ENCYCLOPEDIA OF MATHEMATICS AND ITS APPLICATIONS 77

POLYNOMIALS WITH SPECIAL REGARD TO REDUCIBILITY

ANDRZEJ SCHINZEL

CAMBRIDGE

more information - www.cambridge.org/0521662257

This page intentionally left blank

ENCYCLOPEDIA OF MATHEMATICS AND ITS APPLICATIONS

EDITED BY G.-C. ROTA

Editorial Board

R. S. Doran, M. Ismail, T.-Y. Lam, E. Lutwak, R. Spigler

Volume 77

Polynomials with Special Regard to Reducibility

ENCYCLOPEDIA OF MATHEMATICS AND ITS APPLICATIONS

4 W. Miller, Jr. *Symmetry and separation of variables*
6 H. Minc *Permanents*
11 W. B. Jones and W. J. Thron *Continued fractions*
12 N. F. G. Martin and J. W. England *Mathematical theory of entropy*
18 H. O. Fattorini *The Cauchy problem*
19 G. G. Lorentz, K. Jetter, and S. D. Riemenschneider *Birkhoff interpolation*
21 W. T. Tutte *Graph theory*
22 J. R. Bastida *Field extensions and Galois theory*
23 J. R. Cannon *The one-dimensional heat equation*
25 A. Salomaa *Computation and automata*
26 N. White (ed.) *Theory of matroids*
27 N. H. Bingham, C. M. Goldie, and J. L. Teugels *Regular variation*
28 P. P. Petrushev and V. A. Popov *Rational approximation of real functions*
29 N. White (ed.) *Combinatorial geometries*
30 M. Pohst and H. Zassenhaus *Algorithmic algebraic number theory*
31 J. Aczel and J. Dhombres *Functional equations containing several variables*
32 M. Kuczma, B. Chozewski, and R. Ger *Iterative functional equations*
33 R. V. Ambartzumian *Factorization calculus and geometric probability*
34 G. Gripenberg, S.-O. Londen, and O. Staffans *Volterra integral and functional equations*
35 G. Gasper and M. Rahman *Basic hypergeometric series*
36 E. Torgersen *Comparison of statistical experiments*
37 A. Neumaier *Interval methods for systems of equations*
38 N. Korneichuk *Exact constants in approximation theory*
39 R. A. Brualdi and H. J. Ryser *Combinatorial matrix theory*
40 N. White (ed.) *Matroid applications*
41 S. Sakai *Operator algebras in dynamical systems*
42 W. Hodges *Model theory*
43 H. Stahl and V. Totik *General orthogonal polynomials*
44 R. Schneider *Convex bodies*
45 G. Da Prato and J. Zabczyk *Stochastic equations in infinite dimensions*
46 A. Bjorner, M. Las Vergnas, B. Sturmfels, N. White, and G. Ziegler *Oriented matroids*
47 E. A. Edgar and L. Sucheston *Stopping times and directed processes*
48 C. Sims *Computation with finitely presented groups*
49 T. Palmer *Banach algebras and the general theory of *-algebras*
50 F. Borceux *Handbook of categorical algebra I*
51 F. Borceux *Handbook of categorical algebra II*
52 F. Borceux *Handbook of categorical algebra III*
54 A. Katok and B. Hassleblatt *Introduction to the modern theory of dynamical systems*
55 V. N. Sachkov *Combinatorial methods in discrete mathematics*
56 V. N. Sachkov *Probabilistic methods in discrete mathematics*
57 P. M. Cohn *Skew fields*
58 Richard J. Gardner *Geometric tomography*
59 George A. Baker, Jr. and Peter Graves-Morris *Padé approximants*
60 Jan Krajicek *Bounded arithmetic, propositional logic, and complex theory*
61 H. Gromer *Geometric applications of Fourier series and spherical harmonics*
62 H. O. Fattorini *Infinite dimensional optimization and control theory*
63 A. C. Thompson *Minkowski geometry*
64 R. B. Bapat and T. E. S. Raghavan *Nonnegative matrices and applications*
65 K. Engel *Sperner theory*
66 D. Cvetkovic, P. Rowlinson and S. Simic *Eigenspaces of graphs*
67 F. Bergeron, G. Labelle and P. Leroux *Combinatorial species and tree-like structures*
68 R. Goodman and N. Wallach *Representations of the classical groups*
69 H. Beth, D. Jungnickel, and T. Beth *Design theory I*
70 A. Pietsch and J. Wenzel *Orthonormal systems and Banach space geometry*
71 G. Andrews, R. Askey, and R. Ray *Special functions*
72 R. Ticciati *Quantum-field theory for mathematicians*
73 M. Stern *Semimodular lattices*
76 A. A. Ivanov *Geometry of sporadic groups I*

ENCYCLOPEDIA OF MATHEMATICS AND ITS APPLICATIONS

Polynomials with Special Regard to Reducibility

A. SCHINZEL

CAMBRIDGE UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

© Cambridge University Press 2004

First published in printed format 2000

ISBN 0-511-03370-2 eBook (Adobe Reader)
ISBN 0-521-66225-7 hardback

Contents

Preface

It is my pleasant duty to thank here for all the help I received in the preparation of this book.

Colin Day, Director of the University of Michigan Press has permitted me to reuse material from my book *Selected Topics on Polynomials* published by the Press in question.

Professors Francesco Amoroso, David W. Boyd, Pierre Dèbes, Kálmán Győry, Gerhard Turnwald and Umberto Zannier have on my request read parts of the book, corrected mistakes and suggested many improvements. Chapter 1, Sections 1–3 of Chapter 3 and Section 9 of Chapter 5 have been read by U. Zannier. He has also written a very important appendix 'Proof of Conjecture 1'. Chapter 2 has been read by G. Turnwald, who has also made most useful comments on Appendix A. Section 4 of Chapter 3 has been read by D.W. Boyd, Sections 1, 2, 3 of Chapter 4 by F. Amoroso, Section 4 of Chapter 5 and Sections 1–8 of Chapter 5 by P. Dèbes, finally Chapter 6 by K. Győry. In addition the whole book has been generously proofread by Jadwiga Lewkowicz and Andrzej Mąkowski, and the beginning of Chapter 1 by Andrzej Kondracki. I have also profited by advice from Dr. Michael Zieve concerning Section 5 of Chapter 4, from Professors Dieter Geyer, David Masser and Peter Roquette concerning Section 4 of Chapter 4 and from Professors Zbigniew Ciesielski, Piotr Mankiewicz, Aleksander Pełczyński and Dr. Marcin Kuczma concerning Appendix G.

The typing was patiently done in the Institute of Mathematics of the Polish Academy of Sciences by Joanna Zemła, Katarzyna Szynkiewicz and Anna Poczmańska with the help of Dr. Jan Kowalski.

In the first stage of the work on the book I was supported by grant PB 500/2/91 from the Polish Committee for Scientific Research.

Andrzej Schinzel

Acknowledgments

D. Coppersmith & J. Davenport 1991, Polynomials whose powers are sparse, *Acta Arith.* **58**, 79–87 is reproduced by kind permission of D. Coppersmith.

G. Turnwald 1995, On Schur's conjecture, *J. Austral. Math. Soc. Ser.* A. **58**, 312–357 is reproduced by kind permission of the Australian Mathematical Society.

W. Lawton 1983, A problem of Boyd concerning geometric means of polynomials, *J. Number Theory* **16**, 356–362 is reproduced by kind permission of Academic Press.

W. Ruppert 1986, Reduzibilität ebener Kurven, *J. Reine angew. Math.* **369**, 167–191
and U. Zannier 1993, Ritt's second theorem in arbitrary characteristic, *J. Reine angew. Math.* **445**, 175–203 are reproduced by kind permission of Walter de Gruyter & Co. K G Publishers.

A. Bazylewicz 1976, On the product of the conjugates outside the unit circle of an algebraic integer, *Acta Arith.* **30**, 43–61 is reproduced by kind permission of A. Bazylewicz.

Introduction

This book is an attempt to cover most of the results on reducibility of polynomials over fairly large classes of fields; results valid only over finite fields, local fields or the rational field have not been included. On the other hand, included are many topics of interest to the author that are not directly related to reducibility, e.g. Ritt's theory of composition of polynomials.

Here is a brief summary of the six chapters.

Chapter 1 (Arbitrary polynomials over an arbitrary field) begins with Lüroth's theorem (Sections 1 and 2). This theorem is nowadays usually presented with a short non-constructive proof, due to Steinitz. We give a constructive proof and present the consequences Lüroth's theorem has for subfields of transcendence degree 1 of fields of rational functions in several variables. The much more difficult problem of the minimal number of generators for subfields of transcendence degree greater than 1 belongs properly to algebraic geometry and here only references are given.

The next topic to be considered (Sections 3 and 4) originated with Ritt. Ritt 1922 gave a complete analysis of the behaviour of polynomials in one variable over $\mathbb{C}$ under composition. He called a polynomial prime if it is not the composition of two polynomials of lower degree and proved the two main results:

(i) In every representation of a polynomial as the composition of prime polynomials the number of factors is the same and their degrees coincide up to a permutation.

(ii) If A, H and B, G are polynomials of relatively prime degrees m and n, respectively, and

$$A(G) = B(H), \tag{1}$$

then A, B, G, H can be given explicitly.

Ritt showed also how every representation of a polynomial as the composition of prime polynomials can be obtained from a given one by solving several equations of the form (1), where A and B are prime.

We present an extension of Ritt's result to polynomials over an arbitrary field, for (ii) obtained only recently by Zannier 1993. Ritt's term 'prime' is replaced by 'indecomposable'.

Indecomposability plays an essential role in the next topic: reducibility of polynomials of the form $(f(x) - f(y))/(x - y)$ (Section 5). A necessary and sufficient condition for reducibility over fields of characteristic 0 was proved by Fried 1970. We give a proof of Fried's theorem published recently by Turnwald 1995 and summarize the more recent progress on this topic and the state of knowledge on reducibility of $f(x) - g(y)$, where g, h are polynomials. Section 6 contains results of Kronecker on factorization of polynomials. They include properties of the Kronecker substitution, a theorem of Kronecker once called fundamental and now nearly forgotten, that will be used later, and the theorem of Kronecker and A. Kneser. The latter describes a connection between reducibility of a polynomial $f \in \boldsymbol{k}[x]$ over $\boldsymbol{k}(\eta)$ and that of a polynomial $g \in \boldsymbol{k}[x]$ over $\boldsymbol{k}(\xi)$, where $f(\xi) = g(\eta) = 0$. Section 7 takes again the study of reducibility of polynomials with separated variables. H. Davenport and the author proved in 1963 that a polynomial of the form $F(x, y) + G(z)$ is reducible over a field $\boldsymbol{k}$ of characteristic 0 if and only if $F = H(A(x, y))$, $A, H \in \boldsymbol{k}[t]$ and $H(t) + G(z)$ is reducible over $\boldsymbol{k}$. Section 7 contains a natural generalization of this result and a discussion of the related results of Tverberg and Geyer. After some auxiliary results have been established in Section 8, a connection between irreducibility of a polynomial and of its substitution value after a specialization of some of the variables is treated in Section 9. This topic, connected with the names of Bertini and Hilbert, will be considered again in Chapter 3, Section 3 and Chapter 4, Section 4. The last Section 10 deals with the properties of the Newton polytope of a polynomial in many variables, a natural generalization of the Newton polygon.

Chapter 2 (Lacunary polynomials over an arbitrary field) begins with theorems of Capelli and M. Kneser. Capelli 1898 gave a simple necessary and sufficient condition for reducibility of a binomial $x^n - a$ over a subfield of $\mathbb{C}$. The case of positive characteristic was settled by Rédei 1967. The theorem can also be viewed as a necessary and sufficient condition for an element of a field $\boldsymbol{k}$ to satisfy the equality $[\boldsymbol{k}(\sqrt[n]{a}) : \boldsymbol{k}] = n$. In this aspect the theorem is open to generalization, specifically, one can study the degree $[\boldsymbol{k}(\sqrt[n_1]{a_1}, \sqrt[n_2]{a_2}, \dots, \sqrt[n_l]{a_l}) : \boldsymbol{k}]$. An all encompassing result in this direction for separable extensions has been found by M. Kneser 1975. It is reproduced in Section 1 together with a more immediate extension of Capelli's theorem.

It is an almost immediate consequence of Capelli's theorem that for $a \neq 0$ the polynomial $x^m + y^n + a$ is irreducible over every field of characteristic 0 containing a. This observation is generalized in Section 2 to an easily applicable irreducibility criterion for polynomials in many variables.

Following the work of Ritt 1927, Gourin 1933 proved that for a polynomial $F(x_1, \ldots, x_s)$ with more than two terms, irreducible over $\mathbb{C}$, and for arbitrary positive integers $t_1, \ldots, t_s$, the factorization of $F(x_1^{t_1}, \ldots, x_s^{t_s})$ into irreducible factors can be derived from the factorization of $F(x_1^{t_1}, \ldots, x_s^{t_s})$, where $\langle t_1, \ldots, t_s \rangle$ belongs to a finite set of integral vectors depending only on F. Gourin's proof applies with small modifications to polynomials over an arbitrary algebraically closed field and to integers $t_1, \ldots, t_s$ non-divisible by the characteristic of the field. An extension of the theorem to polynomials over fields no longer algebraically closed is given in Section 3. The only polynomials to which this extension does not apply apart from cx_i are of the form

$$F_0\left(\prod_{i=1}^{s} x_i^{\delta_i}\right) \prod_{i=1}^{s} x_i^{-d \min(0,\delta_i)}, \tag{2}$$

where $F_0(x)$ is a polynomial of degree d and $\delta_1, \ldots, \delta_s$ are integers, possibly negative.

The long Section 4 deals with reducibility of trinomials over any rational function field $\boldsymbol{k}(\boldsymbol{y})$. A necessary and sufficient condition for reducibility is given for any trinomial $x^n + Ax^m + B$ $(n > m > 0)$ such that $A^{-n}B^{n-m} \notin \boldsymbol{k}$ and $nm(n-m)$ is not divisible by the characteristic of $\boldsymbol{k}$. The cases $A \in \boldsymbol{k}$ and $B \in \boldsymbol{k}$ are given special attention. These results are used in Section 5 to characterize reducible quadrinomials depending essentially on at least two variables and such that the exponent vectors are all different modulo the characteristic of the ground field.

Section 6 presents a lower estimate for the number of non-zero coefficients of f^l in terms of l and of the number of non-zero coefficients of a polynomial f in one variable. An upper estimate is also given, valid in infinitely many essentially different cases.

Chapter 3 (Polynomials over an algebraically closed field) begins with the result of E. Noether, according to which a form of degree d in n variables is reducible over an algebraically closed field if and only if its coefficients satisfy a system of algebraic equations depending only on d and n (Section 1). Section 2 presents a theorem of Ruppert in which for $n = 3$ and characteristic 0 a system of equations with the above property is explicitly constructed. Section 3 is devoted to Bertini's theorem on reducibility. This theorem in its original formulation characterizes forms

$$f_0(\boldsymbol{x}) + \lambda_1 f_1(\boldsymbol{x}) + \cdots + \lambda_n f_n(\boldsymbol{x})$$

defined over $\mathbb{C}$ that become reducible over $\mathbb{C}$ for every choice of parameters $\lambda_1, \ldots, \lambda_n$. We present an extension of this result to all algebraically closed fields with a proof due to Krull 1937.

Section 4 differs definitely from the former three in that it concerns exclusively polynomials over $\mathbb{C}$. For such polynomials, in any number of variables, Mahler has introduced a measure M, that is multiplicative, i.e. $M(fg) = M(f)M(g)$. This measure has many interesting properties itself and also helps to describe the behaviour at the multiplication of other measures, e.g. of the length, defined for a polynomial as the sum of the absolute values of its coefficients. Section 4 presents several theorems on the Mahler measure of polynomials over $\mathbb{C}$, some of them quite recent.

Chapter 4 (Polynomials over a finitely generated field) begins with an extension of Gourin's theorem (discussed in Chapter 2, Section 3) to polynomials of the form (2), which is possible for every finitely generated ground field $\boldsymbol{K}$, provided the polynomial F_0 is irreducible over $\boldsymbol{K}$ and has neither 0 nor roots of unity as zeros (Section 1). Section 2 presents the best known lower bound in terms of the degree for the Mahler measure of an irreducible non-cyclotomic polynomial with integer coefficients. This bound is used in Section 3 to the study of the following problem.

Suppose that P, Q are coprime polynomials over a field $\boldsymbol{K}$. Then there exists a number $c(P, Q)$ with the following property. If $P(\xi^{n_1}, \ldots, \xi^{n_k}) = Q(\xi^{n_1}, \ldots, \xi^{n_k}) = 0$ for some integers $n_1, \ldots, n_k$ and some $\xi \neq 0$ in the algebraic closure of $\boldsymbol{K}$ then either $\xi^q = 1$ for a positive integer q or there exist integers $\gamma_1, \ldots, \gamma_k$ such that

$$\sum_{i=1}^{k} \gamma_i n_i = 0 \quad \text{and} \quad 0 < \max_{1 \le i \le k} |\gamma_i| \le c(P, Q).$$

This is established in Section 3 only for $k \le 3$, $\boldsymbol{K}$ arbitrary and for k arbitrary, $\boldsymbol{K}$ of positive characteristic. The result is placed in Chapter 4 rather than in Chapter 2 since the decisive role is played by the field generated over the prime field of $\boldsymbol{K}$ by the coefficients of P and Q.

For $k > 3$, $\boldsymbol{K}$ of zero characteristic, the assertion is established in the appendix written by Umberto Zannier, entitled Proof of Conjecture 1. Indeed, in the first version of Section 3 the assertion in full generality was only conjectured and the name Conjecture has been retained.

Section 4 is devoted to Hilbert's irreducibility theorem. The simplest case of this theorem asserts that if a polynomial $F(x, t)$ is irreducible over $\mathbb{Q}$ as a polynomial in two variables then $F(x, t^*)$ is irreducible over $\mathbb{Q}$ for infinitely many integers t^*. Section 4 presents a much more general form of the theorem, in which in particular $\mathbb{Q}$ is replaced by an arbitrary finitely generated field. In order to prove the theorem in such generality we use a method of Eichler based on some deep properties of equations over finite fields, rather than the more elementary approach sufficient to establish the theorem for number fields.

