Motivation

- Easy: Prove identities in one free variable like

$$A(n) = B(n)$$

- Routine: Prove identities in a fixed number of variables like

$$A(m, n, x, y) = B(m, n, x, y)$$

- **But** we need a human to prove things of the form

$$A(x_1, \ldots, x_n) = B(x_1, \ldots, x_n)$$

  with *arbitrary* n, $x_1, \ldots, x_n$.

- Do we?

- We will present an algorithm for doing it by the computer, for certain kinds of A and B.

Battle Plan

- Motivation
- Fundamentals I: Notions from Difference Algebra
  - Difference Rings, Difference Fields, and so on
  - Comparison to Differential Algebra
- Fundamentals II: A new Algorithm for Proving Univariate Identities
  - Admissible Sequences
  - Example Sequences
  - Algebraization
  - The Algorithm
  - Note on Correctness and Termination
- Application to Arbitrary Many Variables
  - Admissible Sequences
  - Modification of the Algorithm
  - Examples
  - Schoissengeier's Problem

Fundamentals I: Notions from Difference Algebra

- $R = (R, s)$ is called *difference ring* if R is a ring and $s\colon R \to R$ is an automorphism on R.
- A difference ring which is also a field, is called difference field.
- Let K be a difference field, $m \in \mathbb{N}$. The set $K\{t_1, \ldots, t_m\}$ of polynomials in $s^i t_j$ ($i \geq 0, j = 1, \ldots, m$) is called m-fold *polynomial difference ring*
- The quotient field of $K\{t_1, \ldots, t_m\}$ is written $K\langle t_1, \ldots, t_m\rangle$.
- The *order* of a difference polynomial $p \in K\{t_1, \ldots, t_m\}$ in the maximum j such that $s^j t_i$ appears in p for some i.
- $K\{t_1, \ldots, t_m\}_k := \{\, p \in K\{t_1, \ldots, t_m\} : \operatorname{order}(p) \leq k \,\}$ is called *restriction* of $K\{t_1, \ldots, t_m\}$.
- $K\langle t_1, \ldots, t_m\rangle_k$ is the quotient field of $K\{t_1, \ldots, t_m\}_k$.
- If $S \subseteq K\{t_1, \ldots, t_m\} =: R$ then $\langle S\rangle$ is the polynomial ideal generated by S in R, $\langle\!\langle S\rangle\!\rangle$ is the *difference ideal* generated by S in R, $\langle S\rangle_k = \langle S\rangle \cap R_k$, $\langle\!\langle S\rangle\!\rangle_k = \langle\!\langle S\rangle\!\rangle \cap R_k$.

Fundamentals II: Proving Univariate Identities

- Admissible Sequences
- Example Sequences
- Algebraization
- The Algorithm
- Notes on Correctness and Termination

Admissible Sequences

- Define the class $\mathcal{C}$ of sequences as follows:
- If
  - $f_1, \ldots, f_s \in \mathcal{C}$ ($s \geq 0$)
  - f is a sequence, $r > 0$
  - p is a polynomial in

$$\begin{gathered} f_1(n), f_1(n+1), \ldots, f_1(n+r), \\ f_2(n), f_2(n+1), \ldots, f_2(n+r), \\ \vdots \\ f_s(n), f_s(n+1), \ldots, f_s(n+r), \\ f(n), f(n+1), \ldots, f(n+r-1) \end{gathered}$$

  - f obeys a recurrence of the form

$$f(n+r) = p \qquad \text{or} \qquad f(n+r) = 1/p$$

  Then $f \in \mathcal{C}$
- Note: induction base given for $s = 0$

Example Sequences in $\mathcal{C}$

- rational functions in n
- $n \mapsto \alpha^n$ ($\alpha \in \mathbb{Q}$)
- $n \mapsto \alpha^{\beta^n}$ ($\alpha \in \mathbb{Q}, \beta \in \mathbb{N}$)
- $n \mapsto \mathrm{Fib}(n)$
- $n \mapsto 2^{\mathrm{Fib}(n)}$
- $n \mapsto \mathrm{Fib}(2^n)$
- $n \mapsto \Gamma(\alpha n + \beta)$ ($\alpha \in \mathbb{Q}$)
- $n \mapsto \sum_{k=1}^{n} \mathrm{H}(k) \binom{n}{k}$

