A Bayesian Model for Root Computation

Josef Schicho, RICAM

joint work with Hanna K. Pikkarainen, RICAM

Approximate Algebraic Computation

Observation: choose any computational problem in algebra that involves real or complex numbers. Then there are good chances that the solution of the problem is *infinitely sensitive* to perturbations: no matter how small the perturbation is, the answer might be qualitatively different.

Easy Example

The GCD of the polynomials

$$F = x^2 - 2x + 1, G = x^2 - 1$$

is $x - 1$.

The GCD of F and $G + \epsilon$ is 1, for all $\epsilon > 0$.

Why Bother?

Depending on the situation, there are (at least) two reasons to prefer accepting small errors to exact computations:

Problem immanent: either the input is not known exactly, or the output is not needed exactly.

Compution costs: the exact answer may be too big or too hard to compute.

Probably Approximately Correct Algorithms

In the presence of discontinuities (very common in algebraic problems), it cannot be ruled out that small errors may lead to completely wrong results (i.e. division by a quantity that is mistaken to be different from zero).

A "PAC algorithm" is one for which this happens only for few instances. In order to make this precise, one has to introduce a probability distribution over the input instances.

Polynomial Root Finding

In this talk, we are concerned with the following algebraic problem involving real or complex numbers:

Given: a normed polynomial $P \in \mathbb{R}[x]$ or $\mathbb{C}[x]$.

Find: all roots of P.

We are especially interested in structural information, meaning the multiplicity and reality of the roots.

Motivation

Example 1: To compute the singularities of a plane algebraic curve given by $F(x,y) = 0$, one might compute the discriminant $\Delta(x)$ of F with respect to y. Then the x-coordinates of the singularities are among the multiple roots of Δ.

Example 2: If the multiplicity pattern is known, then the computation of the numerical value is numerically robust, and we have efficient and stable algorithms to compute them (Zeng, Math. Comp. 2005).

A Trivial Solution

(complex case, we are only interested in multiplicity but not in reality)

Ignore the input, and return that there are only simple roots.

Whether this algorithm is PAC or not, depends on the chosen distribution of inputs. If we fix the degree and choose a uniform distribution of the coefficient vector in some cube, then the algorithm is PAC.

Bayesian Inference aka Bayesian Inversion

An **Inverse Problem:** for a fixed function $F : A \to B$, we want to compute a when $b = F(a)$ is given approximately.

Prior Distribution: we fix a distribution on a (i.e. on the possible outputs). This will be discussed later.

Likelihood: for a fixed $a \in A$, we fix a distribution of the input $b \in B$. A very common choice is **additive noise**: $b = F(a) + \epsilon e$, where $\epsilon > 0$ is fixed (error level) and e is a Gaussian random variable.

The Bayesian Formula

For given input b, one computes not one output but a distribution of outputs

$$\text{Posterior}_b(a) = \frac{\text{Prior}(a) \cdot \text{Likelihood}_a(b)}{\sum_{a \in A} (\text{Prior}(a) \cdot \text{Likelihood}_a(b))}$$

One may then produce for b various point estimates and spread estimates, such as the **conditional mean** and the **conditional variance**.

The Sum in the Denominator

.. was, of course, an oversimplification. If A and B are infinite, then both the prior and the likelihood are probability measures and the posterior is also a probability measure that is computed by integration. One has to show that the integral exists “almost always”.

The Choice of Prior Distributions

Ideally, one would like to choose a prior which does not give information on the output. A non-example is the uniform distribution in the problem of multiplicity computation: the set of polynomials with multiple roots has measure zero, hence the almost certain answer is: "ONLY SIMPLE ROOTS".

We recognize the trivial algorithm from slide 7.

Root Computation as an Inverse Problem

To see "root finding" as in inverse problem, we fix a degree d and set $A := \mathbb{C}^d/S_d$, a set of roots with multiplicities, and $B \subset \mathbb{C}[z]$ as the set of all normed polynomials of degree d. The function F is defined by

$$F : A \to B, [(a_1, \ldots, a_d)] \mapsto (z - a_1) \ldots (z - a_d).$$

As the likelihood, we choose additive noise.

Our Choice of the Prior

We give each multiplicity/reality pattern equal probability. Within a fixed structural pattern, we assume a fixed upper bound M for the absolute value of the roots and uniform distribution in this polydisk.