Hilbert's theorem in its simplest form stated above is closely related to the following property of diophantine equations. If an algebraic equation $F(x, t) = 0$ is soluble in rational or integer x for a sufficiently large set of integers t, then it is soluble for x in $\mathbb{Q}(t)$ or $\mathbb{Q}[t]$, respectively. A question suggests itself, whether a similar statement holds for equations with a greater number of unknowns and parameters and with $\mathbb{Q}$ replaced by a number field $\boldsymbol{K}$. The bulk (Sections 1–8) of Chapter 5 (Polynomials over a number field) is devoted to the study of this question. Section 1 constitutes an introduction to Sections 2–8, therefore here we only explain the fact that many theorems proved in this section concern polynomials over $\mathbb{C}$ rather than over a number field. Specifically, in every such case the main difficulty lies in proving the theorem for polynomials over $\boldsymbol{K}$ and then the general statement follows by linear algebra.

The result of Section 9 is tantamount to the following theorem. Let $F \in \boldsymbol{K}[x_1, \ldots, x_s]$, where $\boldsymbol{K}$ is a number field, be irreducible over $\boldsymbol{K}$, not a scalar multiple of x_i and not of the form (2), where F_0 has roots of unity as zeros. Then there exists a number $c_0(\boldsymbol{K}, F)$ with the following property. If for some integers $n_1, \ldots, n_s$ the only zeros of $F(x^{n_1}, \ldots, x^{n_s})$ are 0 and roots of unity, then there exist integers $\gamma_1, \ldots, \gamma_k$ such that

$$\sum_{i=1}^{s} \gamma_i n_i = 0 \quad \text{and} \quad 0 < \max |\gamma_i| \leq c_0(\boldsymbol{K}, F).$$

The title of the last chapter 'Polynomials over a Kroneckerian field' itself requires an explanation. By a Kroneckerian field (a term due to K. Győry) we mean a totally real number field or a totally complex quadratic extension of such a field. Among polynomials defined over a Kroneckerian field and prime to the product of the variables, exceptional in several respects are polynomials called self-inversive, i.e. polynomials F that satisfy an identity

$$F(x_1^{-1}, \ldots, x_k^{-1}) \prod_{i=1}^{k} x_i^{d_i} = c\overline{F}(x_1, \ldots, x_k),$$

where d_i is the degree of F with respect to x_i, $c \in \mathbb{C}$ and the bar denotes complex conjugation.

Section 1 presents estimates for the Mahler measure of non-self-inversive polynomials. They are far better than the estimates true in general.

Section 2 shows, for arbitrary integers $n_1, \ldots, n_k$, how all non-self-inversive factors of a polynomial $F(x^{n_1}, \ldots, x^{n_k})$ irreducible over a Kroneckerian field $\boldsymbol{K}$ can be obtained together with their multiplicities from the factorization of finitely many polynomials

$$F\left(\prod_{i=1}^{r} y_i^{\nu_{i1}}, \ldots, \prod_{i=1}^{r} y_i^{\nu_{ik}}\right), \text{ where } \max|\nu_{ij}| \le c(\boldsymbol{K}, F).$$

For $k = 1$ this is a consequence of the result of Chapter 4, Section 1. For $k > 1$ there is an analogy between the two results, but the above result lies much deeper, concerning reducibility of polynomials in one variable. Probably a similar result is true for all factors of $F(x^{n_1}, \ldots, x^{n_k})$ irreducible over $\boldsymbol{K}$ that have neither 0 nor roots of unity as zeros, however this is far from being proved and Section 3 presents only some steps in this direction. As a consequence one obtains for a given algebraic number $a \neq 0, \pm 1$ and a given polynomial $f(x)$ with algebraic coefficients the existence of a polynomial

$$x^n + ax^m + f(x) \text{ irreducible over } \boldsymbol{K}(a, \boldsymbol{f}),$$

where $\boldsymbol{f}$ is the coefficient vector of f. Unfortunately, there is a very restrictive condition that the field $\boldsymbol{K}(a, \boldsymbol{f})$ should be linearly disjoint with all cyclotomic fields.

Section 4, the last one, gives an exposition of the work of Győry on reducibility over Kroneckerian fields of composite polynomials $F(G(x))$.

The choice of material has been dictated by the personal taste of the author; out of 82 theorems, 37 belong to him and out of these 23 (Theorems 23, 24, 52, 54, 56, 58–66, 72, 74–81) have not been published before with the same degree of generality. Also Theorems 17, 29, 43, 50, 51, 55, 57, 67–71 are technically new, although their crucial special cases have been published before. In particular, Theorem 43 is taken from an unpublished and now lost manuscript of the late J. Wójcik.

Theorems proved in the sequel, conjectures and definitions are numbered successively for the whole book except the appendices; lemmas, conventions, remarks, examples and formulae are numbered separately for each section.

The book is not self-contained, the reader is often referred to the following five books:

E. Hecke, *Lectures on the theory of algebraic numbers*,

S. Lang, *Algebra*,

H. Mann, *Introduction to algebraic number theory*,

W. Rudin, *Principles of mathematical analysis*,
W. Rudin, *Real and complex analysis*,

abbreviated as [H], [L], [M], [P], [R]. The definitions and the results needed to follow the exposition, not found in the above books, are collected in 10 appendices: A, B, C, D, E, F, G, I, J, K. The reference Theorem E5, say, means Theorem 5 of Appendix E, the reference Theorem [L] 10.1 means Theorem 10.1 of Lang's book.

At the end of the book there are an index of theorems and an index of definitions and conjectures covering the main part of the book, not the appendices. The index of terms covers the whole book. There is no index of names, but in the bibliography for each reference, except ones listed as standard, there are indicated pages, where this reference is cited.

Notation

The letters $\boldsymbol{k}$ and $\boldsymbol{K}$ are reserved for fields, in Chapters 4–6 the letter $\boldsymbol{K}$ denotes a finitely generated field.

char $\boldsymbol{k}$ is the characteristic of $\boldsymbol{k}$,
$\boldsymbol{k}^*$ is the multiplicative group of the field $\boldsymbol{k}$,
$\overline{\boldsymbol{k}}$ is the algebraic closure of $\boldsymbol{k}$,$\boldsymbol{k}^{\mathrm{sep}}$ the maximal subfield of $\overline{\boldsymbol{k}}$ separable over $\boldsymbol{k}$.
$O_{\boldsymbol{K}}$ is the ring of integers of a number field $\boldsymbol{K}$, disc $\boldsymbol{K}$ is its discriminant, $O_{\boldsymbol{K}}^*$ the group of units. For an extension $\boldsymbol{K}/\boldsymbol{k}$, tr.deg. $\boldsymbol{K}/\boldsymbol{k}$ is the transcendence degree of $\boldsymbol{K}$ over $\boldsymbol{k}$. For a finite extension $\boldsymbol{K}/\boldsymbol{k}$ the symbols $N_{\boldsymbol{K}/\boldsymbol{k}}$ and $\mathrm{Tr}_{\boldsymbol{K}/\boldsymbol{k}}$ denote the norm and the trace, respectively, from $\boldsymbol{K}$ to $\boldsymbol{k}$ or from $\boldsymbol{K}(x_1, \dots, x_n)$ to $\boldsymbol{k}(x_1, \dots, x_n)$, where $x_1, \dots, x_n$ are variables.
$\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ are the fields of rational, real and complex numbers, respectively,
$\mathbb{F}_q$ is the finite field of q elements,
$\mathbb{Z}$ is the ring of rational integers,
$\mathbb{N}$, $\mathbb{N}_0$, $\mathbb{R}_+$ are the sets of positive integers, non-negative integers and non-negative real numbers, respectively,
$\mathfrak{M}_{k,l}(S)$ is the set of all matrices with k rows and l columns and with entries from the set S, ${}^tM$, and rank M are the transpose and the rank of a matrix M, ${}^aM$ and $\det M$ the adjoint and the determinant of a square matrix M, respectively. Vectors are treated as matrices with one row. For a set S of vectors rank S is the number of linearly independent vectors in S.
$GL(\mathbb{Z}, n)$ is the multiplicative group formed by all elements of $\mathfrak{M}_{n,n}(\mathbb{Z})$ with determinant ± 1,
I_n is the identity matrix of order n.

Bold face letters denote fields or vectors; which of the two should be clear from the context; in addition $\boldsymbol{C}(F)$ and $\boldsymbol{M}(F)$ have a special meaning explained in Chapter 1, Section 10 and bold face letters are freely used in Chapter 4, Section 3. If $\boldsymbol{a}$ is a vector, a_i is its ith coordinate; for two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, $\boldsymbol{ab}$ and $\boldsymbol{a} \wedge \boldsymbol{b}$ denote the inner and the external product, respectively. German letters, except $\mathfrak{M}$ with subscripts, denote prime divisors and prime ideals, script letters usually denote groups.

If distinct bold face letters occur as arguments of a polynomial, it is assumed that the coordinates of the relevant vectors are independent variables. For a polynomial $F(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)$ over an integral domain D or a field $\boldsymbol{k}$:

$\partial_{\boldsymbol{x}_i} F$ is the maximum degree of F with respect to x, where x runs over all variables occurring in $\boldsymbol{x}_i$, if $n = 1$, $\partial_{\boldsymbol{x}_1} F =: \partial F$, however $\frac{\partial F}{\partial x}$ is the partial derivative of F with respect to x;

$\deg_{\boldsymbol{x}_i} F$ is the degree of F viewed as a polynomial in $\boldsymbol{x}_i$, if $n = 1$, $\deg_{\boldsymbol{x}_1} F =: \deg F$.
If $f = \frac{F}{G}$, where F, G are coprime polynomials, then $\deg f := \max\{\deg F, \deg G\}$.
If $f, g \in \boldsymbol{k}(\boldsymbol{x})$, $f \underset{\boldsymbol{k}}{\cong} g$ means that $fg^{-1} \in \boldsymbol{k} \setminus \{0\}$ (f, g are scalar multiples of each other) and $f \underset{\boldsymbol{k}}{\ncong} g$ means that the above relation does not hold. Further

$$F(\boldsymbol{x}) \overset{\text{can}}{\underset{D}{=}} \text{const} \prod_{\sigma=1}^{s} F_\sigma(\boldsymbol{x})^{e_\sigma}$$

means that

$$F(\boldsymbol{x}) \prod_{\sigma=1}^{s} F_\sigma(\boldsymbol{x})^{-e_\sigma} \in D \setminus \{0\},$$

the polynomials $F_\sigma \in D[\boldsymbol{x}]$ $(1 \le \sigma \le s)$ are irreducible over the quotient field of D and pairwise relatively prime, $e_\sigma \in \mathbb{N}$.

The leading coefficient of F is the coefficient of the first term of F in the antilexicographic order†. A polynomial with leading coefficient 1 is called monic, the greatest common divisor of non-zero polynomials is assumed to be monic,

$\operatorname{disc}_x F$ is the discriminant of F with respect to the variable x,

$\operatorname{cont} F$ is the content of F defined as the greatest common divisor of the coefficients of F, F is primitive if $\operatorname{cont} F = 1$. For rational functions f and g in one variable we set

$$f \circ g = f(g(x)).$$

For a rational function of the form

$$f(x_1, x_2, \ldots, x_n) = x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_n^{\alpha_n} F(x_1, x_2, \ldots, x_n),$$

† i.e. such a term $a \prod_{i=1}^{n} x_i^{\alpha_i}$ $(a \ne 0)$ that for every other term $b \prod_{i=1}^{n} x_i^{\beta_i}$ $(b \ne 0)$ there is a $k \ge 0$ satisfying $\alpha_i = \beta_i$ $(i \le k)$, $\alpha_{k+1} > \beta_{k+1}$.

where F is a polynomial prime to $x_1x_2\dots x_n$ we set

$$Jf(x_1, x_2, \dots, x_n) = F(x_1, x_2, \dots, x_n)$$

and consider the leading coefficient and the content of F as those of f. A homogeneous polynomial is called a form. A form $F \in \boldsymbol{k}[x, y]$ is called singular if it has a multiple factor over $\overline{\boldsymbol{k}}$, and non-singular otherwise.

$\operatorname{res}\begin{pmatrix} H_1, \dots, H_s \\ x_1, \dots, x_s \end{pmatrix}$ is the resultant of forms $H_1, \dots, H_s$ with respect to variables $x_1, \dots, x_s$.

Braces denote sets, card S is the cardinality of S, S^n is usually the Cartesian nth power of S, but occasionally, when $\boldsymbol{k}$ is a field, $\boldsymbol{k}^n = \{x^n : x \in \boldsymbol{k}\}$ and similarly for groups or rings. For sets A and $B : A \setminus B = \{x \in A : x \notin B\}$, $A - B = \{a - b : a \in A,\ b \in B\}$.

Parenthesis is used as above to denote matrices, but $(abc\dots)$ denotes the cycle $a \to b \to c \dots \to a$;

$(a, b, c, \dots)$ denotes the greatest common divisor of $a, b, c, \dots$, but occasionally $(a, b) = \{x \in \mathbb{R} : a < x < b\}$;

$\boldsymbol{k}(S)$ denotes the least field containing the field $\boldsymbol{k}$ and the set S,

$\boldsymbol{k}((\boldsymbol{x}))$ is the field of Laurent series over $\boldsymbol{k}$ of the variable vector $\boldsymbol{x}$.

Brackets $[a, b, c, \dots]$ denote the least common multiple of $a, b, c, \dots$, but occasionally, $[a, b] = \{x \in \mathbb{R} : a \le x \le b\}$, $[a, b) = \{x \in \mathbb{R} : a \le x < b\}$;

$[\boldsymbol{L} : \boldsymbol{K}]$ or $[\mathcal{H} : \mathcal{G}]$ denotes the degree of extension $\boldsymbol{L}/\boldsymbol{K}$ or the index of the group $\mathcal{G}$ in $\mathcal{H}$, depending on the context;

$D[S]$ denotes the least ring containing the ring D and the set S,

$D[[\boldsymbol{x}]]$ is the ring of power series over D of the variable vector $\boldsymbol{x}$.

For an $x \in \mathbb{R} : \lfloor x \rfloor = \max\{n \in \mathbb{Z} : n \le x\}$, $\lceil x \rceil = \min\{n \in \mathbb{Z} : n \ge x\}$.

Brackets $\langle\ \rangle$ denote vectors, $\mathcal{G}\langle S \rangle$ denotes the least group containing the group $\mathcal{G}$ and the set S, also if S is a set of permutations, $\langle S \rangle$ denotes the least group of permutations containing S.

$|\cdot|$ denotes an absolute value or the Euclidean norm (except in Chapter 1, Section 9), but $|\mathcal{G}|$, where $\mathcal{G}$ is a group, denotes the order of $\mathcal{G}$.

For $z \in \mathbb{C}$, $\overline{z}$ is the complex conjugate of z, Re z and Im z are the real and the imaginary part of z, respectively. For $A = (a_{ij}) \in \mathfrak{M}_{k,l}(\mathbb{C}) : \overline{A} = (\overline{a}_{ij})$, unless stated to the contrary. For $P \in \mathbb{C}[\boldsymbol{x}]$, $\overline{P}$ is the polynomial with the coefficients equal to the complex conjugates of the corresponding coefficients of P.

For $P \in \boldsymbol{k}[x]$, $P' = \dfrac{dP}{dx}$.

ζ_n is a primitive root of unity of order n,

μ is the Möbius function,

φ is the Euler function,
$\mathcal{S}_n$ is the symmetric group on n letters,
$b_n \ll a_n$ means that the sequence $b_n a_n^{-1}$ is bounded,
$O(a_n)$ is any sequnce such that $b_n \ll a_n$,
ord $_p\, a$ is the highest power to which a prime element p of a unique factorization domain or a prime ideal p of a Dedekind domain divides an element a of this domain. $p^\pi \| a$ means that ord $_p\, a = n$.

Here is the list of special symbols used in more than one section of the book, arranged alphabetically, except the last five:

$A_{\nu,\mu}$, $B_{\nu,\mu}$: Chapter 2, Section 4, Table 1,
$A^*_{\nu,\mu}$: Chapter 2, Section 4, Table 3,
$B^*_{\nu,\mu}$: Chapter 2, Section 4, Table 2,
$\boldsymbol{C}(F)$: Chapter 1, Section 10, Definition 9,
$\mathcal{C}_0(\boldsymbol{K}, r, s), \mathcal{C}_1(\boldsymbol{K}, r, s), \mathcal{C}_2(\boldsymbol{K}, r, s), \mathcal{C}_3(\boldsymbol{K}, r, s)$: Chapter 5, Section 1, Definitions 23–26,
$D_n(x, a)$, $D_n(x)$: Chapter 1, Section 4, Definition 3,
D_n: Chapter 2, Section 3, Theorem 24,
$d(\sigma)$: Chapter 5, Section 6, Convention 1,
$d(\mathcal{J})$: Chapter 5, Section 6, Convention 2,
$E(\alpha, \boldsymbol{K})$: Chapter 4, Section 1, Convention 2,
$h(A)$: Chapter 3, Section 4, Definition 13,
$H(f)$: Chapter 3, Section 2, Definition 12,
KF: Chapter 4, Section 3, Definition 20,
$L(f)$: Chapter 3, Section 2, Definition 12,
$L_{\boldsymbol{K}}F$: Chapter 6, Section 2, Definition 30,
$\boldsymbol{M}(F)$: Chapter 1, Section 10, Definition 9,
$M(F)$: Chapter 3, Section 4, Definition 14,
$M(\alpha)$: Chapter 4, Section 2, Definition 19,
$\mu(\boldsymbol{K})$: Chapter 5, Section 9, Convention 1,
$P_{n,d}(z, \boldsymbol{a})$: Chapter 3, Section 1, Convention,
S_d: Chapter 1, Section 6, Definition 5,
$\tau_j(x_1, \dots, x_m)$: Chapter 1, Section 6, Convention 2,
$\sim$: Chapter 1, Section 3, Definition 2,
$\boldsymbol{z}^A$, where A is a matrix: Chapter 3, Section 4, Convention 4,
$\|\ \|$: Chapter 3, Section 4, Definition 14,
$\sqcap$: Chapter 4, Section 2, Definition 19,
$\cong$: Definition A 8.