Some Closure Properties

- Let $f(n), g(n), h(n) \in \mathcal{C}$, $h(n) \neq 0$ for all n.
- $f + g, f \cdot g, f - g, \alpha f, f/h \in \mathcal{C}$
- $n \mapsto \sum_{k=1}^{n} f(n) \in \mathcal{C}$
- $n \mapsto \prod_{k=1}^{n} f(n) \in \mathcal{C}$
- $n \mapsto \mathop{\mathrm{K}}_{k=1}^{n} h(n) = h(1) + \cfrac{1}{h(2) + \cfrac{1}{\cdots + \cfrac{1}{h(n)}}} \in \mathcal{C}$

Algebraic Representation of Elements of $\mathcal{C}$

- Consider difference ideals “generated” by the recurrences
- Let $f_1, \ldots, f_m \in \mathcal{C}$ be such that the definition of f_i depends only on f_j with $j < i$, let $f = f_m$.
- Example: $f(n) := \sum_{k=1}^{n} \frac{k!\,\mathrm{H}(k)}{\mathrm{Fib}(k)}$

Component	Recurrence	Defining relation
$f_1(n) = n$	$f_1(n+1) = f_1(n) + 1$	$st_1 - t_1 - 1$
$f_2(n) = \mathrm{Fib}(n)$	$f_2(n+2) = f_2(n+1) + f_2(n)$	$sst_2 - st_2 - t_2$
$f_3(n) = n!$	$f_3(n+1) = (n+1)f_3(n)$	$st_3 - t_3st_1$
$f_4(n) = 1/n$	$f_4(n) = 1/n$	$t_4t_1 - 1$
$f_5(n) = \mathrm{H}(n) = \sum_{k=1}^{n} \frac{1}{k}$	$f_5(n+1) = f_5(n) + \frac{1}{n+1}$	$st_5 - t_5 - st4$
$f_6(n) = \frac{1}{\mathrm{Fib}(n)}$	$f_6(n) = \frac{1}{\mathrm{Fib}(n)}$	$t_6t_2 - 1$
$f_7(n) = f(n)$	$f_7(n+1) = f_7(n) + \frac{(n+1)!\,\mathrm{H}(n+1)}{\mathrm{Fib}(n+1)}$	$st_7 - t_7 - st_3st_5st_6$

The Proving Algorithm

Input: $f \in \mathcal{C}$ a polynomially recursive function
D defining relations of f and its "subexpressions"

- Assumption: t_m corresponds to $\{f(n)\}_n$;
The highest order in D is of order r.

Output: "correct" or "incorrect" depending on whether $f \equiv 0$ or not

```
function isZero(f, R)
  k := r − 1
  repeat
     k := k + 1
     I_k := ⟨⟨D⟩⟩_k + ⟨t_m, s t_m, ..., s^{k−1} t_m⟩_k
  until s^k t_m ∈ Rad I_k
  // check initial conditions
  if f(1) = f(2) = ··· = f(k) = 0 then
     return "correct"
  else
     return "incorrect"
```

Correctness and Termination

- **Theorem** The algorithm is correct and terminats.

Application to Arbitrary Many Variables

- Admissible Sequences
- Modification of the Algorithm
- Examples
- Schoissengeier's Problem

Admissible Sequences

- Define the class $\mathcal{C}'$ of formal sequences as follows:
- $\boxed{\text{Formal sequences } u_1, u_2, u_3, \ldots \text{ of indeterminates belong to } \mathcal{C}'.}$
- If, as before,
  - $f_1, \ldots, f_s \in \mathcal{C}$ ($s \geq 0$)
  - f is a sequence, $r > 0$
  - p is a polynomial in

$$\begin{gathered} f_1(n), f_1(n+1), \ldots, f_1(n+r), \\ f_2(n), f_2(n+1), \ldots, f_2(n+r), \\ \vdots \\ f_s(n), f_s(n+1), \ldots, f_s(n+r), \\ f(n), f(n+1), \ldots, f(n+r-1) \end{gathered}$$

  - f obeys a recurrence of the form

$$f(n+r) = p \qquad \text{or} \qquad f(n+r) = 1/p$$

Then $f \in \mathcal{C}$

Modification of the Algorithm

- Represent sequences of indeterminates also by some t_i.

