GRÖBNER BASES AND GENERALIZED PADÉ APPROXIMATION

JEFFREY B. FARR AND SHUHONG GAO

Abstract. It is shown how to find general multivariate Padé approximation using Gröbner basis technique. This method is more flexible than previous approaches, and several examples are given to illustrate this advantage. When the number of variables is small compared to the degree of approximation, the Gröbner basis technique is more efficient than the linear algebra methods in the literature.

1. Introduction and Main Result

The classical Padé approximation theory for univariate polynomials says that for any polynomials $f, g \in \mathbb{F}[x]$, where $\mathbb{F}$ is any field and g has degree $t > 1$, and for any positive integers t_1 and t_2 with $t_1 + t_2 = t + 1$, there are polynomials $a \in \mathbb{F}[x]$ of degree $< t_1$ and $b \in \mathbb{F}[x]$ of degree $< t_2$ so that

$$b \cdot f \equiv a \mod g, \tag{1}$$

and the ratio a/b is unique for all the solutions a and b. Furthermore, the extended Euclidean algorithm can be used to find a minimal solution a and b. Is there a parallel theory for multivariate polynomials?

Given a function $f(x_1, \ldots, x_m)$, the generalized Padé approximation problem is to find suitable polynomials $a, b \in \mathbb{F}[x_1, \ldots, x_m]$ so that $f \equiv \frac{a}{b}$ modulo some predetermined conditions. The details of the requirements for a and b vary for different types of problems. A general approach is to consider solutions of the form

$$b \cdot f \equiv a \mod I, \tag{2}$$

where $I \subset \mathbb{F}[x_1, \ldots, x_m]$ is a given ideal. In the univariate case above, I is the ideal generated by g in $\mathbb{F}[x]$. In the multivariate case, the ideal I is more complicated. We shall see that for different choices of ideals, Equation (2) generalizes various forms of approximation that are studied in the literature.

The straightforward approach to finding a suitable Padé approximant is to recognize (2) as a homogeneous linear system (where the coefficients of a and b are

Date: February 5, 2004.

This work was supported in part by National Science Foundation (NSF) under Grant DMS0302549, National Security Agency (NSA) under Grant MDA904-02-1-0067, the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Office of Naval Research (ONR) under Grant N00014-00-1-0565. MITACS also partially supported the first author.

unknowns) and to apply Gauss elimination. This linear algebra approach has cubic complexity in the number of coefficients in a and b. We measure *the degree of approximation* by the total number of coefficients in a and b. Recently, some effort has been made to apply the method of Gröbner bases to the Padé approximation problem [11, 16]. The approaches in these two papers do not work in general. Even in the case that these methods are successful, they indicate that the contribution of Gröbner bases to this problem are primarily theoretical rather than practical since the efficiency of computing a Gröbner basis via Buchberger's algorithm is not comparable to the numerical Gauss elimination methods. However, our recent work in [8] indicates that the Gröbner basis in this problem can be computed efficiently and improves upon the linear algebra technique when the number of variables m is small relative to the degree of approximation.

To fix notation, we denote $\mathbb{F}[\mathbf{x}] = \mathbb{F}[x_1, \ldots, x_m]$, and, for any $\alpha = (\alpha_1, \ldots, \alpha_m) \in \mathbb{N}^m$ where $\mathbb{N}$ is the set of nonnegative integers,

$$\mathbf{x}^\alpha = x_1^{\alpha_1} \cdots x_m^{\alpha_m}.$$

We shall use monomial orders and Gröbner basis theory; see [2, 5, 6] for an excellent introduction. We fix an arbitrary monomial order on $\mathbb{F}[\mathbf{x}]$, and $\mathrm{LT}(g)$ denotes the leading term of a polynomial $g \in \mathbb{F}[\mathbf{x}]$. Define

$$\mathcal{B}(I) = \{\mathbf{x}^\alpha : \alpha \in \mathbb{N}^m \text{ and } \mathbf{x}^\alpha \neq \mathrm{LT}(g) \text{ for all } g \in I\}.$$

