Congratulations Dennis

On the Convexity of the Condition Number in the Condition Metric

Michael Shub, CONICET, IMAS, Universidad de Buenos Aires and CUNY Graduate School

Joint work with Carlos Beltrán, Jean Pierre Dedieu, Gregorio Malajovich, Steve Smale.

June 2, 2011

Ingredients of a Complexity Theory

- A problem and a notion of solution. The problem should have infinitely many problem instances as potential inputs.
- A notion of input size of a problem instance.
- A notion of cost of computation.

Complexity theory measures the cost of finding a solution for a problem instance in terms of the input size. The class of problems **P** are those problems for which there is an algorithm which solves the problem in polynomial cost.

Ingredients of a Complexity Theory

- A problem and a notion of solution. The problem should have infinitely many problem instances as potential inputs.
- A notion of input size of a problem instance.
- A notion of cost of computation.

Complexity theory measures the cost of finding a solution for a problem instance in terms of the input size. The class of problems **P** are those problems for which there is an algorithm which solves the problem in polynomial cost.
Examples:

- 3-Sat: Is a Boolean expression in 3-Conjunctive Normal Form Satisfiable?
- Hilbert's Nullstellensatz: Does a system of n-quadratic equations in n-complex (resp. real) unknowns have a solution (resp. real solution)?

3-Sat and Hilbert's Nullstellensatz are NP-complete problems, but in different contexts. The notions of input size, cost and algorithm are different.

- 3-Sat is considered for classical Turing Machines and the problem is 3-Sat in **P** is Cook's famous problem Does **P** $=$ **NP?** The input size is the bit length and the cost the bit operations (or time on the Turing machine).
- Hilbert's Nullstellensatz is considered for BSS-Machines over the complex numbers i.e. Complex Turing machines, and the question of whether Hilbert's Nullstellensatz is in **P** is equivalent to **P** $=$ **NP?** over $\mathbb{C}$. The input size is the dimension and the cost the number of arithmetic operations and comparisons (or the time on a BSS-machine).

One of our (Blum,Shub,Smale) goals in 1989 was to fashion a theory of complexity which addresses the problems of numerical analysis and scientific computing. Ultimately concerns of error (computational and input) have to be taken into account. This task is till very much in progress. Our model problem was to find the zeros of systems of polynomial equations defined over the complex numbers. We deal with with machines over the real numbers. A great deal of progress has been made here. **Beltran-Pardo** and **Buergisser-Cucker** in particular. Geometry as I will discuss today should paly an important role.

Basic notations

Let $f = (f_1, \ldots, f_n)$ be a system of homogeneous polynomial equations with unknowns $X_0, \ldots, X_n$ and degrees $d_1, \ldots, d_n$. Denote by $\mathcal{H}_{(d)}$ the vector space of such systems, and by

$$V = \{(f, \zeta) \in \mathbb{P}(\mathcal{H}_{(d)}) \times \mathbb{P}(\mathbb{C}^{n+1}) : f(\zeta) = 0\}$$

the solution variety. Let

$$W = \{(f, \zeta) \in V : \; Df(\zeta) \text{ is of maximal rank}\},$$

and

$$\mu(f, \zeta) = \|f\| \left\| \left(Df(\zeta) \mid_{\zeta^\perp} \right)^{-1} Diag\left(d_i^{1/2} \|\zeta\|^{d_i - 1} \right) \right\|$$

be the condition number, defined for $(f, \zeta) \in W$.

Homotopy method

- Let f_1 be a system you want to solve, and let f_0 be a system you can solve.
- Construct a path of systems f_t joining f_0 and f_1.
- Choose some solution ζ_0 of f_0. Let $z_0 = \zeta_0$
- Choose a small step size t_0. Apply Newton's projective method

$$z_1 = N_{f_{t_0}}(z_0) = z_0 - (Df_{t_0}\mid_{z_0^\perp})^{-1} f(z_0)$$

- Continue the process until you are close to f_1. Generate $z_2, z_3, \ldots$.
- Output the last value z_j.

Smale's 17th Problem

Can an approximate root of of a polynomial system be found in average polynomial time in the input size?
Recent Progress by Beltran-Pardo and Buergisser-Cucker.
Homotopy methods play a big role.

Condition number and number of homotopy steps

[S.]
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 by

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

that is the length of the path Γ_t in the condition metric.

Choice of (f_0, ζ_0)

In order to have a specific alorithm we now take the simplest paths possible. Let (f_0, ζ_0) be a known pair of system-solution. For any system f_1, define the path

$$f_t = (1-t)f_0 + tf_1.$$

Then, define the complexity measure:

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

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

Choice of (f_0, ζ_0)

[Beltrán & Pardo]
A randomly chosen initial pair is indeed a good starting point.
That is,

$$\mathrm{E}_{g \ \ a \ \ system}\left[\frac{1}{\mathcal{D}}\sum_{\zeta : g(\zeta)=0} A(g,\zeta)\right] \leq 16\pi n N,$$

where N is the number of monomials of a generic system and $\mathcal{D} = d_1 \cdots d_n$ is the number of solutions of a generic system.

