# A THEORETICAL BASIS FOR PADÉ APPROXIMATION

Patrick Fitzpatrick

**Abstract:** The theory of Gröbner bases of polynomial ideals and modules has opened up new horizons in computational commutative algebra and algebraic geometry. We review this theory briefly and show how it leads to a new interpretation of the construction of (multivariable) Padé approximants as minimal elements in Gröbner bases. One of the more interesting aspects of this interpretation is its application to (1-variable) Padé approximation over a finite field, which is the key step in decoding the well-known classes of BCH and Goppa codes, normally carried out using the Berlekamp-Massey algorithm or the extended Euclidean algorithm. This leads to a new theoretical derivation for a decoding algorithm, which is—in its practical implementation—equivalent to that based on the extended Euclidean algorithm.

## 1. Introduction—Gröbner bases of ideals

The main difficulty in passing from the 1-variable polynomial ring $k[x]$ to the multivariable ring $k[x_1, \dots, x_n]$ is that there is no longer a uniquely specified division algorithm. In fact, it is no longer clear what is meant by a quotient and a remainder and whether or not these are well defined. In $k[x]$, division is based on successive comparison of the leading term of the divisor with that of the dividend/remainder—it is clear what these leading terms are and we implicitly use an ordering of monomials based on degree. In $k[x_1, \dots, x_n]$ many monomial orders (defined more precisely below) are possible and each has its own division algorithm.

This article is the text of a lecture given by the author at the September meeting of the Irish Mathematical Society held at the Regional Technical College, Waterford, September 3–4, 1992.

For example, consider ordering the monomials using *lex* order, that is, lexicographically, and let us take $x > y > z$. Then dividing $x^3 + x^2y^2 + yz$ by $x + y$ (we work over $\mathbf{Q}$ unless otherwise stated) gives

$$x^3 + x^2y^2 + yz = (x+y)(x^2 + xy^2 - xy - y^3 + y^2) \\ + y^4 - y^3 + yz.$$

On the other hand using *gradlex*—or *graduated lexicographic*—order, that is, using total degree first and ordering lexicographically the monomials of the same total degree, we obtain

$$x^2y^2 + x^3 + yz = (x+y)(xy^2 - y^3) + y^4 + x^3 + yz.$$

In both cases the algorithm stops because the leading monomial of the divisor does not divide the leading monomial of the remainder.

**Remark.** In the second case we could continue a little further by moving the $y^4$ term to the remainder and carrying out a further division based on comparison of the leading $x$ of the divisor with the $x^3$ term of the remainder to give

$$(x+y)(xy^2 - y^3 + x^2 - xy + y^2) + y^4 - y^3 + yz.$$

This difficulty is intimately related to the ideal membership problem. In $k[x]$ each ideal $I$ is principal, that is, it can be generated by a single element $g$ say, written $I = (g)$. Thus the division algorithm solves the ideal membership problem: by a simple argument $f \in I$ if and only if the remainder on division of $f$ by $g$ is 0. In $k[x_1, \dots, x_n]$ ideals are not usually principal (although by Hilbert's Basis Theorem they all have finite generating sets which we indicate by writing $I = (g_1, \dots, g_r)$), and the monomial order plays a crucial role. For example, suppose to investigate the membership or otherwise of a polynomial $f$ in the ideal $I$ we divide successively by the generators $g_j$ of $I$, determining the order of division by the leading monomial of $g_j$. Then we can derive seemingly contradictory equations as in the following example.

**Example 1.** In $\mathbf{Q}[x,y,z]$ we have

$$x^3+2xyz+xy+y = x(x^2+yz)+y(xz+x+1)$$

$$\begin{aligned}2xyz+x^3+xy+y &= 2x(yz+x^2)+0(xz+x+1)\\ &\qquad -x^3+xy+y\end{aligned}$$

where the first equation—based on gradlex with $x>y>z$—indicates that the polynomial on the left is in the ideal $(x^2+yz, xz+x+1)$, while the second—based on gradlex with $z>y>x$—seems to imply that it is not.

