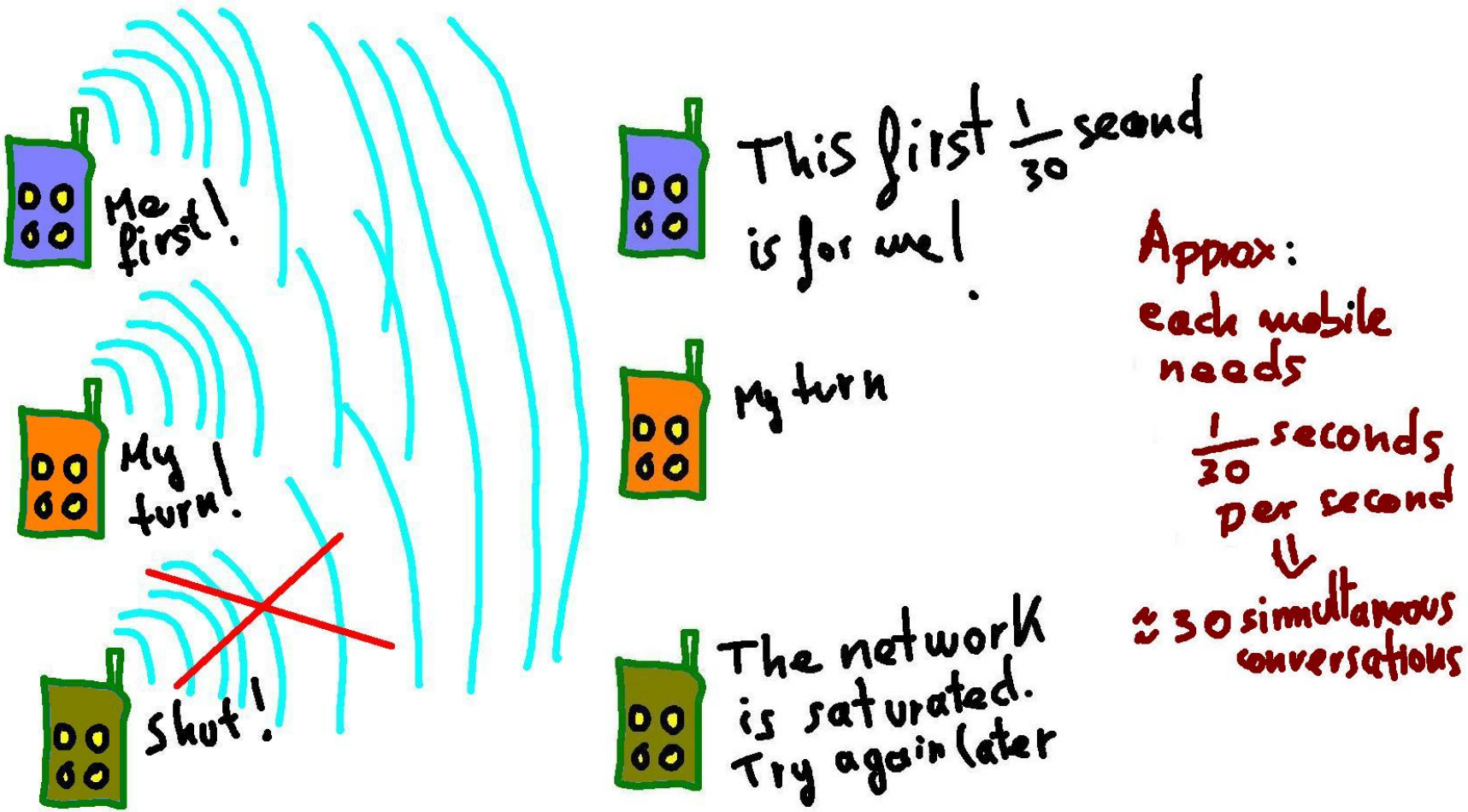
Now let us look at the problem of solving $f : \mathbb{K}^n \to \mathbb{K}^n$

20 people can use their mobiles at the same time in the same room. How is it possible?



Now let us look at the problem of solving $f : \mathbb{K}^n \to \mathbb{K}^n$

20 people can use their mobiles at the same time in the same room. How is it possible?

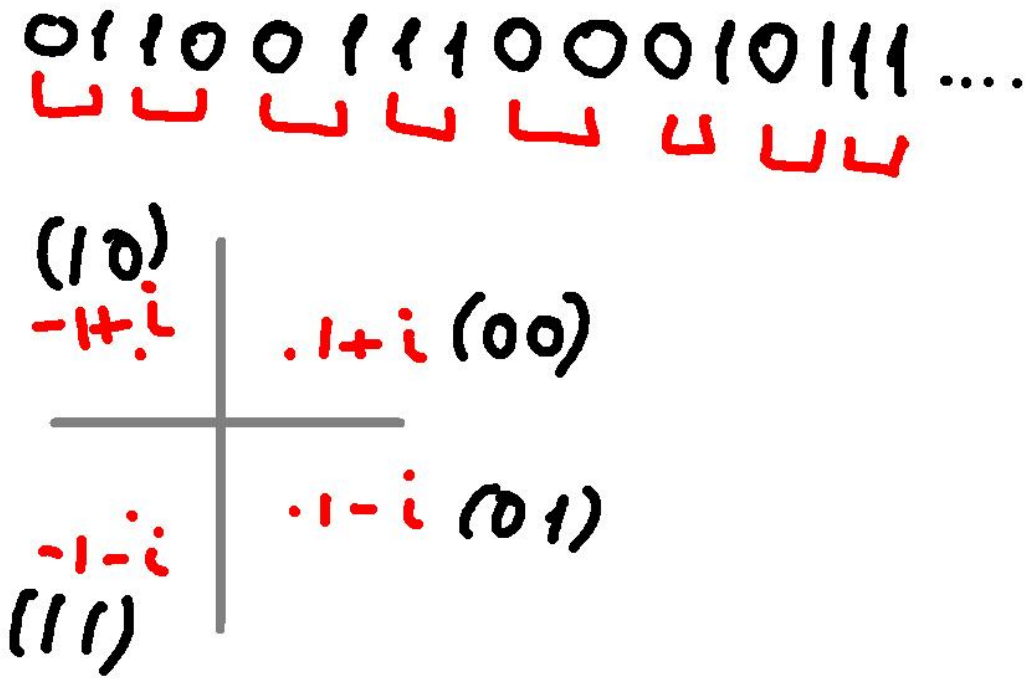


Now let us look at the problem of solving $f : \mathbb{K}^n \to \mathbb{K}^n$

Why $0, 1$ sequences are waves?

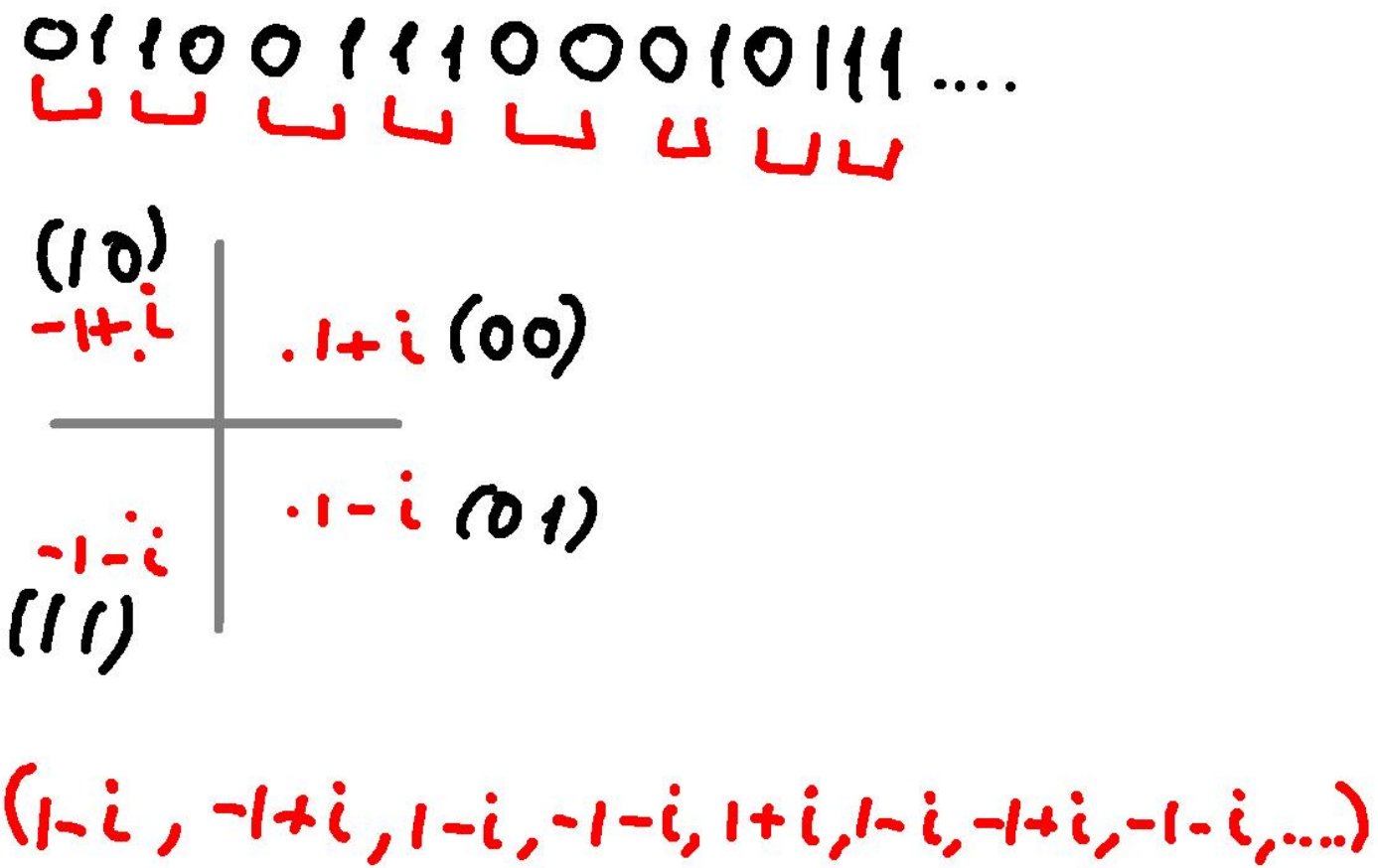
Now let us look at the problem of solving $f : \mathbb{K}^n \to \mathbb{K}^n$

Why $0, 1$ sequences are waves?