1

Arbitrary polynomials over an arbitrary field

1.1 Lüroth's theorem

We first prove

Theorem 1. *If* $\boldsymbol{k} \subset \boldsymbol{K} \subset \boldsymbol{k}(\boldsymbol{x})$, *then* $\boldsymbol{K} = \boldsymbol{k}(g_1, \dots, g_t)$, *where the* g_i *lie in* $\boldsymbol{k}(\boldsymbol{x})$. *If* $\operatorname{char} \boldsymbol{k} = 0$, $t \leq 1 + \operatorname{tr.deg.} \boldsymbol{K}/\boldsymbol{k}$.

Proof. Let $\boldsymbol{x} = \langle x_1, \dots, x_n \rangle$. By Theorem [L] 10.1 we have $\operatorname{tr.deg.} \boldsymbol{k}(\boldsymbol{x})/\boldsymbol{k} = n$, hence $r := \operatorname{tr.deg.} \boldsymbol{K}/\boldsymbol{k} \leq n$. Let $\{g_1, \dots, g_r\}$ be a transcendence basis of $\boldsymbol{K}/\boldsymbol{k}$. By the quoted theorem, one can renumber the xs so that $\{g_1, \dots, g_r, x_{r+1}, \dots, x_n\}$ is a transcendence basis for $\boldsymbol{k}(\boldsymbol{x})/\boldsymbol{k}$. We assert that

$$\begin{aligned}[\boldsymbol{K} \colon \boldsymbol{k}(g_1, \dots, g_r)] &\leq [\boldsymbol{K}(x_{r+1}, \dots, x_n) \colon \boldsymbol{k}(g_1, \dots, g_r, x_{r+1}, \dots, x_n)] \\ &\leq [\boldsymbol{k}(\boldsymbol{x}) \colon \boldsymbol{k}(g_1, \dots, g_r, x_{r+1}, \dots, x_n)] < \infty.\end{aligned}$$

The second and the third inequality are clear. Suppose that the first inequality is not true, so we have $y_1, \dots, y_s \in \boldsymbol{K}$, linearly independent over $\boldsymbol{k}(g_1, \dots, g_r)$, but linearly dependent over $\boldsymbol{k}(g_1, \dots, g_r, x_{r+1}, \dots, x_n)$; thus

$$b_1 y_1 + \cdots + b_s y_s = 0,$$

where

$$b_i = \sum_{\boldsymbol{j} \in \mathbb{N}_0^{n-r}} a_{i\boldsymbol{j}} x_{r+1}^{j_{r+1}} \dots x_n^{j_n}, \quad a_{i\boldsymbol{j}} \in \boldsymbol{k}(g_1, \dots, g_r).$$

We can write this as

$$\sum_{\boldsymbol{j} \in \mathbb{N}_0^{n-r}} x_{r+1}^{j_{r+1}} \dots x_n^{j_n} \sum_{i=1}^{s} a_{i\boldsymbol{j}} y_i = 0,$$

whence $\sum_{i=1}^{s} a_{ij} y_i = 0$ for all $\boldsymbol{j} \in \mathbb{N}_0^{n-r}$. By the assumption $a_{ij} = 0$ for all $i, \boldsymbol{j}$, so $b_i = 0$ for all $i \le s$. Thus our assertion is proved and we take $g_{r+1}, \ldots, g_t$ to be generators of $\boldsymbol{K}$ over $\boldsymbol{k}(g_1, \ldots, g_r)$. If char $\boldsymbol{K} = 0$, we need to add only one generator by Theorem [L] 7.14. $\square$

Remark. More generally, if $\boldsymbol{k} \subset \boldsymbol{K} \subset \boldsymbol{L}$ and $\boldsymbol{L}$ is finitely generated over $\boldsymbol{k}$ then $\boldsymbol{K}$ is finitely generated over $\boldsymbol{k}$.

It follows from Theorem [L] 10.1 that, in the notation of Theorem 1, $t \ge$ tr. deg. $\boldsymbol{K}/\boldsymbol{k}$. Lüroth's theorem states that in the case $n = 1$, we have here an equality.

Theorem 2. *If* $\boldsymbol{k} \subset \boldsymbol{K} \subset \boldsymbol{k}(x)$ *and* $\boldsymbol{K} \ne \boldsymbol{k}$, *then* $\boldsymbol{K} = \boldsymbol{k}(g)$, $g \in \boldsymbol{k}(x) \setminus \boldsymbol{k}$.

Proof. By Theorem 1 we have $\boldsymbol{K} = \boldsymbol{k}(g_1, \ldots, g_s)$, $g_i \in \boldsymbol{k}(x) \setminus \boldsymbol{k}$. Let $g_i = \frac{F_i}{G_i}$, where $F_i, G_i \in \boldsymbol{k}(x)$, $(F_i, G_i) = 1$. Consider the polynomials

$$F_i(t) - g_i G_i(t) \in \boldsymbol{k}(x)[t], \quad (i = 1, \ldots, s),$$

all divisible by $t - x$, and let their highest common factor be $\frac{D(x,t)}{d_0(x)}$, where $D(x, t)$ is primitive as a polynomial in t with the leading coefficient $d_0(x)$. Since $t - x \mid D(x, t)$ we have $D \notin \boldsymbol{k}[t]$. By Gauss's lemma ([L], Ch. V, §6)

$$F_i(t)G_i(x) - F_i(x)G_i(t) = D(x, t)C_i(x, t), \quad \text{where } C_i(x, t) \in \boldsymbol{k}[x, t].$$

Take i such that $\partial g_i = m$ is least. If $\partial_t D(x, t) < m$ then $\partial_t C_i > 0$. Suppose $\partial_x C_i(x, t) = 0$, say $C_i(x, t) = C_i(t)$. Let $F_i(t) \equiv \tilde{F}_i(t)\big(\operatorname{mod} C_i(t)\big)$, $\partial \tilde{F}_i < \partial C_i$, similarly $G_i(t) \equiv \tilde{G}_i(t)\big(\operatorname{mod} C_i(t)\big)$, $\partial \tilde{G}_i < \partial C_i$. We have $\tilde{F}_i(t)G_i(x) - F_i(x)\tilde{G}_i(t) \equiv 0\big(\operatorname{mod} C_i(t)\big)$ and comparing degrees in t we get $\tilde{F}_i(t)G_i(x) = F_i(x)\tilde{G}_i(t)$. But $(F_i, G_i) = 1$, hence either $F_i \in \boldsymbol{k}$ or $\tilde{F}_i(t) = 0$ and either $G_i \in \boldsymbol{k}$ or $\tilde{G}_i(t) = 0$. All four resulting cases are impossible, since $\partial g_i > 0$ and $(F_i, G_i) \not\equiv 0 (\operatorname{mod} C_i)$. Hence C_i depends on both x, t and $\partial_x D < m$. Now $\frac{D(x,t)}{d_0(x)}$ is monic in $\boldsymbol{k}(x)[t]$. Its coefficients belong to $\boldsymbol{K}$, have degree $< m$ and at least one coefficient must be non-constant since $D \notin \boldsymbol{k}[t]$. We add one of the non-constant coefficients to the generators $g_1, \ldots, g_s$ and repeat the whole procedure.

By repeating the procedure with the larger set of generators, we must come to a point where

$$\underset{i \ge 1}{\text{g.c.d.}} \{F_i(t) - g_i(x)G_i(t)\} = c\big(F_\nu(t) - g_\nu(x)G_\nu(t)\big), \quad c \in \boldsymbol{k}(x). \tag{1}$$

Then $g_\nu(x)$ is the required generator. Indeed, for each i

$$F_i(t) - g_i(x)G_i(t) = \big(F_\nu(t) - g_\nu(x)G_\nu(t)\big)C_i(t),\ C_i \in \boldsymbol{k}(x)[t].$$

Now in $\boldsymbol{k}(g_\nu)[t]$ for a given i there exist P, Q, R, S such that

$$\begin{aligned} F_i(t) &= P(t)[F_\nu(t) - g_\nu G_\nu(t)] + Q(t), \quad \partial_t Q < \partial_t[F_\nu(t) - g_\nu G(t)] \\ G_i(t) &= R(t)[F_\nu(t) - g_\nu G_\nu(t)] + S(t), \quad \partial_t S < \partial_t[F_\nu(t) - g_\nu G(t)]. \end{aligned}$$

If $Q = 0$, $F_i(t) = P(t)[F_\nu(t) - g_\nu G_\nu(t)]$ and writing $P(t)$ as $\frac{T(g_\nu,t)}{p(g_\nu)}$, where T, p are polynomials over $\boldsymbol{k}$, we get

$$\begin{aligned} F_i(t) &= \frac{T(g_\nu,t)}{p(g_\nu)}[F_\nu(t) - g_\nu G_\nu(t)], \\ F_i(t)p(g_\nu) &= T(g_\nu,t)[F_\nu(t) - g_\nu G_\nu(t)], \end{aligned}$$

which is impossible, since $F_\nu(t) - g_\nu G_\nu(t)$ does not factor in $\boldsymbol{k}[g_\nu, t]$.

Hence $Q \neq 0$ and similarly $S \neq 0$. Also

$$F_i(t) - g_i G_i(t) = [P(t) - g_i R(t)][F_\nu(t) - g_\nu G_\nu(t)] + Q(t) - g_i S(t).$$

It follows from (1) that $Q(t) = g_i S(t)$. Taking the leading coefficients q_0, s_0 of Q, S respectively we get

$$q_0 = g_i s_0 \in \boldsymbol{k}(g_\nu), \quad \text{so } g_i = \frac{q_0}{s_0} \in \boldsymbol{k}(g_\nu). \qquad \square$$

The above proof is constructive, that is it permits one to find a generator of $\boldsymbol{K}$ given as $\boldsymbol{k}(g_1, \dots, g_s)$ and to express $g_1, \dots, g_s$ in terms of this generator.

Notes. Theorem 1 was proved by E. Noether 1926 and rediscovered by Samuel 1953. The Remark is taken from Ojanguren 1990. Theorem 2 was proved by Lüroth 1876 for $\boldsymbol{k} = \mathbb{C}$, by Steinitz 1910 in general. Steinitz's proof, short but non-constructive, is reproduced in van der Waerden 1967. The proof given above is Ostrowski's 1936 proof, made effective by Chebotarev 1948, and not Netto's 1895 proof, as stated by mistake in [S].

If tr. deg. $\boldsymbol{K}/\boldsymbol{k} = 2$ and $\boldsymbol{k} = \mathbb{C}$ then in analogy with Lüroth's theorem $\boldsymbol{K} = \boldsymbol{k}(g_1, g_2)$ for suitable g_1, g_2 (Castelnuovo 1894). Castelnuovo's proof was simplified by Conforto 1939 (Chapter 7) and by Kodaira (see Algebraic Surfaces 1967, Chap. III), but it remains difficult and non-constructive. The case of algebraically closed fields of positive characteristic is treated by Zariski 1958. If $\boldsymbol{k}$ is not algebraically closed, e.g. if $\boldsymbol{k} = \mathbb{Q}$ or $\mathbb{R}$ the equality $\boldsymbol{K} = \boldsymbol{k}(g_1, g_2)$ need not hold, as shown by Segre 1951 and more recently by Ojanguren 1990. If tr. deg. $\boldsymbol{K}/\boldsymbol{k} = 3$ then, even for $\boldsymbol{k} = \mathbb{C}$, $\boldsymbol{K}/\boldsymbol{k}$ may need four generators (Artin and Mumford 1972, Clemens and Griffiths 1972 and

Iskovskih and Manin 1971, see also Ojanguren 1990, which however is not free from errors).

For an extension of Lüroth's theorem in a different direction see Moh and Heinzer 1979.

1.2 Theorems of Gordan and E. Noether

Theorem 3. *If* $\boldsymbol{k} \subset \boldsymbol{K} \subset \boldsymbol{k}(\boldsymbol{x})$*,* tr. deg. $\boldsymbol{K}/\boldsymbol{k} = 1$*, then* $\boldsymbol{K} = \boldsymbol{k}(g)$*,* $g \in \boldsymbol{k}(\boldsymbol{x})$.

Proof. Let $\boldsymbol{x} = \langle x_1, \ldots, x_n \rangle$. We shall first consider the case of $\boldsymbol{k}$ infinite. By Theorem 1 $\boldsymbol{K} = \boldsymbol{k}(\varphi_1, \ldots, \varphi_t)$. By Theorem [L] 10.1, on renumbering xs one can assume $x_2, \ldots, x_n$ are algebraically independent over $\boldsymbol{K}$. We have

$$\boldsymbol{k}(x_2, \ldots, x_n) \subset \boldsymbol{K}(x_2, \ldots, x_n) \subset \boldsymbol{k}(x_1, \ldots, x_n).$$

By Lüroth's theorem

$$\boldsymbol{K}(x_2, \ldots, x_n) = \boldsymbol{k}(x_2, \ldots, x_n, \eta), \quad \text{where } \eta \in \boldsymbol{k}(x_1, \ldots, x_n).$$

Hence

$$\varphi_i = g_i(\eta, x_2, \ldots, x_n), \quad \text{where } g_i \in \boldsymbol{k}(y_1, \ldots, y_n) \ \ (1 \le i \le t)$$

and

$$\eta = h(\varphi_1, \ldots, \varphi_t, x_2, \ldots, x_n), \quad \text{where } h \in \boldsymbol{k}(y_1, \ldots, y_t, x_2, \ldots, x_n).$$

Therefore

$$\varphi_i = g_i\big(h(\varphi_1, \ldots, \varphi_t, x_2, \ldots, x_n), x_2, \ldots, x_n\big) \ \ (1 \le i \le t) \tag{1}$$

identically over $\boldsymbol{K}$, since $x_2, \ldots, x_n$ are algebraically independent over $\boldsymbol{K}$. Choose values $x_2^*, \ldots, x_n^*$ in $\boldsymbol{k}$ so that after substitution $x_i = x_i^*$ the rational functions on the right hand side of (1) make sense. Now $h(\varphi_1, \ldots, \varphi_t, x_2^*, \ldots, x_n^*)$ is the desired generator for $\boldsymbol{K}/\boldsymbol{k}$, since

$$\varphi_i = g_i\big(h(\varphi_1, \ldots, \varphi_t, x_2^*, \ldots, x_n^*), x_1^*, \ldots, x_n^*\big) \quad \text{for all } i \le t.$$

If $\boldsymbol{k}$ is a finite field the above proof gives only the existence of a finite extension $\boldsymbol{k}_0$ of $\boldsymbol{k}$ such that $\boldsymbol{k}_0\,\boldsymbol{K} = \boldsymbol{k}_0(g_0)$, where g_0 is in $\boldsymbol{k}_0(x_1, \ldots, x_n)$. $\boldsymbol{k}_0$ should be large enough to contain values $x_2^*, \ldots, x_n^*$ with the property required above. Let

$$g_0 = P/Q, \quad \text{where } P,\ Q \in \boldsymbol{k}_0[x_1, \ldots, x_n],\ (P, Q) = 1.$$

Since $g_0 \notin \boldsymbol{k}_0$, there exist monomials M_1 and M_2 such that the coefficients p_i, q_i of M_i in P and Q respectively satisfy $p_1 q_2 - q_1 p_2 \neq 0$.

Now let σ be the substitution, which generates the Galois group of $\boldsymbol{k}_0/\boldsymbol{k}$ (the so-called Frobenius substitution). It operates in the obvious way on $\boldsymbol{k}_0[x_1,\dots,x_n]$ and we have $g_0^\sigma = P^\sigma/Q^\sigma$. On the other hand

$$\boldsymbol{k}_0(g_0^\sigma) = \boldsymbol{k}_0^\sigma(g_0^\sigma) = \big(\boldsymbol{k}_0(g_0)\big)^\sigma = (\boldsymbol{k}_0\boldsymbol{K})^\sigma = \boldsymbol{k}_0\boldsymbol{K} = \boldsymbol{k}_0(g_0),$$

hence (see [L], Chapter V, Exercise 9)

$$g_0^\sigma = \frac{ag_0+b}{cg_0+d} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} g_0, \quad \text{where } a,b,c,d \in \boldsymbol{k}_0 \text{ and } \begin{vmatrix} a & b \\ c & d \end{vmatrix} \neq 0.$$

Since $\frac{aP+bQ}{cP+dQ} = \frac{P^\sigma}{Q^\sigma}$ and $(P^\sigma, Q^\sigma) = 1 = (aP+bQ, cP+dQ)$, we have for suitable $e \in \boldsymbol{k}_0$ that $aP+bQ = eP^\sigma$, $cP+dQ = eQ^\sigma$. Comparing the coefficients of the monomial M_i on both sides we obtain

$$ap_i + bq_i = ep_i^\sigma,\ cp_i + dq_i = eq_i^\sigma,$$

which gives

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} p_1 & p_2 \\ q_1 & q_2 \end{pmatrix} = \begin{pmatrix} p_1^\sigma & p_2^\sigma \\ q_1^\sigma & q_2^\sigma \end{pmatrix}\begin{pmatrix} e & 0 \\ 0 & e \end{pmatrix}.$$

Putting

$$g = \begin{pmatrix} p_1 & p_2 \\ q_1 & q_2 \end{pmatrix}^{-1} g_0$$

we find

$$\begin{aligned} g^\sigma &= \begin{pmatrix} p_1^\sigma & p_2^\sigma \\ q_1^\sigma & q_2^\sigma \end{pmatrix}^{-1} g_0^\sigma = \begin{pmatrix} p_1^\sigma & p_2^\sigma \\ q_1^\sigma & q_2^\sigma \end{pmatrix}^{-1} \begin{pmatrix} a & b \\ c & d \end{pmatrix} g_0 \\ &= \begin{pmatrix} p_1^\sigma & p_2^\sigma \\ q_1^\sigma & q_2^\sigma \end{pmatrix}^{-1} \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} p_1 & p_2 \\ q_1 & q_2 \end{pmatrix} g = \begin{pmatrix} e & 0 \\ 0 & e \end{pmatrix} g. \end{aligned}$$

Hence $g \in \boldsymbol{K}$ and since $\boldsymbol{k}_0\boldsymbol{K} = \boldsymbol{k}_0(g_0) = \boldsymbol{k}_0(g)$ and $[\boldsymbol{k}_0\boldsymbol{K}:\boldsymbol{K}] = [\boldsymbol{k}_0:\boldsymbol{k}]$ we get $[\boldsymbol{k}_0(g):\boldsymbol{K}] = [\boldsymbol{k}_0:\boldsymbol{k}] = [\boldsymbol{k}_0(g):\boldsymbol{k}(g)]$, hence $\boldsymbol{K} = \boldsymbol{k}(g)$. □

Theorem 4. *If, under the assumption of Theorem 3, $\boldsymbol{K}$ contains a non-constant polynomial over $\boldsymbol{k}$, then $\boldsymbol{K}$ has a generator which is a polynomial over $\boldsymbol{k}$.*

We recall that for a polynomial F in one variable ∂F is the degree of F.