These difficulties were resolved by B. Buchberger [1] by the introduction of what he called *Gröbner bases* of polynomial ideals (in honour of his supervisor W. Gröbner who had suggested to him the problem of finding constructively a multiplication table for the quotient ring $k[x_1,...,x_n]/I$ and indicated a possible avenue of exploration). The existence of such bases—although not their construction—had already been discovered independently a year earlier by H. Hironaka [8] who called them *standard bases.* Since the early '70s their theory and applications have received wide attention and Gröbner basis routines are now implemented in all the major computer algebra packages.

Essentially, Buchberger focussed on the set of leading terms of the ideal $I$ in question, where the leading term $\mathrm{Lt}(p)$ of a polynomial $p$ is the greatest monomial of $p$ under the chosen monomial order. This monomial order $<$ can be varied—and different Gröbner bases of $I$ will result—but it must have certain properties, namely, it must be compatible with the multiplication so that if $\alpha,\beta,\gamma$ are monomials and $\alpha<\beta$ then $\alpha\gamma<\beta\gamma$, and also it must be a well-ordering (equivalently, $1<\alpha$ for every monomial $\alpha$). The set of leading terms of (non-zero) polynomials in $I$ is denoted $\mathrm{Lt}(I)$ and it generates an ideal $(\mathrm{Lt}(I))$. The existence of a finite basis for $(\mathrm{Lt}(I))$ may be established using Dickson's Lemma (cf. [3]) so there exist $p_1, \ldots, p_s \in I$ such that $(\mathrm{Lt}(I)) = (\mathrm{Lt}(p_1), \ldots, \mathrm{Lt}(p_s))$. Now it is clear that if $\{g_1, \ldots, g_r\}$ is a basis of $I$ then $(\mathrm{Lt}(g_1), \ldots, \mathrm{Lt}(g_r)) \subseteq (\mathrm{Lt}(I))$ but the reverse inclusion is not always true as the example above shows. There—using

gradlex with $x>y>z$—we have $I=(g_1,g_2)=(x^2+yz,\ xz+x+1)$ so $(\mathrm{Lt}(g_1),\mathrm{Lt}(g_2))=(x^2,xz)$, whereas $(\mathrm{Lt}(I))$ contains $\mathrm{Lt}(zg_1-xg_2)=\mathrm{Lt}(yz^2-x^2-x)=yz^2$ which is not in $(x^2,xz)$. The definition of a Gröbner basis is precisely that this reverse inclusion should hold, that is, $\{g_1, \ldots, g_r\}$ is a *Gröbner basis* of $I$ if $(\mathrm{Lt}(g_1), \ldots, \mathrm{Lt}(g_r)) = (\mathrm{Lt}(I))$. Moreover, it can be shown that if $\{g_1, \ldots, g_r\}$ is a subset of $I$ such that $(\mathrm{Lt}(g_1), \ldots, \mathrm{Lt}(g_r)) = (\mathrm{Lt}(I))$ then indeed $\{g_1, \ldots, g_r\}$ is a basis—*a fortiori* a Gröbner basis—of $I$. (In this approach Hilbert's Basis Theorem is derived as a corollary of Dickson's Lemma.)

Henceforth we write GB for Gröbner basis. The ideal membership problem is solved completely by GBs: $f\in I$ if and only if $f$ has remainder 0 under division by a GB of $I$. By division here we mean successive reduction of $f$ by multiples of the generators based on comparison of the leading terms of the GB with the leading terms of the dividend/remainder. The defining property of the GB ensures that such a reduction is always possible when the remainder is in $I$. In the example above with gradlex and $z>y>x$, $(yz+x^2, xz+x+1, x^3-2xy-y)$ is a GB for $(yz+x^2, xz+x+1)$ and the division algorithm now gives

$$\begin{aligned}2xyz+x^3+xy+y &= 2x(yz+x2)+0(xz+x+1)\\ &\qquad -1(x^3-xy-y)\end{aligned}$$

showing the polynomial on the left hand side to be in the ideal as required.

The construction of GBs—more about this later—is (unfortunately!) computationally complex in the general case and a great deal of research has gone into finding improvements to Buchberger's original algorithm, for example by studying the effects of changing the monomial order used. Examples are known however, that, no matter what refinements are introduced, will always take up large amounts of time and/or space because of expansion in the degrees of the polynomials in the basis or in the coefficients of the polynomials involved in the intermediate computations. This has not deterred the use of GBs in practice since it is believed that the constructions are "on the average" (and particularly when only two or three variables are involved) much less complex than the worst case.