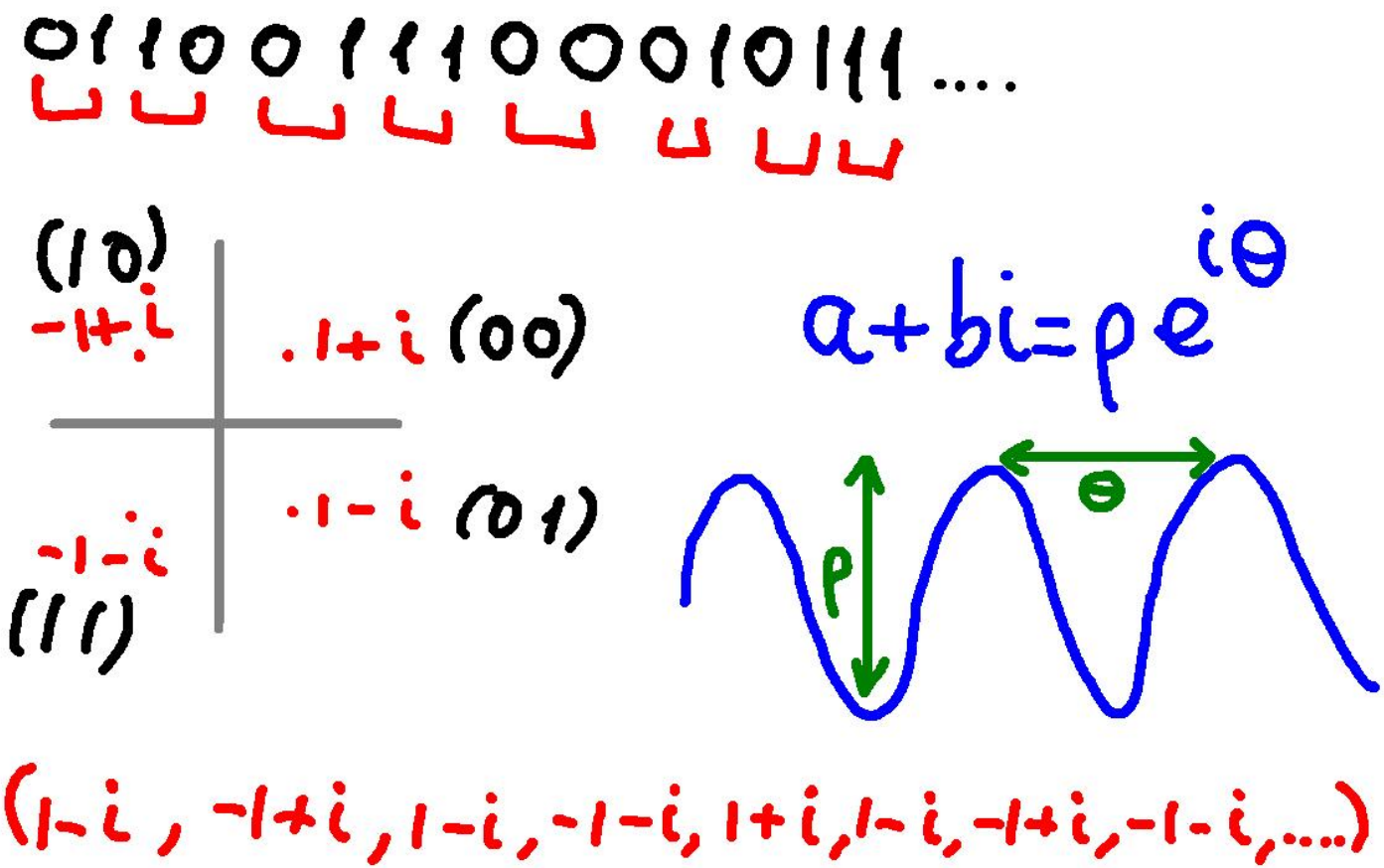
Now let us look at the problem of solving $f : \mathbb{K}^n \to \mathbb{K}^n$

Why $0, 1$ sequences are waves?



Now let us look at the problem of solving $f : \mathbb{K}^n \to \mathbb{K}^n$

Why $0, 1$ sequences are waves?

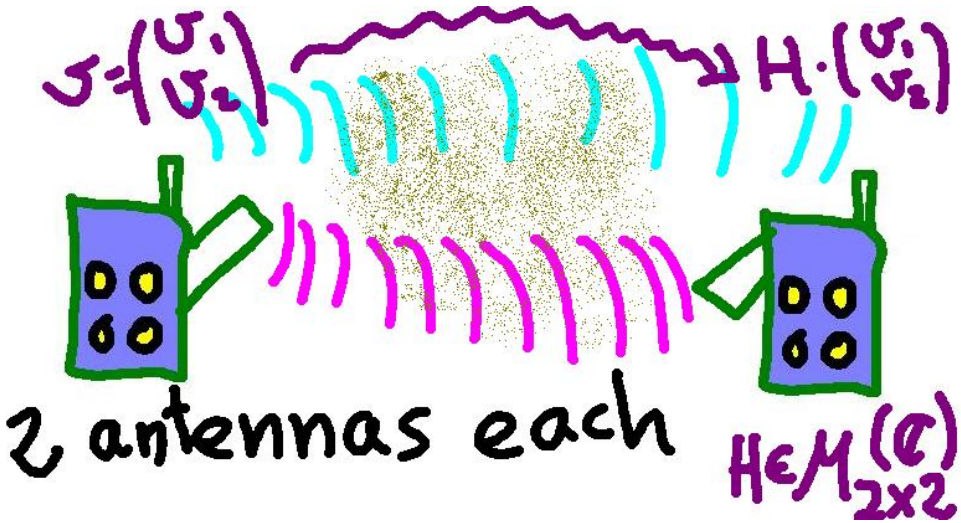


Now let us look at the problem of solving $f : \mathbb{K}^n \to \mathbb{K}^n$

So you can send a vector in $\mathbb{C}^N$, N the number of "antennas", and your friend receives a linear modification of it

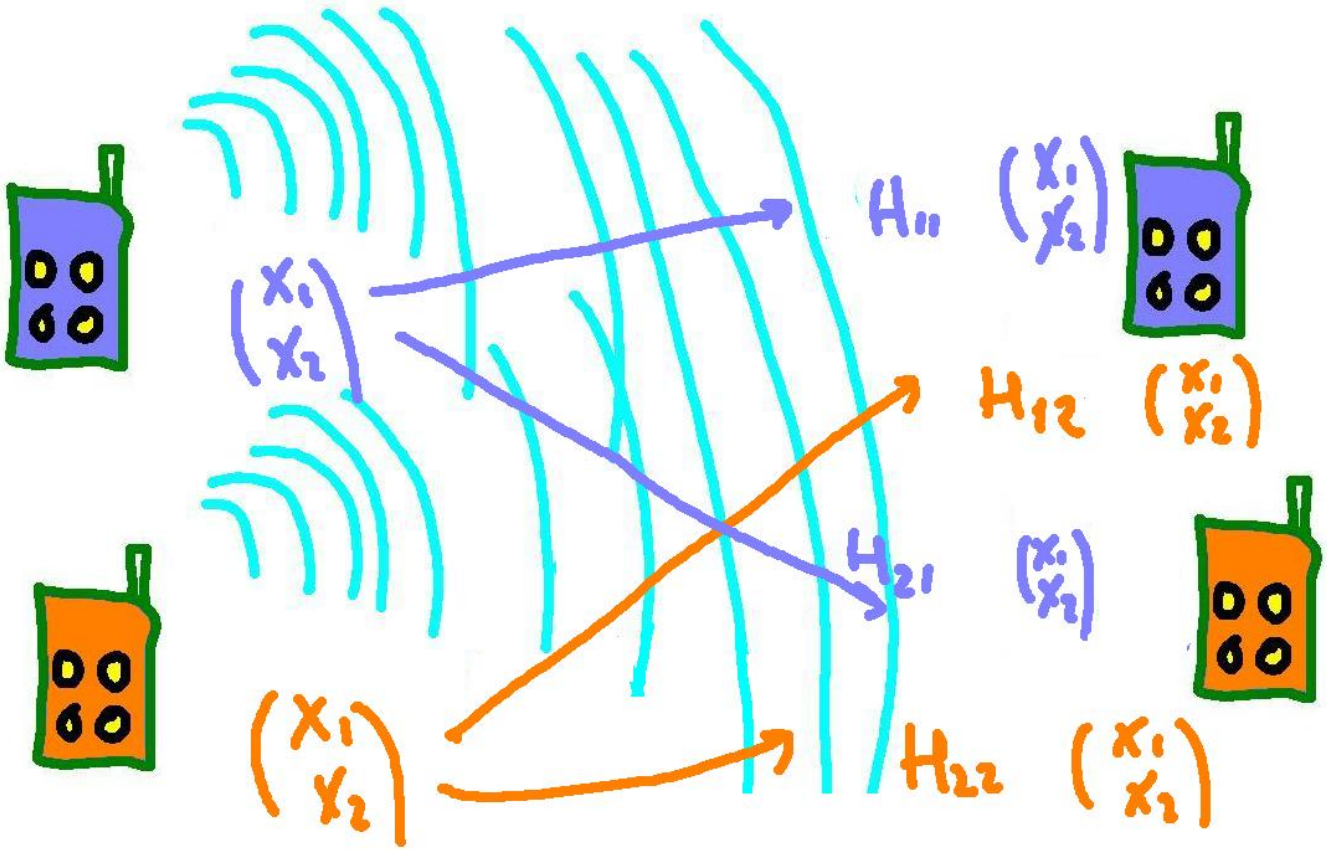
Now let us look at the problem of solving $f : \mathbb{K}^n \to \mathbb{K}^n$

So you can send a vector in $\mathbb{C}^N$, N the number of "antennas", and your friend receives a linear modification of it



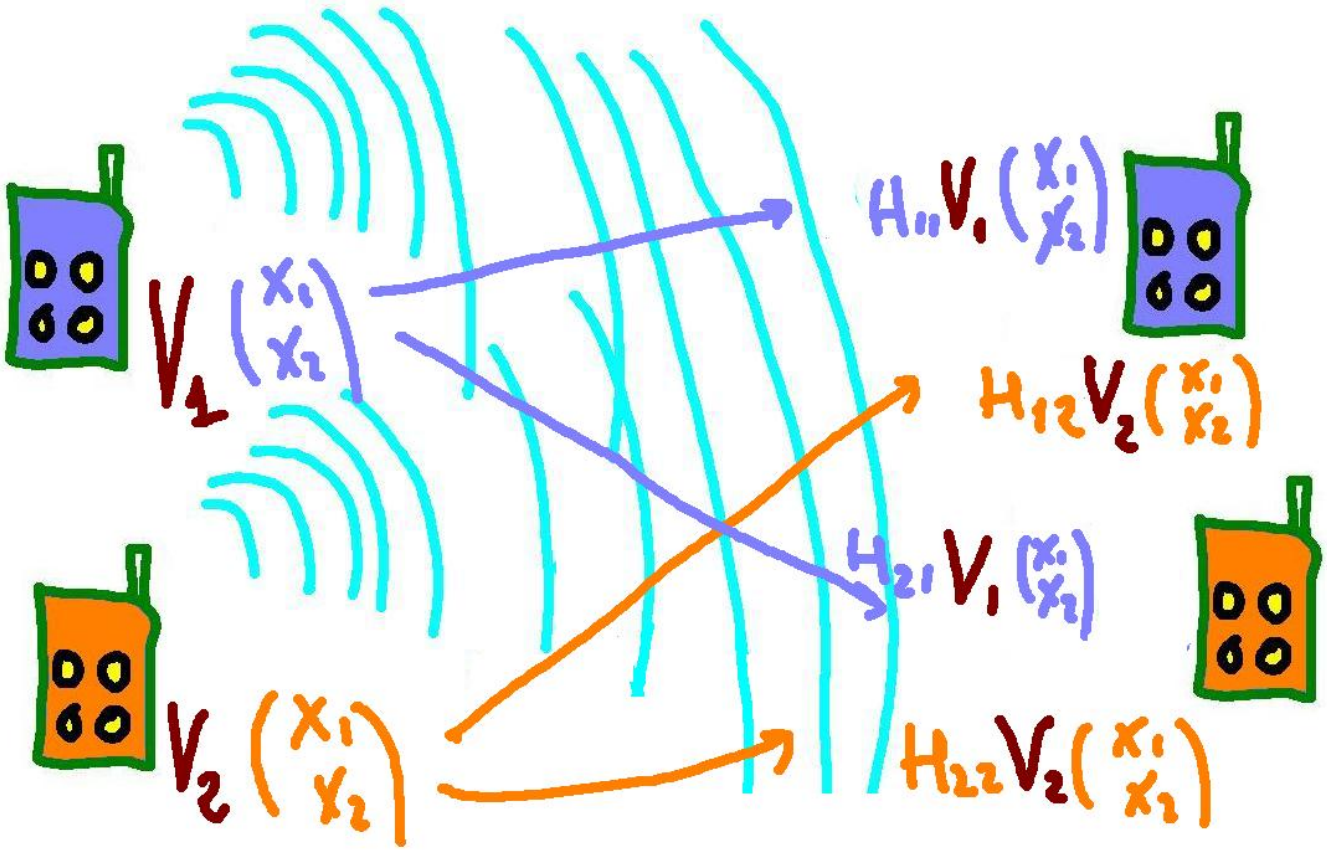
Now let us look at the problem of solving $f : \mathbb{K}^n \to \mathbb{K}^n$

