Solving Trinomials Quickly over $\mathbb{R}$

Erick Boniface

Texas A&M University

July 27, 2021

Outline

Big Picture

- We want to *solve* systems of polynomial equations *quickly*.

Big Picture

- We want to *solve* systems of polynomial equations *quickly*.
- This is important problem that arises in numerous scientific and engineering applications.

Big Picture

- We want to *solve* systems of polynomial equations *quickly*.
- This is important problem that arises in numerous scientific and engineering applications.
- But in order to solve the multivariate case with several polynomials, we should at least be able to settle the univariate case.

Big Picture

- We want to *solve* systems of polynomial equations *quickly*.
- This is important problem that arises in numerous scientific and engineering applications.
- But in order to solve the multivariate case with several polynomials, we should at least be able to settle the univariate case.
- This research settles the trinomial case.

Solve?

What do we mean by *solving*?

Solve?

What do we mean by *solving*?

Definition (Approximate Root ([2]))

Let f be a polynomial with $f(\zeta) = 0$. We say z is an *approximate root* of f provided that the sequence given by $z_0 = z$ and $z_{i+1} = z_i - f(z_i)/f'(z_i)$ for all $i \in \mathbb{N}$ satisfies

$$|z_i - \zeta| \leq \left(\frac{1}{2}\right)^{2^i - 1} |z - \zeta|.$$

We call ζ the *associated root*.

This notion provides an efficient encoding of an approximation that can be quickly tuned to any desired accuracy.

Quickly?

If our algorithm takes I bit operations, we want $I \leq Cs^n$ where C and n are positive constants, and s is the "input size" of our polynomial. In other words, we want to find a $O(s^n)$ algorithm.

Definition

Let $f(x) = \sum_{i=1}^{t} c_i x^{a_i}$. We define the *size* of our polynomial as the sum $\sum_{i=1}^{t} \log((|c_i| + 2)(|a_i| + 2))$.

We will develop an algorithm that requires at most $\log^4(dH)$ bit operations where d is the degree and all coefficients absolute value are at most H.

Problem Statement

Problem

Given

$$f(x_1) = c_1 + c_2 x_1^{a_2} + c_3 x_1^{a_3} \in \mathbb{Z}[x_1]$$

with $c_1 c_2 c_3 \neq 0$, $d := a_3 > a_2 \geq 1$, *and* $|c_i| \leq H$, *devise an algorithm that finds an approximate root of* f *using* $\log^{O(1)}(dH)$ *bit operations.*

Problem Statement

Problem

Given

$$f(x_1) = c_1 + c_2 x_1^{a_2} + c_3 x_1^{a_3} \in \mathbb{Z}[x_1]$$

with $c_1 c_2 c_3 \neq 0$, $d := a_3 > a_2 \geq 1$, *and* $|c_i| \leq H$, *devise an algorithm that finds an approximate root of* f *using* $\log^{O(1)}(dH)$ *bit operations.*

Why trinomials? Monomials and binomials are well understood and such algorithms for them already exist. We run into problems extending this to tetranomials, which we will later discuss.

Our approach

1. Via rescaling, we can reduce finding the roots of f to finding the roots of the polynomial

$$g(x_1) = 1 + cx_1^m + x_1^n \in \mathbb{C}[x_1]$$

where $c \neq 0$, $0 < m < n$, and $\gcd(m, n) = 1$.

Our approach

1. Via rescaling, we can reduce finding the roots of f to finding the roots of the polynomial

$$g(x_1) = 1 + cx_1^m + x_1^n \in \mathbb{C}[x_1]$$

where $c \neq 0$, $0 < m < n$, and $\gcd(m, n) = 1$.

2. We can use $\mathcal{A}$-hypergeometric series to efficiently find an approximate root of g.

Simplifying the problem

Consider the equation $f(x_1) = c_1 + c_2 x_1^{a_2} + c_3 x_1^{a_3} = 0$.