- **Key idea I:** Omit the defining relation for them.

- Given $f_1, \ldots, f_m \in \mathcal{C}'$ where
  - $f_1, \ldots, f_i$ are sequences of indeterminates and
  - $f_{i+1}, \ldots, f_m$ are given by recurrences, each f_j depending only on $f_{j'}$ with $j' < j$

- The original algorithm is applicable, but does not (always) terminate.

- **Key idea II:** Put indeterminates into the ground field, compute in $K\langle t_1, \ldots, t_i\rangle\{t_{i+1}, \ldots, t_m\}$

- The original algorithm is applicable and the termination proof goes through.

Examples

- **Any** family of orthogonal polynomials, defined by

$$P_n(x) = (x - c_n)P_{n-1}(x) - \lambda_n P_{n-2}(x), \qquad P_{-1}(x) = 0, \quad P_0(x) = 1$$

  for **arbitrary** $c_1, c_2, \ldots$ and $\lambda_1, \lambda_2, \ldots$ satisfies the *Christoffel-Darboux-formula*

$$\sum_{k=0}^{n} \frac{P_k(x)P_k(u)}{\prod_{i=1}^{k+1} \lambda_i} = \frac{P_{n+1}(x)P_n(u) - P_n(x)P_{n+1}(u)}{(x-u)\prod_{i=1}^{n+1} \lambda_i}$$

- Proof: less than a second

- Continued fraction identity: Write

$$C_n := x_0 + \cfrac{y_1}{x_1 + \cfrac{y_2}{\ddots + \cfrac{y_n}{x_n}}} = \frac{K(y_1, \ldots, y_n; x_0, \ldots, x_n)}{K(y_2, \ldots, y_n; x_1, \ldots, x_n)}$$

  for **arbitrary** $x_1, \ldots, x_n$, $y_1, \ldots, y_n$ and n, and prove, in less than a second,

$$C_n = \frac{K(y_1, \ldots, y_n; x_0, \ldots, x_n)}{K(y_2, \ldots, y_n; x_1, \ldots, x_n)} = x_0 + \sum_{k=1}^{n} (-1)^{k+1} \frac{\prod_{i=1}^{k} y_i}{K(y_2, \ldots, y_{k-1}; x_1, \ldots, x_{k-1})K(y_2, \ldots, y_k; x_1, \ldots, x_k)}$$

Examples II

- Also easy (less than one second):

$$\sum_{k=1}^{n} \frac{\prod_{i=1}^{k-1}(x_i+\alpha)}{\prod_{i=1}^{k} x_k} = \frac{1}{\alpha}\Big(\prod_{k=1}^{n}\frac{x_k+\alpha}{x_k} - 1\Big)$$

$$\sum_{k=1}^{n} \mathrm{H}(k) \sum_{i=1}^{k} x_i = (-n + (1+n)\,\mathrm{H}(n)) \sum_{i=1}^{n} x_i - \sum_{k=1}^{n} (\mathrm{H}(k)-1) k x_k$$

$$\sum_{r=1}^{n} \frac{\sum_{l=1}^{r}\prod_{k=1}^{l}\frac{x_k}{1+x_k}}{1+x_r} = \sum_{k=1}^{n}\frac{\prod_{i=1}^{k}\frac{x_i}{1+x_i}}{1+x_k} + \sum_{i=1}^{n}\frac{1}{1+x_i}\sum_{i=1}^{n}\prod_{k=1}^{i}\frac{x_k}{1+x_k} - \sum_{k=1}^{n}\Big(\prod_{i=1}^{k}\frac{x_i}{1+x_i}\Big)\Big(\sum_{i=1}^{k}\frac{1}{1+x_i}\Big)$$

Schoissengeier's Problem

- Appears in number theoretic context
- An identitity with 27 summation signs, two sequences of indeterminates.
- Not of the appropriate form.
- but can be converted to an appropriate form.
- At the cost of expression blow up: The result contains 2000 summation signs.
- Elementary simplification gives 100 summation signs.
- Compute a Gröbner basis of 250 polynomials in 350 variables. Game over.