Interference Alignment: an idea of Jafar's and Khandani's research groups



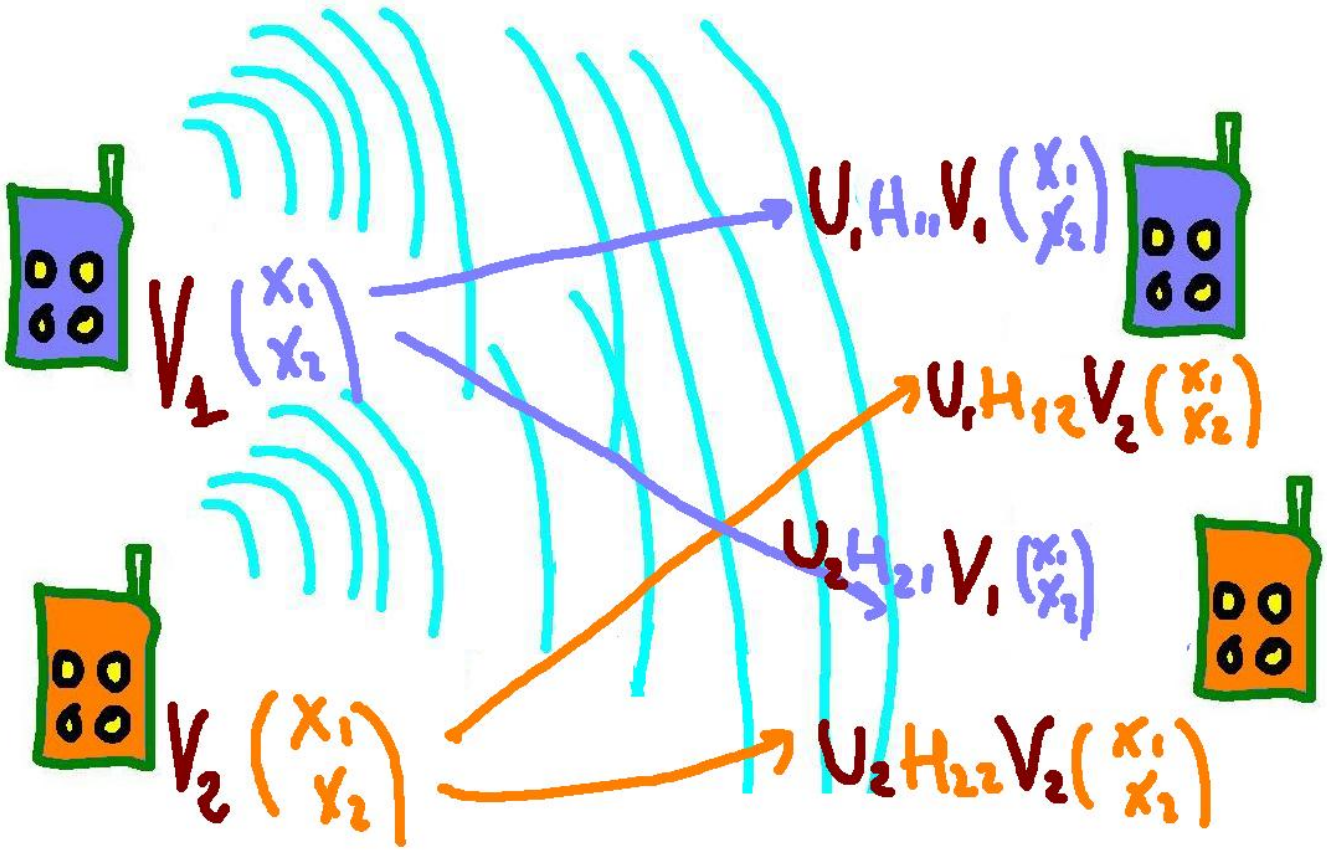
Now let us look at the problem of solving $f : \mathbb{K}^n \to \mathbb{K}^n$

Interference Alignment: an idea of Jafar's and Khandani's research groups



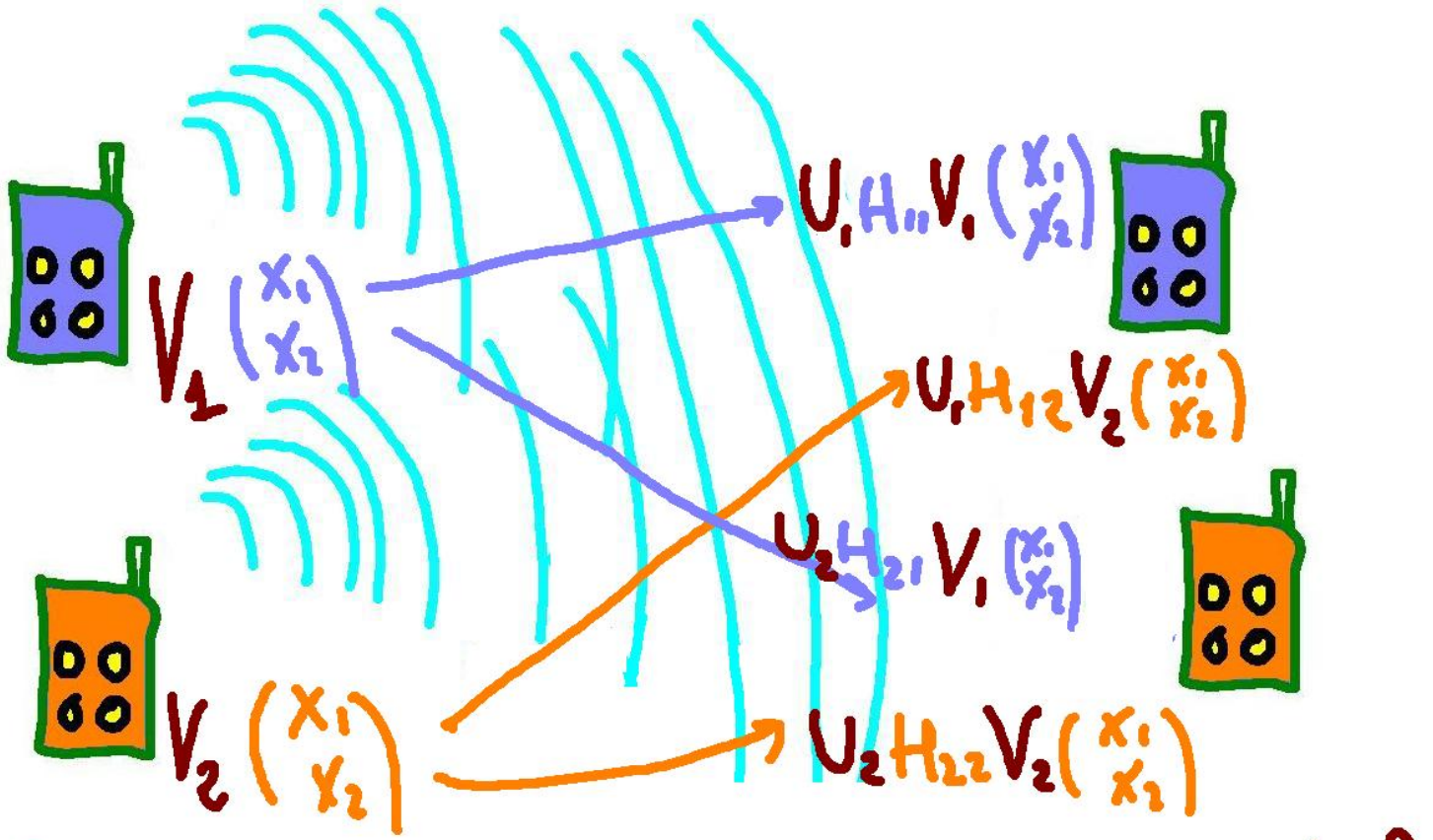
Now let us look at the problem of solving $f : \mathbb{K}^n \to \mathbb{K}^n$

Interference Alignment: an idea of Jafar's and Khandani's research groups



Now let us look at the problem of solving $f : \mathbb{K}^n \to \mathbb{K}^n$

Interference Alignment: an idea of Jafar's and Khandani's research groups



Require: $U_k H_{kl} V_l = 0, \forall k \neq l$

Now let us look at the problem of solving $f : \mathbb{K}^n \to \mathbb{K}^n$

After engineering considerations have been taken into

Let K be the number of transmitters/receivers. Let

$$\Phi = \{(k, \ell) : \text{transmitter } \ell \text{ interfers receiver } k\} \subseteq \{1, \ldots, K\}^2.$$

Let transmitter ℓ have M_ℓ antennas, receiver k have N_k antennas. Let $d_j \leq \min\{M_j, N_j\}$, $1 \leq j \leq K$, and let

$$H_{k\ell} \in \mathcal{M}_{N_k \times M_\ell}(\mathbb{C})$$

be fixed (known). Compute $U_k \in \mathcal{M}_{M_k \times d_k}(\mathbb{C})$, $1 \leq k \leq K$ and $V_\ell \in \mathcal{M}_{N_\ell \times d_\ell}(\mathbb{C})$, $1 \leq \ell \leq K$ such that

$$U_k^T H_{k\ell} V_\ell = 0 \in \mathcal{M}_{d_k \times d_\ell}(\mathbb{C}), \quad k \neq \ell.$$

This is a system of **many** polynomial equations (degree 2) in **many** variables.

Three facts

- Whenever a fast algorithm for *some* problem is devised, **many** other problems are attacked by reducing them to the already solved one. This was the case with linear algebra. It is also becoming standard to reduce many problems to just solving a system of polynomial equations $f : \mathbb{K}^n \to \mathbb{K}^n$ (i.e. wavelet analysis by Bank, Lehmann and coworkers, or the study of the Stuart platform by Giusti, Schost and coworkers, databases analysis by Heintz and coworkers, and much more!).

Three facts

- Whenever a fast algorithm for *some* problem is devised, **many** other problems are attacked by reducing them to the already solved one. This was the case with linear algebra. It is also becoming standard to reduce many problems to just solving a system of polynomial equations $f : \mathbb{K}^n \to \mathbb{K}^n$ (i.e. wavelet analysis by Bank, Lehmann and coworkers, or the study of the Stuart platform by Giusti, Schost and coworkers, databases analysis by Heintz and coworkers, and much more!).
- In the problem of linear algebra, the condition number controls the precision, the stability w.r.t. changes in the problem input, and the complexity of some of the most important iterative algorithms (i.e. conjugate gradient).

Three facts

- Whenever a fast algorithm for *some* problem is devised, **many** other problems are attacked by reducing them to the already solved one. This was the case with linear algebra. It is also becoming standard to reduce many problems to just solving a system of polynomial equations $f : \mathbb{K}^n \to \mathbb{K}^n$ (i.e. wavelet analysis by Bank, Lehmann and coworkers, or the study of the Stuart platform by Giusti, Schost and coworkers, databases analysis by Heintz and coworkers, and much more!).
- In the problem of linear algebra, the condition number controls the precision, the stability w.r.t. changes in the problem input, and the complexity of some of the most important iterative algorithms (i.e. conjugate gradient).
- In the problem of the numerical solution of systems $f : \mathbb{K}^n \to \mathbb{K}^n$ this will also happen.

Polynomial systems $f : \mathbb{C}^n \to \mathbb{C}^n$

Let $f = (f_1, \ldots, f_n)$ where $f_i \in \mathbb{C}[X_1, \ldots, X_n]$, $\deg(f_i) = d_i$ is a system of equations. What do we do to find its solution set $V(f)$? Different approaches:

- Find a Groebner basis ([Buchberger]) of the ideal $I = (f_1, \ldots, f_n)$, with respect to lexicographical order. Possibly best implementation by Faugére.
- Find a "Kronecker solution", that is a projection $\pi : V(f) \to L$ for some line L, a polynomial $p(T)$ such that its zeros correspond to points in $\pi(V(f))$, and rational functions which lift those zeros back to $V(f)$. TERA team, leaded by B. Bank, M. Giusti, J. Heintz, L.M. Pardo. Possibly best implementation by G. Lécerf.

Polynomial systems $f : \mathbb{C}^n \to \mathbb{C}^n$

Let $f = (f_1, \ldots, f_n)$ where $f_i \in \mathbb{C}[X_1, \ldots, X_n]$, $\deg(f_i) = d_i$ is a system of equations. What do we do to find its solution set $V(f)$? Different approaches:

- Direct search by the size of boxes containing approximate zeros. Cucker, Krick, Malajovich & Wschebor. See the course by Gregorio Malajovich.
- The method of moments (Lasserre, Laurent & Rostalski).
- The use of polar varieties (Bank, Heintz, Lehmann, Mbakop, & Pardo), see the talk by Marc Giusti.