Simplifying the problem

Consider the equation $f(x_1) = c_1 + c_2 x_1^{a_2} + c_3 x_1^{a_3} = 0$.

1. Multiply f and/or the variable x_1 by ± 1 so to reduce the special case of approximating the positive roots where $c_3 > 0$.

Simplifying the problem

Consider the equation $f(x_1) = c_1 + c_2 x_1^{a_2} + c_3 x_1^{a_3} = 0$.

1. Multiply f and/or the variable x_1 by ± 1 so to reduce the special case of approximating the positive roots where $c_3 > 0$.
2. Using rescaling, simplify to the polynomial

$$1 + cx^m + x^n$$

where $c \neq 0$, $0 < m < n$ and $\gcd(m, n) = 1$.

Rescaling

Consider the equation $f(x_1) = c_1 + c_2 x_1^{a_2} + c_3 x_1^{a_3} = 0.$

Rescaling

Consider the equation $f(x_1) = c_1 + c_2 x_1^{a_2} + c_3 x_1^{a_3} = 0$.

- We can express a root of f as a function $x(c_1, c_2, c_3)$. Note that for any non-zero scalar λ,

Rescaling

Consider the equation $f(x_1) = c_1 + c_2 x_1^{a_2} + c_3 x_1^{a_3} = 0$.

- We can express a root of f as a function $x(c_1, c_2, c_3)$. Note that for any non-zero scalar λ,
  - $x(\lambda c_1, \lambda c_2, \lambda c_3) = x(c_1, c_2, c_3)$.

Rescaling

Consider the equation $f(x_1) = c_1 + c_2 x_1^{a_2} + c_3 x_1^{a_3} = 0$.

- We can express a root of f as a function $x(c_1, c_2, c_3)$. Note that for any non-zero scalar λ,
  - $x(\lambda c_1, \lambda c_2, \lambda c_3) = x(c_1, c_2, c_3)$.
  - $x(c_1, \lambda^{a_2} c_2, \lambda^{a_3} c_3) = \lambda^{-1} x(c_1, c_2, c_3)$.

Rescaling

Consider the equation $f(x_1) = c_1 + c_2 x_1^{a_2} + c_3 x_1^{a_3} = 0$.

- We can express a root of f as a function $x(c_1, c_2, c_3)$. Note that for any non-zero scalar λ,
  - $x(\lambda c_1, \lambda c_2, \lambda c_3) = x(c_1, c_2, c_3)$.
  - $x(c_1, \lambda^{a_2} c_2, \lambda^{a_3} c_3) = \lambda^{-1} x(c_1, c_2, c_3)$.
- Choose complex constants λ_0 and λ_1 satisfying

$$\lambda_0 \lambda_1^0 = c_1^{-1} \quad \text{and} \quad \lambda_0 \lambda_1^{a_3} = c_3^{-1}$$

Rescaling

Consider the equation $f(x_1) = c_1 + c_2 x_1^{a_2} + c_3 x_1^{a_3} = 0$.

- We can express a root of f as a function $x(c_1, c_2, c_3)$. Note that for any non-zero scalar λ,
    - $x(\lambda c_1, \lambda c_2, \lambda c_3) = x(c_1, c_2, c_3)$.
    - $x(c_1, \lambda^{a_2} c_2, \lambda^{a_3} c_3) = \lambda^{-1} x(c_1, c_2, c_3)$.
- Choose complex constants λ_0 and λ_1 satisfying

$$\lambda_0 \lambda_1^0 = c_1^{-1} \quad \text{and} \quad \lambda_0 \lambda_1^{a_3} = c_3^{-1}$$

- Consider $\lambda_0 f(\lambda_1 x_1) = 1 + c_2 \lambda_0 \lambda_1^{a_2} x^{a_2} + x_1^{a_3}$. If ζ is a root of $\lambda_0 f(\lambda_1 x_1)$, then $\lambda_1 \zeta$ is a root of $f(x_1)$.

