Computational Learning Theory

Learning and mathematical theorems

Akihiro Yamamoto 山本 章博
http://www.iip.ist.i.kyoto-u.ac.jp/member/akihiro/
akihiro@i.kyoto-u.ac.jp

Contents

- GCD and the Euclidian Algorithm
- Math of Rings and Polynomials
- Hilbert’s basis theorem
- Relation to Machine Learning

GCD and the Euclidian Algorithm

- In mathematics, the greatest common divisor (GCD) of two or more integers, which are not all zero, is the largest positive integer that divides each of the integers. For example, the gcd of 8 and 12 is 4.   [Wikipedia]
- By the Euclidian algorithm, gcd(m, n) can be computed *efficiently*.

$$\gcd(m, 0) = m$$

$$\gcd(m, n) = \gcd(n, m \bmod n)$$

Example

gcd(34, 21)=gcd(21, 13)=gcd(13, 8)=gcd(8, 5)=gcd(5, 3)
=gcd(3, 2)=gcd(3, 2)=gcd(2, 1)=gcd(1, 1)=gcd(1, 0)=0

GCD and Learning

A class of languages in $\mathbf{N}$:

$$\mathsf{C} = \{L(m) \mid m \in \mathbf{N}\}$$

$$L(m) = \{0\underbrace{1\ldots1}_{n}0 \mid n \bmod m = 0\}$$

$$L(m) = \{n \in \mathbf{N} \mid n \bmod m = 0\}$$

A class of languages in $\mathbf{Z}$:

$$\mathsf{C} = \{L(m) \mid m \in \mathbf{N}\}$$

$$L(m) = \{\underbrace{1\ldots1}_{n} \mid n \bmod m = 0\} \cup \{0\underbrace{1\ldots1}_{n} \mid n \bmod m = 0\}$$

$$L(m) = \{n \in \mathbf{Z} \mid |n| \bmod m = 0\}$$

GCD and Learning

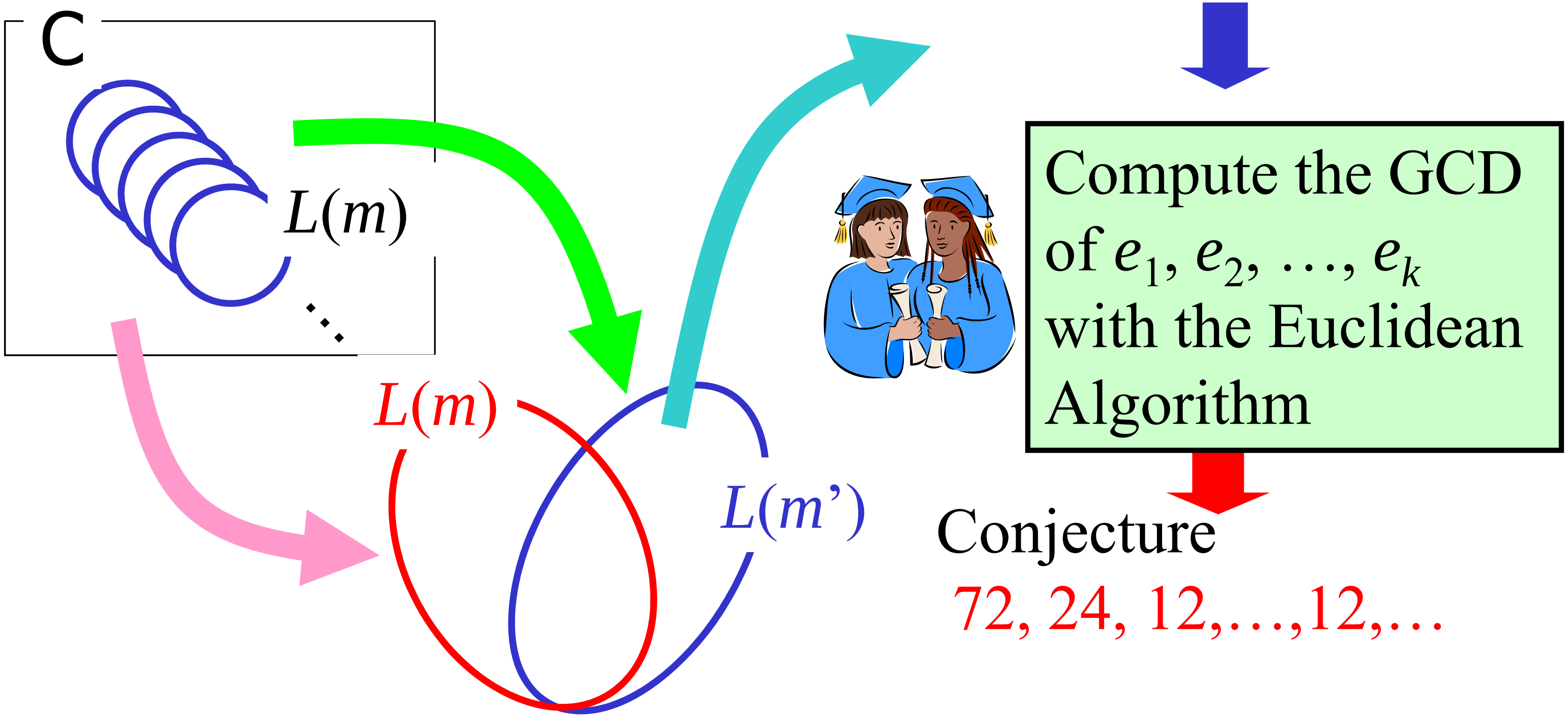
- A class of languages :