Polynomial systems $f : \mathbb{C}^n \to \mathbb{C}^n$

Let $f = (f_1, \ldots, f_n)$ where $f_i \in \mathbb{C}[X_1, \ldots, X_n]$, $\deg(f_i) = d_i$ is a system of equations. What do we do to find its solution set $V(f)$? Different approaches:

These algorithms give us a lot of information, but take exponential running time. The topic of this part of the course is: can we get just a little information, but fast?

Ask for less information and you might be faster

A simple question: can we approximate just one zero, but guaranteeing polynomial running time?

Ask for less information and you might be faster

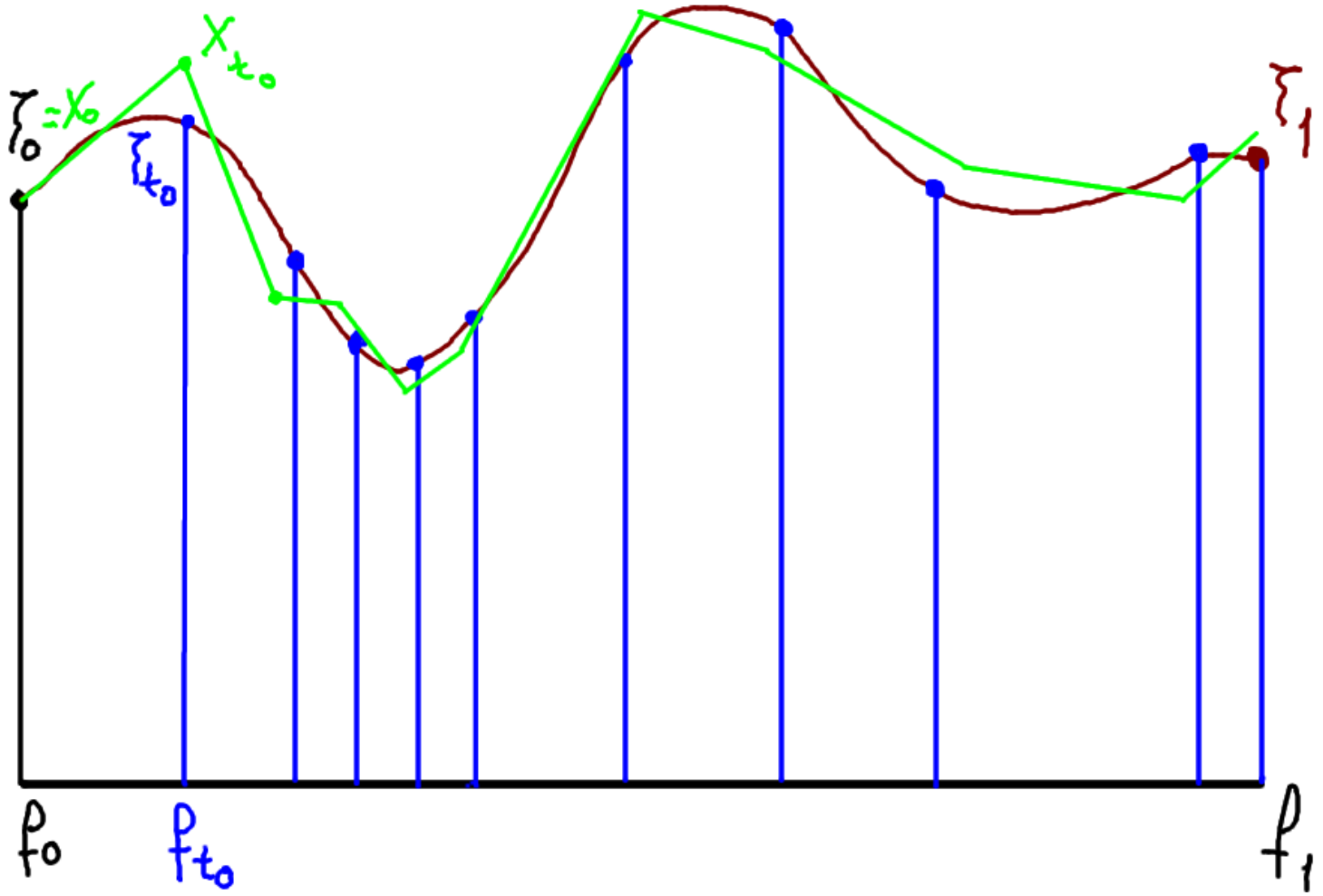
A simple question: can we approximate just one zero, but guaranteeing polynomial running time?
Smale's 17th Problem:

Can a zero of n complex polynomial equations in n unknowns be found approximately, on the average, in polynomial time with a uniform algorithm?

Stephen Smale, *Mathematical problems for the next century*.
Mathematics: frontiers and perspectives.
American Mathematical Society, 2000.

Homotopy method.

Modern usage based on Shub & Smale's fundamental work.



Homotopy method

Modern usage based on Shub & Smale's fundamental work.

An intuitive idea: why does this work?

Because we have reduced our original problem to a well–known problem!

Differentiate the expression

$$f_t(\zeta_t) = 0,$$

with respect to time, to get

$$\dot{f}_t(\zeta_t) + D(f_t)(\zeta_t)\dot{\zeta}_t.$$

That is, we have:

$$\begin{cases} \dot{\zeta}_t = -(D(f_t)(\zeta_t))^{-1}\dot{f}_t(\zeta_t) \\ \zeta_0 \text{ known} \end{cases}$$

This is an ODE system (Cauchy problem)!!! There are many methods for something like this...

Homogeneous systems of equations

Instead of solving g we can just solve f:

$$g = \begin{cases} x_1x_2 - x_2 - 7 = 0 \\ x_1^3 + 7x_2 - 9 = 0 \end{cases} \qquad f = \begin{cases} x_1x_2 - x_2x_0 - 7x_0^2 = 0 \\ x_1^3 + 7x_2x_0^2 - 9x_0^3 = 0 \end{cases}$$

Homogeneous systems of equations

Instead of solving g we can just solve f:

$$g = \begin{cases} x_1x_2 - x_2 - 7 = 0 \\ x_1^3 + 7x_2 - 9 = 0 \end{cases} \qquad f = \begin{cases} x_1x_2 - x_2x_0 - 7x_0^2 = 0 \\ x_1^3 + 7x_2x_0^2 - 9x_0^3 = 0 \end{cases}$$

Note that the solutions of f are points in $\mathbb{P}(\mathbb{C}^3)$, that is if (x_0, x_1, x_2) is a solution then so is $(\lambda x_0, \lambda x_1, \lambda x_2)$ for any $\lambda \in \mathbb{C}$. Besides,

(x_0, x_1, x_2) is a solution of f, $x_0 \neq 0 \Rightarrow (x_1/x_0, x_2/x_0)$ is a solution of g.

$(1, a, b)$ is a solution of f $\Leftarrow (a, b)$ is a solution of g.

Homogeneous systems of equations

Instead of solving g we can just solve f:

$$g = \begin{cases} x_1x_2 - x_2 - 7 = 0 \\ x_1^3 + 7x_2 - 9 = 0 \end{cases} \qquad f = \begin{cases} x_1x_2 - x_2x_0 - 7x_0^2 = 0 \\ x_1^3 + 7x_2x_0^2 - 9x_0^3 = 0 \end{cases}$$

Note that the solutions of f are points in $\mathbb{P}(\mathbb{C}^3)$, that is if (x_0, x_1, x_2) is a solution then so is $(\lambda x_0, \lambda x_1, \lambda x_2)$ for any $\lambda \in \mathbb{C}$. Besides,

(x_0, x_1, x_2) is a solution of f, $x_0 \neq 0 \Rightarrow (x_1/x_0, x_2/x_0)$ is a solution of g.

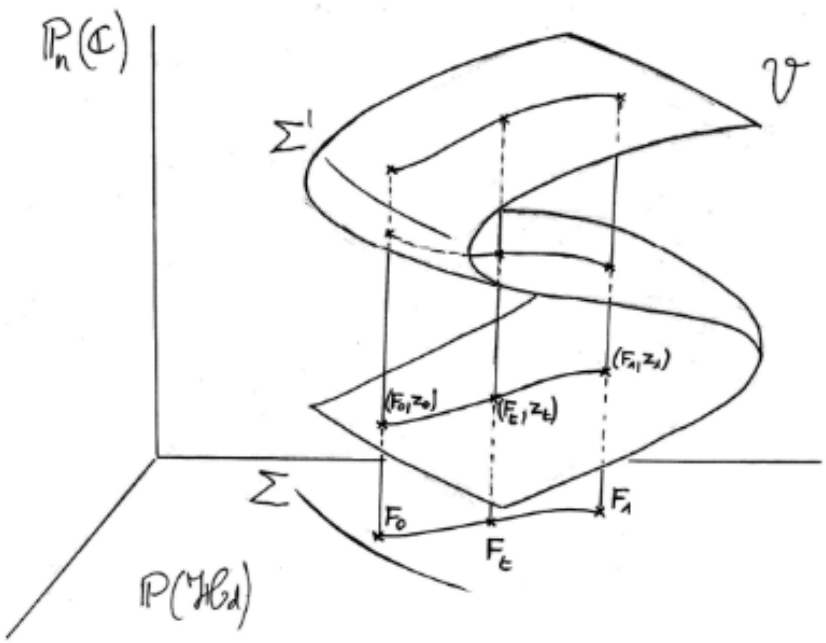
$$(1, a, b) \text{ is a solution of f} \Leftarrow (a, b) \text{ is a solution of g.}$$

Thus we consider just homogeneous systems $f : \mathbb{C}^{n+1} \to \mathbb{C}^n$, and look for zeros in $\mathbb{P}(\mathbb{C}^{n+1})$. Let $\mathbb{P}(\mathcal{H}_{(d)})$ be the projective of the vector space of homogeneous sytems.

The solution variety $\mathcal{V} = \{(f, \zeta) : f(\zeta) = 0\}$

Is an algebraic variety and a differential submanifold of $\mathbb{P}(\mathcal{H}_{(d)}) \times \mathbb{P}(\mathbb{C}^{n+1})$

This sketch is courtesy of Jean Pierre Dedieu.



$$\Sigma' = \{(f, \zeta) \in \mathcal{V} : rank(Df(\zeta)) < n\}, \qquad \Sigma = \pi_1(\Sigma').$$

Homotopy method

- Convergence results for Newton's Method (Kantorovich, Kim, Smale, Shub & Smale, Dedieu & Malajovich, Wang & Hang, Giusti, Lecerf, Salvy & J.-C. Yakoubsohn). See the talk by Jean Claude Yakoubsohn for more.

Homotopy method

- Convergence results for Newton's Method (Kantorovich, Kim, Smale, Shub & Smale, Dedieu & Malajovich, Wang & Hang, Giusti, Lecerf, Salvy & J.-C. Yakoubsohn). See the talk by Jean Claude Yakoubsohn for more.
- Precise statements about the length of the homotopy steps and the complexity by Renegar, Shub & Smale, B. & Pardo, Burgisser & Cucker.

Homotopy method

- Convergence results for Newton's Method (Kantorovich, Kim, Smale, Shub & Smale, Dedieu & Malajovich, Wang & Hang, Giusti, Lecerf, Salvy & J.-C. Yakoubsohn). See the talk by Jean Claude Yakoubsohn for more.
- Precise statements about the length of the homotopy steps and the complexity by Renegar, Shub & Smale, B. & Pardo, Burgisser & Cucker.
- Required precision in operations and best choice of Newton's method by Malajovich.

Homotopy method

- Convergence results for Newton's Method (Kantorovich, Kim, Smale, Shub & Smale, Dedieu & Malajovich, Wang & Hang, Giusti, Lecerf, Salvy & J.-C. Yakoubsohn). See the talk by Jean Claude Yakoubsohn for more.
- Precise statements about the length of the homotopy steps and the complexity by Renegar, Shub & Smale, B. & Pardo, Burgisser & Cucker.
- Required precision in operations and best choice of Newton's method by Malajovich.
- Different techniques to take advantage of sparsity, detect singularities...