Example

Let $f(x_1) = 2 + 3x_1^2 + 5x_1^3$.

- $f(x_1)$ has only one negative real root. So we consider $\tilde{f}(x_1) = -f(-x_1) = -2 - 3x_1^2 + 5x_1^3$, which has one positive real root and $5 > 0$.

Example

Let $f(x_1) = 2 + 3x_1^2 + 5x_1^3$.

- $f(x_1)$ has only one negative real root. So we consider $\tilde{f}(x_1) = -f(-x_1) = -2 - 3x_1^2 + 5x_1^3$, which has one positive real root and $5 > 0$.
- We then solve for λ_0 and λ_1 so that

$$\lambda_0 \lambda_1^0 = -\frac{1}{2} \quad \text{and} \quad \lambda_0 \lambda_1^3 = \frac{1}{5}$$

Example

Let $f(x_1) = 2 + 3x_1^2 + 5x_1^3$.

- $f(x_1)$ has only one negative real root. So we consider $\tilde{f}(x_1) = -f(-x_1) = -2 - 3x_1^2 + 5x_1^3$, which has one positive real root and $5 > 0$.
- We then solve for λ_0 and λ_1 so that

$$\lambda_0 \lambda_1^0 = -\frac{1}{2} \quad \text{and} \quad \lambda_0 \lambda_1^3 = \frac{1}{5}$$

- Hence

$$\lambda_0 \tilde{f}(\lambda_1 x) = -\lambda_0 f(-\lambda_1 x) = \boxed{1 - \left(\frac{3}{2}\left(\frac{2}{5}\right)^{2/3}\right) x^2 + x^3}$$

Hypergeometric Solution

Now that we've simplified, how can we solve?

Hypergeometric Solution

Now that we've simplified, how can we solve?

Theorem (Passare and Tsikh [3, 1])

Consider the equation

$$a_0 + a_1 x + a_2 x^2 + \cdot\cdot + x^p + \cdot\cdot + x^q + \cdot\cdot + a_{n-1} x^{n-1} + a_n x^n = 0$$

The solution $x(a_0, .., [p], .., [q], .., a_n)$ *may be expressed as*

$$\sum_{k \in \mathbb{N}^{n-1}}^{\infty} \frac{\varepsilon^{-\langle \beta_q, k\rangle + 1}}{(q-p)k!} \frac{\Gamma\left((-\langle \beta_q, k\rangle + 1)/(q-p)\right)}{\Gamma\left(1 + (\langle \beta_p, k\rangle + 1)/(q-p)\right)} a_0^{k_0} a_1^{k_1} \cdot\cdot [p] \cdot\cdot [q] \cdot\cdot a_n^{k_n}$$

Hypergeometric Solution

Theorem (Trinomial case)

Consider the equation $1 + cx^m + x^n = 0$ *with* $c \neq 0$, $0 < m < n$, $\gcd(m,n) = 1$. *Let* $r_{m,n} := \frac{n}{m^{\frac{m}{n}}(n-m)^{\frac{n-m}{n}}}$

- *If* $|c| < r_{m,n}$, $x(c) = \nu_n \left[1 + \sum_{k=1}^{\infty} \left(\frac{\nu_n^{mk}}{kn^k} \cdot \prod_{j=1}^{k-1} \frac{1+km-jn}{j} \right) c^k \right]$ *where* ν_n *is any n-th root of* -1.
- *If* $|c| > r_{m,n}$,
  $x_{low}(c) = \frac{\nu_m}{|c|^{1/m}} \left[1 + \sum_{k=1}^{\infty} \left(\frac{\nu_m^{nk}}{km^k} \cdot \prod_{j=1}^{k-1} \frac{1+kn-jm}{j} \right) \left(\frac{1}{|c|^{n/m}} \right)^k \right]$
  and $x_{hi}(c) = \nu_{n-m}|c|^{1/(n-m)} \left[1 - \sum_{k=1}^{\infty} \left(\frac{\nu_{n-m}^{-nk}}{k(n-m)^k} \cdot \prod_{j=1}^{k-1} \frac{km+j(n-m)-1}{j} \right) \left(\frac{1}{|c|^{n/(n-m)}} \right)^k \right]$
  where ν_m *and* ν_{n-m} *are any m-th and* $n-m$*-th root of* -1.