  C = $\{L(m) \mid m \in \mathbf{N}\}$

  $L(m) = \{01\ldots10 \mid n \bmod m = 0\}$

Positive presentation

72, 48, 60, …,12,…



Compute the GCD of e_1, e_2, …, e_k with the Euclidean Algorithm

Conjecture

72, 24, 12,…,12,…

C4: Finite thickness

- A class C of languages has the finite thickness property if for all $w \in \Sigma^*$ only a finite number of languages contains w,

Theorem [Angluin] If a class C of languages has the finite thickness, C is identifiable in the limit from positive data.

Note : Even if C has the finite thickness property, $\cup^N U$ might not have the same property.

Identification in the limit [Gold]

$e_1, e_2, e_3, \ldots$ →

$g_1, g_2, g_3, \ldots$

- A learning algorithm A EX-identifies $L(g)$ in the limit from positive presentations if
for any positive presentation $\sigma = e_1, e_2, e_3, \ldots$ of $L(g)$ and the output sequence $g_1, g_2, g_3, \ldots$ of A, there exists N such that for all $n > N$ $g_n = g'$ and $L(g') = L(g)$
- A learning algorithm A EX-identifies a class C of languages in the limit from positive presentations if
A EX-identifies every language in C in the limit from positive presentations.

Proving that C is identifiable

- From the finite thickness condition:

  The class C $= \{L(m) \mid m \in \mathbf{N}\}$ has the finite thickness property.

- From a property of natural numbers and the property

  $$\mathrm{GCD}(e_1, e_2, \ldots, e_k) \geq \mathrm{GCD}(e_1, e_2, \ldots, e_k, e_{k+1})$$

  The property is :

  Let $a_1, a_2, \ldots, a_n, \ldots$ be a infinite sequence of natural numbers satisfying that

  $$a_n \geq a_{n+1} \quad \text{for all } n \geq 1.$$

  Then there is $N \geq 1$ such that $a_n = a_{n+1}$ for all $n \geq N$.

Generalizing the Setting

- The class C = $\{L(m) \mid m \in \mathbf{N}\}$ is defined with multiplication
- The Euclidean Algorithm as the learning machine is based on

computing remainders of two integers,

which can be constructed of division and subtraction, and division of two integers can be implemented with multiplication and subtraction. So the Euclidean Algorithm can be implemented with

multiplication, subtraction (the inverse of addition).

Generalizing the Setting

- The class C = $\{L(m) \mid m \in \mathbf{N}\}$ is defined with multiplication
- The Euclidean Algorithm as the learning machine is based on

computing remainders of two integers,

which can be constructed of division and subtraction, and division of two integers can be implemented with multiplication and subtraction. So the Euclidean Algorithm can be implemented with

multiplication, subtraction (the inverse of addition).

Monomimial Case

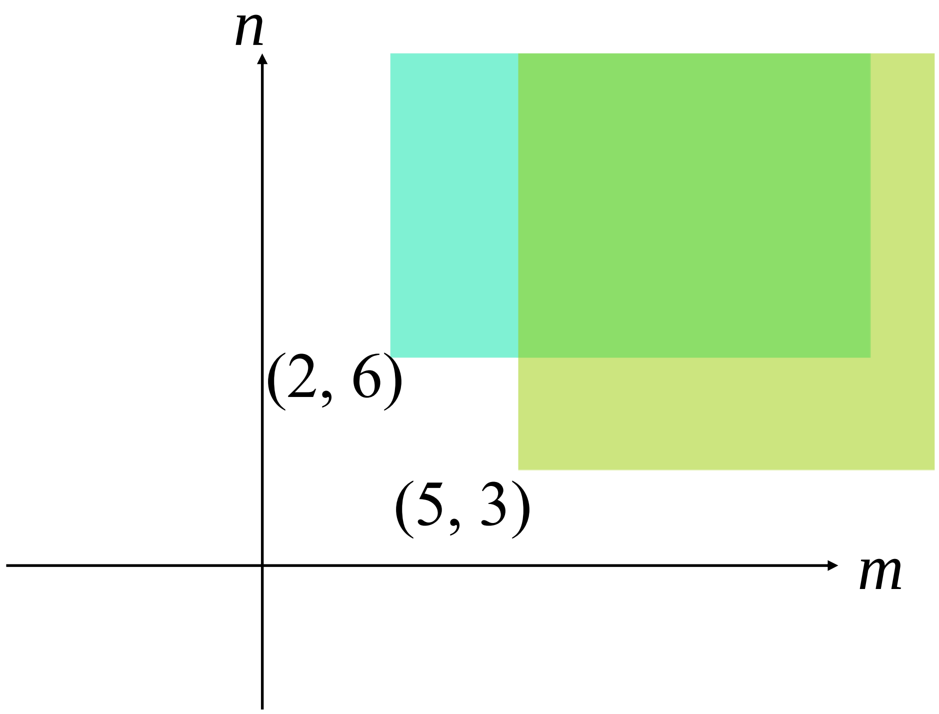
- Consider the case U = $\{x^m y^n \mid m, n \in \mathbf{N}\}$

  $L(x^k y^l) = \{ x^{k+a} y^{l+b} \mid a, b \in \mathbf{N} \}$

  Then the class C = $\{ L(x^m y^n) \mid m, n \in \mathbf{N}\}$ is identifiable from positive representation.

