Book Review

Algebraic Function Fields and Codes

H. Stichtenoth
Springer-Verlag 1993, x+260 pp.
ISBN 3-540-56489-6
Price $34.00.

Reviewed by Gary McGuire

The purpose of this book is two-fold. Firstly, to give an exposition of the basic theory of algebraic function fields using a purely algebraic approach. Secondly, to give the applications of this theory to the theory of error-correcting codes. We refer to the book under review as [S]. Before we begin the review, we shall briefly discuss the topics covered in the book.

1. Algebraic Function Fields

Let K be any field. An algebraic function field F/K of one variable over K is an extension field F of K such that F contains an element x which is transcendental over K, and F is an algebraic extension of finite degree over $K(x)$.

The algebraic approach to studying such extensions F/K was first taken by Dedekind and Weber [4], with K the complex numbers. Chevalley [3] treated arbitrary K with this approach, and discussed geometry only with K the complex numbers. Algebraic geometry enters the picture as soon as one considers the plane algebraic curve C arising from F/K. This is defined by a polynomial equation $f(x,y) = 0$, where $F = K(x,y)$ and f has coefficients in K. Conversely, given a curve C defined by some irreducible polynomial $f \in K[x,y]$, the quotient field of the domain $K[x,y]/(f)$ is an algebraic function field of one variable, usually denoted $K(C)/K$. The geometric approach has been taken by many authors: Noether [13], Severi [14], Weil [19], and more recently one may consult Shafarevich [15], Hartshorne [7].

Two classics are Fulton [5] and Walker [18]. For a more sketchy presentation but with all the ideas, see Abhyankar [1] or Moreno [12].

In his review of [3], Weil [20] almost chastises Chevalley for the lack of geometry:

Here is algebra with a vengeance; ... if it were not for a few hints ... one might never suspect him of ever having heard of algebraic curves or of taking any interest in them.

He later concedes, it should be pointed out, that "this is a valuable and useful book". Not the least of the reasons for this is the strong analogy between the algebraic approach to algebraic functions (Chevalley) and the theory of algebraic numbers, *viz.* primes and irreducible polynomials, rational numbers and rational functions. For a simultaneous treatment of algebraic functions and algebraic numbers, see Artin [2].

Of course the "geometric" approach is through algebraic geometry, and involves a nontrivial amount of algebra itself. It would seem that this approach is the more popular. It is in reality a pleasant mixture of both algebra and geometry. For example, there is a one-to-one correspondence between points on a nonsingular curve C and places (maximal ideals of valuation rings) of $K(C)/K$.

A cornerstone of either approach is the Riemann-Roch Theorem (see Chapter I). A divisor A is a formal sum $\sum_P n_P P$ where the n_P are integers and only finitely many are nonzero. The sum is over all points on a curve, or all places of a function field, depending on one's point of view. The degree of a divisor, $deg(A)$, is $\sum_P n_P$. Assuming the existence of something called a canonical divisor, W, and the divisor of any $f \in F$, denoted (f), the Riemann-Roch theorem states that

$$\ell(A) - \ell(W - A) = deg(A) + 1 - g$$

where g is the genus of the curve C, or the function field F/K, and $\ell(A)$ is the dimension of the K-vector space

$$\mathcal{L}(A) = \{f \in F : (f) \geq -A\} \cup \{0\}.$$

The Riemann-Roch Theorem has many important consequences. For example, if f and h are two curves of degrees m and n over the complex numbers (let us say), then by Bézout's Theorem [1] f and h intersect in mn points (counted properly). Conversely, given f and mn points, does there exist a curve h of degree n which intersects f in precisely those mn points? Algebraically, we might ask: given specified poles and zeroes with multiplicities, does there exist $f \in K(x)$ with exactly those poles and zeroes? The Riemann-Roch Theorem provides answers.

2. Error-Correcting Codes

A code C over an alphabet Q is a subset of Q^n. Elements of C are called codewords, and n is called the length of the code. Usually Q is taken to be $\mathbf{F}_q$, the finite field of q elements. A linear code is a subspace of $\mathbf{F}_q^n$, and we assume linearity from now on. A code is called a code and not a subspace because of interest in a rather non-algebraic property, its minimum distance d. For $x, y \in Q^n$, the Hamming distance between x and y, $d(x,y)$, is defined to be the number of coordinates where x and y differ. For example, the distance between 110101 and 111100 ($q = 2$) is 2. Then d is defined by

$$d := min\{d(x,y) : x, y \in C, x \neq y\}.$$

If C is a k-dimensional subspace of $\mathbf{F}_q^n$, we say that C is a q-ary $[n,k,d]$ code. If $e = \lfloor \frac{d-1}{2} \rfloor$, C is an e-error-correcting code. This is because in practice, codewords are transmitted over a channel to a receiver. Due to noise there may be errors introduced during transmission, but if there are not more than e errors, the receiver can correct them and decode the received vector to the unique nearest codeword. Error-correcting codes are used every day in compact disc players, and have been used by NASA to receive data from space probes such as Mariner and Voyager. For an introduction to the theory of error-correcting codes, see [9] or [21].