A number of other fundamental problems in commutative algebra and algebraic geometry may be solved algorithmically using GBs (cf. [2]). Among these are the determination of whether or not a system of polynomial equations has finitely or infinitely many solutions (or none at all) and the constructive evaluation of these solutions in the finite case, the construction of the elimination ideals $I \cap k[x_1, \dots, x_j]$, $1 < j < n$, the implicitization problem (elimination of parameters), and the construction of bases of syzygy modules. It is this latter application that interests us here.

## 2. Gröbner bases of modules, syzygies and Padé approximants

We consider submodules $M$ of the free module $R^r$ where $R = k[x_1, \dots, x_n]$. Each such module has a finite basis and the theory of GBs can be extended in a natural way. The set of *terms of length* $r$ (replacing the monomials) is

$$T_r = \{(0, \dots, 0, \alpha_j, 0, \dots, 0) : \alpha_j \text{ is a monomial}\}.$$

If $<$ is a monomial order then we define a *term order* $<$ on $T_r$ by $(0, \dots, \alpha_j, \dots, 0) < (0, \dots, \alpha_l, \dots, 0)$ if $\alpha_j < \alpha_l$ or if $\alpha_j = \alpha_l$ and $j < l$. In fact, we require something slightly more general, namely, let $w = (\psi_1, \dots, \psi_r)$ be any *weight vector* where the $\psi_j$ are monomials and let $<$ be a monomial order. Then the term order $<_w$ on $T_r$ *induced* from $<$ and $w$ is defined by the relation $(0, \dots, \alpha_j, \dots, 0) < (0, \dots, \alpha_l, \dots, 0)$ if $\psi_j \alpha_j < \psi_l \alpha_l$ or if $\psi_j \alpha_j = \psi_l \alpha_l$ and $j < l$. The terms form a vector space basis of $R^r$. Henceforth, for definiteness, we shall use gradlex with $x_1 < \dots < x_n$ as our chosen monomial order.

We are particularly interested in modules of syzygies: given a set of polynomials $G = \{g_1, \dots, g_r\}$, the module of syzygies of $G$ is defined as

$$\operatorname{syz}(G) = \left\{ (h_1, \dots, h_r) \subseteq R^r : \sum_{j=1}^{r} h_j g_j = 0 \right\}.$$

In fact the construction of a GB from a given basis

$$G = \{g_1, \dots, g_r\}$$

of the ideal $I \subseteq R$ proceeds by calculating certain "S(yzygy)-polynomials" $\alpha g_i + \beta g_j$, namely, those that eliminate the leading terms of the pairs of polynomials $g_i$, $g_j$. These are then appended to the given basis and it was Buchberger's original contribution to prove that this procedure eventually terminates with a basis $G'$ in which all these S-polynomials may be expressed with certain restrictions on the coefficients. This property is equivalent to the defining property of a GB given above and thus it turns out that the construction of the GB $G' = \{g_1, \dots, g_r, g_{r+1}, \dots, g_t\}$ for $I = (g_1, \dots, g_r)$ actually produces, in addition, a GB for $\operatorname{syz}(G)$ under the term order *induced* from the monomial order in $R$ and the weight vector $(\mathrm{Lt}(g_1), \dots, \mathrm{Lt}(g_r))$. For more details see Möller and Mora [12].

We need one final piece of terminology: if $\phi$ is a monomial and $I$ is an ideal then $\phi$ is said to be *reduced* modulo $I$ if $\phi \notin (\mathrm{Lt}(I))$. Also, a polynomial $p$ is reduced modulo $I$ if each of its monomials is reduced modulo $I$. It is easy to see that if $G$ is a GB for $I$ then each polynomial $f \in R$ can be reduced using $G$ to a polynomial $p$ which is reduced modulo $I$. This is just the remainder on division of $f$ by $G$ provided that the division algorithm is extended—as in the remark at the beginning of section 1—to allow reduction *as far as possible* by every element $g$ of $G$, by comparing the leading term of $g$ with every monomial in the remainder rather than just the leading monomial.