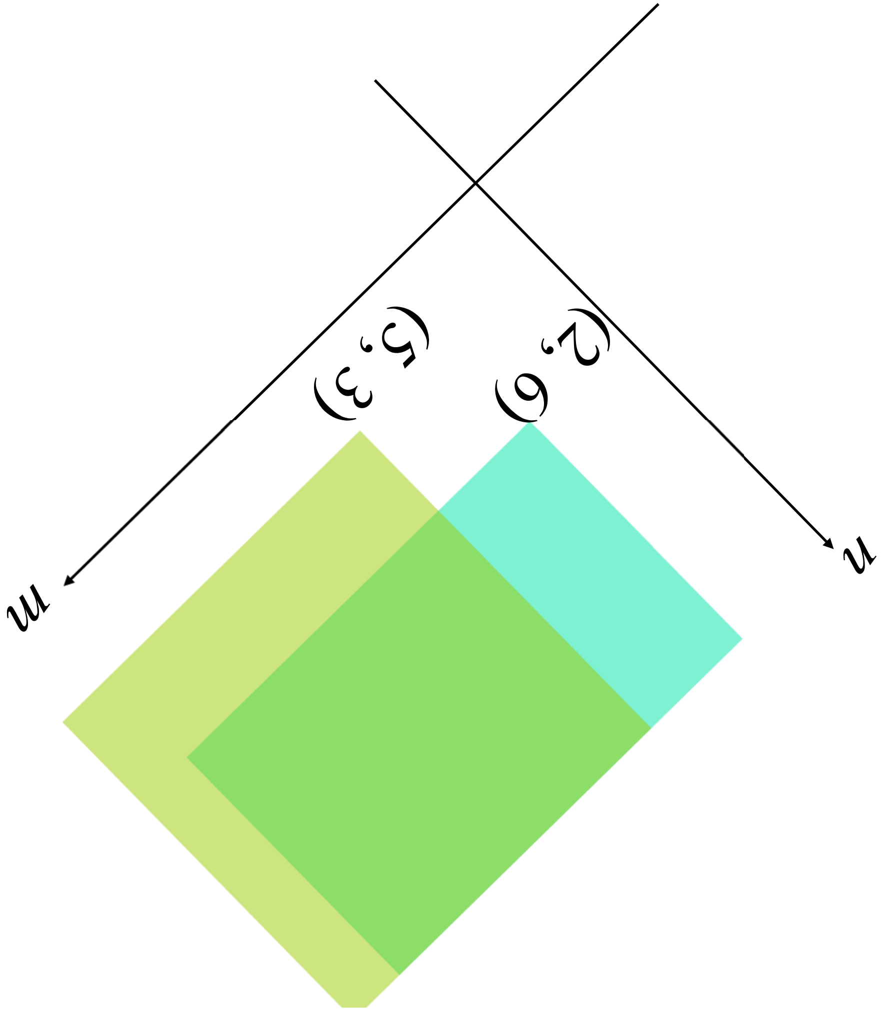
Monomial and Hasse Diagram

$$x^m y^n \Rightarrow (m, n)$$



Dicksonの補題(2)

$$x^m y^n \Rightarrow (m, n)$$



Rings

Ring R : an algebraic structure with operations of addition + and multiplication ·.

- The addition must satisfy the followings:
  - $(a + b) + c = a + (b + c)$ for all a, b, c in R (+ is associative).
  - $a + b = b + a$ for all a, b in R (+ is commutative).
  - There is an element 0 in R such that $a + 0 = a$ for all a in R (0 is the additive identity).
  - For each a in R there exists $-a$ in R such that $a + (-a) = 0$ ($-a$ is the additive inverse of a).
- The multiplication must satisfy the followings:
  - $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ for all a, b, c in R (· is associative).
  - There is an element 1 in R such that $a \cdot 1 = a$ and $1 \cdot a = a$ for all a in R (1 is the multiplicative identity).

Rings (cnt'd)

- The addition and multiplication must satisfy the followings:
  - $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$ for all a, b, c in R (left distributivity).
  - $(b + c) \cdot a = (b \cdot a) + (c \cdot a)$ for all a, b, c in R (right distributivity).

- If $a \cdot b = b \cdot a$ for all a, b in R ($\cdot$ is commutative), R is called a commutative ring.

Examples other than the set of integers **Z**:

- the set of all n-by-n matrices whose elements are **Q** (or **R**).
- **Q**$[x_1, x_2,\ldots, x_n]$ (**R**$[x_1, x_2,\ldots, x_n]$) : the set of all polynomials of variables $x_1, x_2,\ldots, x_n$ whose coefficients are in **Q** (or **R**).

Polynomials(1)

- Polynomials containing only one variable *x* (with coefficients in **Q**) have similar properties to integers.
  - **Q**[*x*] is a ring, moreover, division and reminder of two polynomials are defined, the GCD of two polynomials is defined and computed by Euclidian algorithm.