For a fixed n (and q), a central problem in coding theory is to find codes which maximize both k and d. Unfortunately, the Singleton bound (trivial to prove) says

$$k + d \leq n + 1,$$

and so these aims are contradictory.

Much work has been done on bounds relating n, k, d and q. For asymptotic bounds, applicable for large n, the simplest results are obtained when the rate $R = k/n$ is plotted as a function of $\delta = d/n$. Clearly

$$R + \delta \leq 1 + \frac{1}{n}.$$

In fact, in Shannon's Theorem, the desirable codes whose existence is proven have very large n. However, constructing such codes is another matter.

Following [9], a family of codes over $\mathbf{F}_q$ (q fixed) is said to be *good* if it contains an infinite sequence of codes C_i, where C_i is an $[n_i, k_i, d_i]$ code, such that both the rate $R_i = k_i/n_i$ and $\delta_i = d_i/n_i$ approach a nonzero limit as $i \to \infty$.

Examples of classical families of codes are Hamming codes, BCH codes, Reed-Solomon codes and Reed-Muller codes. These codes have nice algebraic constructions and properties. It turns out that all these families are bad. Construction of good families became a problem. J.L. Massey said

... good codes just might be messy.

Justesen (1972) constructed an infinite family of good binary codes, see [9].

That good codes exist was never in doubt: the Gilbert-Varshamov lower bound states that if R is fixed, $0 \leq R \leq 1$, then there exist binary $[n,k,d]$ codes with $k/n \geq R$ and $d/n \geq H_2^{-1}(1-R) > 0$ where $H_2^{-1}(x)$ is the inverse of the entropy function $H_2(x) = -xlog_2(x) - (1-x)log_2(1-x)$.

There is an analogous statement of the Gilbert-Varshamov lower bound for any q, which can be translated into a lower bound for a function $\alpha_q(\delta)$ (which we leave undefined; see [S], Chapter VII).

3. Algebraic Function Fields and Codes

The main idea is that algebraic function fields can be used to construct codes which lead to an improved lower bound for $\alpha_q(\delta)$. It was thought for over thirty years that the Gilbert-Varshamov

lower bound would prove to be exact. Hence the improved lower bound caused a sensation in the field.

The improvement came in two stages. First came a construction from Goppa (1981) – after many years of trying to generalize his earlier work – of codes from algebraic function fields. We summarize this construction; it is fully described in [S], Chapter II. Let $F/\mathbf{F}_q$ be an algebraic function field of genus g and let $P_1, \ldots, P_n$ be pairwise distinct places of $F/\mathbf{F}_q$ of degree one. Let D be the divisor $P_1 + \cdots + P_n$ and let G be a divisor of $F/\mathbf{F}_q$ with disjoint support from D. The geometric Goppa code $C(D, G)$ (also called an algebraic geometry code [16]) is the image of the linear map $\beta : \mathcal{L}(G) \to \mathbf{F}_q^n$ defined by

$$\beta(f) = (f(P_1), \ldots, f(P_n)).$$

The dimension and a bound on the minimum distance are found by using the Riemann-Roch Theorem.

Asymptotic values of the ratio of the number of places of degree one to the genus (as $g \to \infty$) are related to whether geometric Goppa codes are good. Hence bounds on these asymptotic values (from algebraic geometry) can be related to the Gilbert-Varshamov lower bound.

The definition above can be phrased in terms of nonsingular curves, which is how Goppa first described it.

Choosing $F = \mathbf{F}_q(t)$, these codes are the Reed-Solomon codes mentioned earlier. Hence geometric Goppa codes are a natural generalization of Reed-Solomon codes.