Turning now to the problem of constructing Padé approximants we observe that this is a special case of solving for the pair $(a, b)$ the congruence

$$a \equiv bh \bmod I \qquad (*)$$

where $h$ is a given polynomial and $I$ is a given ideal. For the purposes of this exposition we restrict to the case that $I$ is a *monomial ideal* (that is, generated by monomials). The polynomial $h$ is derived by various means (such as Taylor expansion) from some more or less known function $u$ and the classical 1-variable

Padé approximation problem is to derive $(a, b)$ such that the quotient $a/b$ agrees with the expansion $h$ of $u$ as far as terms of a certain degree $m-1$ say, where restrictions are placed on $\deg(a)$ and $\deg(b)$ so that $\deg(a)+\deg(b)<m$. This may be interpreted as the solution of $(*)$ where $I=(x^m)$. In the classical theory a great deal of attention is (justifiably) paid to questions of convergence, but here we ignore such considerations altogether and deal only with the construction problem. One of the most interesting aspects of $(*)$ from our point of view is that in the 1-variable case it may be solved using the following theorem.

**Theorem 1** (cf. McEliece [11], Theorem 8.5, p.177). *Let $a$, $b$, $h$ be polynomials satisfying*

$$a \equiv bh \bmod x^m$$

*and suppose that $\deg(a)+\deg(b)<m$. Then in the extended Euclidean algorithm applied to $h$ and $x^m$ giving a sequence of remainders $r_j$, two sequences of auxiliary polynomials $u_j$, $v_j$, and a sequence of equations*

$$u_j h + v_j x^m = r_j$$

*there is a unique index $k$ and a polynomial $c$ such that*

$$a = cr_k, \quad b = cu_k.$$

Thus the construction of Padé approximants is completely solved in this case by the extended Euclidean algorithm. Of course, this does not make sense unless computations in the field $k$ are exact—so, for example, it makes no sense to consider using this method for Padé approximation using a machine representation of the real numbers. (For example what is the degree of the polynomial $10^{-10}x+1$, if the computer only has 8 decimal places of precision?) However, in another case of interest, congruence $(*)$ arises in the context of decoding BCH, Reed-Solomon and Goppa error correcting codes: $h$ is the *syndrome* polynomial, $b$ is the *error locator* polynomial and $a$ is the *error evaluator* polynomial

(for Goppa codes $x^m$ is replaced by the Goppa polynomial), and there the computations—over a finite field—*are* exact.

An alternative to this method of solution in the 1-variable case is to use the Berlekamp-Massey algorithm (cf. [10]); for the relationships between the extended Euclidean algorithm, the Berlekamp-Massey algorithm and linear recurring sequences see [4], [6]). The Berlekamp-Massey algorithm was generalized to $n$ variables by Sakata [13].

In [7], we gave a generalization of the Euclidean algorithm method by interpreting the solution of $(*)$—for arbitrary $I$—as a minimal element in a GB of a certain syzygy module. We outline this method in the next section, noting that because of the relative complexity of computing GBs this provides a *theoretical "basis"* for Padé approximation rather than a new practical method. However, in the 1-variable case our method turns out *in practice* to lead to an algorithm equivalent to that based on the extended Euclidean algorithm—we shall return to this point in Section 4. Moreover, in the context of multivariable codes and Goppa geometric codes, there are grounds for believing that our techniques may be valuable in the search for a general decoding algorithm alternative to that based on Sakata's extension of the Berlekamp-Massey algorithm.

## 3. Changing the term order

Further details for this section may be found in [7].

Let $\{g_1, \ldots, g_r\}$ be a GB for $I$ and consider the set

$$F = \{-1, h, g_1, \ldots, g_r\}$$

which is clearly a GB for $R$ (since it contains a scalar multiple of 1). We may assume that $h$ is reduced modulo $I$. Each solution of $(*)$ corresponds to an equation

$$a(-1) + bh + \sum_{j=1}^{r} c_j g_j = 0,$$

in other words to a syzygy on $F$. The algorithm for constructing a GB (in this case verifying that the set is a GB) gives a basis for $\mathrm{syz}(F) \subseteq R^{r+2}$ relative to the term order on $T_{r+2}$ induced by the chosen monomial order $<$ and the weight vector $(1, \mathrm{Lt}(h), \mathrm{Lt}(g_1), \ldots, \mathrm{Lt}(g_r))$. This basis consists of the elements $\{(h, 1, 0, \ldots, 0), (g_j, 0, \ldots, 0, 1, 0, \ldots 0), 1 < j < r\}$, where the second vector has a 1 in the $j+2$ place. These are just the "obvious" syzygies that one would write down immediately; what is important is that they form a GB.