Example In **Q**[*x*] the GCD of $x^3 + 2x^2 + x$ and $x^2 + 3x + 2$ is computed as follows:

$$(x^3 + 2x^2 + x) \div (x^2 + 3x + 2) = x \ \ldots - x^2 - x$$

$$(x^2 + 3x + 2) \div (- x^2 - x) = 1 \ \ldots \ 2x + 2$$

$$(- x^2 - x) \div (2x + 2) = - x /2 \ \ldots 0$$

$$\mathrm{GCD}(x^3 + 2x^2 + x, x^2 + 3x + 2) = k(x + 1)$$

Polynomials(2)

- For polynomials containing more than one variables (with coefficients in **Q**) we cannot have simple extension of the operations for **Q**[x].

Example In **Q**[x, y] the division of two polynomials is not unique:

$$(x^2y^3 + x^2) \div (xy + y^3) = xy^2 - y^4 \ \ldots\ x^2 + y^7$$

$$(x^2y^3 + x^2) \div (y^3 + xy) = x^2 \ \ldots\ - x^3y + x^2$$

- To keep the uniqueness of the division and reminder, we fix the ordering of monomials.

Example Both of the expressions $x^2y^3 + x^2$ and $xy + y^3$ follow the lexicographic order , but $y^3 + xy$ does not.

Ideals (1)

- In algebra, the set of integers

$$L(m) = \{n \in \mathbf{Z} \mid |\, n \,| \bmod m = 0\}$$
$$= \{\, km \mid \; k \in \mathbf{Z} \,\}$$

is called the ideal generated by m and denoted by $\langle m \rangle$

- This can be defined for any commutative ring R:

$$\langle a \rangle = \{\, ra \mid \; r \in R \,\}$$

and extended as

$$\langle a_1, a_2, \ldots, a_n \rangle = \{\, r_1 a_1 + r_2 a_2 + \ldots + r_n a_n \mid r_1, r_2, \ldots, r_n \; \in R \,\},$$

which is called the ideal generated by $\{a_1, a_2, \ldots, a_n\}$.

Ideals (2)

- The set $I = \langle a_1, a_2,\ldots, a_n\rangle$ has the following properties:

  $0 \in I$

  If $b_1 \in I$ and $b_2 \in I$, $b_1 + b_2 \in I$

  If $b \in I$ and $r \in R$, $r\, b \in I$
- Any subset of R which satisfies the property above is called an ideal.
- For an ideal I, if there exists a set $\{b_1, b_2,\ldots, b_n\} \subset R$ such that $I = \langle b_1, b_2,\ldots, b_n\rangle$, the set is called a basis of I.
  - The basis of I is not unique.

Ideals (3)

- Example For $\mathbf{Z}$ and $\mathbf{Q}[x]$, it is known that

$$\langle m_1, m_2, \ldots, m_n \rangle = \langle \mathrm{GCD}(m_1, m_2, \ldots, m_n) \rangle$$

- This example means that GCD is the standard basis of $\langle m_1, m_2, \ldots, m_n \rangle$.

Note on Groebner Bases

- The concept Groebner bases is one of foundations of mathematical systems, e.g. Mathematica, Maple, …
- The notion of standard basis was independently invented by Hironaka.

Reduced Groebner Bases

- Consider $\mathbf{Q}[x_1, x_2,\ldots, x_n]$ and assume the ordering of monomials is fixed.
- Buchberger invented a transformation of any finite set of polynomials $\{f_1, f_2,\ldots, f_m\}$ into $\{ g_1, g_2,\ldots, g_k\}$ so that $\langle f_1, f_2,\ldots, f_m \rangle = \langle g_1, g_2,\ldots, g_k\rangle$. The set is determined uniquely and called the reduced Groebner bases.