The second stage of the improvement came about by finding certain suitable function fields (curves) $F/\mathbf{F}_q$. This was done by Ihara [8], and independently by Tsfasman, Vladut and Zink [17], although work of Manin [10] (and presumably unpublished) paved the way. Manin and Vladut [11] gave a proof using Drinfeld modules. The improved lower bound is valid for $q \geq 49$ and square, so in particular we are still without any codes beating the Gilbert-Varshamov bound in the binary case. The suitable curves that were used are Shimura curves [8], or Drinfeld curves [16], first suggested in [10]. [S] does not describe this second stage.

4. Review

Chapter I contains Weil's proof of the Riemann-Roch theorem using adeles. The author defines a Weil differential as a linear map on the space of adeles, vanishing on some translate of F (embedded). As the namesake of these differentials says,

This rather abstract concept of differential is of course what makes possible such a brief proof of the Riemann-Roch Theorem.

Later in Chapter IV these Weil differentials are identified with our "usual" notion of a differential. Chapter I also contains the Strong Approximation Theorem (crucial to many proofs in the book), Weierstrass gaps, and local components of Weil differentials (later to become residues of our usual differentials). This chapter is self-contained, requiring only basic graduate algebra. A useful appendix is provided with a summary of field theory.

In Chapter II the reader will find an introduction to coding theory and the definition of geometric Goppa codes. The dual code of $C(D, G)$ is also defined using local components of Weil differentials (later residues), and that it is the dual is proved using what is "really" the residue theorem (although not called such). Here we see perhaps the disadvantage to the algebraic approach. BCH and classical Goppa codes are constructed from geometric Goppa codes as subfield subcodes.

Chapter III (Extensions of Algebraic Function Fields) is the longest and most technical in the book. The presentation of many important ideas goes straight to the key theorems, and the proofs are concise but complete. As in Chapter I, however, the reader must make up his or her own examples in all sections except III.7. Topics covered include extensions and ramification, the different and the Hurwitz Genus formula, constant field extensions, Galois extensions (Kummer and Artin-Schreier), wild ramification, inseparability, and Castelnuovo's Inequality for the genus. A knowledge of algebraic number theory is useful (for familiarity purposes) but certainly not essential.

Chapter IV defines differentials via derivations and proves a one-to-one correspondence with Weil differentials from Chapter I.

We also find the P-adic completion of F/K with respect to a place P, giving us P-adic power series, analogous to complex power series over $\mathbf{C}$. Here lies the residue theorem, another cornerstone of the theory.

The length of the code $C(D, G)$ is limited by how many places of degree one the extension $F/\mathbf{F}_q$ has. The Hasse-Weil Theorem (also known as the Riemann hypothesis for function fields over finite fields, proved by Weil) tells us approximately how many places of degree one we can expect. This theorem is famous and has many implications, both inside and outside this book. Chapter V defines the zeta function of $F/\mathbf{F}_q$ and presents Bombieri's short elementary proof of Hasse-Weil. It uses only the Riemann-Roch Theorem. As Manin [10] points out, the proof of the upper bound is "quite code-theoretic in spirit". Again the presentation is faultless. Improvements to the Hasse-Weil bound are given with proofs, including the asymptotic lower bound due to Drinfeld-Vladut. This bound was proved to be tight (for $q \geq 49$ and square) by Ihara and Tsfasman-Vladut-Zink (by constructing the suitable curves to give equality).

The reader may heave a sigh of relief upon seeing the title of Chapter VI – Examples of Algebraic Function Fields. The author does say (page 30) that "we defer such examples to Chapter VI at which point we will have better methods at hand for calculating the genus". Indeed, results from all previous chapters are drawn on to give a thorough treatment of elliptic function fields. It is a pleasure to see characteristic 2 not excluded. Next are hyperelliptic function fields, and more generally, function fields $F = K(x, y)$ defined by $y^n = f(x)$. Examples done are Fermat ($ax^n + by^n = c$) and Hermitian ($x^{q+1} + y^{q+1} = 1$) function fields.

Chapter VII concerns the Gilbert-Varshamov bound and the story recounted earlier. Automorphism groups of geometric Goppa codes are discussed and applied to Hermitian codes (from the Hermitian function field). A decoding algorithm for geometric Goppa codes due to Skorobogatov and Vladut (following Justesen) is presented. However these codes are still a long way from being used in practice.

The final chapter (VIII) discusses the trace code of a q^m-ary code C of length n, which is defined as $Tr_q^{q^m}(C) \subseteq \mathbf{F}_q^n$, with trace taken componentwise. In certain special cases the minimum distance and perhaps all the weights in these codes can be found, or at least bounded. The bounds can be tight. Again results from previous chapters (especially chapter III) are used regularly.