How many terms are enough?

In the case when $|c| > r_{m,n}$,

Theorem (x_{low})

For any integer $\ell \geq 2$,

$$\left| \frac{\nu_m}{c^{1/m}} \sum_{k=\ell+1}^{\infty} \left(\frac{\nu_m^{nk}}{km^k} \cdot \prod_{j=1}^{k-1} \frac{1+kn-jm}{j} \right) \left(\frac{1}{c^{n/m}} \right)^k \right|$$

$$\leq \frac{\nu_m}{c^{1/m}} \cdot \frac{\left(\frac{n}{n-m}\right)^{\frac{1+n+\ell n}{m}} (n-m)^\ell \nu_m^n}{\ell \left(c^{n/m} - n \left(\frac{n}{n-m}\right)^{\frac{n-m}{m}} \nu_m^n \right) \left(\frac{c^{n/m} m}{\nu_m^n} \right)^\ell}.$$

For any integer $\ell \geq 2$,

Theorem (x_{hi})

$$\left|\nu_{n-m} c^{1/(n-m)} \sum_{k=\ell+1}^{\infty} \left(\frac{\nu_{n-m}^{-nk}}{k(n-m)^k} \cdot \prod_{j=1}^{k-1} \frac{km + j(n-m) - 1}{j} \right) \left(\frac{1}{c^{n/(n-m)}} \right)^k \right|$$

$$\leq \nu_{n-m} c^{1/(m-n)} \frac{n^{\ell} \left(\frac{n}{m}\right)^{\frac{-1+m+\ell m}{n-m}} \left(\frac{c^{\frac{n}{m-n}} \nu_{n-m}^{-n}}{n-m} \right)^{\ell}}{\ell \left(n \left(\frac{n}{m}\right)^{\frac{m}{n-m}} + c^{\frac{n}{n-m}} (m-n) \nu_{n-m}^{n} \right)}.$$

How many terms?

- The prior bounds give a useful metric to determine how quickly the $\mathcal{A}$-hypergeometric series converge, but how many terms are necessary to be an approximate root?

How many terms?

- The prior bounds give a useful metric to determine how quickly the $\mathcal{A}$-hypergeometric series converge, but how many terms are necessary to be an approximate root?
- We've found that $\log(dH)$ many terms work through numerical testing, but we've yet to formulate a proof.

How many terms?

- The prior bounds give a useful metric to determine how quickly the $\mathcal{A}$-hypergeometric series converge, but how many terms are necessary to be an approximate root?
- We've found that $\log(dH)$ many terms work through numerical testing, but we've yet to formulate a proof.
- We suspect that the results provided in Rojas and Ye [4] will be particularly useful in finding this.

Example

Proceeding from our prior example, consider
$-\lambda_0 f(-\lambda_1 x) = 1 - \left(\frac{3}{2}\left(\frac{2}{5}\right)^{2/3}\right) x^2 + x^3.$

Example

Proceeding from our prior example, consider $-\lambda_0 f(-\lambda_1 x) = 1 - \left(\frac{3}{2}\left(\frac{2}{5}\right)^{2/3}\right) x^2 + x^3$.
The solution to $-\lambda_0 f(-\lambda_1 x) = 0$ is given by

$$x = (-1)\left[1 + \sum_{k=1}^{\infty}\left(\frac{(-1)^{2k}}{k3^k} \cdot \prod_{j=1}^{k-1} \frac{1+2k-3j}{j}\right)\left(\frac{3}{2}\left(\frac{2}{5}\right)^{2/3}\right)^k\right]$$