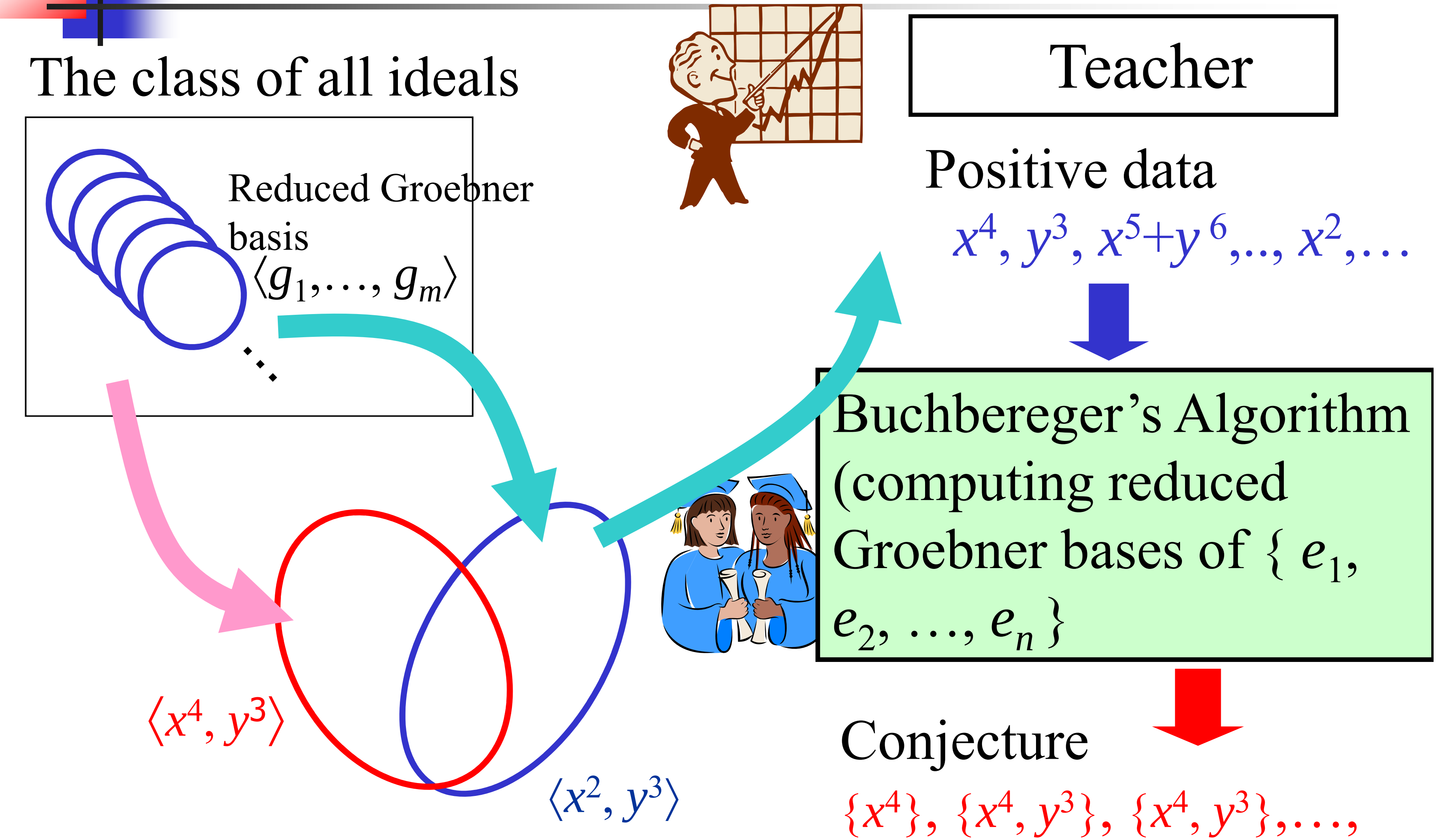
  Moreover he invented an algorithm which takes a pair of a set $\{f_1, f_2,\ldots, f_n\}$ and a polynomial g as an input and outputs .

  - The concept Groebner bases are one of foundations of mathematical systems
  - The notion of standard basis was independently invented by Hironaka

Hilbert’s bases theorem

Theorem[Hilbert] For any ideal I in $\mathbf{Q}[x_1,\ldots, x_n]$ there exist $f_1, f_2, \ldots , f_m \in \mathbf{Q}[x_1,\ldots, x_n]$ such that $I = \langle f_1, f_2,\ldots, f_m \rangle$.

Learning Ideals of Q[x,…, x_n]



The class of all ideals

Reduced Groebner basis $\langle g_1,\ldots, g_m\rangle$

$\langle x^4, y^3\rangle$

$\langle x^2, y^3\rangle$

Teacher

Positive data

$x^4, y^3, x^5+y^6,\ldots, x^2,\ldots$

Buchbereger’s Algorithm (computing reduced Groebner bases of { e_1, e_2, …, e_n }

Conjecture

$\{x^4\}, \{x^4, y^3\}, \{x^4, y^3\},\ldots,$

$\{x^2, y^3\}$,…

Basis as Characteristic Sets

- In the context of learning from positive examples, the reduced Groebner basis $\{g_1, g_2, \ldots, g_m\}$ works as a characteristic set of an ideal.

C2: The Characteristic Set Property

- A subset $C(g)$ of a language of $L(g)$ is a characteristic set of $L(g)$ in L(H) if