Lemma. *Let $P,\ Q \in \boldsymbol{k}[\boldsymbol{x}]$, $R,\ S \in \boldsymbol{k}[x]$, $(P,Q) = (R,S) = 1$, $R(x,y) = y^{\partial R} R\big(\frac{x}{y}\big)$, $S(x,y) = y^{\partial S} S\big(\frac{x}{y}\big)$.*

Then $Q,\ R(P,Q),\ S(P,Q)$ are prime in pairs.

Proof. We write $RU + SV = 1$, where $U,\ V \in \boldsymbol{k}[x]$, $\partial U < \partial S$, $\partial V < \partial R$ and we obtain

$$R(x, y)U(x, y) + S(x, y)V(x, y) = y^{\partial(RU)}.$$

Now we substitute $x = P$, $y = Q$ and obtain

$$R(P, Q)U(P, Q) + S(P, Q)V(P, Q) = Q^{\partial(RU)}.$$

The lemma follows since $\big(R(P, Q), Q\big) = \big(S(P, Q), Q\big) = 1$. □

Proof of Theorem 4. Let the generator g of $\boldsymbol{K}$ have the form

$$g = \frac{P}{Q}, \quad P,\ Q \in \boldsymbol{k}[\boldsymbol{x}],\ (P, Q) = 1.$$

By hypothesis there is a polynomial F in $\boldsymbol{K}$

$$F = \frac{R}{S}(g) = \frac{R(P/Q)}{S(P/Q)} = \frac{R(P, Q)}{S(P, Q)} Q^{s-r}, \quad r = \partial R,\ s = \partial S.$$

By the lemma, $S(P, Q) \in \boldsymbol{k}$, $Q \in \boldsymbol{K}$ or $s \geq r$. Factoring $S(P, Q)$ we obtain

$$S(P, Q) = \alpha(P - \xi_1 Q)(P - \xi_2 Q) \dots (P - \xi_s Q) \in \boldsymbol{k}^*,$$

hence

$$P - \xi_i Q = \gamma_i \in \bar{\boldsymbol{k}}.$$

Now $S(x)$ cannot have two different roots, since otherwise $(\xi_1 - \xi_2)Q = -\gamma_1 + \gamma_2$ implies successively $Q \in \bar{\boldsymbol{k}}$, $P \in \bar{\boldsymbol{k}}$, $g \in \bar{\boldsymbol{k}}$, which is impossible. Thus

$$S(P, Q) = \alpha(P - \xi Q)^s, \quad P - \xi Q = \gamma.$$

If $Q \notin \boldsymbol{k}$ then $\xi \in \boldsymbol{k}$, $\gamma \in \boldsymbol{k}$, $g = \frac{P}{Q} = \xi + \frac{\gamma}{Q}$ is expressed as a rational function of Q and we take Q to be a generator. If $Q \in \boldsymbol{k}$, we may take P to be a generator. □

Notes. Theorem 3 was proved by Gordan 1887 for $\boldsymbol{k} = \mathbb{C}$, by Igusa 1951 in general. The proof given above is due to Samuel 1953 for $\boldsymbol{k}$ infinite, to Laubie and Schinzel 1982 for $\boldsymbol{k}$ finite.

Theorem 4 was proved by E. Noether 1915 for char $\boldsymbol{k} = 0$, by Schinzel 1963b in general, and the latter proof is given above.

1.3 Ritt's first theorem

Convention. Ordinary capital letters denote polynomials in one variable.

Theorem 5. *If $\boldsymbol{k}(F) \cap \boldsymbol{k}(G)$ contains a polynomial H such that $\partial H \not\equiv 0 \bmod \operatorname{char} \boldsymbol{k}$, then*

$$[\boldsymbol{k}(F) : \boldsymbol{k}(F) \cap \boldsymbol{k}(G)] = \frac{[\partial F, \partial G]}{\partial F},$$
$$[\boldsymbol{k}(F, G) : \boldsymbol{k}(F)] = \frac{\partial F}{(\partial F, \partial G)}.$$

Lemma 1. *If $H \in \boldsymbol{k}(F)$, then $H = A(F)$.*

Proof. If $H = \frac{R}{S}(F)$, $(R, S) = 1$, then $\big(R(F), S(F)\big) = 1$ and hence $S \in \boldsymbol{k}$. □

Lemma 2. $[\boldsymbol{k}(x) : \boldsymbol{k}(F)] = \partial F$.

Proof. If $F = a_0 x^n + \cdots + a_n$, then $G = a_0 X^n + \cdots + a_n - F$ is an irreducible polynomial in $\boldsymbol{k}[F, X]$ because it is linear in F, whence it is irreducible over $\boldsymbol{k}(F)$ with x as a zero. □

Lemma 3. *If $A \in \boldsymbol{k}[x]$ is monic and $\partial A = r$, $r \not\equiv 0 (\operatorname{mod} \operatorname{char} \boldsymbol{k})$, then there exists a monic polynomial $C \in \boldsymbol{k}[x]$ such that $\partial C = n$, $\partial(A - C^r) < n(r-1)$.*

Proof. For each non-negative $i \le n$ there exists $C_i \in \boldsymbol{k}[x]$ such that $\partial C_i = n$, $\partial(A - C_i^r) < nr - i$. We prove this by induction on i. If $i = 0$, $C_0 = x^n$. Suppose the statement proved for $i - 1$, where $0 < i \le n$. Hence we have a polynomial C_{i-1} of degree n such that $\partial(A - C_{i-1}^r) < nr - i + 1$. We look for C_i of the form

$$C_i = C_{i-1} + \xi x^{n-i}.$$

We have

$$C_i^r = C_{i-1}^r + r C_{i-1}^{r-1} \xi x^{n-i} + \binom{r}{2} C_{i-1}^{r-2} \xi^2 x^{2(n-i)} + \cdots.$$

The degree of the third and latter terms is at most $n(r-2) + 2(n-i) = nr - 2i < nr - i$. Consider $A - C_{i-1}^r - r\xi C_{i-1}^{r-1} x^{n-i}$. We have $\partial(C_{i-1}^{r-1} x^{n-i}) = nr - i$. Select ξ so that the terms of degree $nr - i$ cancel each other and then $\partial(A - C_i^r) < nr - i$. Since C_0 is monic the construction ensures all C_i are monic. □

Proof of Theorem 5. By Lemma 1 and the hypothesis we have $H = A(F) = B(G)$. Without loss of generality we may assume H, F, G, A, B all monic. If $\partial F = n = d\nu$, $\partial G = m = d\mu$, where $(\mu, \nu) = 1$ we have $\partial H = rd\mu\nu \not\equiv 0(\text{mod char } \boldsymbol{k})$. Also $\partial A = r\mu$, $\partial B = r\nu$.

By Theorem 4, there exists a polynomial generating $\boldsymbol{k}(F) \cap \boldsymbol{k}(G)$; by Lemma 1 we may assume it without loss of generality to be H. We shall prove $r = 1$. By Lemma 3 there exist monic polynomials $C,\ D \in \boldsymbol{k}[x]$ such that $\partial C = \mu$, $\partial D = \nu$,

$$\partial(A - C^r) < \mu(r-1), \quad \partial(B - D^r) < \nu(r-1).$$

Hence

$$\partial\big(A(F) - C^r(F)\big) < d\mu\nu(r-1),$$

$$\partial\big(B(G) - D^r(G)\big) < d\mu\nu(r-1),$$

$$\partial\big(C^r(F) - D^r(G)\big) < d\mu\nu(r-1).$$

But $C^r(F) - D^r(G) = \big(C(F) - D(G)\big)\big(C^{r-1}(F) + \cdots + D^{r-1}(G)\big)$. Since C, D, F, G are monic and $r \not\equiv 0(\text{mod char } \boldsymbol{k})$, the second factor has degree $(r-1)\,d\mu\nu$ and therefore $C(F) = D(G) \in \boldsymbol{k}(F) \cap \boldsymbol{k}(G) = \boldsymbol{k}(H)$. Then $\partial C(F) \geq \partial H$, i.e. $d\mu\nu \geq r\,d\mu\nu$, where $r = 1$. Hence

$$\begin{aligned} [\boldsymbol{k}(F){:}\,\boldsymbol{k}(F) \cap \boldsymbol{k}(G)] &= [\boldsymbol{k}(F){:}\,\boldsymbol{k}(H)] = [\boldsymbol{k}(F){:}\,\boldsymbol{k}(A(F))] \\ &= \partial A = \mu = \frac{[\partial F, \partial G]}{\partial F}. \end{aligned}$$

Similarly

$$[\boldsymbol{k}(G){:}\,\boldsymbol{k}(F) \cap \boldsymbol{k}(G)] = \frac{[\partial F, \partial G]}{\partial G}$$

and since the right hand sides of the above equalities are coprime

$$[\boldsymbol{k}(F, G){:}\,\boldsymbol{k}(F) \cap \boldsymbol{k}(G)] = \frac{[\partial F, \partial G]}{(\partial F, \partial G)}.$$

The theorem follows. □

The following examples show that the assumption $\partial H \not\equiv 0(\text{mod char } \boldsymbol{k})$ cannot be omitted.

Example 1. $\boldsymbol{k} = \mathbb{F}_2$, $F = x^2$, $G = x^2 + x$, $H = x^4 + x^2 = F^2 + F = G^2$; $\boldsymbol{k}(F) \cap \boldsymbol{k}(G) = \boldsymbol{k}(H)$, $\boldsymbol{k}(F, G) = \boldsymbol{k}(x)$.

Example 2. $\boldsymbol{k} = \mathbb{F}_3$, $F = x^2$, $G = x^2+x$, $H = x^6+x^4+x^2 = F^3+F^2+F = G^3 + G^2$; $\boldsymbol{k}(F) \cap \boldsymbol{k}(G) = \boldsymbol{k}(H)$, $\boldsymbol{k}(F, G) = \boldsymbol{k}(x)$.

Corollary 1. $[\boldsymbol{k}(F){:}\,\boldsymbol{k}(F)\cap\boldsymbol{k}(G)] = [\boldsymbol{k}(F,G){:}\,\boldsymbol{k}(G)]$, *if* $\partial H \not\equiv 0 (\mathrm{mod\ char}\ \boldsymbol{k})$.

The second example given above shows that the assumption $\partial H \not\equiv 0 (\mathrm{mod\ char}\ \boldsymbol{k})$ cannot be omitted here either.

Definition 1. A polynomial F is *indecomposable* over $\boldsymbol{k}$ if $F = F_1 \circ F_2$, $F_1, F_2 \in \boldsymbol{k}[x]$ implies $\partial F_1 = 1$ or $\partial F_2 = 1$.

Corollary 2. *If F is indecomposable over $\boldsymbol{k}$, the same is true for $L \circ F$ and $F \circ L$, where L is a linear function.*

Proof. Clear. □

Corollary 3. *$F \in \boldsymbol{k}[x]$ is indecomposable over $\boldsymbol{k}$ if and only if the extension $\boldsymbol{k}(x)/\boldsymbol{k}(F)$ is primitive, i.e. if and only if $\boldsymbol{k}(F) \subset \boldsymbol{K} \subset \boldsymbol{k}(x)$ implies $\boldsymbol{K} = \boldsymbol{k}(F)$ or $\boldsymbol{K} = \boldsymbol{k}(x)$.*

Proof. Suppose $\boldsymbol{k}(F) \subset \boldsymbol{K} \subset \boldsymbol{k}(x)$. Then by Theorem 4, $\boldsymbol{K} = \boldsymbol{k}(G)$ and hence by Lemma 1 $F = H(G)$. Thus $\boldsymbol{K}$ is primitive if and only if the above equality implies $\partial H = 1$ or $\partial G = 1$, which means that F is indecomposable. □

Theorem 6. *If $\partial F \not\equiv 0 (\mathrm{mod\ char}\ \boldsymbol{k})$ and F is indecomposable over $\boldsymbol{k}$, then it is indecomposable over any extension of $\boldsymbol{k}$.*

Proof. Let $F = F_1 \circ F_2$ be a decomposition of F over some extension $\boldsymbol{K}$ of $\boldsymbol{k}$, $\partial F_1 = r$, $\partial F_2 = n$.

Assume without loss of generality that F is monic. If $F_1 = a_0 x^r + a_1 x^{r-1} + \cdots + a_r$ we can write $F = \tilde{F}_1 \circ \tilde{F}_2$, where $\tilde{F}_1(x) = F_1\big(x - \frac{a_1}{a_0 r}\big)$, $\tilde{F}_2(x) = F_2(x) + \frac{a_1}{a_0 r}$ and the coefficient of x^{r-1} in $\tilde{F}_1(x)$ is 0.

By Lemma 3 there exists $C \in \boldsymbol{k}[x]$ such that $\partial C = n$ and $\partial(F - C^r) < n(r-1)$, so $\partial(\tilde{F}_1 \circ \tilde{F}_2 - C^r) < n(r-1)$. It follows that $\partial(a_0 \tilde{F}_2^r - C^r) < n(r-1)$. However

$$a_0 \tilde{F}_2^r - C^r = a_0 \prod_{\nu=1}^{r} (\tilde{F}_2 - \zeta_r^\nu a_0^{-1/r} C)$$

and at most one factor has degree $< n$.

It follows that $\tilde{F}_2 = \zeta_r^\nu a_0^{-1/r} C$ for some $\nu \le r$. Setting $\tilde{F}_1(x) = a_0 x^r + \sum_{i=1}^{r} \tilde{a}_i x^{r-i}$ we infer from $F = \tilde{F}_1 \circ \tilde{F}_2$ by induction on i that $\tilde{a}_i \zeta_r^{-\nu i} a_0^{\frac{i}{r}-1} \in \boldsymbol{k}$, whence $\tilde{F}_1(\zeta_r^\nu a_0^{-1/r} x) \in \boldsymbol{k}[x]$. But then F is decomposable over $\boldsymbol{k}$. □

Example 3. Let $\boldsymbol{k} = \mathbb{F}_2$. Then $F(x) = x^4 + x^2 + x = (x^2 + \alpha x)^2 + \alpha^{-1}(x^2 + \alpha x)$ where $\alpha^2 - \alpha + 1 = 0$, $\alpha \in \mathbb{F}_4$ shows that the assumption $\partial F \not\equiv 0(\text{mod char } \boldsymbol{k})$ cannot be omitted.

Definition 2. Two decompositions of F, say $F = F_1 \circ F_2 \circ \cdots \circ F_r$ and $F = G_1 \circ G_2 \circ \cdots \circ G_r$ are *equivalent*, symbolically $\langle F_1, \ldots, F_r \rangle \sim \langle G_1, \ldots, G_r \rangle$ or $\langle F_i \rangle_{i \le r} \sim \langle G_i \rangle_{i \le r}$ if either $r = 1$, $F_1 = G_1$ or $r \ge 2$ and there exist linear functions $L_1, \ldots, L_{r-1}$, such that $G_1 = F_1 \circ L_1$, $G_j = L_{j-1}^{-1} \circ F_j \circ L_j (1 < j < r)$, $G_r = L_{r-1}^{-1} \circ F_r$.

Corollary 4. *The relation $\sim$ is an equivalence.*

Corollary 5. *If $\langle F_i \rangle_{i \le r} \sim \langle G_i \rangle_{i \le r}$ then for any H*

$$\langle F_i, \ldots, F_r, H \rangle \sim \langle G_1, \ldots, G_r, H \rangle.$$

Theorem 7. *If $\partial F \not\equiv 0(\text{mod char } \boldsymbol{k})$, and $F = G_1 \circ G_2 \circ \cdots \circ G_r = H_1 \circ H_2 \circ \cdots \circ H_s$, where G_i, H_i are indecomposable of degree > 1, then $r = s$, and the sequences $\langle \partial G_i \rangle_{i \le r}$, $\langle \partial H_i \rangle_{i \le r}$ are permutations of each other. Moreover, there exists a finite chain of decompositions $F = F_1^{(j)} \circ \cdots \circ F_r^{(j)}$ $(j \le n)$, such that*

$$\langle F_i^{(1)} \rangle_{i \le r} = \langle G_i \rangle_{i \le r}, \quad \langle F_i^{(n)} \rangle_{i \le r} \sim \langle H_i \rangle_{i \le r}$$

and

for each $j < n$, $\langle F_i^{(j)} \rangle_{i \le r}$ and $\langle F_i^{(j+1)} \rangle_{i \le r}$ differ only by having two consecutive terms with the same composition and reversed coprime degrees. (1)

Proof by induction on ∂F. For $\partial F = 1$ the theorem holds. Assume it is true for polynomials of degree $< \partial F$ and let $F = G_1 \circ G_2 \circ \cdots \circ G_r = H_1 \circ H_2 \circ \cdots \circ H_s$, where G_i, H_i are as above.

Case 1. $\boldsymbol{k}(G_r) = \boldsymbol{k}(H_s)$. Then $H_s = L \circ G_r$, $\partial L = 1$,

$$G_1 \circ G_2 \circ \cdots \circ G_{r-1} \circ G_r = H_1 \circ H_2 \circ \cdots \circ H_{s-1} \circ L \circ G_r.$$

If $r = 1$, then also $s = 1$ and we take $n = 1$, $F_1^{(1)} = G_1 = H_1$. If $r > 1$ then by Corollary 2 also $s > 1$. On the other hand $A \circ B = C \circ B$, $\partial B > 0$ implies $A = C$. Hence $G_1 \circ G_2 \circ \cdots \circ G_{r-1} = H_1 \circ H_2 \circ \cdots \circ (H_{s-1} \circ L)$.