Example 1: A Cubic Polynomial

Let $P(z) = z^3 - az - b$ with $a, b \in \mathbb{R}$. The possible patterns are:

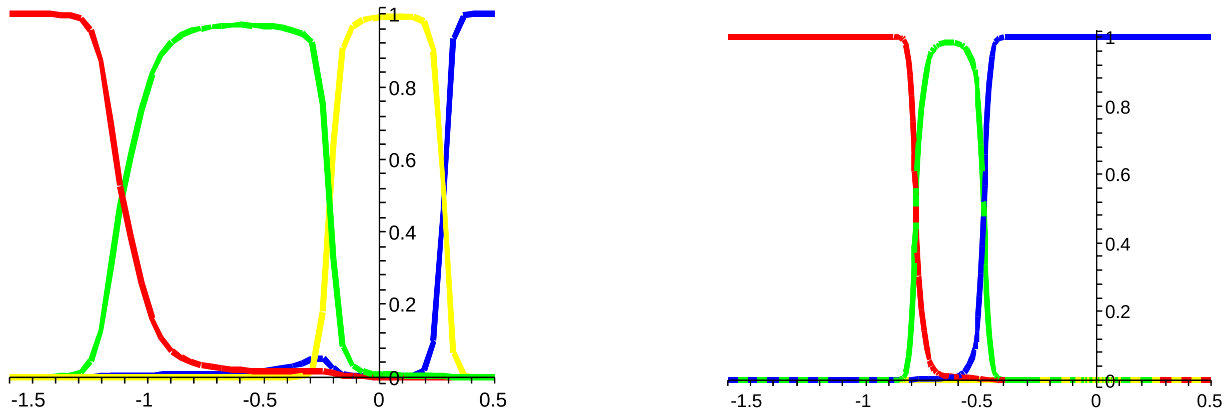
$4a^3 - 27b^2 > 0$: three real single roots,

$4a^3 - 27b^2 = 0, a \neq 0$: a real double root and a real single root,

$a = b = 0$: a real triple root, and

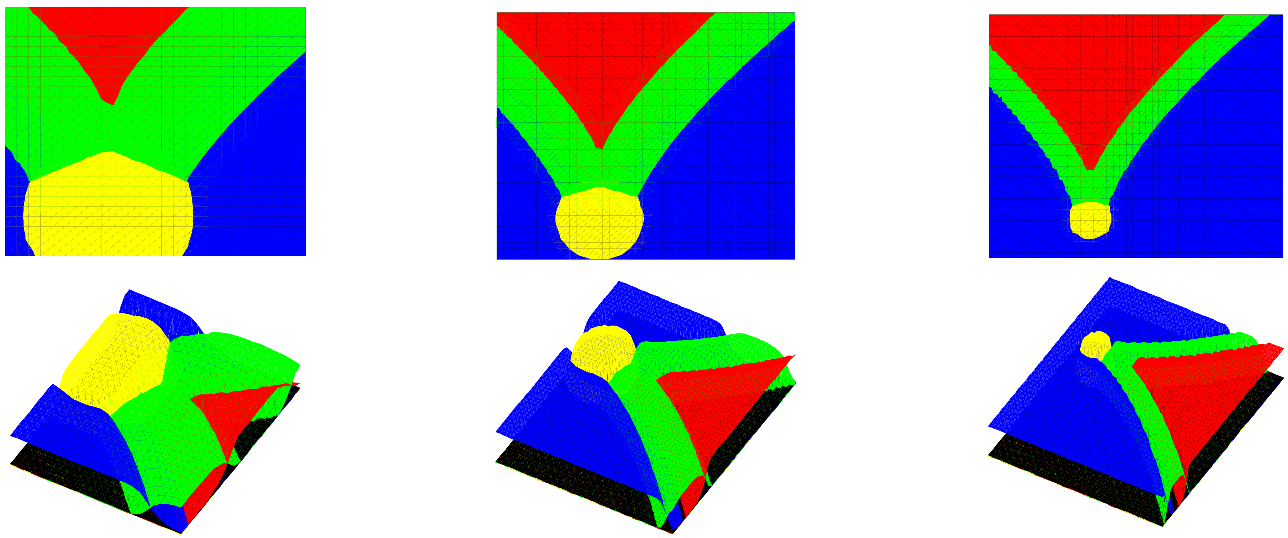
$4a^3 - 27b^2 < 0$: a complex conjugate pair and a real single root.

A Close Miss



Here are the posterior probabilities of different multiplicity patterns for $b = 0.2$ and upper bound $M = 25$, when $\epsilon = 0.1$ (left) and $\epsilon = 0.05$ (right): three real roots, a double root, a triple root, and two complex roots.