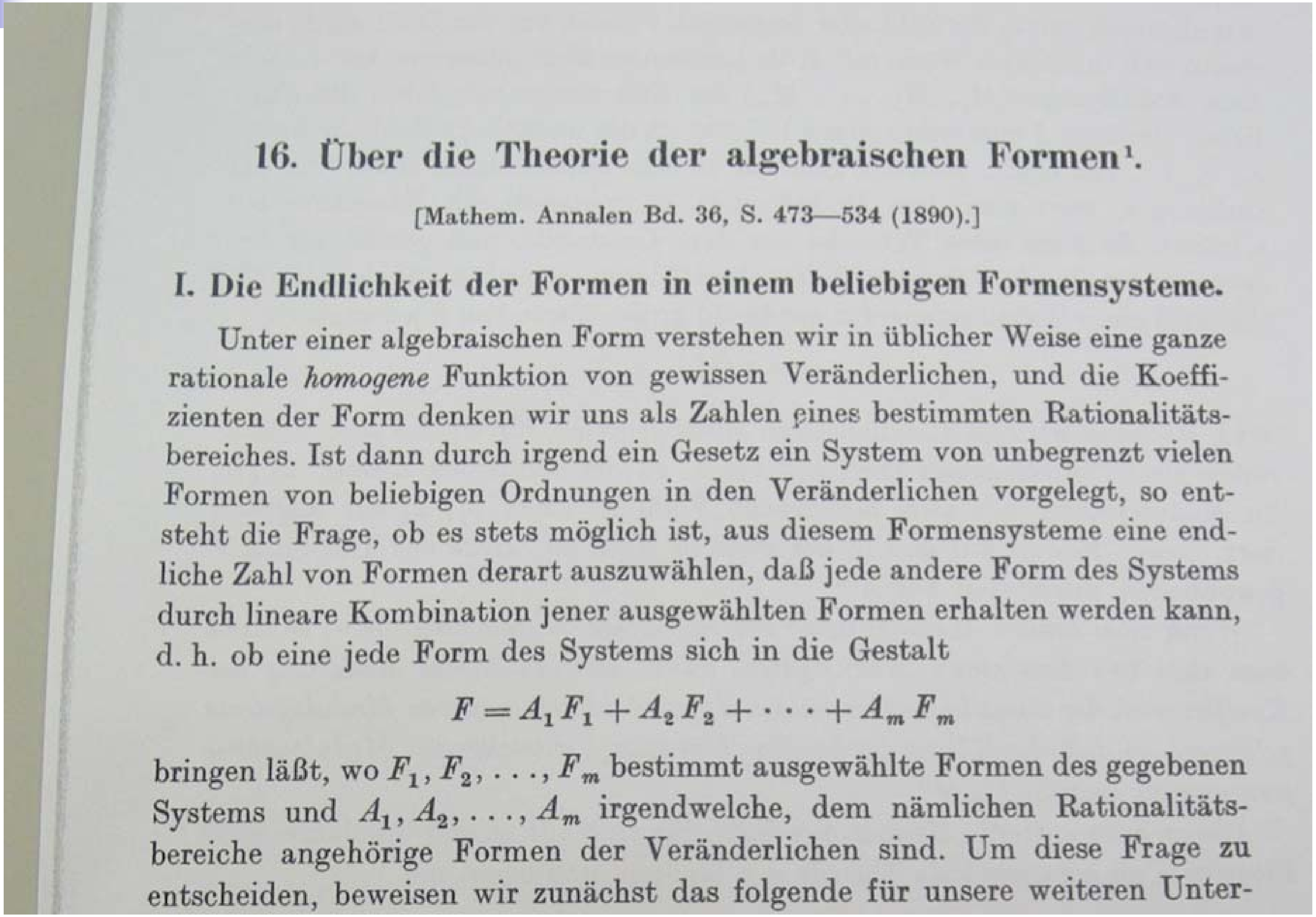
  (1) $C(g)$ is a finite set and

  (2) for every $L(g') \in$ L(H) $C(g) \subseteq L(g')$ implies

  $L(g) \subseteq L(g')$

Theorem [Kobayashi] A class L(H) of languages is identifiable in the limit from positive presentation if every language $L(g)$ in L(H) has a characteristic set $C(g)$ in L(H).

Hilbert's original paper



16. Über die Theorie der algebraischen Formen[1].

[Mathem. Annalen Bd. 36, S. 473—534 (1890).]

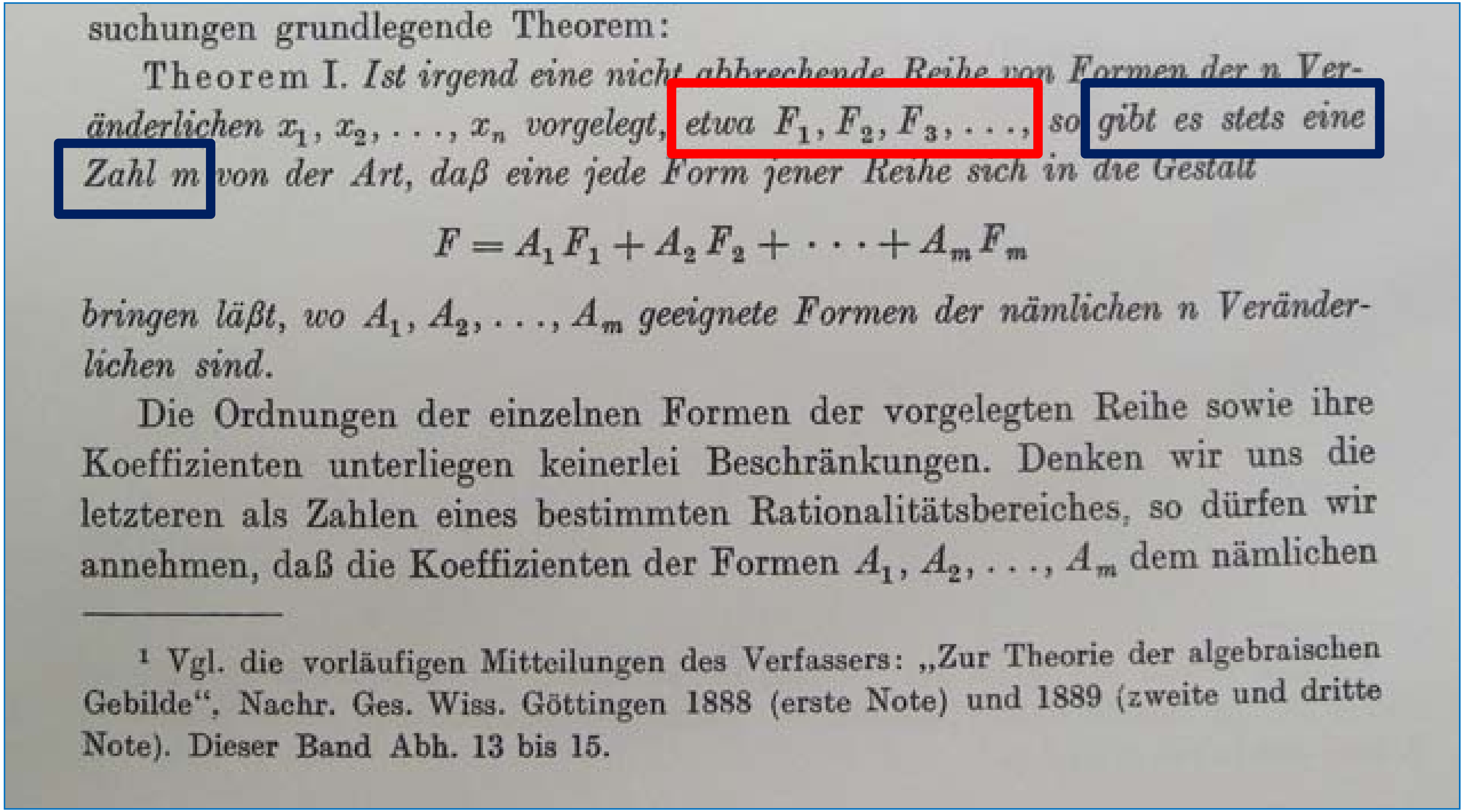
I. Die Endlichkeit der Formen in einem beliebigen Formensysteme.