By Corollary 2 and by the inductive assumption $r-1 = s-1$, $r = s$. Moreover, there exists a chain of decompositions $\langle F_i^{(j)} \rangle_{i \le r-1}$ $(j \le n)$ satisfying (1) with r replaced by $r-1$, such that

$$\langle F_i^{(1)} \rangle_{i \le r-1} = \langle G_i \rangle_{i \le r-1}, \ \langle F_i^{(n)} \rangle_{i \le r-1} \sim \langle H_1, \dots, H_{r-1} \circ L \rangle.$$

We set $F_r^{(j)} = G_r$ $(1 \le j \le n)$, find that the new chain satisfies (1) and by Corollary 5

$$\langle F_i^{(n)} \rangle_{i \le r} \sim \langle H_1, \dots, H_{r-2}, H_{r-1} \circ L, G_r \rangle \sim \langle H_i \rangle_{i \le r}$$

whence by Corollary 4 $\langle F_i^{(n)} \rangle_{i \le r} \sim \langle H_i \rangle_{i \le r}$.

Case 2. $\boldsymbol{k}(G_r) \ne \boldsymbol{k}(H_s)$. Then $\boldsymbol{k}(x) \supset \boldsymbol{k}(G_r, H_s) \supsetneqq \boldsymbol{k}(G_r)$, thus by Corollary 3 $\boldsymbol{k}(G_r, H_s) = \boldsymbol{k}(x)$. By Corollary 1 (recall $F \in \boldsymbol{k}(G_r) \cap \boldsymbol{k}(H_s)$)

$$[\boldsymbol{k}(G_r) : \boldsymbol{k}(G_r) \cap \boldsymbol{k}(H_s)] = [\boldsymbol{k}(G_r, H_s) : \boldsymbol{k}(G_r)] = [\boldsymbol{k}(x) : \boldsymbol{k}(H_s)] = \partial H_s.$$

Since $F \in \boldsymbol{k}(G_r) \cap \boldsymbol{k}(H_s)$, by Theorem 4 the intersection $\boldsymbol{k}(G_r) \cap \boldsymbol{k}(H_s)$ is generated by some polynomial P, hence $P = A \circ G_r$, $\partial A = \partial H_s$ and $P = B \circ H_s$, $\partial B = \partial G_r$. Suppose $A = A_1 \circ A_2$. Since $\boldsymbol{k}(G_r) \cap \boldsymbol{k}(H_s) = \boldsymbol{k}(P)$, $P \in \boldsymbol{k}(A_2 \circ G_r) \cap \boldsymbol{k}(H_s)$ implies $\boldsymbol{k}(A_2 \circ G_r) \cap \boldsymbol{k}(H_s) = \boldsymbol{k}(P)$. On the other hand $\boldsymbol{k}(H_s) \subset \boldsymbol{k}(H_s, A_2 \circ G_r) \subset \boldsymbol{k}(x)$. Therefore either $\boldsymbol{k}(H_s) = \boldsymbol{k}(A_2 \circ G_r)$ or $\boldsymbol{k}(H_s, A_2 \circ G_r) = \boldsymbol{k}(x)$. In the first case $\boldsymbol{k}(P) = \boldsymbol{k}(A_2 \circ G_r)$, hence $\partial A_1 = 1$. In the second case by Corollary 1

$$\begin{aligned}
&[\boldsymbol{k}(A_2 \circ G_r) : \boldsymbol{k}(A_2 \circ G_r) \cap \boldsymbol{k}(H_s)] = [\boldsymbol{k}(x) : \boldsymbol{k}(H_s)] = \partial H_s = \partial A, \\
&[\boldsymbol{k}(A_2 \circ G_r) : \boldsymbol{k}(A_1 \circ A_2 \circ G_r)] = \partial A,
\end{aligned}$$

but the above degree also equals ∂A_1; $\partial A_1 = \partial A$, thus $\partial A_2 = 1$. It follows that A is indecomposable and by symmetry so is B.

We have now $F = C \circ P$. If $\partial C = 1$ we have

$$F = \begin{cases} C \circ A \circ G_r = G_1 \circ \cdots \circ G_{r-1} \circ G_r, \\ C \circ B \circ H_s = H_1 \circ \cdots \circ H_{s-1} \circ H_s, \end{cases}$$

hence $C \circ A = G_1 \circ \cdots \circ G_{r-1}$, $C \circ B = H_1 \circ \cdots \circ H_{s-1}$ and by Corollary 2, $r-1 = 1 = s-1$, $r = s = 2$, $\partial G_1 = \partial A = \partial H_2$, $\partial H_1 = \partial B = \partial G_2$. Besides, by Theorem 5 $(\partial H_2, \partial G_2) = 1$. Thus the chain $\langle F_1^{(1)}, F_2^{(1)} \rangle = \langle G_1, G_2 \rangle$, $\langle F_1^{(2)}, F_2^{(2)} \rangle = \langle H_1, H_2 \rangle$ satisfies the condition (1) for $r = 2$. Assume now that $\partial C > 1$ and let $C = C_1 \circ \cdots \circ C_t$, where C_j are indecomposable, $\partial C_j > 1$. We have

$$F = \begin{cases} C_1 \circ \cdots \circ C_t \circ A \circ G_r = G_1 \circ \cdots \circ G_{r-1} \circ G_r, \\ C_1 \circ \cdots \circ C_t \circ B \circ H_s = H_1 \circ \cdots \circ H_{s-1} \circ H_s, \end{cases}$$

hence $C_1 \circ \cdots \circ C_t \circ A = G_1 \circ \cdots \circ G_{r-1}$,

$$C_1 \circ \cdots \circ C_t \circ B = H_1 \circ \cdots \circ H_{s-1} \tag{2}$$

and by the inductive assumption $r - 1 = t + 1 = s - 1$; $r = s$. Moreover, there exists a chain of decompositions satisfying (1) with r replaced by $r - 1$ and such that

$$\begin{aligned} \langle F_i^{(1)} \rangle_{i \le r-1} &= \langle G_i \rangle_{i \le r-1}, \\ \langle F_i^{(n)} \rangle_{i \le r-1} &\sim \langle C_1, \ldots, C_{r-2}, A \rangle. \end{aligned}$$

It follows that for some linear function L

$$\begin{aligned} &F_1^{(n)} \circ \cdots \circ F_{r-2}^{(n)} = C_1 \circ \cdots \circ C_{r-2} \circ L^{-1}, \quad F_{r-1}^{(n)} = L \circ A, \\ &F_1^{(n)} \circ \cdots \circ F_{r-2}^{(n)} \circ (L \circ B) = C_1 \circ \cdots \circ C_{r-2} \circ B. \end{aligned} \tag{3}$$

On the other hand by (2) and (3), we have a chain of decompositions $\langle F_i^{(j)} \rangle_{i \le r-1}$ $(n < j \le n + m)$ satisfying (1) with r replaced by $r - 1$, where

$$\begin{aligned} \langle F_1^{(n+1)}, \ldots, F_{r-1}^{(n+1)} \rangle &= \langle F_1^{(n)}, \ldots, F_{r-2}^{(n)}, L \circ B \rangle, \\ \langle F_1^{(n+m)}, \ldots, F_{r-1}^{(n+m)} \rangle &\sim \langle H_1, \ldots, H_{s-1} \rangle. \end{aligned}$$

Define

$$F_r^{(j)} = \begin{cases} G_r & \text{if } j \le n, \\ H_s & \text{if } n < j \le n + m. \end{cases}$$

The new chain satisfies all conditions since by Theorem 5, $(\partial G_r, \partial H_s) = 1$. □

Without the assumption $\partial F \not\equiv 0 (\text{mod char } \boldsymbol{k})$ Theorem 7 is not true in general, as it is shown by the following

Example 4.

$$\begin{aligned} F(x) &= x^{p+1} \circ (x^p + x) \circ (x^p - x) = (x^{p^2} - x)^{p+1} \\ &= (x^{p^2} - x^{p^2-p+1} - x^p + x) \circ x^{p+1}. \end{aligned}$$

Notes. Theorem 5 is due to Engstrom 1941 for char $\boldsymbol{k} = 0$, to Fried & MacRae 1969 in general. These authors also proved Theorem 6. Theorem 7 was proved by Ritt 1922 for $\boldsymbol{k} = \mathbb{C}$, by Engstrom 1941 for char $\boldsymbol{k} = 0$, in general the first part was proved by Fried & MacRae 1969, the second part in [S]. Example 2 is due to Bremner & Morton 1978, Example 4 to Dorey & Whaples 1974.

1.4 Ritt's second theorem

Ritt's second theorem deals with the case to which Theorem 7 reduces the problem of decomposition of polynomials, i.e. with the equation

$$G \circ A = H \circ B, \text{ where } \partial G = \partial B \text{ and } \partial A = \partial H \text{ are coprime.}$$

We put char $\boldsymbol{k} = \pi \geq 0$.

Definition 3. *Dickson's polynomials* $D_n(x, a)$ are given by the recurrence formulae:

$$D_0(x, a) = 2,\ D_1(x, a) = x,\ D_n(x, a) = xD_{n-1}(x, a) - aD_{n-2}(x, a).$$

We put $D_n(x, 1) = D_n(x)$.

Corollary 1. $D_n(x, a) = \sum_{i=0}^{\lfloor n/2 \rfloor} \frac{n}{n-i}\binom{n-i}{i}(-a)^i x^{n-2i}$ $(n \geq 1)$.

Corollary 2. $D_n(x + \frac{a}{x}, a) = x^n + (\frac{a}{x})^n$ $(n \geq 1)$.

Corollary 3. $D_n(x, a) = \sqrt{a}^n D_n(\frac{x}{\sqrt{a}})$ $(a \neq 0)$.

Corollary 4. *If* $\pi = 2$, $n \geq 2$,

$$D_{n+2}(x) = x^2 D_n(x) + D_{n-2}(x).$$

Proofs are by induction on n.

Theorem 8. *Let* $A, B, G, H \in \boldsymbol{k}[x]$, $\partial G = \partial B = m > 1$, $\partial H = \partial A = n > 1$, $(m, n) = 1$, $m > n$ *and* $G'H' \neq 0$. *The equation* $G(A) = H(B)$ *holds if and only if there exist linear functions* $L_1, L_2 \in \boldsymbol{k}[x]$ *such that either*

(i) $\langle L_1 \circ G, A \circ L_2 \rangle \sim \langle x^r P(x)^n, x^n \rangle$,
$\langle L_1 \circ H, B \circ L_2 \rangle \sim \langle x^n, x^r P(x^n) \rangle$, *where* $P \in \boldsymbol{k}[x]$, $r = m - n\partial P \in \mathbb{N}$

or

(ii) $\langle L_1 \circ G, A \circ L_2 \rangle \sim \langle D_m(x, a^n), D_n(x, a) \rangle$,
$\langle L_1 \circ H, B \circ L_2 \rangle \sim \langle D_n(x, a^m), D_m(x, a) \rangle$, *where* $a \in \boldsymbol{k}$.

Lemma 1. *The conditions are sufficient.*

Proof. (i) implies

$$L_1 \circ G \circ A \circ L_2 = x^{rn} P(x^n)^n = L_1 \circ H \circ B \circ L_2,$$

hence $G \circ A = H \circ B$.

(ii) implies

$$L_1 \circ G \circ A \circ L_2 = D_m(D_n(x, a), a^n).$$

Now, by Corollary 2

$$\begin{aligned} D_m(D_n(x + \frac{a}{x}, a), a^n) &= D_m(x^n + (\frac{a}{x})^n, a^n) = x^{mn} + \frac{a^{mn}}{x^{mn}} \\ &= D_n(D_m(x + \frac{a}{x}, a), a^m). \end{aligned}$$

Hence $D_m(D_n(x, a), a^n) = D_n(D_m(x, a), a^m) = L_1 \circ H \circ B \circ L_2$ and $G \circ A = H \circ B$. □

Lemma 2. *If the conditions are necessary for the field $\bar{\boldsymbol{k}}$ they are necessary for $\boldsymbol{k}$.*

Proof. Consider first the condition (i) and let $\bar{L}_1, \bar{L}_2, \bar{L}_3, \bar{L}_4, \bar{P} \in \bar{\boldsymbol{k}}[x]$ be such that

$$\begin{aligned} \bar{L}_1 \circ G = x^r \bar{P}(x)^n \circ \bar{L}_3, &\quad A \circ \bar{L}_2^{-1} = \bar{L}_3^{-1} \circ x^n, \\ \bar{L}_1 \circ H = x^n \circ \bar{L}_4, &\quad B \circ \bar{L}_2^{-1} = \bar{L}_4^{-1} \circ x^r \bar{P}(x^n). \end{aligned}$$

Put $\bar{L}_i = \lambda_i(x + \mu_i)$ $(i \le 4)$. We shall prove that $\mu_i \in \boldsymbol{k}$. Indeed we have

$$\lambda_1(H + \mu_1) = L_4^n = \lambda_4^n(x + \mu_4)^n,$$

and, comparing the coefficients of x^n and of x^{n-1}, $\lambda_1^{-1}\lambda_4^n \in \boldsymbol{k}$, $n\mu_4 \in \boldsymbol{k}$. Since $H' \ne 0$ we have $n \not\equiv 0 (\text{mod } \pi)$, hence $\mu_4 \in \boldsymbol{k}$, $H + \mu_1 = \lambda_1^{-1}\lambda_4^n(x + \mu_4)^n$, $\lambda_1^{-1}\lambda_4^n\mu_4^n - \mu_1 = H(0) \in \boldsymbol{k}$, hence $\mu_1 \in \boldsymbol{k}$.

Similarly from

$$\lambda_3(A + \mu_3) = \bar{L}_3 \circ A = \bar{L}_2^n = \lambda_2^n(x + \mu_2)^n$$

we infer that $\mu_2, \mu_3 \in \boldsymbol{k}$, $\lambda_3\lambda_2^{-n} \in \boldsymbol{k}$.

Define

$$\begin{aligned} &L_1 = \lambda_1\lambda_4^{-n}(x + \mu_1), \quad L_2 = x + \mu_2, \quad L_3 = \lambda_3\lambda_2^{-n}(x + \mu_3), \\ &L_4 = x + \mu_4, \quad P(x) = \lambda_4^{-1}\lambda_2^r\bar{P}(\lambda_2^n x). \end{aligned}$$

We obtain

$$L_4 \circ B \circ L_2^{-1} = \lambda_4^{-1}\bar{L}_4 \circ B \circ \bar{L}_2^{-1}(\lambda_2 x) = \lambda_4^{-1}(\lambda_2 x)^r \bar{P}(\lambda_2^n x^n) = x^r P(x^n),$$

since $L_2^{-1} = \bar{L}_2^{-1}(\lambda_2 x)$. Hence $P \in \boldsymbol{k}[x]$. Moreover,

$$B \circ L_2^{-1} = L_4^{-1} \circ x^r P(x^n).$$

We check

$$
\begin{aligned}
L_1 \circ H &= \lambda_4^{-n} \bar{L}_1 \circ H = \lambda_4^{-n} \bar{L}_4^n = (x+\mu_4)^n = x^n \circ L_4,\\
A \circ L_2^{-1} &= A \circ \bar{L}_2^{-1}(\lambda_2 x) = \bar{L}_3^{-1} \circ x^n \circ (\lambda_2 x)\\
&= \bar{L}_3^{-1} \circ (\lambda_2 x)^n = \lambda_3^{-1}\lambda_2^n x^n - \mu_3 = L_3^{-1} \circ x^n,\\
L_1 \circ G \circ L_3^{-1} \circ x^n &= L_1 \circ G \circ A \circ L_2^{-1}\\
&= L_1 \circ H \circ B \circ L_2^{-1} = x^{rn} P(x^n)^n,\\
L_1 \circ G \circ L_3^{-1} &= x^r P(x)^n,\\
L_1 \circ G &= x^r P(x)^n \circ L_3.
\end{aligned}
$$

Hence

$$\langle L_1 \circ G, A \circ L_2^{-1}\rangle \sim \langle x^r P(x)^n, x^n\rangle, \langle L_1 \circ H, B \circ L_2^{-1}\rangle \sim \langle x^n, x^r P(x^n)\rangle.$$

Consider now the condition (ii). If this condition is satisfied over $\bar{\boldsymbol{k}}$ there exist by Corollary 3 linear functions $\bar{L}_i \in \boldsymbol{k}[x]$ $(i \le 4)$ such that

$$
\begin{aligned}
\bar{L}_1 \circ G &= D_m \circ \bar{L}_3, \quad & A \circ \bar{L}_2^{-1} &= \bar{L}_3^{-1} \circ D_n,\\
\bar{L}_1 \circ H &= D_n \circ \bar{L}_4, \quad & B \circ \bar{L}_2^{-1} &= \bar{L}_4^{-1} \circ D_m.
\end{aligned}
$$

Let $\bar{L}_i = \lambda_i(x+\mu_i)$. In the first of the above equations the quotient of the first two coefficients on the left is in $\boldsymbol{k}$, on the right we have $D_m(\lambda_3(x+\mu_3))$, so we obtain $m\mu_3 \in \boldsymbol{k}$. Since $G' \neq 0$ we have $D'_m \neq 0$, hence

$$D'_m\Big(x+\frac{1}{x}\Big)\Big(1-\frac{1}{x^2}\Big) \neq 0 \text{ and, by Corollary 2, } m \not\equiv 0 (\operatorname{mod} \pi).$$

Thus $\mu_3 \in \boldsymbol{k}$. It follows similarly that all $\mu_i \in \boldsymbol{k}$. Let g_0 be the leading coefficient of G. From $\bar{L}_1 \circ G = D_n \circ \bar{L}_3$ we obtain $\lambda_1 g_0 = \lambda_3^m$, $\lambda_1\lambda_3^{-m} \in \boldsymbol{k}$. Similarly we have $\lambda_3\lambda_2^{-n} \in \boldsymbol{k}$. In the identity

$$\lambda_1(G+\mu_1) = D_m(\lambda_3(x+\mu_3))$$

substitute $x-\mu_3$ for x. We obtain

$$G(x-\mu_3)+\mu_1 = \lambda_1^{-1} D_m(\lambda_3 x).$$

The third coefficient on the right (see Corollary 1) is $-m\lambda_3^{m-2}\lambda_1^{-1} \in \boldsymbol{k}$, thus $\lambda_3^2 \in \boldsymbol{k}$. Similarly $\lambda_2^2 \in \boldsymbol{k}$. We also obtain

$$\lambda_1^{-1}\lambda_3^m,\ \lambda_3^{-1}\lambda_2^n,\ \lambda_1^{-1}\lambda_2^{mn} \in \boldsymbol{k};\ \lambda_4\lambda_2^{-m} \in \boldsymbol{k}.$$

Put now

$$
\begin{aligned}
a = \lambda_2^{-2}, \quad L_1 &= \lambda_1\lambda_2^{-mn}(x+\mu_1), \quad & L_2 &= x+\mu_2,\\
L_3 &= \lambda_3\lambda_2^{-n}(x+\mu_3), \quad & L_4 &= \lambda_4\lambda_2^{-m}x+\mu_4.
\end{aligned}
$$

We have $a \in \boldsymbol{k}$, $L_i \in \boldsymbol{k}[x]$. Moreover,

$$L_1 \circ G = D_m(x, a^n) \circ L_3, \ \ A \circ L_2^{-1} = L_3^{-1} \circ D_n(x, a),$$
$$L_1 \circ H = D_n(x, a^m) \circ L_4, \ \ B \circ L_2^{-1} = L_4^{-1} \circ D_m(x, a),$$

hence (ii) is satisfied with L_2 replaced by L_2^{-1}. □

From now on we assume $\boldsymbol{k}$ algebraically closed, but not till Lemma 16 inclusive that $m > n$.