Then $\mathcal{B}(I)$ forms a basis for the quotient ring $\mathbb{F}[\mathbf{x}]/I$ as a vector space over $\mathbb{F}$; see [5]. The basis $\mathcal{B}(I)$ is called *the monomial basis* of I with respect to the monomial order used. Actually, we can define $\mathcal{B}(g_1, \ldots, g_s)$ for any set of polynomials in I by

$$\mathcal{B}(g_1, \ldots, g_s) = \{\mathbf{x}^\alpha : \alpha \in \mathbb{N}^m \text{ and } \mathbf{x}^\alpha \text{ is not divisible by any } \mathrm{LT}(g_i),\ 1 \leq i \leq s\}.$$

For the Padé approximation problem in (2), I will be a zero-dimensional ideal in $\mathbb{F}[\mathbf{x}]$, so the quotient ring $\mathbb{F}[\mathbf{x}]/I$ is finite dimensional as a vector space over $\mathbb{F}$. We call this dimension *the degree* of I. The degree of I corresponds to the degree of approximation mentioned above. The Padé approximation problem is to find certain "minimal" solutions a and b that satisfy (2) where "minimal" may mean total degree or any other conditions. We shall denote by M_f the set of all solutions, that is,

$$M_f = \{(a, b) \in \mathbb{F}[\mathbf{x}]^2 \ : \ a \text{ and } b \text{ satisfy (2)}\}.$$

One can check that M_f is closed under addition (of vectors) and if $(a, b) \in M_f$ then $h \cdot (a, b) = (ha, hb) \in M_f$ for all $h \in \mathbb{F}[\mathbf{x}]$. Hence, M_f forms a module over the ring $\mathbb{F}[\mathbf{x}]$.

Theorem 1. *Let I be a zero-dimensional ideal in $\mathbb{F}[\mathbf{x}] = \mathbb{F}[x_1, \ldots, x_m]$ of degree t. Fix a monomial order on $\mathbb{F}[\mathbf{x}]$, and denote the corresponding monomial basis by*

$$\mathcal{B}(I) = \{1 = \mathbf{x}^{\alpha_1}, \mathbf{x}^{\alpha_2}, \ldots, \mathbf{x}^{\alpha_t}\},$$

ordered in increasing order. Then for any $f \in \mathbb{F}[\mathbf{x}]$ and any positive integers t_1 and t_2 with $t_1 + t_2 = t + 1$, there is a pair of polynomials of the form

$$a = \sum_{i=1}^{t_1} a_i \mathbf{x}^{\alpha_i}, \qquad b = \sum_{i=1}^{t_2} b_i \mathbf{x}^{\alpha_i}, \tag{3}$$

not both zero, that satisfy (2). Further, there is a pair (a, b) of the above form that is contained in the reduced Gröbner basis for the module M_f under a certain term order.

The main issue in the above theorem is to define an appropriate term order on M_f for any given monomial order on $\mathbb{F}[\mathbf{x}]$. We shall deal with this issue in Section 2 and prove Theorem 1 in Section 3.

To see how Theorem 1 generalizes the approaches in the literature, let us consider the equation lattice approach as described in [7] where the main focus is on bivariate polynomials. In this approach, the shape of the numerator a is controlled by a set $N \subset \mathbb{N}^m$, and that of the denominator b by a set $D \subset \mathbb{N}^m$. That is,

$$a = \sum_{\gamma \in N} a_\gamma \mathbf{x}^\gamma \qquad \text{and} \qquad b = \sum_{\beta \in D} b_\beta \mathbf{x}^\beta, \tag{4}$$