Unter einer algebraischen Form verstehen wir in üblicher Weise eine ganze rationale *homogene* Funktion von gewissen Veränderlichen, und die Koeffizienten der Form denken wir uns als Zahlen eines bestimmten Rationalitätsbereiches. Ist dann durch irgend ein Gesetz ein System von unbegrenzt vielen Formen von beliebigen Ordnungen in den Veränderlichen vorgelegt, so entsteht die Frage, ob es stets möglich ist, aus diesem Formensysteme eine endliche Zahl von Formen derart auszuwählen, daß jede andere Form des Systems durch lineare Kombination jener ausgewählten Formen erhalten werden kann, d. h. ob eine jede Form des Systems sich in die Gestalt

$$F = A_1 F_1 + A_2 F_2 + \cdots + A_m F_m$$

bringen läßt, wo $F_1, F_2, \ldots, F_m$ bestimmt ausgewählte Formen des gegebenen Systems und $A_1, A_2, \ldots, A_m$ irgendwelche, dem nämlichen Rationalitätsbereiche angehörige Formen der Veränderlichen sind. Um diese Frage zu entscheiden, beweisen wir zunächst das folgende für unsere weiteren Unter-

Hilbert's original paper



suchungen grundlegende Theorem:

Theorem I. *Ist irgend eine nicht abbrechende Reihe von Formen der n Veränderlichen $x_1, x_2, \ldots, x_n$ vorgelegt, etwa $F_1, F_2, F_3, \ldots,$ so gibt es stets eine Zahl m von der Art, daß eine jede Form jener Reihe sich in die Gestalt*

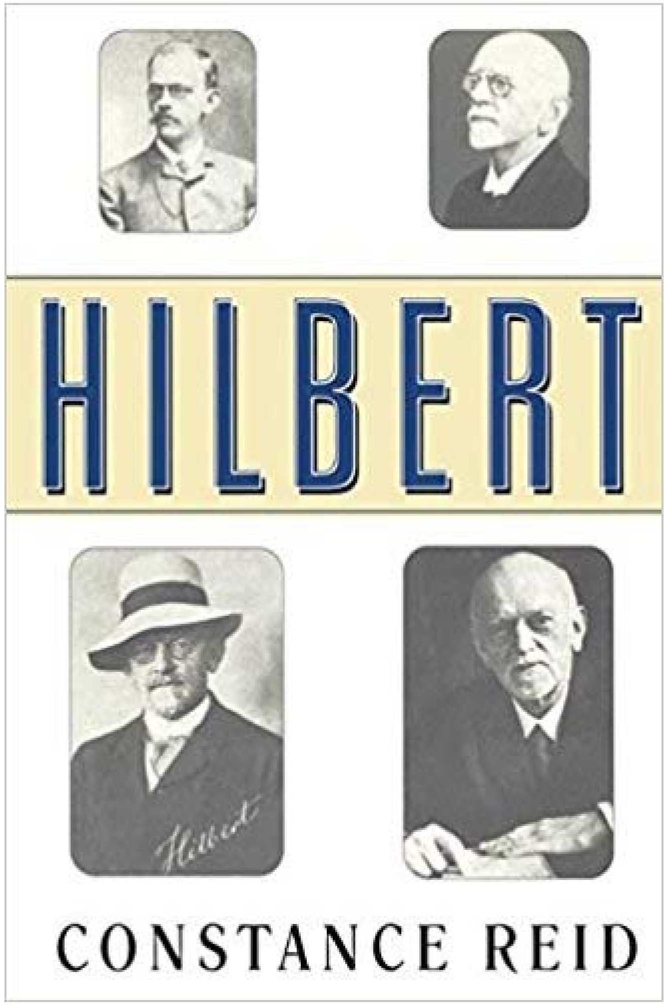
$$F = A_1 F_1 + A_2 F_2 + \cdots + A_m F_m$$

bringen läßt, wo $A_1, A_2, \ldots, A_m$ geeignete Formen der nämlichen n Veränderlichen sind.