Now write $M$ for the submodule of $R^2$ formed by the solutions $(a, b)$ of $(*)$. Then by projection on the first two places we find that $M$ has a GB $\{(h, 1), (g_j, 0), 1 < j < r\}$ under the term order on $T_2$ induced by $<$ and $w = (1, \mathrm{Lt}(h))$. Moreover, it can be shown that $(h, 1)$ is the unique element of least leading term (namely, $(0, 1)$) under this order. Again to simplify the exposition we now concentrate on the case that $I$ is generated by all the monomials of total degree $m$. Thus $I = (x_1^m, x_1^{m-1}x_2, \ldots, x_{n-1}x_n^{m-1}, x_n^m)$ and we observe that the given basis is a GB of $I$. Let the total degree $\tau(p)$ of a polynomial $p$ be defined as the maximum of the total degrees of its monomials. One example of the sort of restriction that may be placed on $a$, $b$ is the following *total degree condition:*

$$\tau(a) < k, \ \tau(b) < l,$$

where $k, l$ are non-negative integers and $k + l < m$. Then the following theorem is a special case of [7], Theorem 2.4.

**Theorem 2.** *Suppose that $(*)$ with $I$ generated by monomials of total degree $m$ has a reduced solution $(a, b)$ with $a, b$ relatively prime and satisfying the total degree condition above and let $w = (x_n^l, x_n^k)$. Then $(a, b)$ is the minimal reduced solution relative to the term order induced by $<$ and $w$ (uniquely defined up to a scalar multiple). A scalar multiple of $(a, b)$ appears in any GB of $M$ under this order.*

(Here a *reduced* solution is one in which both $a$ and $b$ are reduced modulo $I$ and a *minimal* solution is one of least leading term. Thus to calculate the required solution $(a, b)$ it is only necessary to convert the known GB $\{(h, 1), (g_j, 0)\}$ to a GB relative to the term order $<_w$ and pick out the minimal element.)

We end this section with two examples. The first is a 1-variable calculation derived from Knuth [9], Exercise 4, p. 515, while the second shows the method at work in $\mathbf{F}_2[x, y]$ where $\mathbf{F}_2$ is the field of 2 elements.

**Example 2.** Let $h = 7x^3 + 3x^2 + x + 1$ in $\mathbf{Q}[x]$. Then there are essentially four Padé approximants $(a, b)$ to $h$ modulo $I = (x^4)$, namely,

$$(h,\ 1), \quad (-2x^2 + 4x - 3,\ 7x - 3), \quad (x - 1,\ x^2 + 2x - 1),$$
$$(1,\ -2x^3 - 2x^2 - x + 1).$$

According to the Theorem these may be determined by finding minimal elements in GBs of the module $M$ generated by $\{(h, 1), (x^4, 0)\}$ relative to the term orders adapted to the weight vectors $(1, x^3)$, $(x, x^2)$ (equivalently $(1, x)$), $(x^2, x)$ (equivalently $(x, 1)$) and $(x^3, 1)$, respectively.

**Example 3.** Let $h = xy^3 + x^4 + y^3 + xy^2 + x^3 + x^2 + y + x$ in $\mathbf{F}_2[x, y]$. Suppose that $(a, b)$ exists as the Padé approximant for $h$ relative to the ideal $I$ generated by the monomials of total degree 5, where $a$ and $b$ are restricted to have total degree (at most) 2. Then we seek a GB for $M$ under the term order $<_w$ induced from $<$ and $w = (y^2, y^2)$ (equivalently $(1, 1)$). This is just the well-known *term-order-position* order (note $x < y$)

$$(1, 0) <_w (0, 1) <_w (x, 0) <_w (0, x) <_w \ldots$$

Converting the basis $\{(h, 1), (g_j, 0), 1 < j < r\}$ to a basis relative to this order we obtain

$$\{\, (x^3y + x^2y + x^3,\ x^3 + x^2),\ (0,\ x^4),\ (x^4 + x^2y + x^3,\ x^2),$$
$$(xy + y + x,\ y^2 + x + 1),\ (xy^2 + x^3,\ x^3 + xy + x^2),\ (0,\ x^3y),$$
$$(y^3 + x^2y + xy + y + x,\ y^3 + x^2y + xy + x + 1)\,\}$$

in which it is clear that the fourth element is the desired minimum. Hence

$$\frac{xy + y + x}{y^2 + x + 1} \equiv xy^3 + x^4 + y^3 + xy^2 + x^3 + x^2 + y + x \bmod I.$$