The result of Buergisser-Cucker

Buergisser and Cucker almost solve Smale's problem, but at higher average cost. A particular feature of their approach is to produce a deterministic starting point which for bounded degrees answers Smale's question positively. For large degrees (compared to n) symbolic methods already do the job.

In particular:

There is a uniform algorithm which which finds an approximate zero of a system of n homogeneous quadratic equations in $n+1$ unknowns in polynomial time on the average.

Choice of (f_0, ζ_0)

Three ways to choose the initial pair:
1) Choose (f_0, ζ_0) at random, which guarantees average number of Newton steps $O(nN)$.
2) Use the "most simple" ie best conditioned (system,root) pair:

$$g = \begin{cases} d_1^{\frac{1}{2}} X_0^{d_1 - 1} X_1 = 0, \\ \cdots \\ d_n^{\frac{1}{2}} X_0^{d_n - 1} X_n = 0, \end{cases} \qquad e_0 = (1, 0, \ldots, 0)$$

Conjectured by [S. & Smale] to satisfy $A(g, e_0) \leq$ "Small".
3)

$$h = \begin{cases} X_0^{d_1} - X_1^{d_1} = 0, \\ \cdots \\ X_0^{d_n} - X_n^{d_n} = 0, \end{cases} \qquad e_0 = (1, \zeta_1, \ldots, \zeta_n)$$

Where ζ_i is any one of the d_i roots of unity. (Buergisser-Cucker)

Back to the condition metric.

How well can homotopy methods do?
[Beltrán & S.]
The distance in the condition metric from the (g, e_0) to any system (f, ζ) is bounded by $O(nN \log \mu(f, \zeta))$. The average number of steps following geodesics for the condition number, is at most

$$O(n^2 \log(N)).$$

Thus, much faster average than the linear homotopy estimate $O(nN)$.

What are the geodesics like? μ is comparable to the distance in V to the degenerate (system,root) pairs. Is the condition number maximized at the endpoints? (Quasi-convexity) or even:
Consider W with the condition metric. Let γ be a geodesic. Is the function

$$t \mapsto \log \mu(\gamma(t))$$

convex? We shall say "μ is a self-convex function in W".

[Beltrán &Dedieu &Malajovich &S.]

Convexity aspects of μ

Let $\mathbb{GL}_{m,n}$ be linear space of m by n matrices with the condition metric (here, the condition number of a matrix A is $||A^{\dagger}||$). Then, the answer to the question above is Yes: $||A^{\dagger}||$ is self-convex in $\mathbb{GL}_{m,n}$.

The same is true for the condition number $\kappa(A) = ||A||_F||A^{\dagger}||$ in the projective set of matrices $\mathbb{P}(\mathbb{GL}_{m,n})$.

The same is true in the solution variety for the linear case, i.e. $W = \{(A, \zeta) \in \mathbb{P}(\mathbb{GL}_{n,n+1}) \times \mathbb{P}(\mathbb{C}^{n+1}) | A(\zeta) = 0\}$.

Is the condition number self convex on the solution variety W in the non-linear case?

Let $M, <,>_x$ be a smooth Riemannian Manifold of Class C^2. Let $\alpha : M \to \mathbb{R}_+$ and consider the new Riemann Structure $<,>_{\alpha,x}= \alpha(x)^2 <,>_x$ We say α is self-convex if $\log(\alpha(\gamma(t)))$ is convex for all geodesics $\gamma(t)$ in the new metric (Parameterized by arc length).

Proposition. If α is C^2 self-convexity is equivalent to

$$2\alpha(x)D^2\alpha(x)(v,v)+ \parallel D\alpha(x) \parallel_x \parallel v \parallel_x -4(D\alpha(x)v)^2 \geq 0$$

for all $(x, v) \in TM$.

Self-convexity for the distance function to a C^2 submanifold of $\mathbb{R}^j$

Let $N \subset \mathbb{R}^j$ be a C^2 submanifold without boundary. Let us denote by

$$\rho(x) = d(x, N) = \min_{y \in N} \|x - y\| \quad \text{and } \alpha(x) = \frac{1}{\rho(x)}.$$

Let $\mathcal{U}$ be the largest open set in $\mathbb{R}^j$ such that, for any $x \in \mathcal{U}$, there is a unique closest point in to x. When $\mathcal{U}$ is equipped with the new metric $\alpha(x)^2 \langle ., . \rangle$ we have the following theorem.

Theorem

The function $\alpha : \mathcal{U} \setminus N \to \mathbb{R}$ *is self-convex.*

What about other Riemannnian manifolds?
Good. But the functions we are interested in are only Lipschitz and there are points in the complement of $\mathcal{U}$.

Properties of Lipschitz (conformal) metrics

- Geodesics are $C^{1,1}$ and satisfy differential inclusions.
- Initial value problem may not have unique solutions.
- When are geodesics between two points locally unique? (in general no.)
- If the Lipschitz conformal scaling function is a) Clarke regular and b) C^2 and self-convex when restricted to a countable union of submanifolds whose union is M and c) the second divided differences do not take on the value $-\infty$ then it is self-convex on M. (Piecing together).
- Behaviour when there is a group of symmetries. (Relations with moment maps).