A useful feature of this book is the two appendices, one containing a summary of field theory and the other explaining how to switch between the algebraic and geometric approaches, i.e. function fields and curves. The reader interested in curves can consult this second appendix and translate the results in the text to results about curves. Curves are not mentioned at all during the text, in keeping with the author's promise of an algebraic exposition. The reader may also find it helpful to glance at Chapter VI while reading the earlier chapters (especially I, III and IV), in order to see some examples.

The book is approximately 250 pages long and reasonably priced. It is typeset with some form of TEX, and one can have few complaints about that. A minor quibble concerns the letters of "Gal" and "Aut" (page 109), and "Der" (page 137), which are too close together. "Aut" has been corrected by page 209. The only typographical error this reviewer found (apart from a trivial one on page 144) is on page 243, where "monom" should be monomial. There are no exercises.

The exposition in this book is clean and tight. The quickest proofs of all the theorems are given, with no time wasted. The book is entirely self-contained. The author accomplishes what he set out to do with simplicity. It is recommended for anyone interested in algebraic function fields and their applications to codes.

References

[S] H. Stichtenoth, Algebraic Function Fields and Codes. Springer-Verlag: Berlin, 1993.

[1] S.S. Abhyankar, Algebraic Geometry for Scientists and Engineers. Mathematical Surveys and Monographs, Vol. 35, American Mathematical Society: New York, 1990.

[2] E. Artin, Algebraic Numbers and Algebraic Functions. Gordon and Breach: New York, 1967.

[3] C. Chevalley, Introduction to the Theory of Algebraic Functions of One Variable. Mathematical Surveys and Monographs, Vol. 6, American Mathematical Society: New York, 1951.

[4] R. Dedekind and H. Weber, *Theorie der algebraischen functionen einer veränderlichen*, Crelle J. **92** (1882), 181-290.

[5] W. Fulton, Algebraic Curves. Benjamin: New York, 1969.

[6] V.D. Goppa, Geometry and Codes. Mathematics and Its Applications, Vol. 24, Kluwer Academic Publishers: Dordrecht, 1988.

[7] R. Hartshorne, Algebraic Geometry. Graduate Texts in Mathematics, Vol. 52, Springer-Verlag: New York, 1977.

[8] Y. Ihara, *Some remarks on the number of rational points of algebraic curves over finite fields*, J. Fac. Sci. Univ. Tokyo **28** (1981), 721-724.

[9] F. J. McWilliams and N. J. A. Sloane, The Theory of Error-Correcting Codes. North-Holland: Amsterdam, 1977.

[10] Yu. I. Manin, *What is the maximum number of points on a curve over* $\mathbf{F}_2$?, J. Fac. Sci. Univ. Tokyo **28** (1981), 715-720.

[11] Yu. I. Manin and S. G. Vladut, *Linear codes and modular curves*, J. Soviet Math. **30** (1985), 2611-2643.

[12] C. J. Moreno, Algebraic Curves over Finite Fields. Cambridge Tracts in Mathematics, Vol. 97, Cambridge University Press: Cambridge, 1991.

[13] M. Noether, *Über einen satz aus der theorie der algebraischen functionen*, Math. Ann. **6** (1873), 351-359.

[14] F. Severi, Vorlesungen über Algebraische Geometrie. Teubner: Leipzig, 1921.

[15] I. R. Shafarevich, Basic Algebraic Geometry. Grundlehren der Mathematischen Wissenschaften, Vol. 213, Springer-Verlag: Berlin, 1977.

[16] M. A. Tsfasman and S. G. Vladut, Algebraic-Geometric Codes. Kluwer: Dordrecht, 1991.

[17] M. A. Tsfasman, S. G. Vladut and T. Zink, *Modular curves, Shimura curves and Goppa codes, better than Varshamov-Gilbert bound*, Math. Nachr. **109** (1982) 21-28.

[18] R. J. Walker, Algebraic Curves. Princeton University Press: Princeton, 1950.

[19] A. Weil, Sur les courbes algébriques et les variétés qui s'en déduisent. Hermann: Paris, 1948.

[20] A. Weil, Review of "Introduction to the Theory of Algebraic Functions of One Variable", Bull. A.M.S. **57** (1951), 384-398.

[21] J. H. van Lint, Introduction to Coding Theory. Graduate Texts in Mathematics, Vol. 86, Springer-Verlag, Second Ed.: Berlin, 1992.

Gary McGuire,
Department of Mathematics,
California Institute of Technology,
Pasadena, CA 91125,
USA.