Lemma 3. *The polynomial $f(x, y) = G(y) - H(x)$ is irreducible over $\boldsymbol{k}$.*

Proof. Suppose that $f(y, x) = f_1(y, x) f_2(y, x)$, where $f_i \in \boldsymbol{k}[y, x]$, $\deg f_i > 0$ $(i = 1, 2)$. Let us give x the weight m, y the weight n. The part of the greatest weight of f, $ay^m - bx^n$, must be the product of the parts of the greatest weight of $f_1(x, y)$, $f_2(x, y)$. Hence these two are of the form $a_i y^{\mu_i} + \cdots + b_i x^{\nu_i}$ $(i = 1, 2)$, where $\mu_i n = \nu_i m$ and $0 < \mu_i < m$, $0 < \nu_i < n$. However in view of $(m, n) = 1$ this is impossible. □

Lemma 4. *If $\pi \neq 2$ the equation*

$$(Q(t) - q_1)(Q(t) - q_2) = (t - \xi_1)(t - \xi_2) R^2(t), \qquad Q, R \in \boldsymbol{k}[t],$$

$q_1, q_2, \xi_1, \xi_2 \in \boldsymbol{k}$, $q_1 \neq q_2$, $\xi_1 \neq \xi_2$, $d = \partial Q$ implies

$$Q(t) = L \circ D_d \circ M^{-1},$$

where for a suitable $\varepsilon = \pm 1$

$$L(t) = \varepsilon \frac{q_1 - q_2}{4} t + \frac{q_1 + q_2}{2}, \quad M(t) = \frac{\xi_1 - \xi_2}{4} t + \frac{\xi_1 + \xi_2}{2}.$$

Proof. Without loss of generality we may assume that one of the following holds:

$$Q(t) - q_i = (t - \xi_i) R_i^2 \qquad (i = 1, 2) \tag{1}$$

or

$$Q(t) - q_1 = (t - \xi_1)(t - \xi_2) R_3^2(t), \ \ Q(t) - q_2 = R_4(t)^2, \tag{2}$$

or

$$Q(t) - q_1 = R_3^2(t), \ \ Q(t) - q_2 = (t - \xi_1)(t - \xi_2) R_4(t)^2, \tag{3}$$

where $R_i \in \boldsymbol{k}(t)$. Put

$$P(t) = L^{-1} \circ Q \circ M.$$

In the case (1) we obtain

$$\frac{q_1 - q_2}{4}(P(t) \pm 2) = \frac{\xi_1 - \xi_2}{4}(t \pm 2) R^2_{\frac{3\pm 1}{2}}(M(t)), \tag{4}$$

in the case (2) or (3) for the upper or lower sign, respectively,

$$\begin{aligned} \frac{q_1 - q_2}{2}(P(t) \mp 2\varepsilon) &= \left(\frac{\xi_1 - \xi_2}{4}\right)^2 (t^2 - 4) R_3^2(M(t)), \\ \frac{q_1 - q_2}{2}(P(t) \pm 2\varepsilon) &= R_4(M(t))^2. \end{aligned} \tag{5}$$

Choose now ε so that $\mp 2\varepsilon = -2$ and substitute $t = z + z^{-1}$. From both (4) and (5) we obtain

$$P(z + z^{-1}) - 2 = z^{-\partial P} S_1(z)^2, \quad P(z + z^{-1}) + 2 = z^{-\partial P} S_2(z)^2$$

and $S_1(1) = 0$. Thus

$$4z^{\partial P} = S_2^2 - S_1^2 = (S_2 - S_1)(S_2 + S_1).$$

Since $\pi \neq 2$, $\max\{\partial(S_2 - S_1), \partial(S_2 + S_1)\} = \partial P$, hence $\min\{\partial(S_2 - S_1), \partial(S_2 + S_1)\} = 0$ and for a suitable sign $S_2 \pm S_1 = s \in \bar{\boldsymbol{k}}$. Then $s(s \mp 2S_1) = 4z^{\partial P}$ and on substituting $z = 1$ we obtain $s^2 = 4$.

Now

$$S_1 = \pm \frac{2}{s}(1 - z^{\partial P}),$$

and

$$P(z + z^{-1}) = 2 + z^{-\partial P} S_1(z)^2 = 2 + \frac{4}{s^2} z^{-\partial P}(z^{\partial P} - 1)^2 = z^{\partial P} + z^{-\partial P}.$$

Hence by Corollary 2 $P(t) = D_{\partial P}(t)$, which proves the lemma since $\partial P = \partial Q = d$. □

Lemma 5. *Let $n > 1$ and assume $\pi \nmid n$ and*

$$D_n(at + b) + d = cD_n(t), \tag{6}$$

where $a, c \in \boldsymbol{k}^*$, $b, d \in \boldsymbol{k}$.

Then $b = 0$ and either $n = 2$ or $d = 0$, $a = \pm 1$, $c = a^n$.

Proof. On comparing the coefficients of t^n and t^{n-1} on both sides of (6) and using Corollary 1 we find

$$a^n = c, \quad na^{n-1}b = 0,$$

hence $b = 0$. For $n > 2$ on comparing the coefficients of t^{n-2} we find

$$-na^{n-2} = -nc = -na^n,$$

hence $a^2 = 1$, $a = \pm 1$ and

$$d = cD_n(t) - D_n(at) = a^n D(t) - a^n D(t) = 0. \qquad \square$$

Convention 1. $\boldsymbol{F} = \boldsymbol{k}(x, y)$, where $G(y) - H(x) = 0$. For a prime divisor v of $\boldsymbol{F}/\boldsymbol{k}$ we shall denote ord_v again by v and for $f \in \boldsymbol{F}$ with $v(f) \geq 0$ we shall denote by $f(v)$ the element a of $\boldsymbol{k}$ such that $v(f - a) > 0$. Similar convention applies to prime divisors of $\boldsymbol{k}(x)/\boldsymbol{k}$ and of $\boldsymbol{k}(y)/\boldsymbol{k}$.

Convention 2. If $a \in \boldsymbol{k}$ we shall denote by w_a, w_a^* the prime divisor of $\boldsymbol{k}(x)/_{\boldsymbol{k}}, \boldsymbol{k}(y)/_{\boldsymbol{k}}$, respectively, such that $w_a(x - a) > 0$, $w_a^*(y - a) > 0$, respectively. By w_∞, w_∞^* we shall denote the prime divisor of $\boldsymbol{k}(x)/_{\boldsymbol{k}}, \boldsymbol{k}(y)/_{\boldsymbol{k}}$ such that $w_\infty(x^{-1}) > 0$, $w_\infty(y^{-1}) > 0$, respectively. By $S_{a,b}$ we shall denote the set of prime divisors of $\boldsymbol{F}/\boldsymbol{k}$ lying simultaneously above w_a, w_b^*.

Corollary 5. *If $a, b \neq \infty$, then*

$$S_{a,b} = \{v | v(x - a) > 0,\ v(y - b) > 0\},$$

where v runs through the prime divisors of F.

Lemma 6. *Let I be a prime ideal in $R = \boldsymbol{k}[X_1, \ldots, X_n]$, $\boldsymbol{K}$ be the quotient field of R/I and $\langle a_1, \ldots, a_n\rangle \in \boldsymbol{k}^n$ be such that $p(a_1, \ldots, a_n) = 0$ for $p \in I$. Then there exists a valuation of $\boldsymbol{K}$ trivial on $\boldsymbol{k}$ with the maximal ideal $\mathfrak{M}$ of the valuation ring such that $\overline{X_i - a_i} \in \mathfrak{M}$ for all $i \leq n$, where $\overline{X_i - a_i}$ is the residue class of $X_i - a_i$ mod I.*

Proof. Put in the Corollary to Theorem [L] 9.1 : $A = R/I$, $K = \boldsymbol{K}$, $L = \boldsymbol{k}$ and define φ by the formula $\varphi(\bar{a}) = a$ for $a \in \boldsymbol{k}$, $\varphi(\bar{X}_i) = a_i (1 \leq i \leq n)$. Then by the corollary, the maximal subring B of $\boldsymbol{K}$ to which φ may be prolonged as homomorphism into $\boldsymbol{k}$ has the property that if $x \in \boldsymbol{K}^*$ then either $x \in B$ or $x^{-1} \in B$. Let $U = \{x \in B : x^{-1} \in B\}$. The factor group $\boldsymbol{K}^*/U$ can be ordered (see [L], Chapter XII, § 4), hence the map assigning to each element $x \in \boldsymbol{K}^*$ the coset xU and to $x = 0$ the element 0 is a valuation of $\boldsymbol{K}$. Since $\varphi(\bar{a}) = a$ for $a \in \boldsymbol{k}$, the valuation is trivial on $\boldsymbol{k}$. Since, by the definition of B, $\varphi(\overline{X_i - a_i}) = 0$, we have $(\overline{X_i - a_i})^{-1} \notin B$, hence

$$\overline{X_i - a_i} \in B \setminus U$$

and as shown in [L], Chapter XII, § 4, $B \setminus U$ is the maximal ideal of B. $\square$

Lemma 7. *w_∞ is totally ramified in $\boldsymbol{F}$, so there is just one prime divisor of $\boldsymbol{F}$, denoted by v_∞, above w_∞. If t_∞ is a local parameter at v_∞ we have*

$$v_\infty\Big(\frac{dx}{dt_\infty}\Big) = -m - 1 + n(m - 1 - \delta),$$

where $\delta = \partial G'$.

Proof. Write $am + bn = 1$ with integers a, b. Set $u = x^a y^b \in \boldsymbol{F}$ and $t = x^{-n} y^m$. We obtain

$$x = u^m t^{-b}, \qquad y = u^n t^a. \tag{7}$$

Now

$$t = \frac{y^m}{G(y)} \frac{H(x)}{x^n} = \frac{H^*\big(\frac{1}{x}\big)}{G^*\big(\frac{1}{y}\big)},$$

where H^*, G^* are polynomials with $H^*(0)G^*(0) \neq 0$.

Since clearly both $1/x$ and $1/y$ are zero at every prime divisor v of F above w_∞ we see that t is a unit at each such prime divisor.

From the first equation of (7) we thus derive the initial part of the lemma. Moreover, we see that $1/u$ is a local parameter at v_∞.

If $\pi \nmid m$ we have $\delta = m - 1$. Also, Theorem A5 (ii) combined with the first equation of (7) again shows that

$$v_\infty\Big(\frac{dx}{d(1/u)}\Big) = -m - 1,$$

so Lemma 7 holds in this case.

Suppose now that $\pi \mid m$. Directly from the definition of u we have

$$\frac{du}{u} = \frac{a}{x}dx + \frac{b}{y}dy = \Big(\frac{a}{x} + \frac{bH'(x)}{yG'(y)}\Big)dx. \tag{8}$$

Also, from (7)

$$v_\infty(y) = -n, \;\; v_\infty(G'(y)) = -n\delta, \;\; v_\infty(H'(x)) = -m(n-1)$$

since $\partial H' = n - 1$, as $\pi \nmid n$. But $\pi \nmid b$ also, for $am + bn = 1$ so $v_\infty\Big(\frac{bH'(x)}{yG'(y)}\Big) = n(1 + \delta - m) + m < m$, since $\delta \leq m - 2$ in this case.

But $v_\infty(a/x) \geq v_\infty(1/x) = m$, so (8) implies

$$v_\infty\Big(\frac{du}{dx}\Big) = v_\infty(u) + n(1 + \delta - m) + m = -1 + n(1 + \delta - m) + m,$$

whence finally

$$v_\infty\Big(\frac{du}{d(1/u)}\Big) = v_\infty\Big(\frac{dx}{du}\Big) - 2 = -m - 1 + n(m - 1 - \delta),$$

as required. □

We now deal with the splitting of finite prime divisors of $\boldsymbol{k}(x)$.

Lemma 8. *Let $r \geq 1$, $s \geq 1$, $(r, s) = d$, $r = dr'$, $s = ds'$, $p, q \in \boldsymbol{k}[t]$. The ideal I of $\boldsymbol{k}[X, Y, T]$ generated by the polynomials*

$$F_1 = X^{s'}T - Y^{r'}, \quad F_2 = T^d q(Y) - p(X)$$

is a prime ideal, provided $y^r q(y) - x^s p(x)$ is irreducible over $\boldsymbol{k}$ and $p(0)q(0) \neq 0$.

Proof. Put $f(X, Y) = Y^r q(Y) - X^s p(X)$. Assume that $gh \in I$, where $g, h \in \boldsymbol{k}[X, Y, T]$. Then clearly the rational function $g\Big(X, Y, \frac{Y^{r'}}{X^{s'}}\Big)h\Big(X, Y, \frac{Y^{r'}}{X^{s'}}\Big) \in \boldsymbol{k}[X, X^{-1}, Y]$ has a numerator divisible by $f(X, Y)$. Since this is irreducible it divides the numerator of, say, $g\Big(X, Y, \frac{Y^{r'}}{X^{s'}}\Big)$. We have, after division by $T - \frac{Y^{r'}}{X^{s'}}$ in $\boldsymbol{k}[X, X^{-1}, Y][T]$, the equation

$$g(X, Y, T) = g\Big(X, Y, \frac{Y^{r'}}{X^{s'}}\Big) + F_1 g_1(X, Y, T), \tag{9}$$

where $g_1 \in \boldsymbol{k}[X, X^{-1}, Y][T]$.

Since

$$0 \equiv X^s F_2 = (F_1 + Y^{r'})^d q(Y) - X^s p(X) \equiv f (\mathrm{mod}\ I)$$

we have $f \in I$ and we see by (9) that if a is a sufficiently large integer, $X^a g \in I$. It suffices now in order to show $g \in I$ to prove that $Xg \in I$ implies $g \in I$ for any $g \in \boldsymbol{k}[X, Y, T]$.

Write $Xg = \alpha F_1 + \beta F_2$. Then $\alpha(0, Y, T)Y^{r'} = \beta(0, Y, T)(T^d q(Y) - p(0))$, whence

$$\alpha(0, Y, T) = \rho(Y, T)(T^d q(Y) - p(0)), \quad \beta(0, Y, T) = \rho(Y, T)Y^{r'}$$

for some $\rho \in \boldsymbol{k}(Y, T)$ and so, clearly

$$\alpha(X, Y, T) = \rho F_2 + X\gamma, \quad \beta(X, Y, T) = -\rho F_1 + X\delta,$$

where $\gamma, \delta \in \boldsymbol{k}[X, Y, T]$. So $\alpha F_1 + \beta F_2 = \rho F_2 F_1 + X\gamma F_1 - \rho F_1 F_2 + X\delta F_2$. Finally $g = \gamma F_1 + \delta F_2 \in I$, as required. □

Lemma 9. *Let $G(y) = y^r p(y)$, $H(x) = x^s q(x)$, where $r \geq 1$, $s \geq 1$, $p, q \in \boldsymbol{k}[X]$, $p(0)q(0) \neq 0$. Put $r = dr'$, $s = ds'$, $d = (r, s)$ and let a, b be any integers satisfying $ar' + bs' = 1$, also write $d = d_* \pi^\mu$, when $\pi \nmid d_* \in \mathbb{Z}$,*

$$t = x^{-s'} y^{r'}, \quad u = x^a y^b. \tag{10}$$

We have

(i) *If $v \in S_{0,0}$, then*

$$\frac{r}{(r,s)} | e(v|w_0).$$

(ii) *The function t is a unit at each $v \in S_{0,0}$.*
Also

$$\mathrm{card}\,\{t(v) \colon v \in S_{0,0}\} = (r, s)_*.$$

(iii) $\displaystyle\sum_{v \in S_{0,0}} e(v|w_0) = r.$

(iv) $(r, s)_* \leq \mathrm{card}\ S_{0,0} \leq (r, s).$

Proof. Observe that

$$t^d q(y) = p(x) \tag{11}$$

and that

$$x = u^{r'} t^{-b}, \quad y = u^{s'} t^a. \tag{12}$$

That t is a unit at each prime divisor $v \in S_{0,0}$ follows from (11), since $p(0)q(0) \neq 0$, so we have the first part of (ii). This fact combined with the first half of (10) proves (i). We now prove the second half of (ii). Consider the ideal I of $\boldsymbol{k}[X, Y, T]$ described in Lemma 8. By that lemma and Lemma 3 I is a prime ideal, hence the quotient field $\boldsymbol{F}_*$ of $\boldsymbol{k}[X, Y, T]/I$ is well defined. Let x_*, y_*, t_* be the images of X, Y, T in $\boldsymbol{F}_*$. Then clearly, since $t_* = x_*^{-s'} y_*^{r'}$ $\boldsymbol{F}_* = \boldsymbol{k}(x_*, y_*)$, where $f(x_*, y_*) = 0$. Since f is irreducible $\boldsymbol{F}_*$ is isomorphic to $\boldsymbol{F}$.