Homotopy method

- Convergence results for Newton's Method (Kantorovich, Kim, Smale, Shub & Smale, Dedieu & Malajovich, Wang & Hang, Giusti, Lecerf, Salvy & J.-C. Yakoubsohn). See the talk by Jean Claude Yakoubsohn for more.
- Precise statements about the length of the homotopy steps and the complexity by Renegar, Shub & Smale, B. & Pardo, Burgisser & Cucker.
- Required precision in operations and best choice of Newton's method by Malajovich.
- Different techniques to take advantage of sparsity, detect singularities...
- More theory, applications, heuristics, software by Bates, Hauenstein, Hubber, Kearfott, Leykin, Lee, Li, Rojas, Sommese, Sottile, Sturmfels, Tsai, Van der Hoeven, Verschelde, Wampler, Xing, Zhao...

Homotopy method

- Convergence results for Newton's Method (Kantorovich, Kim, Smale, Shub & Smale, Dedieu & Malajovich, Wang & Hang, Giusti, Lecerf, Salvy & J.-C. Yakoubsohn). See the talk by Jean Claude Yakoubsohn for more.
- Precise statements about the length of the homotopy steps and the complexity by Renegar, Shub & Smale, B. & Pardo, Burgisser & Cucker.
- Required precision in operations and best choice of Newton's method by Malajovich.
- Different techniques to take advantage of sparsity, detect singularities...
- More theory, applications, heuristics, software by Bates, Hauenstein, Hubber, Kearfott, Leykin, Lee, Li, Rojas, Sommese, Sottile, Sturmfels, Tsai, Van der Hoeven, Verschelde, Wampler, Xing, Zhao...
- Enormous systems are solved this way.

Approximate zero theory of Shub and Smale

Let f be a system and ζ a projective zero of f. Let z be a projective point. We say that z is an approximate zero of f with associated zero ζ if for every $k \geq 0$

$$distance(N_f^k(z), \zeta) \leq \frac{1}{2^{2^k}} distance(z, \zeta),$$

where N_f^k is the result of applying k times the Newton iteration.

Condition number $\mu(f,\zeta)$ and approximate zeros

Let f be a system with a zero ζ and let z be a projective point. Assume that

$$distance(z,\zeta) \leq \frac{3-\sqrt{7}}{2d^{3/2}\mu(f,\zeta)}.$$

Then, z is an approximate zero of f with associate zero ζ.

Condition number $\mu(f,\zeta)$ and approximate zeros

Let f be a system with a zero ζ and let z be a projective point. Assume that

$$distance(z,\zeta) \leq \frac{3-\sqrt{7}}{2d^{3/2}\mu(f,\zeta)}.$$

Then, z is an approximate zero of f with associate zero ζ.Here,

$$\mu(f,\zeta) = \|Diag(\sqrt{d_1},\ldots,\sqrt{d_n})\left(Df(\zeta)\mid_{\zeta^\perp}\right)^{-1}\|$$

is the **condition number** of f at ζ (the formula assumes $\|f\| = \|\zeta\| = 1$).

Thank you for your attention

If we're still in the first class... then we're done for now!

We recall the condition number $\mu(f, \zeta)$

Let f be a system with a zero ζ and let z be a projective point. Assume that

$$distance(z, \zeta) \leq \frac{3 - \sqrt{7}}{2d^{3/2}\mu(f, \zeta)}.$$

Then, z is an approximate zero of f with associate zero ζ.

We recall the condition number $\mu(f,\zeta)$

Let f be a system with a zero ζ and let z be a projective point. Assume that

$$distance(z,\zeta) \leq \frac{3-\sqrt{7}}{2d^{3/2}\mu(f,\zeta)}.$$

Then, z is an approximate zero of f with associate zero ζ.Here,

$$\mu(f,\zeta) = \|Diag(\sqrt{d_1},\ldots,\sqrt{d_n})\left(Df(\zeta)\mid_{\zeta^\perp}\right)^{-1}\|$$

is the **condition number** of f at ζ (the formula assumes $\|f\| = \|\zeta\| = 1$).

What do we mean by $\|f\|$?

One can think on $\|f\|^2$ as simply taking all the coefficients of the monomials in f, compute the square of their modulus, and add it all up...

What do we mean by $\|f\|$?

One can think on $\|f\|^2$ as simply taking all the coefficients of the monomials in f, compute the square of their modulus, and add it all up...

This is almost the Bombieri–Weyl product, in which each coefficient is actually multiplied by a number: the coefficient of $X_0^{\alpha_0} \cdots X_n^{\alpha_n}$ is multiplied by

$$\frac{\alpha_0! \cdots \alpha_n!}{(\alpha_0 + \cdots + \alpha_n)!}.$$

What do we mean by $\|f\|$?

One can think on $\|f\|^2$ as simply taking all the coefficients of the monomials in f, compute the square of their modulus, and add it all up...

This is almost the Bombieri–Weyl product, in which each coefficient is actually multiplied by a number: the coefficient of $X_0^{\alpha_0} \cdots X_n^{\alpha_n}$ is multiplied by

$$\frac{\alpha_0! \cdots \alpha_n!}{(\alpha_0 + \cdots + \alpha_n)!}.$$

This is not an arbitrary or capricious choice!!! Indeed:

- It comes from a vector product (the same way that the usual norm comes from the usual vector product)
- It satisfies a key property, physically meaningful and mathematically helpful:

$$\|f \circ U\| = \|f\|,$$

for any unitary $(n+1) \times (n+1)$ matrix U.

We recall the condition number $\mu(f,\zeta)$

There's no need to assume $\|f\| = \|\zeta\| = 1$, we just need to write down a more complicated formula:

$$\mu(f,\zeta) = \|f\| \| Diag(\sqrt{d_1}\|\zeta\|^{d_1-1}, \ldots, \sqrt{d_n}\|\zeta\|^{d_n-1}) \left(Df(\zeta)\mid_{\zeta^\perp}\right)^{-1} \|.$$

We recall the condition number $\mu(f,\zeta)$

There's no need to assume $\|f\| = \|\zeta\| = 1$, we just need to write down a more complicated formula:

$$\mu(f,\zeta) = \|f\| \| Diag(\sqrt{d_1}\|\zeta\|^{d_1-1}, \ldots, \sqrt{d_n}\|\zeta\|^{d_n-1}) \left(Df(\zeta)\mid_{\zeta^\perp}\right)^{-1} \|.$$

This number controls the stability of solutions, just as κ does in the case of linear system solving: let G be the inverse function associated to π_1 near to (f,ζ). Then, $\|DG(f)\| \leq \mu(f,\zeta)$, where $\|DG(f)\|$ is the norm of the derivative of G at f. This means that if we have a smooth curve (f_t,ζ_t), $f_t(\zeta_t) = 0$, then,

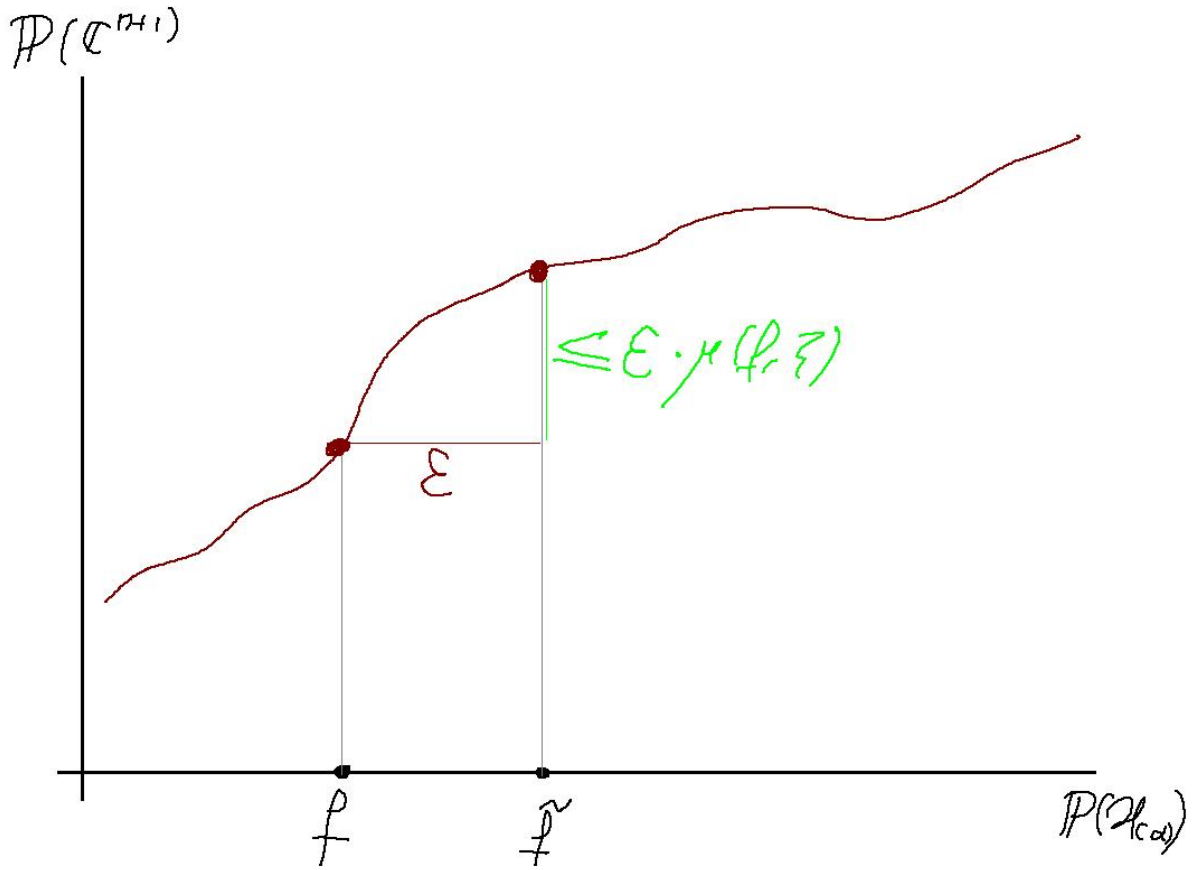
$$\|\dot{\zeta_0}\| \leq \mu(f_0,\zeta_0)\|\dot{f_0}\|.$$

That is, *if* $f(\zeta) = 0$ *and we change* f *to* $\tilde{f}$ *with* $distance(f,\tilde{f}) < \varepsilon$ *then the zero* $\tilde{\zeta}$ *of* $\tilde{f}$ *satisfies*

$$distance(\zeta,\tilde{\zeta}) \lesssim \varepsilon\mu(f,\zeta).$$

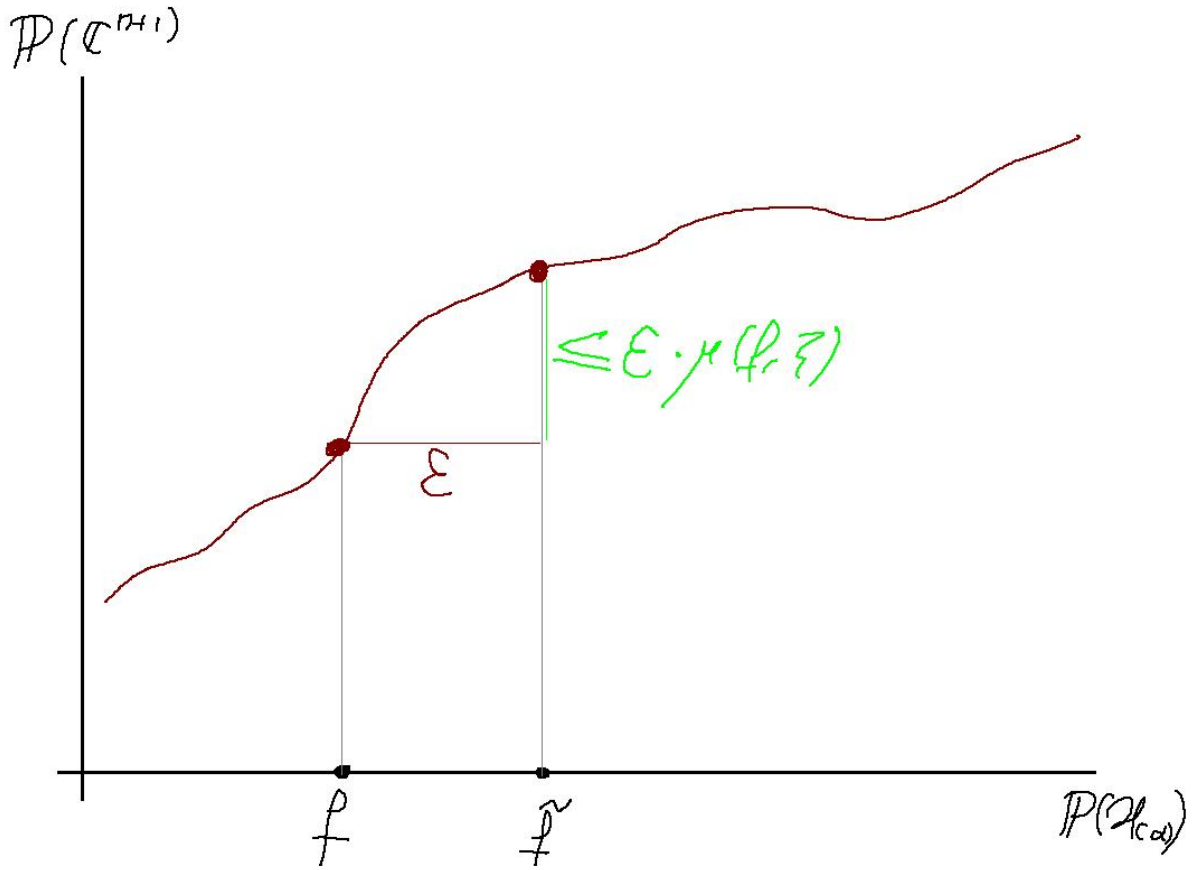
A geometric definition of the condition number