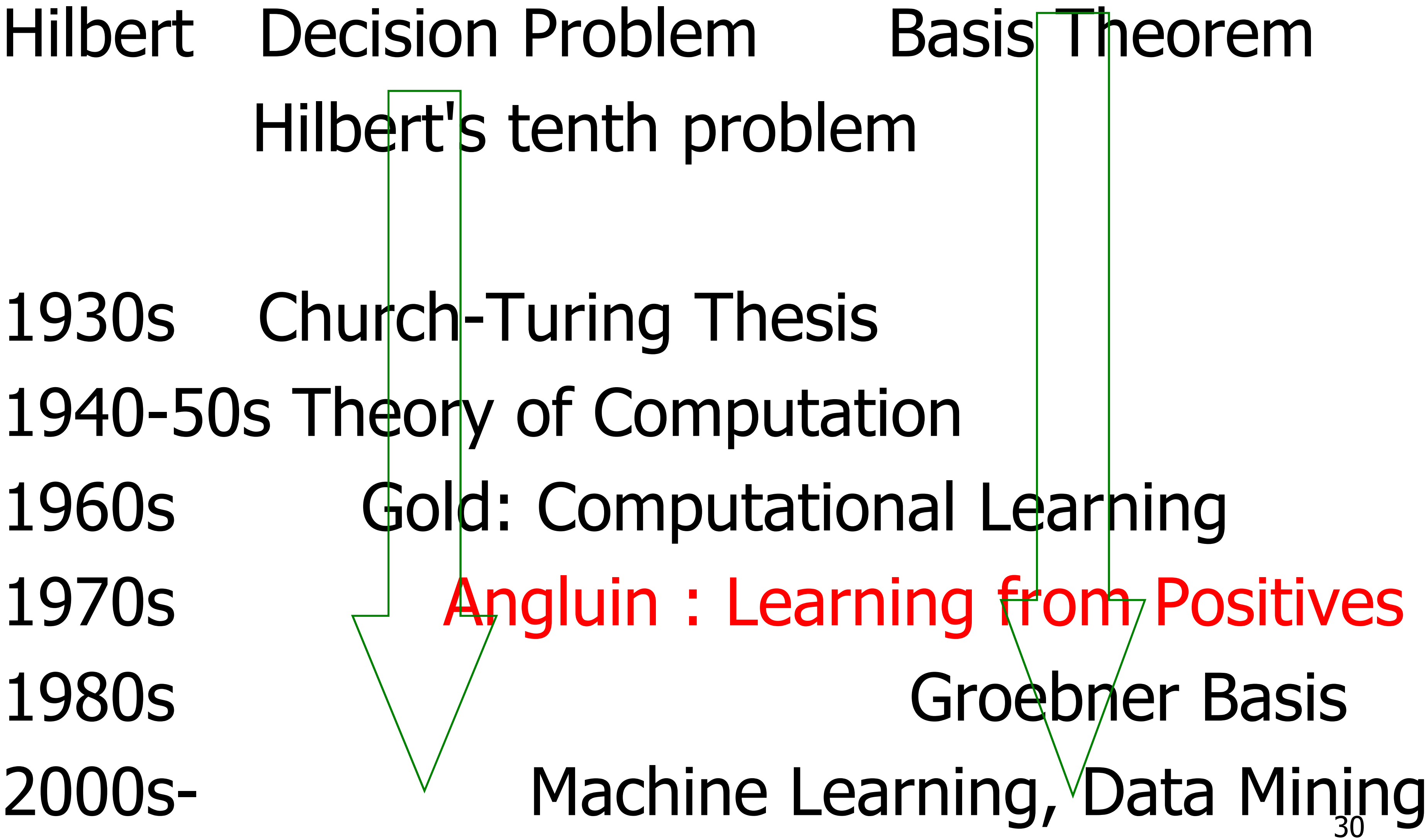
Die Ordnungen der einzelnen Formen der vorgelegten Reihe sowie ihre Koeffizienten unterliegen keinerlei Beschränkungen. Denken wir uns die letzteren als Zahlen eines bestimmten Rationalitätsbereiches, so dürfen wir annehmen, daß die Koeffizienten der Formen $A_1, A_2, \ldots, A_m$ dem nämlichen

[1] Vgl. die vorläufigen Mitteilungen des Verfassers: „Zur Theorie der algebraischen Gebilde", Nachr. Ges. Wiss. Göttingen 1888 (erste Note) und 1889 (zweite und dritte Note). Dieser Band Abh. 13 bis 15.

History of Mathmatics



History of Math and CS



Hilbert Decision Problem Basis Theorem

Hilbert's tenth problem

1930s Church-Turing Thesis

1940-50s Theory of Computation

1960s Gold: Computational Learning

1970s Angluin : Learning from Positives

1980s Groebner Basis

2000s- Machine Learning, Data Mining