By Lemma 6 and by the fact that every valuation of F trivial on $\boldsymbol{k}$ is discrete (see [L], Chapter XII, § 4, Example) each point $\langle 0, 0, a\rangle \in \boldsymbol{k}^3$, where $a^d q(0) = p(0)$ corresponds to at least one prime divisor v of $\boldsymbol{F}$ such that $x(v) = y(v) = 0$, $t(v) = a$, so in particular $v \in S_{0,0}$.

On the other hand, if $v \in S_{0,0}$ clearly $t^d(v)q(0) = p(0)$. But the equation $z^d = \frac{p(0)}{q(0)}$ has exactly d_* distinct solutions in $\boldsymbol{k}$, so (ii) is completely proved.

To prove (iii) we use Theorem A2 and factor $l^{-1} f = l^{-1}(Y^r q(Y) - X^s p(X))$ (l is the leading coefficient of q) over $\boldsymbol{k}((X))$, obtaining

$$l^{-1} f = P_1(Y, X) \ldots P_h(Y, X),$$

where P_i are elements of $\boldsymbol{k}[[X]][Y]$ monic in Y and irreducible over $\boldsymbol{k}((X))$. If the valuation v_i corresponds to the factor P_i, and if moreover $v_i \in S_{0,0}$, i.e. $v_i(y) > 0$, then, by Theorem A2 and Corollary A6, $P_i(Y, 0) = Y^{e_i}$ and conversely.

So (iii) follows on comparing the greatest power of Y dividing the sides of the equation

$$l^{-1}Y^r q(Y) = P_1(Y, 0) \dots P_h(Y, 0).$$

Now (iv) is trivial, the lower bound following from (ii), the upper bound from (i) and (iii). □

Convention 3. We put

$$c(x_0, y_0) = \sum_{v \in S_{x_0, y_0}} v\left(\frac{dx}{dt_v}\right), \tag{13}$$

$$G(y) - G(y_0) = (y - y_0)^{r(y_0)} Q_{y_0}(y), \quad \text{where } Q_{y_0}(y_0) \neq 0, \tag{14}$$

$$H(x) - H(x_0) = (x - x_0)^{s(x_0)} P_{x_0}(x), \quad \text{where } P_{x_0}(x_0) \neq 0, \tag{15}$$

$$\mu(y_0) = \operatorname{ord}_{y-y_0} Q'_{y_0} \quad \text{in the case that } \pi | r(y_0), \tag{16}$$

$$\Gamma = \{\langle x_0, y_0\rangle \in \boldsymbol{k}^2 | G(y_0) = H(x_0)\}. \tag{17}$$

Lemma 10. *For $\langle x_0, y_0\rangle \in \Gamma$ we have*

(i) $c(x_0, y_0) \geq r(y_0) - (r(y_0), s(x_0))$.

If there is equality then

(a) card $S_{x_0,y_0} = (r(y_0), s(x_0))$.
(b) *For all $v \in S_{x_0,y_0}$ we have that $\pi \nmid e(v|w_{x_0}) = \frac{r(y_0)}{(r(y_0),s(x_0))}$.*

(ii) *If $\pi \mid r(y_0)$, but $\pi \nmid s(x_0)$ then*

$$c(x_0, y_0) \geq r(x_0) - (r(y_0), s(x_0)) + s(x_0)(1 + \mu(y_0)).$$

Proof. For each $\langle x_0, y_0\rangle \in \Gamma$ the polynomials $\tilde{G}(y) = G(y + y_0) - G(y_0)$, $\tilde{H}(x) = H(x+x_0) - H(x_0)$ satisfy the assumptions of Lemma 9, and denoting the parameters corresponding to them by $\tilde{S}$, $\tilde{c}$, we have

$$S_{x_0,y_0} = \tilde{S}_{0,0}, \quad c(x_0, y_0) = \tilde{c}_{0,0}, \quad r(y_0) = r,$$

$$Q_{y_0}(y + y_0) = q(y), \quad s(x_0) = s, \quad P_{x_0}(x + x_0) = p(x).$$

Therefore, we may at once suppose that $x_0 = y_0 = 0$. G, H satisfy the assumptions of Lemma 9 and $\mu(0) = \operatorname{ord}_y q'(y) = \mu$.

By Theorem A5 (ii) we have

$$v(dx/dt_v) \geq e(v|w_0) - 1 \text{ for all } v \text{ above } w_0,$$

with equality if and only if $\pi \nmid e(v|w_0)$, so

$$c(0,0) \geq \sum_{v \in S_{0,0}} e(v|w_0) - \text{card } S_{0,0} = r - \text{card } S_{0,0}$$

by (iii) of Lemma 9. If equality holds then $\pi \nmid e(v|w_0)$ for all $v \in S_{0,0}$. Now part (i) follows at once from this inequality combined with (i) and (iv) of Lemma 9.

To prove (ii) observe that $\pi \nmid (r,s)$ implies, by Lemma 9 again, that card $S_{0,0} = (r,s)$ and $e(v|w_0) = \frac{r}{(r,s)} = r'$ for all $v \in S_{0,0}$.

Also, the equation $x = u^{r'} t^{-b}$ implies that u is a local parameter at each such v, where t, u are defined by (10).

To calculate $v(dx/du)$ we argue as in the proof of Lemma 7 and differentiate the equation $u = x^a y^b$ obtaining

$$\frac{du}{u} = a\frac{dx}{x} + b\frac{dy}{y},$$

or

$$\frac{du}{dx} = u\left(\frac{a}{x} + \frac{b}{y}\frac{dy}{dx}\right). \tag{18}$$

Since $y^r q(y) = x^s p(x)$ and since $\pi | r, \pi \nmid s$ we obtain

$$y^r q'(y) dy = x^{s-1}(sp(x) + xp'(x))dx$$

and

$$rv(y) + v(q'(y)) + v\left(\frac{dy}{dx}\right) = (s-1)v(x). \tag{19}$$

In fact $v(sp(x) + xp'(x)) = 0$ since $p(0) \neq 0$ and since $\pi \nmid s$.

On the other hand $v(y) = s'$, $v(x) = r'$, by (12). Since $rv(y) = rs' = sr' = sv(x)$ equation (19) gives

$$v\left(\frac{dy}{dx}\right) = -v(x) - \mu v(y) \leq -v(x). \tag{20}$$

But the equation $ar' + bs' = 1$ implies $\pi \nmid b$, so $v((b\,dy)/y\,dx) = -v(y) + v(dy/dx) \leq -v(y) - v(x) < -v(x) \leq v(a/x)$.

In conclusion (18) gives

$$v\left(\frac{dx}{du}\right) = -v(u) - v\left(\frac{b}{y}\frac{dy}{dx}\right) = v(x) + v(y)(1+\mu) - 1 = r' + s'(1+\mu) - 1.$$

Summing over $v \in S_{0,0}$ we obtain (ii). □

Lemma 11. *Assume that the curve $G(y) = H(x)$ has genus 0 and that if r is a prime number then for all $\lambda \in \boldsymbol{k}$ neither $G - \lambda$ nor $H - \lambda$ is the rth power of a polynomial. Then either for some linear functions L_1, M_1 and M_2*

$$L_1 \circ G \circ M_1 = D_m, \quad L_1 \circ H \circ M_2 = D_n \tag{21}$$

or

$$G(y) = (y - \eta)Q^r(y) + \lambda^*, \quad H(x) = (x - \xi)P^r(x) + \lambda^*,$$

where $Q(\eta)P(\xi) \neq 0$, P, Q have only simple zeros and $\pi \mid r$. Moreover,

$$\text{card } S_{x_0,y_0} = (r(y_0), s(x_0)) \quad \text{for all } \langle x_0, y_0 \rangle \in \Gamma.$$

Proof. We use Theorem A5 (i) applied with $z = x$ (separability is guaranteed by $G' \neq 0$) and $g = 0$. We split the summation over v as follows

$$-2 = \sum_{\langle x_0, y_0 \rangle \in \Gamma} c(x_0, y_0) + v_\infty \left(\frac{dx}{dt_\infty} \right).$$

This is permissible since at each prime divisor v above w_{x_0} the value $y(v)$ of the function y clearly satisfies $G(y(v)) = H(x_0)$.

Using the value for the last term obtained in Lemma 7 we obtain, after a short calculation

$$\delta = \sum_{\langle x_0, y_0 \rangle \in \Gamma} c(x_0, y_0) + (n - 1)(m - 1 - \delta), \tag{22}$$

where $\delta = \partial G'$.

Define now, for $y_0 \in \boldsymbol{k}$, $\delta(y_0)$ by

$$\delta(y_0) = 1 + \mu(y_0) \quad \text{if } \pi \mid r(y_0), \quad 0 \text{ otherwise}, \tag{23}$$

where $\mu(y_0)$ has been defined by (16).

If $\pi \nmid r(y_0)$ we have

$$r(y_0) - 1 + \delta(y_0) = \text{ord}_{\,y - y_0}\, G'(y). \tag{24}$$

If $\pi | r(y_0)$, differentiating (15) we find

$$G'(y) = (y - y_0)^{r_0} Q'_{y_0}(y)$$

and (22) holds again, thus it is true generally. In particular

$$\delta = \sum_{y_0 \in \boldsymbol{k}} (r(y_0) - 1 + \delta(y_0)).$$

By Lemma 10 we have, for given $y_0 \in \boldsymbol{k}$

$$\sum_{x_0, \langle x_0, y_0\rangle \in \Gamma} c(x_0, y_0) \geq \sum_{x_0, \langle x_0, y_0\rangle \in \Gamma} (r(y_0) - (r(y_0), s(x_0)) + s(x_0)\delta(x_0, y_0)), \tag{25}$$

where

$$\delta(x_0, y_0) = 1 + \mu(y_0) \ \text{ if } \pi | r(y_0),\ \pi \not| \ s(x_0),\ \ 0 \text{ otherwise.} \tag{26}$$

Using (22), (24) and (25) we thus obtain

$$\sum_{y_0 \in \boldsymbol{k}} \{r(y_0) - 1 + \delta(y_0)\} \geq \sum_{y_0 \in \boldsymbol{k}} \sigma(y_0) + (n-1)(m-1-\delta), \tag{27}$$

where

$$\sigma(y_0) = \sum_{x_0, \langle x_0, y_0\rangle \in \Gamma} (r(y_0) - (r(y_0), s(x_0)) + s(x_0)\delta(x_0, y_0)).$$

We proceed to estimate the terms $\sigma(y_0)$.

First observe that, if $r(y_0) > 1$, then $r(y_0)$ cannot divide $s(x_0)$ for all x_0 such that $\langle x_0, y_0\rangle \in \Gamma$, for otherwise $H(x) - G(y_0)$ would be an $r(y_0)$th power contrary to the assumption. We have thus two possibilities for given $r(y_0) > 1$, namely

Case 1. There exist two values of x_0 with $\langle x_0, y_0\rangle \in \Gamma$ and $r(y_0) \not| \ s(x_0)$.
Case 2. There is just one value x_0^* with $\langle x_0^*, y_0\rangle \in \Gamma$ and $r(y_0) \not| \ s(x_0^*)$.

We shall consider these cases successively.

Case 1. Since for the values in question $r(y_0) - (r(y_0), s(x_0)) \geq \frac{r(y_0)}{2}$, we have

$$\sigma(y_0) \geq r(y_0) + \sum_{x_0, \langle x_0, y_0\rangle \in \Gamma} s(x_0)\delta(x_0, y_0) \geq r(y_0) + \delta(y_0). \tag{28}$$

In fact $\pi \not| \ s(x_0^*)$ for at least one x_0^* with $\langle x_0^*, y_0\rangle \in \Gamma$, whence $s(x_0^*)\delta(x_0^*, y_0) = s(x_0^*)\delta(y_0) \geq \delta(y_0)$.

Case 2. Now clearly $(r(y_0), s(x_0^*))$ divides $s(x_0)$ for all relevant x_0, whence $H(x) - G(y_0)$ is an $(r(y_0), s(x_0^*))$th power. By the assumption $(r(y_0), s(x_0^*)) = 1$, whence

$$\sigma(y_0) \geq r(y_0) - 1 + \delta(y_0). \tag{29}$$

The same inequality clearly holds also if $r(y_0) = 1$, so using (27), (28) and (29) we see that Case 1 cannot occur, and that moreover $(n-1)(m-1-\delta) = 0$, so

$$\delta = m - 1, \quad \text{i.e. } \pi \nmid m \tag{30}$$

as $n > 1$.
Also, all the inequalities involved in (27) and (28) must be equalities for all $y_0 \in \boldsymbol{k}$, so in particular

$$\sum_{x_0, \langle x_0, y_0\rangle \in \Gamma} s(x_0)\delta(x_0, y_0) = \delta(y_0). \tag{31}$$

Assume there exist at least two values $y_1 \neq y_2$ with $r(y_i) > 1$ for $i = 1, 2$ and, say, $\lambda_1 = G(y_1) \neq G(y_2) = \lambda_2$.

Since we always end up in Case 2, producing (29) above, we have if $r(y_i) \nmid s(x_i)$ for certain x_i such that $\langle x_i, y_i\rangle \in \Gamma$

$$H(x) - \lambda_i = (x - x_i)^{s(x_i)} H_i^{r(y_i)}(x) \quad i = 1, 2. \tag{32}$$

Differentiating we find that $H'(x)$, (which is $\neq 0$), is divisible by both the polynomials $(x - x_i)^{s(x_i)-1} H_i^{r(y_i)-1}$, which are coprime, since $\lambda_1 \neq \lambda_2$. So, in particular

$$s(x_1) + s(x_2) - 2 + r(y_1)\partial H_1 + r(y_2)\partial H_2 - \partial H_1 - \partial H_2 \leq \partial H' \leq n - 1,$$

which gives

$$n - 1 \leq \partial H_1 + \partial H_2.$$

But, since $s(x_i) \geq 1$, $r(y_i) \geq 2$, (32) implies that $\partial H \leq \frac{n-1}{2}$, so we have in fact always equality, i.e. $s(x_1) = s(x_2) = 1$, $r(y_1) = r(y_2) = 2$, $\partial H_1 = \partial H_2 = \frac{n-1}{2}$ and finally $\partial H' = n - 1$, or equivalently, $\pi \nmid n$.

Also, $\pi \neq 2$, for otherwise, in view of (32) H_i^2 would divide H' for $i = 1, 2$, whence in particular $2(n - 1) \leq n - 1$, which is impossible. So we may apply Lemma 4 to the equation

$$(H(x) - \lambda_1)(H(x) - \lambda_2) = (x - x_1)(x - x_2)(H_1 H_2)^2,$$

which follows from (32) and the subsequent remarks. We obtain

$$H(x) \circ M_1^{-1} = \left(\frac{\lambda_1 - \lambda_2}{4}x + \frac{\lambda_1 + \lambda_2}{2}\right) \circ D_n(x) \tag{33}$$

for a suitable linear M_1.

Now, if there exists y_3 with $r(y_3) > 1$, while $\lambda_3 = G(y_3) \neq \lambda_i$ for $i = 1, 2$, we have similarly

$$H(x) - \lambda_3 = (x - x_3)H_3^2(x).$$

But then H' would be divisible by $H_1 H_2 H_3$, whence $\frac{3}{2}(n-1) \le n-1$, which is impossible.

So we may assume that

$$\text{if } r(y_0) > 1 \text{ then } G(y_0) = \lambda_i \text{ for } i = 1 \text{ or } i = 2. \tag{34}$$

Moreover, we have seen that necessarily $r(y_0) = 1$ or 2 in any case, and that $\pi \neq 2$.

Write $G'(y) = \alpha(y - \xi_1)\dots(y - \xi_{m-1})$. ξ_i are distinct, for otherwise $r(\xi_i) > 2$ for some i. So if, say

$$G(\xi_1) = \dots = G(\xi_h) = \lambda_1, \ G(\xi_{h+1}) = \dots = G(\xi_{m-1}) = \lambda_2,$$

we must have $m - 1 \ge \max\{2h, 2(m-1-h)\}$: in fact $G(y) - \lambda_1$ has the roots $\xi_1, \dots, \xi_h$ with multiplicity 2, and at least one root (otherwise it would be a square, contrary to the assumption), so $m \ge 1 + 2h$, and similarly for $G(y) - \lambda_2$. So necessarily $h = \frac{m-1}{2}$ and, for $i = 1, 2$

$$G(y) - \lambda_i = (y - \eta_i)(y - \xi_{1+h(i-1)})^2 \dots (y - \xi_{hi})^2,$$

say. Again Lemma 4 applies, so, for a suitable linear M_2

$$G(y) \circ M_2^{-1} = \left(\frac{\lambda_1 - \lambda_2}{4} y + \frac{\lambda_1 + \lambda_2}{2}\right) \circ D_m(y).$$

We thus end up in the case (21).

On excluding (34), where $\lambda_1 \neq \lambda_2$ there remains the only possibility

$$r(y_0) > 1 \text{ implies } G(y_0) = \lambda_1. \tag{35}$$

By symmetry we may assume

$$s(x_0) > 1 \text{ implies } H(x_0) = \lambda_2. \tag{36}$$

Equivalently

$$G'(y_0) = 0 \text{ implies } G(y_0) = \lambda_1, \quad H'(x_0) = 0 \text{ implies } H(x_0) = \lambda_2. \tag{37}$$

Write

$$G(y) - \lambda_1 = (y - y_1)^{r(y_1)} \dots (y - y_h)^{r(y_h)} V^{\pi}(y),$$

where $V(y_i) \neq 0$ for $i = 1, \dots, h$ and where $\pi \nmid r(y_1) \dots r(y_h)$.

We find

$$G'(y) = (y - y_1)^{r(y_1)-1} \dots (y - y_h)^{r(y_h)-1} V^{\pi}(y) U(y),$$

where $\deg U = h - 1$ and where $U(y_i) \neq 0$.