μ is the (essentially) smallest quantity that satisfies:



A geometric definition of the condition number

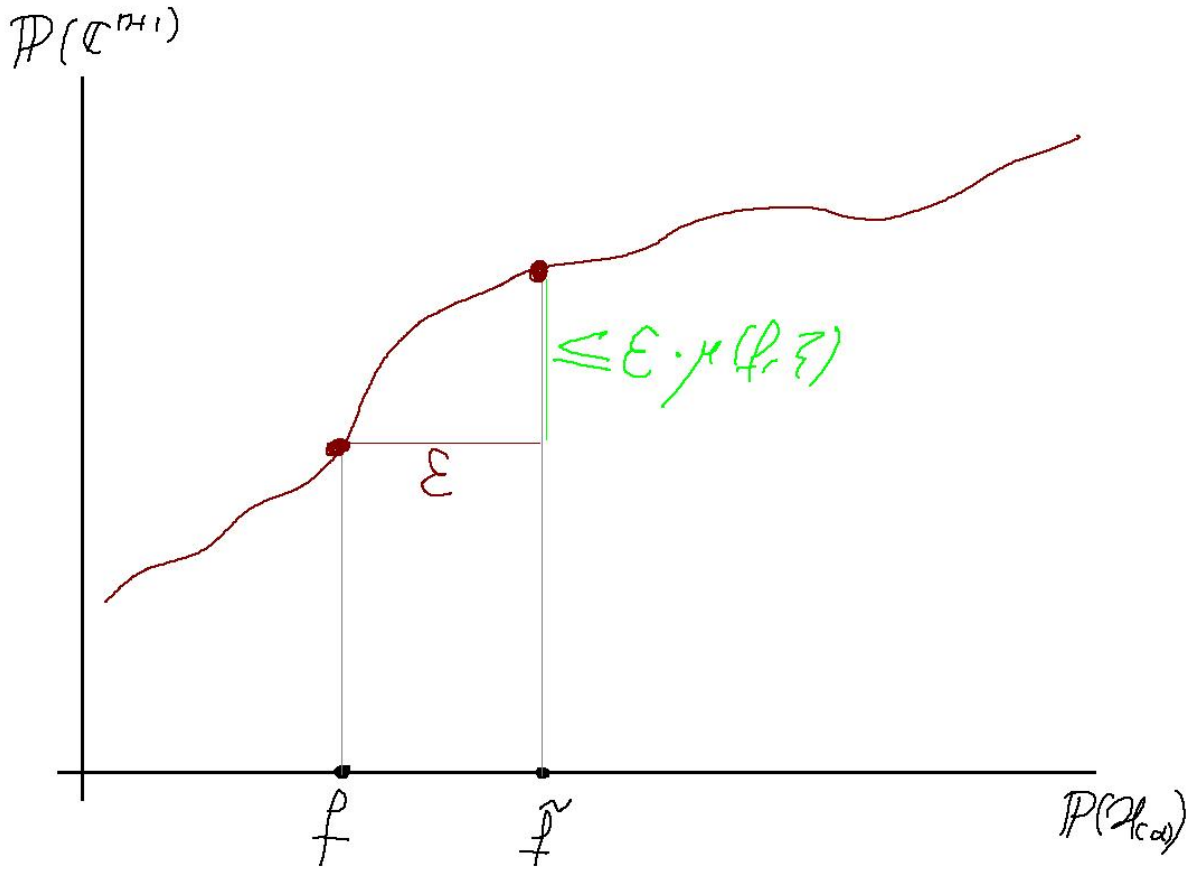
μ is the (essentially) smallest quantity that satisfies:



Indeed, this is valid as far as $\varepsilon < c/\mu(f,\zeta)^2$.

A geometric definition of the condition number

μ is the (essentially) smallest quantity that satisfies:



If I have to move along a path f_t keeping track of ζ_t, then, **the biggest the condition number, the slower I will need to go**.

Let us put things together:

- If z is $c/\mu(f,\zeta)$–close to an actual root ζ of f, then the Newton iterates converge to ζ.

Let us put things together:

- If z is $c/\mu(f,\zeta)$–close to an actual root ζ of f, then the Newton iterates converge to ζ.
- If we have a curve f_t and if f_ε is $c/\mu(f_0,\zeta_0)^2$–close to f, then the zeros of f_ε and f_0 are very similar, and so are $\mu(f_0,\zeta_0)$ and $\mu(f_\varepsilon,\zeta_\varepsilon)$.

Let us put things together:

- If z is $c/\mu(f,\zeta)$–close to an actual root ζ of f, then the Newton iterates converge to ζ.
- If we have a curve f_t and if f_ε is $c/\mu(f_0,\zeta_0)^2$–close to f, then the zeros of f_ε and f_0 are very similar, and so are $\mu(f_0,\zeta_0)$ and $\mu(f_\varepsilon,\zeta_\varepsilon)$.
- Thus, if z is $c/\mu(f,\zeta)$–close to an actual root ζ of f and if f_ε is $c/\mu(f_0,\zeta_0)^2$–close to f, then z is $c/\mu(f_\varepsilon,\zeta_\varepsilon)$–close to ζ_ε... which means that the Newton iterates based on f_ε with initial point z, rapidly converge to ζ_ε.

Let us put things together:

- If z is $c/\mu(f,\zeta)$–close to an actual root ζ of f, then the Newton iterates converge to ζ.
- If we have a curve f_t and if f_ε is $c/\mu(f_0,\zeta_0)^2$–close to f, then the zeros of f_ε and f_0 are very similar, and so are $\mu(f_0,\zeta_0)$ and $\mu(f_\varepsilon,\zeta_\varepsilon)$.
- Thus, if z is $c/\mu(f,\zeta)$–close to an actual root ζ of f and if f_ε is $c/\mu(f_0,\zeta_0)^2$–close to f, then z is $c/\mu(f_\varepsilon,\zeta_\varepsilon)$–close to ζ_ε... which means that the Newton iterates based on f_ε with initial point z, rapidly converge to ζ_ε.
- This gives a better approximation z_1 to ζ_ε than z. And allows us to repeat the process. Using induction, if $\mu(f_t,\zeta_t)<\infty$ for all t, at the end we reach an approximation of the zero of f_{end}.

This is a "Newton–based homotopy", it does not actually use ODE solvers!

Let us put things together:

- If z is $c/\mu(f,\zeta)$–close to an actual root ζ of f, then the Newton iterates converge to ζ.
- If we have a curve f_t and if f_ε is $c/\mu(f_0,\zeta_0)^2$–close to f, then the zeros of f_ε and f_0 are very similar, and so are $\mu(f_0,\zeta_0)$ and $\mu(f_\varepsilon,\zeta_\varepsilon)$.
- Thus, if z is $c/\mu(f,\zeta)$–close to an actual root ζ of f and if f_ε is $c/\mu(f_0,\zeta_0)^2$–close to f, then z is $c/\mu(f_\varepsilon,\zeta_\varepsilon)$–close to ζ_ε... which means that the Newton iterates based on f_ε with initial point z, rapidly converge to ζ_ε.
- This gives a better approximation z_1 to ζ_ε than z. And allows us to repeat the process. Using induction, if $\mu(f_t,\zeta_t)<\infty$ for all t, at the end we reach an approximation of the zero of f_{end}.

This suggests us that the first step should be chosen as $c\|\dot{f}_t\|/\mu^2(f_0,\zeta_0)$.

Condition number and number of homotopy steps

A precedent by Shub & Smale related condition number and complexity. This foundational work was later improved by Shub:

The number of Newton homotopy steps necessary to follow a homotopy path $\Gamma_t = (f_t, \zeta_t),\ 0 \leq t \leq 1$ is bounded above by

$$\textit{Constant}\ d^{3/2} \int_0^1 \mu(f_t, \zeta_t) \|(\dot{f}_t, \dot{\zeta}_t)\|\ dt$$

This is the **length of the curve** (f_t, ζ_t) **in the condition metric, see the talk by Mike Shub.**

Condition number and number of homotopy steps

A precedent by Shub & Smale related condition number and complexity. This foundational work was later improved by Shub:

The number of Newton homotopy steps necessary to follow a homotopy path $\Gamma_t = (f_t, \zeta_t),\ 0 \leq t \leq 1$ is bounded above by

$$\textit{Constant}\ d^{3/2} \int_0^1 \mu(f_t, \zeta_t) \|(\dot{f}_t, \dot{\zeta}_t)\|\ dt$$

This is the **length of the curve** (f_t, ζ_t) **in the condition metric, see the talk by Mike Shub.**
Thus, the condition number μ not only controls stability: it also bounds complexity!

Condition number and number of homotopy steps

A precedent by Shub & Smale related condition number and complexity. This foundational work was later improved by Shub:

The number of Newton homotopy steps necessary to follow a homotopy path $\Gamma_t = (f_t, \zeta_t)$, $0 \leq t \leq 1$ is bounded above by

$$\textit{Constant } d^{3/2} \int_0^1 \mu(f_t, \zeta_t) \|(\dot{f}_t, \dot{\zeta}_t)\| \, dt$$

This is the **length of the curve** (f_t, ζ_t) **in the condition metric, see the talk by Mike Shub.**
Thus, the condition number μ not only controls stability: it also bounds complexity!
The previous slice may have suggested $\int_0^1 \mu^2(f_t, \zeta_t) \|\dot{f}_t\| \, dt$. This is also valid, but is less precise because $\|\dot{\zeta}_t\| \leq \mu(f_t, \zeta_t) \|\dot{f}_t\|$.

Condition number and number of homotopy steps

- Constructive versions of Shub's theoretical result by [B.,Cucker & Bürgisser, Dedieu–Malajovich–Shub] (assuming exact arithmetic computations).

Condition number and number of homotopy steps

- Constructive versions of Shub's theoretical result by [B.,Cucker & Bürgisser, Dedieu–Malajovich–Shub] (assuming exact arithmetic computations).
- An algorithm performing this task (floating point, double precission) [B., Leykin] implemented is included in the Numerical Algebraic Geometry package of Macaulay 2, NAG4M2.