## 4. Solution of the key equation

We return now to the 1-variable version of $(*)$ taking $I = (x^{2t})$. The congruence now takes the form of what Berlekamp called the "key equation" for decoding a $t$-error correcting BCH (or Reed-Solomon, or Goppa) code. By its construction the solution module $M$ contains an element $(\omega, \sigma)$ where $\sigma$ (the error evaluator polynomial) and $\omega$ (the error locator polynomial) are relatively prime and $\sigma(0) = 1$, $\delta\sigma \leq t$, $\delta\omega < \delta\sigma$. Since we have a total degree condition this element is just the minimal element (unique up to a scalar multiple) in a GB of $M$ under the term order induced by $<$ (ordinary degree ordering among the monomials) and the weight vector $(x^{2t}, x^{2t-1})$ (equivalently, $(x, 1)$). The calculations to convert the known basis $\{(h, 1), (x^{2t}, 0)\}$ to a GB under this term order are identical to those which would be carried out in the Euclidean algorithm applied to $h$ and $x^{2t}$, which means that we have derived a new theoretical foundation for this algorithm.

It is in fact possible (cf. [5]) to develop the theory in this 1-variable case, without using the full machinery of GBs and thus to derive a justification for the algorithmic solution of the key equation which is (in our opinion) more intuitive and natural than those based on the Berlekamp-Massey or extended Euclidean algorithm.

## References

[1] B. Buchberger, An algorithm for finding a basis for the residue class ring of a zero-dimensional polynomial ideal (German). Ph. D. Thesis, Univ. Innsbruck (Austria), 1965.

[2] B. Buchberger, *Gröbner bases: an algorithmic method in polynomial ideal theory* in Multidimensional Systems Theory, N. K. Bose (ed.), Reidel: Dordrecht, 1985, 184–216.

[3] L. E. Dickson, *Finiteness of the odd perfect and primitive abundant numbers with n distinct prime factors,* Amer. J. of Math. **35** (1913), 413–426.

[4] J. L. Dornstetter, *On the equivalence between Berlekamp's and Euclid's algorithms,* IEEE Trans. Info. Thy. **IT-33** (1987), 428–431.

[5] P. Fitzpatrick, *A new algorithm for decoding BCH and Goppa codes using Gröbner bases of polynomial modules* (1992) (submitted for publication).

[6] P. Fitzpatrick and G. H. Norton, *The Berlekamp-Massey algorithm and linear recurring sequences over a factorial domain* (1990) (submitted for publication).

[7] P. Fitzpatrick and J. Flynn, *A Gröbner basis technique for Padé approximation,* J. Symbolic Computation **13** (1992), 133–138.

[8] H. Hironaka, *Resolution of singularities of an algebraic variety over a field of characteristic zero I, II,* Annals of Math. **79** (1964), 109–326.

[9] D. E. Knuth, The Art of Computer Programming: Vol. 2 Seminumerical Algorithms (2nd edn.). Addison-Wesley: Reading, Mass., 1981.

[10] S. Lin and D. J. Costello Jr., Error control coding: fundamentals and applications. Prentice-Hall: Englewood Cliffs, NJ, 1983.

[11] R. J. McEliece, The theory of information and coding. Addison-Wesley: Reading, Mass., 1987.

[12] H. M. Möller and F. Mora, *New constructive methods in classical ideal theory,* J. Algebra **100** (1986), 138–178.

[13] S. Sakata, *Extension of the Berlekamp-Massey algorithm to n dimensions,* Information and Computation **84** (1990), 207–239.

Patrick Fitzpatrick,
Department of Mathematics,
University College,
Cork.