Example

Proceeding from our prior example, consider
$-\lambda_0 f(-\lambda_1 x) = 1 - \left(\frac{3}{2}\left(\frac{2}{5}\right)^{2/3}\right)x^2 + x^3$.
The solution to $-\lambda_0 f(-\lambda_1 x) = 0$ is given by

$$x = (-1)\left[1 + \sum_{k=1}^{\infty}\left(\frac{(-1)^{2k}}{k3^k} \cdot \prod_{j=1}^{k-1}\frac{1+2k-3j}{j}\right)\left(\frac{3}{2}\left(\frac{2}{5}\right)^{2/3}\right)^k\right]$$

Evaluating $\log(dH) \approx 3$ (where $d = 3$ and $H = 5$) terms of the series yields $x \approx -1.3584$, so $-\lambda_1 x \approx -1.0009$ is an approximate root of our input polynomial.

A special case

What if $|c| = r_{m,n}$?

A special case

What if $|c| = r_{m,n}$? Then we have a *degenerate root*, a root with multiplicity greater than 1. How do we solve?

A special case

What if $|c| = r_{m,n}$? Then we have a *degenerate root*, a root with multiplicity greater than 1. How do we solve?

Suppose $f(x) = 1 + cx^m + x^n$ has a degenerate root ζ. Then $f(\zeta) = f'(\zeta) = 0$, which implies $f(\zeta) = \zeta f'(\zeta) = 0$. So we have the following system,

$$
\begin{aligned}
1 + c\zeta^m + \zeta^n &= 0 \\
0 + cm\zeta^m + n\zeta^n &= 0.
\end{aligned}
$$

A special case

What if $|c| = r_{m,n}$? Then we have a *degenerate root*, a root with multiplicity greater than 1. How do we solve?

Suppose $f(x) = 1 + cx^m + x^n$ has a degenerate root ζ. Then $f(\zeta) = f'(\zeta) = 0$, which implies $f(\zeta) = \zeta f'(\zeta) = 0$. So we have the following system,

$$\begin{aligned} 1 + c\zeta^m + \zeta^n &= 0 \\ 0 + cm\zeta^m + n\zeta^n &= 0. \end{aligned}$$

This implies that

$$c\zeta^m = \frac{n}{m-n} \quad \text{and} \quad \zeta^n = \frac{m}{n-m}$$

Solving either of those binomial equations will yield our degenerate root ζ.

Algorithm

Given a polynomial $f(x_1) = c_1 + c_2 x_1^{a_2} + c_3 x_1^{a_3}$,

Algorithm

Given a polynomial $f(x_1) = c_1 + c_2 x_1^{a_2} + c_3 x_1^{a_3}$,

1. Using rescaling and multiplying by ± 1, consider the real roots of
$$\lambda_0 f(\lambda_1 x) = 1 + cx^m + x^n$$
where $c \neq 0$, $0 < m < n$, and $\gcd(m, n) = 1$.

Algorithm

Given a polynomial $f(x_1) = c_1 + c_2 x_1^{a_2} + c_3 x_1^{a_3}$,

1. Using rescaling and multiplying by ± 1, consider the real roots of

$$\lambda_0 f(\lambda_1 x) = 1 + cx^m + x^n$$

   where $c \neq 0$, $0 < m < n$, and $\gcd(m, n) = 1$.

2. Compute $r_{m,n} = \frac{n}{m^{\frac{m}{n}} (n-m)^{\frac{n-m}{n}}}$.
   1. If $|c| < r_{m,n}$, compute $\log(dH)$ terms of $\nu_n \left[1 + \sum_{k=1}^{\infty} \left(\frac{\nu_n^{mk}}{kn^k} \cdot \prod_{j=1}^{k-1} \frac{1+km-jn}{j} \right) c^k \right]$.
   2. If $|c| > r_{m,n}$, compute $\log(dH)$ terms of