Condition number and number of homotopy steps

- Constructive versions of Shub's theoretical result by [B.,Cucker & Bürgisser, Dedieu–Malajovich–Shub] (assuming exact arithmetic computations).
- An algorithm performing this task (floating point, double precission) [B., Leykin] implemented is included in the Numerical Algebraic Geometry package of Macaulay 2, NAG4M2.
- Using only rational arithmetic, in case the input systems and initial zero have rational coordinates, the homotopy method can be carried out in rational arithmetic in running time:
  - Linear in $\int_0^1 \mu(f_t, \zeta_t)\|(\dot{f}_t, \dot{\zeta}_t)\|\, dt$

Condition number and number of homotopy steps

- Constructive versions of Shub's theoretical result by [B.,Cucker & Bürgisser, Dedieu–Malajovich–Shub] (assuming exact arithmetic computations).
- An algorithm performing this task (floating point, double precission) [B., Leykin] implemented is included in the Numerical Algebraic Geometry package of Macaulay 2, NAG4M2.
- Using only rational arithmetic, in case the input systems and initial zero have rational coordinates, the homotopy method can be carried out in rational arithmetic in running time:
  - Linear in $\int_0^1 \mu(f_t, \zeta_t)\|(\dot{f}_t, \dot{\zeta}_t)\|\, dt$
  - Polynomial in the dimension of $\mathcal{H}_{(d)}$ and the size of rationals,
  - Polynomial in the logarithm of the max. of μ along the path.

  (B. & Leykin). This algorithm is also implemented in Macaulay 2, NAG4M2. Previous work by Malajovich.

Thank you for your attention

If we're still in the second class... then we're done for now!

Let us recall our last claim

Assume that we know a zero ζ_0 of some system f_0. The complexity (number of arithmetic operations) of following a straight–line path $f_t = (1-t)f_0 + tf_1$ for finding a zero ζ_1 of f_1 is at most a small quantity (polynomial in the size of the input) times

$$\int_0^1 \mu(f_t, \zeta_t) \|(\dot{f}, \dot{\zeta})\| \, dt,$$

and as far as $\mu(f_t, \zeta_t) < \infty$ for $t \in [0, 1]$ the homotopy algorithm *always gives an answer*.

Let us recall our last claim

Assume that we know a zero ζ_0 of some system f_0. The complexity (number of arithmetic operations) of following a straight–line path $f_t = (1-t)f_0 + tf_1$ for finding a zero ζ_1 of f_1 is at most a small quantity (polynomial in the size of the input) times

$$\int_0^1 \mu(f_t, \zeta_t)\|(\dot{f}, \dot{\zeta})\|\, dt,$$

and as far as $\mu(f_t, \zeta_t) < \infty$ for $t \in [0, 1]$ the homotopy algorithm *always gives an answer*.
Recall Smale's 17th problem: can a zero of n complex polynomial equations in n unknowns be found approximately, on the average, in polynomial time with a uniform algorithm?

Let us recall our last claim

Assume that we know a zero ζ_0 of some system f_0. The complexity (number of arithmetic operations) of following a straight–line path $f_t = (1-t)f_0 + tf_1$ for finding a zero ζ_1 of f_1 is at most a small quantity (polynomial in the size of the input) times

$$\int_0^1 \mu(f_t, \zeta_t)\|(\dot{f}, \dot{\zeta})\|\, dt,$$

and as far as $\mu(f_t, \zeta_t) < \infty$ for $t \in [0, 1]$ the homotopy algorithm *always gives an answer*.

Recall Smale's 17th problem: can a zero of n complex polynomial equations in n unknowns be found approximately, on the average, in polynomial time with a uniform algorithm?

That is, we would like to design a way to choose the initial pair (f_0, ζ_0) in such a way that the *average* value of the integral above is small.

A complexity measure for a initial pair (f_0, ζ_0)

Define the complexity measure:

$$A(f_0, \zeta_0) = \mathrm{E}_{f \in \mathbb{P}(\mathcal{H}_{(d)})} \left[\int_0^1 \mu(f_t, \zeta_t) \|(\dot{f}_t, \dot{\zeta}_t)\| \, dt \right].$$

We say that (f_0, ζ_0) is a **good starting pair** for the homotopy if $A(f_0, \zeta_0)$ is "small".

A complexity measure for a initial pair (f_0, ζ_0)

Define the complexity measure:

$$A(f_0, \zeta_0) = \mathrm{E}_{f \in \mathbb{P}(\mathcal{H}_{(d)})} \left[\int_0^1 \mu(f_t, \zeta_t) \|(\dot{f}_t, \dot{\zeta}_t)\| \, dt \right].$$

We say that (f_0, ζ_0) is a **good starting pair** for the homotopy if $A(f_0, \zeta_0)$ is "small".
Note that another option would be to bound

$$\mathrm{E}_{f \in \mathbb{P}(\mathcal{H}_{(d)})} \left[\int_0^1 \mu(f_t, \zeta_t)^2 \, dt \right]$$

A seemingly more complicated integral

Let us take another integral here, that is compute the average value of the function we want to compute:

$$\mathrm{E}_{f_0 \in \mathbb{P}(\mathcal{H}_{(d)})} \left(\sum_{\zeta_0 : f_0(\zeta_0)=0} \mathrm{E}_{f \in \mathbb{P}(\mathcal{H}_{(d)})} \left[\int_0^1 \mu(f_t, \zeta_t)^2 \, dt \right] \right)$$

A seemingly more complicated integral

Let us take another integral here, that is compute the average value of the function we want to compute:

$$\mathrm{E}_{f_0 \in \mathbb{P}(\mathcal{H}_{(d)})} \left(\sum_{\zeta_0 : f_0(\zeta_0)=0} \mathrm{E}_{f \in \mathbb{P}(\mathcal{H}_{(d)})} \left[\int_0^1 \mu(f_t, \zeta_t)^2 \, dt \right] \right)$$

We can use Fubini's theorem: this last integral equals

$$\mathrm{E}_{f_0, f \in \mathbb{P}(\mathcal{H}_{(d)})} \left(\sum_{\zeta_0 : f_0(\zeta_0)=0} \int_0^1 \mu(f_t, \zeta_t)^2 \, dt \right).$$

A seemingly more complicated integral

Let us take another integral here, that is compute the average value of the function we want to compute:

$$\mathrm{E}_{f_0 \in \mathbb{P}(\mathcal{H}_{(d)})} \left(\sum_{\zeta_0 : f_0(\zeta_0)=0} \mathrm{E}_{f \in \mathbb{P}(\mathcal{H}_{(d)})} \left[\int_0^1 \mu(f_t, \zeta_t)^2 \, dt \right] \right)$$

We can use Fubini's theorem: this last integral equals

$$\mathrm{E}_{f_0, f \in \mathbb{P}(\mathcal{H}_{(d)})} \left(\sum_{\zeta_0 : f_0(\zeta_0)=0} \int_0^1 \mu(f_t, \zeta_t)^2 \, dt \right).$$

Because unless $f_t \cap \Sigma \neq \emptyset$, which happens with probability 0, the zeros of f_0 and those of f are in one–to–one correspondence, this last integral equals

$$\mathrm{E}_{f_0, f \in \mathbb{P}(\mathcal{H}_{(d)})} \left(\int_0^1 \sum_{\zeta : f_t(\zeta)=0} \mu(f_t, \zeta)^2 \, dt \right).$$

A typical result from Integral Geometry

We'd like to compute

$$\mathrm{E}_{f_0, f \in \mathbb{P}(\mathcal{H}_{(d)})} \left(\int_0^1 \sum_{\zeta : f_t(\zeta) = 0} \mu(f_t, \zeta)^2 \, dt \right).$$

That is like *the expected value of a function along the points falling in a randomly chosen line*. Does this sound like Santaló?

A typical result from Integral Geometry

We'd like to compute

$$\mathrm{E}_{f_0, f \in \mathbb{P}(\mathcal{H}_{(d)})} \left(\int_0^1 \sum_{\zeta : f_t(\zeta) = 0} \mu(f_t, \zeta)^2 \, dt \right).$$

That is like *the expected value of a function along the points falling in a randomly chosen line*. Does this sound like Santaló? Here is an intuitive fact: there is **no reason why any system $f \in \mathbb{P}(\mathcal{H}_{(d)})$ should weight more than any other for computing this integral**.

A typical result from Integral Geometry

We'd like to compute

$$\mathrm{E}_{f_0, f \in \mathbb{P}(\mathcal{H}_{(d)})} \left(\int_0^1 \sum_{\zeta : f_t(\zeta)=0} \mu(f_t, \zeta)^2 \, dt \right).$$

That is like *the expected value of a function along the points falling in a randomly chosen line*. Does this sound like Santaló? Here is an intuitive fact: there is **no reason why any system $f \in \mathbb{P}(\mathcal{H}_{(d)})$ should weight more than any other for computing this integral**. Thus, the expected value of the function is (up to a constant) the same as the expected value of μ^2 in $\mathbb{P}(\mathcal{H}_{(d)})$. That is:

$$\mathrm{E}_{f \in \mathbb{P}(\mathcal{H}_{(d)})} \left(\mu(f_t, \zeta)^2 \right).$$

Here is the average value of the condition number of linear algebra

It turns out that, because μ is essentially the norm of the inverse of the derivative, this quantity can be computed exactly (bound by Shub & Smale, exact value by B. & Pardo):

$$\mathrm{E}_{f\in\mathbb{P}(\mathcal{H}_{(d)})}\left(\mu(f_t,\zeta)^2\right)=\mathcal{D}N\left(n\left(1+\frac{1}{n}\right)^{n+1}-2n-1\right)\leq nN\mathcal{D},$$

where $\mathcal{D}$ is the product of the degrees and N the dimension of $\mathbb{P}(\mathcal{H}_{(d)})$. This is done by reducing the computation to the linear case:

$$\mathrm{E}_{f\in\mathbb{P}(\mathcal{H}_{(d)})}\left(\mu(f_t,\zeta)^\alpha\right)=\frac{\mathcal{D}\Gamma(N+1)\Gamma(n^2+n-\alpha/2)}{\Gamma(N+1-\alpha/2)\Gamma(n^2+n)}\times$$

$$\mathrm{E}_{M \text{ and } n\times(n+1) \text{ matrix}, \|M\|_F=1}(\kappa(M)^\alpha).$$

Conclusion

Recall we defined

$$A(f_0, \zeta_0) = \mathrm{E}_{f \in \mathbb{P}(\mathcal{H}_{(d)})} \left[\int_0^1 \mu(f_t, \zeta_t) \|(\dot{f}_t, \dot{\zeta}_t)\| \, dt \right].$$

We say that (f_0, ζ_0) is a **good starting pair** for the homotopy if $A(f_0, \zeta_0)$ is "small".
Then, we have proved:

$$\mathrm{E}_{f_0 \in \mathbb{P}(\mathcal{H}_{(d)})} \left(\sum_{\zeta_0 : f_0(\zeta_0) = 0} \mathrm{E}_{f \in \mathbb{P}(\mathcal{H}_{(d)})} [A(f_0, \zeta_0)] \right) nN\mathcal{D}.$$

Conclusion

Recall we defined

$$A(f_0, \zeta_0) = \mathrm{E}_{f \in \mathbb{P}(\mathcal{H}_{(d)})} \left[\int_0^1 \mu(f_t, \zeta_t) \|(\dot{f}_t, \dot{\zeta}_t)\| \, dt \right].$$

We say that (f_0, ζ_0) is a **good starting pair** for the homotopy if $A(f_0, \zeta_0)$ is "small".
Then, we have proved:

$$\mathrm{E}_{f_0 \in \mathbb{P}(\mathcal{H}_{(d)})} \left(\sum_{\zeta_0 : f_0(\zeta_0) = 0} \mathrm{E}_{f \in \mathbb{P}(\mathcal{H}_{(d)})} [A(f_0, \zeta_0)] \right) nN\mathcal{D}.$$

That is to say: **let f_0 be chosen at random, let ζ_0 be a zero of f, chosen at random among the $\mathcal{D}$ zeros of f. Then, the expected value of $A(f_0, \zeta_0)$ is less than nN.**

Conclusion

Recall we defined

$$A(f_0,\zeta_0) = \mathrm{E}_{f\in\mathbb{P}(\mathcal{H}_{(d)})}\left[\int_0^1 \mu(f_t,\zeta_t)\|(\dot{f}_t,\dot{\zeta}_t)\|\,dt\right].$$

We say that (f_0,ζ_0) is a **good starting pair** for the homotopy if $A(f_0,\zeta_0)$ is "small".
Then, we have proved:

$$\mathrm{E}_{f_0\in\mathbb{P}(\mathcal{H}_{(d)})}\left(\sum_{\zeta_0:f_0(\zeta_0)=0}\mathrm{E}_{f\in\mathbb{P}(\mathcal{H}_{(d)})}[A(f_0,\zeta_0)]\right) nN\mathcal{D}.$$

That is to say: **let f_0 be chosen at random, let ζ_0 be a zero of f, chosen at random among the $\mathcal{D}$ zeros of f. Then, the expected value of $A(f_0,\zeta_0)$ is less than nN.** In particular, such a randomly chosen pair is a good initial pair.