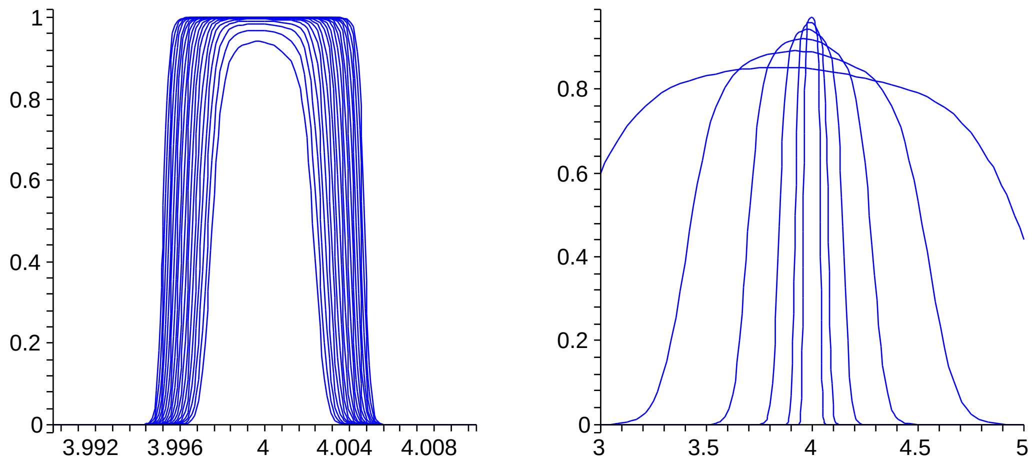
The Most Probably Pattern



Here are the posterior probabilities of different multiplicity $(a, b) \in [-1, 4] \times [-4, 2]$ and $\epsilon = 0.5$ (left), $\epsilon = 0.25$ (middle) and $\epsilon = 0.125$ (right): three real roots, a double root, a triple root, and two complex roots.

The Effects of the Parameters

Let $P(z) = z^2 + az + 4$, where $a \in \mathbb{R}$. It has a double root for $a = 4$.
Here is the posterior probability of a double root:



Left: $\epsilon = 10^{-3}$ and $M = 2^k$ for $k = 1, \ldots, 20$.
Right: $M = 2^4$ and $\epsilon = 2^{-k}$ for $k = 1, \ldots, 6$.

An Example from Zeng

Let $P_a(z) = (z-1+a)^{20}(z-1)^{20}(z+0.5)^5$, where $a = 10^{-l},\ l = 1, 2, 3, \ldots$. It has two nearby roots of multiplicity 20 and one distant root of multiplicity 5.

The number of possible multiplicity patterns of a univariate polynomial of degree 45 is 1353106. We do not test all patterns because this would be too time consuming.

The most interesting patterns are $(5, 20, 20)$ and $(5, 40)$.

Probabilities Confirm Numerical Observations

	$\epsilon = 10^{-5}$	$\epsilon = 10^{-6}$	$\epsilon = 10^{-7}$	$\epsilon = 10^{-8}$
$a = 10^{-3}$	$\mu_2 \approx 10^{-226}$	$\mu_2 < 10^{-999}$	$\mu_2 < 10^{-999}$	$\mu_2 < 10^{-999}$
$a = 10^{-4}$	$\mu_1 \approx 10^{-3}$	$\mu_1 = 0.068$	$\mu_2 \approx 10^{-225}$	$\mu_2 < 10^{-999}$
$a = 10^{-5}$	$\mu_1 \approx 10^{-3}$	$\mu_1 \approx 10^{-3}$	$\mu_1 \approx 10^{-4}$	$\mu_1 = 0.0072$
$a = 10^{-6}$	$\mu_1 \approx 10^{-3}$	$\mu_1 \approx 10^{-3}$	$\mu_1 \approx 10^{-4}$	$\mu_1 \approx 10^{-4}$
$a = 10^{-7}$	$\mu_1 \approx 10^{-3}$	$\mu_1 \approx 10^{-3}$	$\mu_1 \approx 10^{-4}$	$\mu_1 \approx 10^{-4}$

Here are the posterior probabilities of the multiplicity patterns $(5, 20, 20)$ and $(5, 40)$ for the polynomial $P_a(z)$, for different values of a and ϵ. Only the smaller value of $\mu_1 = \mu_{\text{post}}(5, 20, 20)$ and $\mu_2 = \mu_{\text{post}}(5, 40)$ is written since $\mu_1 = 1 - \mu_2$.

Conclusion

It is certainly possible to do reasonable computations on algebraic problems using approximate data (see also the vast literature on numerical algebraic computation, books by Stetter, Sommese/Wampler).

Often, the computation depends on good estimates of qualitative structure of the result (or of intermediate results). Within a fixed structure, the computations are often numerically stable, even if the whole problem is numerically ill-posed.

Bayesian inference can be used to to compute these estimates and to give justification for them.