If $\pi \partial V = 0$ then, letting $U(y_0) = 0$ we would have $G'(y_0) = 0$, $G(y_0) \neq \lambda_1$. The existence of y_0 would therefore contradict (37). Thus $\pi \partial V = 0$

implies $h = 1$, or $G(y) - \lambda_1 = \alpha(y - y_1)^{r(y_1)}$, contrary to the assumption. Therefore, $\pi \partial V > 0$.

Let $V(y_0) = 0$. Then $\pi | r(y_0)$. Also by (31)

$$\sum_{x_0, \langle x_0, y_0 \rangle \in \Gamma} s(x_0)\delta(x_0, y_0) = s + \mu(y_0)$$

or

$$\sum_{x_0, \langle x_0, y_0 \rangle \in \Gamma,\ \pi \nmid s(x_0)} s(x_0)(1 + \mu(y_0)) = 1 + \mu(y_0).$$

We conclude that there is exactly one x_0^* such that $\langle x_0^*, y_0 \rangle \in \Gamma$, $\pi \nmid s(x_0^*)$, and that moreover $s(x_0^*) = 1$.

This means that

$$H(x) - \lambda_1 = (x - x_0^*)Z^\pi(x).$$

Necessarily $\partial Z > 0$, so by (36), $\lambda_1 = \lambda_2 = \lambda_1^*$, say.

By symmetry we have also $h = 1$ and $r(y_1) = 1$. So we may write

$$G(y) = (y - y_1)V^\pi(y) + \lambda^*, \quad H(x) = (x - x_0^*)Z^\pi(x) + \lambda^*$$

for some non-constant polynomials V, Z such that $V(y_1)Z(x_0^*) \neq 0$.

Recall that, for each $\langle x_0, y_0 \rangle \in \Gamma$ we must end up in the case producing (29), i.e. every multiplicity of every zero of V^π must divide the multiplicity of every zero, but one, of $H(x) - \lambda^*$, so it must divide the multiplicity of every zero of Z^π, and by symmetry, also the converse holds. So all multiplicities involved must be equal and we may write

$$G(y) = (y - \eta)Q^r(y) + \lambda^*, \quad H(x) = (x - \xi)H^r(x) + \lambda^*, \tag{38}$$

where $Q(\eta)P(\xi) \neq 0$, P, Q have only simple zeros and $\pi | r$.

Recall also that all the inequalities involved in (27) and (29) must be equalities. In particular

$$\text{card } S_{x_0, y_0} = (r(y_0), s(x_0)) \tag{39}$$

for all $\langle x_0, y_0 \rangle \in \Gamma$. □

Lemma 12. *If $r > 2$ the second term of the alternative in Lemma 11 is impossible. If $r = 2$ we have $y_0^3 Q'^2(y_0 + 1) = x_0^3 P'^2(x_0 + 1)$ for all x_0, y_0 satisfying $Q(y_0 + \eta) = P(x_0 + \xi) = 0$.*

Proof. Assume that the second term of the alternative holds. After a translation on x, y if necessary we may write the equation $G(y) = H(x)$ as

$$yQ^r(y) = xP^r(x), \tag{40}$$

where $P(0)Q(0) \neq 0$. Let $P(x_0) = Q(y_0) = 0$. The formula for card S_{x_0,y_0} now reads card $S_{x_0,y_0} = r$, so there are r prime divisors in S_{x_0,y_0} each necessarily unramified above w_{x_0} by (iii) of Lemma 9. Since each prime divisor $v \in S_{x_0,y_0}$ is unramified above w_{x_0}, a local parameter at each such v is $x - x_0 = u$, say. Let $y = S(u)$ be the power series expansion of y at $v \in S_{x_0,y_0}$. Write

$$Q(y) = (y - y_0)Q_1(y), \quad P(x) = (x - x_0)P_1(x),$$

where $Q_1(y_0)P_1(x_0) \neq 0$.

We have, say, $S(u) = y_0 + c_1u + c_2u^2 + \cdots$, so by (40)

$$(y_0 + c_1u + \cdots)(c_1 + c_2u + \cdots)^r = (x_0 + u)\left(\frac{P_1(x_0 + u)}{Q_1(S(u))}\right)^r. \tag{41}$$

Comparing constant terms we have

$$y_0c_1^r = x_0\left(\frac{P_1(x_0)}{Q_1(y_0)}\right)^r. \tag{42}$$

Write now

$$\frac{P_1(x_0 + u)}{Q_1(S(u))} = T(u) = \frac{P_1(x_0)}{Q_1(y_0)} + t_1u + \cdots.$$

Since $\pi | r$ we have

$$(T(u))^r \equiv \left(\frac{P_1(x_0)}{Q_1(y_0)}\right)^r \pmod{u^\pi}.$$

Also

$$(c_1 + c_2u + \cdots)^r \equiv c_1^r \pmod{u^\pi},$$

whence, by (41), comparing coefficients of u we find

$$c_1^{r+1} = \left(\frac{P_1(x_0)}{Q_1(y_0)}\right)^r. \tag{43}$$

Since $x_0y_0P_1(x_0)Q_1(y_0) \neq 0$ we may combine (42) and (43) to obtain

$$c_1 = \frac{y_0}{x_0}. \tag{44}$$

On the other hand c_1 is the value at v of the function $t = \frac{y-y_0}{x-x_0}$, so this value is uniquely determined by x_0, y_0 and otherwise independent of $v \in S_{x_0,y_0}$.

The present function t coincides with the one introduced in Lemma 9: in fact we now have, with the notation of Lemma 9, $r = s$, $d = r$, $r' = s' = 1$.

By (ii) of Lemma 9 we have that $(r(y_0), s(x_0))_* = 1$, i.e. $r_* = 1$, so r is a power of π, $r = \pi^\mu$, say.

We now show that, provided $r > 2$, the series $S(u)$ is uniquely determined by x_0, y_0, so at most one prime divisor in S_{x_0,y_0} is unramified above w_{x_0} in contradiction to what was shown at the beginning of the proof.

From (40) and the fact that $r = \pi^\mu$ we may write

$$y = (x_0 + u)S_1(u^r)$$

for a certain $S_1 \in \boldsymbol{k}[[u]]$.

Put $u^r = z$ and $S_1(u) = s_0 + s_1 u + s_2 u^2 + \cdots$. We have $s_0 = \frac{x_0}{y_0}$ and, from (40)

$$S_1(z)(y - y_0)^r Q_2((y - y_0)^r) = zP_2(z), \tag{45}$$

say, for certain Q_2, $P_2 \in \boldsymbol{k}[T]$, which depend only on x_0, y_0 and satisfy $Q_2(0)P_2(0) \neq 0$.

Put

$$Q_2(T) = \gamma_1 + \gamma_2 T + \cdots, \quad P_2(T) = \delta_1 + \delta_2 T + \cdots.$$

Now $y - y_0 = s_0 u + (x + u)s_1 z + (x_0 + u)s_2 z^2 + \cdots$, whence

$$(y - y_0)^r = s_0^r z + (x_0^r + z)s_1^r z^r + (x_0^r + z)s_2^r z^{2r} + \cdots. \tag{46}$$

Assume $s_0, \ldots, s_{h-1}$ given, where $h \geq 1$.

Let us consider the coefficient Γ_h of z^{h+1} on both sides of (45). On the left hand side write

$$(y - y_0)^r Q_2((y - y_0)^r) = A_1 z + A_2 z^2 + \cdots.$$

By (46) the coefficients of $1, z, z^2, \ldots, z^{h+1}$ in the series for $(y - y_0)^r$ depend only on the s_i with $i \leq \frac{h+1}{r}$, so we may write, for $j \leq h + 1$

$$A_j = A_j(s_0, s_1, \ldots, s_\nu) \in \boldsymbol{k}[s_0, \ldots, s_\nu],$$

where $\nu = \lfloor \frac{h+1}{r} \rfloor$. We have

$$\Gamma_h = s_h A_1 + s_{h-1} A_2 + \cdots + s_0 A_{h+1} = \delta_{h+1}.$$

But, since $A_1 = s_0^r \gamma_1 \neq 0$, we see that, provided $h > \nu$, s_h is uniquely determined by $s_0, \ldots, s_{h-1}, x_0, y_0$. Now $r > 2$, so, for $h \geq 1$, we have $(r - 1)h \geq r - 1 > 1$ and $h > \frac{h+1}{2} \geq \nu$. Since $s_0 = \frac{y_0}{x_0}$ depends only on x_0, y_0, induction shows that the same holds for all the s_h, as required.

This proves the above contention about the uniqueness of the power series for y and concludes the proof for $r > 2$. If $r = 2$ we combine (43) and (44) to obtain

$$y_0^3 Q_1^2(y_0) = x_0^3 P_1^2(x_0).$$

Recall that this equation must hold for every x_0, y_0 satisfying $Q(y_0) =$

$P(x_0) = 0$. Also observe that $Q_1(y_0) = Q'(y_0)$ and $P_1(x_0) = P'(x_0)$ for each such x_0, y_0. $\square$

Lemma 13. *If $\pi = 2$, for every positive integer n there is at most one solution of the equation*

$$tA^2 + B^2 + 1 + tAB = 0, \quad A, B \in \boldsymbol{k}[t] \tag{47}$$

with $\partial(tA^2 + B^2) = n$.

Proof. The equation (47) can be written as

$$(B + 1)^2 = tA(A + B).$$

Since $(A, B) = 1$ only two cases may arise, namely

$$\begin{aligned} &\text{Case 1}: A = tC^2,\ A + B = D^2,\ B + 1 = tCD \\ \text{or}\quad &\text{Case 2}: A = D^2,\ A + B = tC^2,\ B + 1 = tCD, \end{aligned}$$

where C, D are suitable polynomials in $\boldsymbol{k}[t]$.

In both cases, we obtain, eliminating A, B

$$tC^2 + D^2 + 1 + tCD = 0, \quad B = tC^2 + D^2.$$

We now proceed to prove the lemma by induction on n. If $n = 0$ then $A = 0$ and $\partial B = 0$ so $B = 1$. Assume that the lemma holds with n replaced by m, where $m < n$.

If $n = 2m$, then $\partial B = m$, hence by the inductive assumption C, D are uniquely determined and so is B. Now of the two polynomials A satisfying (47) at most one has degree $< \partial B$, thus the condition $n = \partial(tA^2 + B^2)$ also determines A uniquely.

If $n = 2m + 1$, then $m = \partial A \geq \partial B$. On the other hand in both cases considered above $\partial A \equiv \partial(A + B) + 1 \pmod 2$, whence $\partial B = \partial A = m$, by the inductive assumption C, D are uniquely determined and so is B. Of the two polynomials A satisfying (47) at most one has degree $= \partial B$, thus A is also uniquely determined. $\square$

Lemma 14. *If n is such that a solution A, B of (47) exists with $n = \partial(tA^2 + B^2)$, put*

$$R_n = tA^2 + B^2.$$

R_n satisfies the differential equation

$$t^3 R'^2(t) + 1 = R^2(t) + t^2 R(t) R'(t). \tag{48}$$

Proof. We have $R'(t) = A^2$ and we find

$$t^3 R'^2(t) + 1 - R^2(t) - t^2 R(t) R'(t) = (tA^2 + B^2 + tAB + 1)^2 = 0.$$

□

Lemma 15. *If R is a polynomial of degree n satisfying* (48) *then $R = R_n$.*

Proof. Write $R = tA^2 + B^2$, A, $B \in \boldsymbol{k}[t]$. (48) gives in view of $R' = A^2$,

$$1 + t^3 A^4 = t^2 A^4 + B^4 + t^3 A^4 + t^2 A^2 B^2,$$

i.e.

$$1 + t^2 A^4 + B^4 + t^2 A^2 B^2 = 0.$$

But this is just the square of (47). It now suffices to apply Lemma 13. □

Lemma 16. *Let*

$$R_0^* = 1, \; R_1^* = t + 1, \; R_n^* = tR_{n-1}^* + R_{n-2}^* \quad \textit{for} \;\; n \geq 2. \tag{49}$$

Then $R_n^* = R_n$.

Proof. Put

$$R_n^* = tA_n^{*2} + B_n^{*2}.$$

We have $\deg\, R_n^* = n$. Also (49) easily implies, for $n \geq 2$

$$A_n^* = B_{n-1}^* + A_{n-2}^*, \quad B_n^* = tA_{n-1}^* + B_{n-2}^*.$$

Hence by induction

$$tA_m^* A_{m-1}^* + B_m^* B_{m-1}^* = 1$$

and

$$tA_n^{*2} + B_n^{*2} + tA_n^* B_n^* + 1 = 0$$

which in view of Lemma 13 implies the lemma. □

Lemma 17. $tR_n^2 = D_{2n+1}$.

Proof. According to Lemma 16 we have $R_n = tR_{n-1} + R_{n-2}$, whence after squaring $tR_n^2 = t^3 R_{n-1}^2 + tR_{n-2}^2$. Thus setting $U_n = tR_n^2$ we find

$$U_0 = t, \quad U_1 = t^3 + t, \quad U_n = t^2 U_{n-1} + U_{n-2}.$$

However, according to Corollary 4 the polynomials D_{2n+1} satisfy the same recurrence formula and since $D_1 = U_0$, $D_3 = U_1$ by inspection, we have $D_{2n+1} = U_n$ for all n. □

Lemma 18. *If $\pi = 2$ a polynomial $R \in \boldsymbol{k}[t]$ has at least one simple zero, satisfies $R(0) \neq 0$ and $t_0^3 R'^2(t_0) = \lambda$ whenever $R(t_0) = 0$, then $tR^2 = \lambda D_{2n+1}\left(\frac{t}{\gamma}\right)$, where $\gamma \in \boldsymbol{k}^*$, $n = \partial R$.*

Proof. If t_1 is a simple zero of R then $t_1 \neq 0$ and $R'(t_1) \neq 0$, so $\lambda \neq 0$ and all zeros of R are simple. So

$$t^3 R'^2(t) = \lambda + R(t)V(t), \tag{50}$$

where $V \in \boldsymbol{k}[t]$.

We clearly have

$$\partial V = 3 + 2\partial R' - \partial R \equiv \partial R + 1 (\mathrm{mod}\ 2).$$

Also, differentiating (50) we find

$$t^3 R'^2(t) = R'(t)V(t) + R(t)V'(t) \tag{51}$$

whence, since $(R, R') = 1$, we have that

$$R' \text{ divides } V'. \tag{52}$$

Now if ∂R is even then ∂V is odd, whence $\partial V' = \partial V - 1 = 2 + 2\partial R' - \partial R$. But $\partial R' \leq \partial R - 2$ in this case, so $\partial V' \leq \partial R'$.

If, on the other hand, ∂R is odd then $\partial R' = \partial R - 1$ and ∂V is even, whence $\partial V' \leq \partial V - 2 = 1 + 2\partial R' - \partial R = \partial R'$.

So $\partial V' \leq \partial R'$ in any case, whence by (52)

$$V' = \gamma R' \tag{53}$$

for some $\gamma \in \boldsymbol{k}$. Actually $\gamma \neq 0$, for otherwise $t | V(t)$, by (51), whence $\lambda = 0$, a contradiction.

Plugging (53) into (51) we find

$$V(t) = t^2 R'(t) + \gamma R(t)$$

so

$$t^3 R'^2(t) + \lambda = \gamma R^2(t) + t^2 R(t) R'(t). \tag{54}$$

Set $R_1(t) = \frac{1}{\alpha} R(\gamma t)$, where $\alpha^2 = \frac{\lambda}{\gamma}$. Then $R(t) = \alpha R_1(t/\gamma)$ and substituting into (54) we obtain

$$\frac{\alpha^2}{\gamma^2} t^3 R_1'^2(t/\gamma) + \lambda = \gamma\alpha^2 R_1^2(t/\gamma) + \frac{\alpha^2}{\gamma} t^2 R_1(t/\gamma) R_1'(t/\gamma).$$

Change t into γt to find

$$\alpha^2\gamma t^3 R_1'^2(t) + \lambda = \alpha^2\gamma R_1^2(t) + \alpha^2\gamma t^2 R_1(t) R_1'(t),$$

since $\alpha^2\gamma = \lambda \neq 0$ we see that R_1 satisfies (48) and by Lemma 17

$$t R_1^2(t) = D_{2n+1}, \quad \text{where } n = \partial R.$$

Hence

$$t R^2(t) = \lambda D_{2n+1}(t/\gamma). \qquad \square$$

Lemma 19. *If $r = 2$ the second term of the alternative in Lemma* 11 *gives* (21).

Proof. The second term of the alternative in Lemma 11 gives for $r = 2$

$$G(y) = (y - \eta) Q^2(y) + \lambda^*, \quad H(x) = (x - \xi) P^2(x) + \lambda^*,$$

where $Q(\eta)$, $P(\xi) \neq 0$, p, $Q = 2$. By Lemma 12 we have

$$y_0^3 Q'(y_0 + \eta) = x_0^3 P'^2(x_0 + \xi) = \lambda$$

for all x_0, y_0 satisfying $Q(y_0 + \eta) = P(x_0 + \xi) = 0$. Hence polynomials $Q(t + \eta)$, $P(t + \xi)$ satisfy the assumptions of Lemma 18 and by that lemma

$$t Q(t + \eta)^2 = \lambda D_m(t/\gamma), \quad t P_n(t + \xi)^2 = \lambda D_n(t/\beta),$$

where $\beta, \gamma \in \boldsymbol{k}^*$. Thus (21) holds with

$$L_1^{-1} = \lambda t + \lambda^*, \quad M_1^{-1} = \frac{t - \eta}{\gamma}, \quad M_2^{-1} = \frac{t - \xi}{\beta}. \qquad \square$$

Lemma 20. *Let $G, H \in \boldsymbol{k}[t]$ have coprime degrees m, n, respectively. (We no longer assume $m, n > 1$.) Assume that both derivatives G', H' are non-zero, and that the curve $G(y) = H(x)$ has genus* 0*. Then there exist linear functions L_1, M_1, M_2 such that one of the following cases holds*

$$L_1 \circ G \circ M_1 = t^r P^n(t), \quad L_1 \circ H \circ M_2 = t^n \tag{55a}$$

(here P is a suitable polynomial, while $r \in \mathbb{N}$),

$$\textit{the same as } (55a), \textit{ but with } G, H \textit{ and } m, n \textit{ interchanged}, \tag{55b}$$