A probabilistic problem

- This does not yield an algorithm for we must solve a random f_0. And this is precisely our goal!
- But we have turned the complexity problem into a probabilistic problem: Generate (algorithmically) a random pair system-solution (f_0, ζ_0).

Naive strategy

- GOAL: Let f_0 be chosen at random. Then, find all of the solutions of f_0 and choose one at random.

Naive strategy

- GOAL: Let f_0 be chosen at random. Then, find all of the solutions of f_0 and choose one at random.
- INSTEAD: Let ζ_0 be chosen at random. Then, choose a random f_0 such that $f_0(\zeta_0) = 0$.
  - This can be done: For fixed ζ_0, the set $\{f_0 : f_0(\zeta_0) = 0\}$ is a vector space.
  - Unfortunately, the probability distribution is not the same!

Generate a random pair (f_0, ζ_0)

[B., Pardo] The following is the correct way to do this:

- Choose a random matrix M, with n rows and $n+1$ columns.
- Solve M, call ζ_0 the solution.
- Construct a random system with linear part equal to M and solution ζ_0.

This yields an Average Las Vegas procedure to solve: Input f_1, choose random (f_0, ζ_0) and follow that homotopy. Total complexity is $\tilde{O}(N^2)$.

A “random” choice of starting pair is good

Summary

[Shub & Smale] Foundational results for an average complexity analysis.

A "random" choice of starting pair is good

Summary

[Shub & Smale] Foundational results for an average complexity analysis.

[B. & Pardo] A randomly chosen initial pair *w.r.t. a particular probability distribution* is indeed a good starting point: expected number of homotopy steps is $O(nN)$. A Las Vegas algorithm for Smale's 17th problem.

A "random" choice of starting pair is good

Summary

[Shub & Smale] Foundational results for an average complexity analysis.

[B. & Pardo] A randomly chosen initial pair *w.r.t. a particular probability distribution* is indeed a good starting point: expected number of homotopy steps is $O(nN)$. A Las Vegas algorithm for Smale's 17th problem.

[B. & Shub] Not only the expected complexity is polynomial in the size of the input, also the variance and other higher moments.

What about a deterministic algorithm?

Still an open problem, but it may be just around the corner!

It is natural to demand a deterministic algorithm, i.e. an algorithm which does not need to invoke random choices. The most promising approach is simply finding some (collection, for every n and list of degrees of) (f_0, ζ_0) such that $\mathcal{A}(f_0, \zeta_0) \leq p(N)$, for some fixed polynomial p.

In 1994, Shub & Smale conjectured that the following pair satisfies this claim:

$$f_{good} = \begin{cases} d_1^{1/2} x_0^{d_1-1} x_1 = 0 \\ \vdots \\ d_n^{1/2} x_0^{d_n-1} x_n = 0 \end{cases}, \qquad \zeta_{good} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Note that f_{good} is just a homogeneization of the identity.

What about a deterministic algorithm?

Still an open problem, but it may be just around the corner!

It is natural to demand a deterministic algorithm, i.e. an algorithm which does not need to invoke random choices. The most promising approach is simply finding some (collection, for every n and list of degrees of) (f_0, ζ_0) such that $\mathcal{A}(f_0, \zeta_0) \leq p(N)$, for some fixed polynomial p.

In 2010 Burgïser & Cucker proved that for any pair (f_0, ζ_0) we have

$$\mathcal{A}(f_0, \zeta_0) \leq p(N) \cdot q(\max_{\zeta : f_0(\zeta)=0} \mu(f_0, \zeta)),$$

p and q polynomials. Unfortunately, no f_0 is known such that $\max_{\zeta : f_0(\zeta)=0} \mu(f_0, \zeta)$ is small, so this does not give a polynomial time algorithm...

What about a deterministic algorithm?

Still an open problem, but it may be just around the corner!

It is natural to demand a deterministic algorithm, i.e. an algorithm which does not need to invoke random choices. The most promising approach is simply finding some (collection, for every n and list of degrees of) (f_0, ζ_0) such that $\mathcal{A}(f_0, \zeta_0) \leq p(N)$, for some fixed polynomial p.

But it turns out that the system whose zeros are the roots of unity:

$$f_{total} = \begin{cases} x_0^{d_1} - x_1^{d_1} = 0 \\ \vdots \\ x_0^{d_n} - x_n^{d_n} = 0 \end{cases}, \qquad \zeta_{total} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix},$$

satisfies $\max_{\zeta : f_0(\zeta)=0} \mu(f_0, \zeta) \leq 2(n+1)^d$, where d is the maximum of $d_1, \ldots, d_n$.

What about a deterministic algorithm?

Still an open problem, but it may be just around the corner!

It is natural to demand a deterministic algorithm, i.e. an algorithm which does not need to invoke random choices. The most promising approach is simply finding some (collection, for every n and list of degrees of) (f_0, ζ_0) such that $\mathcal{A}(f_0, \zeta_0) \leq p(N)$, for some fixed polynomial p.

So, for "small" values of the degrees the pair $(f_{total}, \zeta_{total})$ is a good starting pair...

What about a deterministic algorithm?

Still an open problem, but it may be just around the corner!

It is natural to demand a deterministic algorithm, i.e. an algorithm which does not need to invoke random choices. The most promising approach is simply finding some (collection, for every n and list of degrees of) (f_0, ζ_0) such that $\mathcal{A}(f_0, \zeta_0) \leq p(N)$, for some fixed polynomial p.

So, for "small" values of the degrees the pair $(f_{total}, \zeta_{total})$ is a good starting pair... And it turns out that for high degrees $d_1, \ldots, d_n$, a symbolic-numeric algorithm designed by James Renegar in the 80's gives a polynomial time procedure! The combination of the homotopy with starting pair $(f_{total}, \zeta_{total})$ and Renegar's algorithm yields average complexity which can be bounded above by

$$N^{O(\log\log(N))}, \text{ that is almost polynomial.}$$

What about a deterministic algorithm?

Still an open problem, but it may be just around the corner!

It is natural to demand a deterministic algorithm, i.e. an algorithm which does not need to invoke random choices. The most promising approach is simply finding some (collection, for every n and list of degrees of) (f_0, ζ_0) such that $\mathcal{A}(f_0, \zeta_0) \leq p(N)$, for some fixed polynomial p.

So, the state of the art in Smale's 17th problem is:

- Solved using a Las–Vegas algorithm, quadratic running time $\tilde{O}(N^2)$.
- No deterministic algorithm is known working in polynomial time, but one working in $N^{O(\log\log(N))}$ exists.

Table with average number of homotopy steps

[B., Leykin]

system	#sol.	#steps/path (C)	#steps/path (H)
$Random_{(2,2)}$	4	198.5	31
$Random_{(2,2,2)}$	8	370.125	23
$Random_{(2,2,2,2)}$	16	813.812	44.375
$Random_{(2,2,2,2,2)}$	32	1542.5	48.5312
$Random_{(2,2,2,2,2,2)}$	64	2211.58	58.5312
$Katsura_3$	4	569.5	25.75
$Katsura_4$	8	1149.88	41.5
$Katsura_5$	16	1498.38	39.0625
$Katsura_6$	32	2361.81	55.5625

Table with average number of homotopy steps

good: homogeneization of the identity [Shub & Smale]; *random*: random pair [B.& Pardo];*total*: usual total homotopy (roots of unity) [Burgïser & Cucker]

[B., Leykin]

Generate 1000 random degree 2 systems for $n = 4, 5, 6, 7, 8$ and measure average running time.

n	4	5	6	7	8
E_{good}	634.674	1001.25	1452.57	2007.84	2622.45
$\#fail_{good}$	3	3	12	10	22
E_{total}	825.927	1373.76	2028.24	2832.46	3966.77
$\#fail_{total}$	1	3	5	13	16
E_{rand}	1075.58	1777.03	2603.78	3714.34	5013.25
$\#fail_{rand}$	2	1	7	16	26

Note that the third row is the only one with proven polynomial running time! Yet, the two other ones are in this experiment slightly faster, more or less as:

$$E_{good} \leq E_{total} \leq E_{rand} \leq 2E_{good}.$$

Theorem (Ehresmann 1951)

Let X, Y be smooth manifolds with Y connected. Let $U \subseteq X$ be a nonempty open subset of X, and let $\pi : U \to Y$ satisfy:

Theorem (Ehresmann 1951)

Let X, Y be smooth manifolds with Y connected. Let $U \subseteq X$ be a nonempty open subset of X, and let $\pi : U \to Y$ satisfy:

- *π is a submersion.*
- *π is proper, i.e. $\pi^{-1}(compact) = compact$.*

Then, $\pi : U \to Y$ is a fiber bundle. In particular, it is surjective.

Theorem (Ehresmann 1951)

Let X, Y be smooth manifolds with Y connected. Let $U \subseteq X$ be a nonempty open subset of X, and let $\pi : U \to Y$ satisfy:

- *π is a submersion.*
- *π is proper, i.e. $\pi^{-1}(compact) = compact$.*

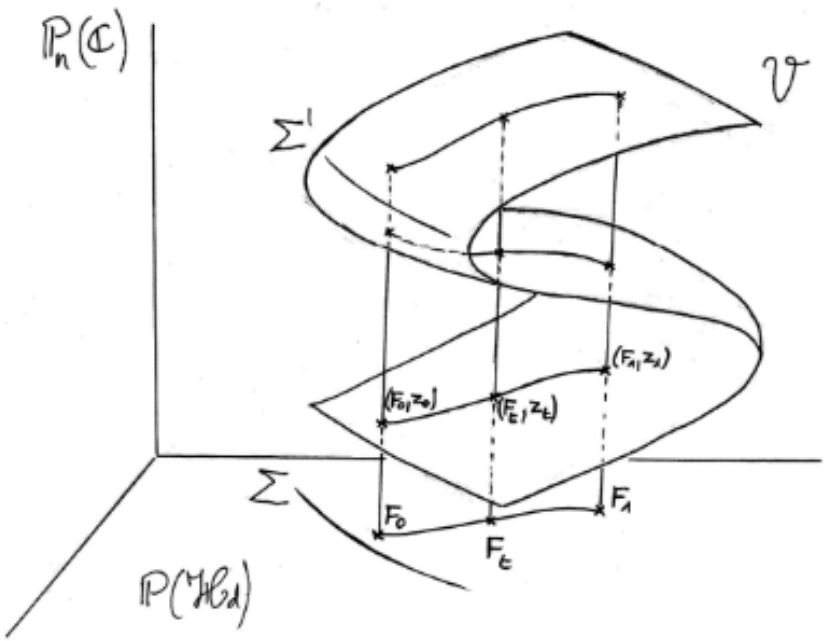
Then, $\pi : U \to Y$ is a fiber bundle. In particular, it is surjective.

Corollary

If additionally we assume $\dim(X) = \dim(Y)$ *then π is a covering map. In particular, the number of preimages of every $y \in Y$ is finite and constant.*

A proof of the Fundamental Theorem of Algebra

Let us see if for J.P.'s sketch: $\pi_1 : U \to \mathbb{P}(\mathcal{H}_{(d)}) \setminus \Sigma$



It is easy to see that, if we remove Σ from $\mathbb{P}(\mathcal{H}_{(d)})$ and let $U = \pi_1^{-1}(\Sigma)$, we are under the conditions of Ehresmann's theorem! In particular, every $f \in \mathbb{P}(\mathcal{H}_{(d)})$ with no singular zeros has the same number of zeros, equal to $\mathcal{D} = d_1, \cdots d_n$. By continuity and compactness, **every** $f \in \mathbb{P}(\mathcal{H}_{(d)})$ **has at least one zero**.

That's it!

Thanks for your attention