$$x_{\text{low}} = \frac{\nu_m}{|c|^{1/m}} \left[1 + \sum_{k=1}^{\infty} \left(\frac{\nu_m^{nk}}{km^k} \cdot \prod_{j=1}^{k-1} \frac{1+kn-jm}{j} \right) \left(\frac{1}{|c|^{n/m}} \right)^k \right] \text{ or}$$

$$x_{\text{hi}}(c) = \nu_{n-m} |c|^{1/(n-m)} \left[1 - \sum_{k=1}^{\infty} \left(\frac{\nu_{n-m}^{-nk}}{k(n-m)^k} \cdot \prod_{j=1}^{k-1} \frac{km+j(n-m)-1}{j} \right) \left(\frac{1}{|c|^{n/(n-m)}} \right)^k \right].$$

   3. If $|c| = r_{m,n}$, use one the following binomial equations to solve for a root: $c\zeta^m = \frac{n}{m-n}$ or $\zeta^n = \frac{m}{n-m}$

A natural question arises: why do we only consider the trinomial case instead of tetranomials and beyond?

A natural question arises: why do we only consider the trinomial case instead of tetranomials and beyond?

Because the techniques of $\mathcal{A}$-hypergeometric series are not as easily applied.

- Consider all possible rescaled trinomials of the form $g(x) = 1 + cx^m + x^n$. It turns out the radius of convergence of the $\mathcal{A}$-hypergeometric series corresponding to the roots of g relate to the discriminant of g.

- Consider all possible rescaled trinomials of the form $g(x) = 1 + cx^m + x^n$. It turns out the radius of convergence of the $\mathcal{A}$-hypergeometric series corresponding to the roots of g relate to the discriminant of g.
- In particular,

$$\Delta = 0 \iff |c| = \frac{n}{m^{m/n}(n-m)^{(n-m)/n}} = r_{m,n}.$$

- Consider all possible rescaled trinomials of the form $g(x) = 1 + cx^m + x^n$. It turns out the radius of convergence of the $\mathcal{A}$-hypergeometric series corresponding to the roots of g relate to the discriminant of g.
- In particular,

$$\Delta = 0 \iff |c| = \frac{n}{m^{m/n}(n-m)^{(n-m)/n}} = r_{m,n}.$$

- Hence, the two families of $\mathcal{A}$-hypergeometric series that solve g correspond to two regions of $\mathbb{R}$, each with its own *known* hypergeometric solution.

- For a rescaled tetranomial, $g(x) = 1 + cx^l + dx^m + x^n$, we have that the discriminant breaks up $\mathbb{R}^2$ into 8 distinct regions.

- For a rescaled tetranomial, $g(x) = 1 + cx^l + dx^m + x^n$, we have that the discriminant breaks up $\mathbb{R}^2$ into 8 distinct regions.
- However, these regions are not convex, and a hypergeometric series solution for each region is not known.

- For a rescaled tetranomial, $g(x) = 1 + cx^l + dx^m + x^n$, we have that the discriminant breaks up $\mathbb{R}^2$ into 8 distinct regions.
- However, these regions are not convex, and a hypergeometric series solution for each region is not known.
- In a future paper, we will investigate this further.

Acknowledgments

I would like to thank Dr. Maurice Rojas, Weixun Deng, and Joshua Goldstein for their help and guidance throughout this project. I would also like to thank Texas A&M University and the National Science Foundation for this opportunity.

References

Richard Birkeland, *Über die auflösung algebraischer gleichungen durch hypergeometrische funktionen*, Mathematische Zeitschrift **26** (1927), no. 1, 566–578.

Lenore Blum, Felipe Cucker, Michael Shub, and Steve Smale, *Complexity and real computation*, Springer Science & Business Media, 1998.

Mikael Passare and August Tsikh, *Algebraic equations and hypergeometric series*, The legacy of Niels Henrik Abel, Springer, 2004, pp. 653–672.

J Maurice Rojas and Yinyu Ye, *On solving univariate sparse polynomials in logarithmic time*, Journal of Complexity **21** (2005), no. 1, 87–110.