where $a_\gamma, b_\beta \in \mathbb{F}$. In most applications, N and D are subsets of a delta set $E \subset \mathbb{N}^m$. Note that a subset E of $\mathbb{N}^m$ is called a *delta set* if $\beta \in E$ and $\alpha \in \mathbb{N}^m$ with $\alpha \leq \beta$ (componentwise) then $\alpha \in E$. As observed in [16], when E is a delta set, then the set

$$\{\mathbf{x}^\alpha : \alpha \in \mathbb{N}^m \text{ with } \alpha \notin E\}$$

is closed under multiplication and generates a monomial ideal, denoted by I_E, in $\mathbb{F}[\mathbf{x}]$. Furthermore, for any monomial order, the corresponding monomial basis of I_E is the same, namely

$$\{\mathbf{x}^\alpha : \alpha \in E\}.$$

Note that I_E has only one common zero, namely $(0, \ldots, 0)$, but with multiplicity equal to the cardinality of E. Hence, the equation (2) corresponds to approximating the Taylor expansion of f at the origin by a rational function a/b, and the degree of approximation is controlled by the cardinality of E, *i.e.*, the degree of the ideal I_E. We shall see in Section 4 how the shapes of a and b in Theorem 1 may vary when we vary the monomial order.

Monomial ideals are only one extreme case of general ideals I where I has only one common zero (but with multiplicity). Another extreme case is for I to be a radical ideal, so I has distinct zeros, and in this case the approximation problem is to find a rational function that interpolates the values of f at distinct points. Of course most ideals fall somewhere between these extremes and correspond to the interpolation of the Taylor expansions of f at different points with possibly different multiplicities. We shall demonstrate this by examples in Section 4.

Rational function approximation has applications including coding theory (decoding algebraic geometry codes, *e.g.*, Berlekamp-Welch [3], Fitzpatrick [10], Guruswami and Sudan [13]) and numerical analysis. Cuyt [7] provides a survey of progress over the past 30 years in attacking multivariate Padé approximation. Additionally, her extensive bibliography includes much of the work in this area by the numerical analysis community. Finally, multivariate polynomial interpolation is necessary in many disciplines, and Gasca and Sauer [12] provide an excellent survey of the varied approaches to this problem.

2. Term Orders for Modules

Most of the background material for Gröbner bases for modules can be found in [15] or in chapter 3 of [2]. Robbiano [19] proved that any valid monomial order on $\mathbb{F}[x_1, \ldots, x_m]$ may be described by a matrix $W \in \mathbb{R}^{\ell \times m}$ with $\ell \leq m$. Under such an order, two monomials are compared using the first row of W as a weighted degree; if the result is a tie, then the next row is used, then the third row, *etc.* Most computer algebra packages actually require the entries of W to be rational or integral. In so doing they admittedly give up some generality; for example, the $(1, \sqrt{2})$-*wdeg* order on x and y cannot be expressed by any 2×2 matrix over $\mathbb{Q}$.

A module over a ring is the analogue of a vector space over a field. We shall consider the ring $A = \mathbb{F}[x_1, \ldots, x_m]$. Let $r \geq 1$, and let M be any free module of rank r over A. This means that there are elements $\mathbf{z}_1, \ldots, \mathbf{z}_r \in M$ such that

$$M = A \cdot \mathbf{z}_1 + \cdots + A \cdot \mathbf{z}_r,$$

and every element $h \in M$ can be written uniquely as

$$h = h_1 \mathbf{z}_1 + \cdots + h_r \mathbf{z}_r, \quad \text{where } h_1, \ldots, h_r \in A.$$

The elements $\mathbf{z}_1, \ldots, \mathbf{z}_r$ thus form a *basis* for M over A. For a fixed basis, an element h can also be written as $(h_1, \ldots, h_r)$, hence M is isomorphic to A^r under this correspondence. In practice, we often work with a submodule of M. For example, the module M_f defined in the last section is a submodule of A^2. Note that submodules of a free module may not be free, *i.e.*, may not have a basis in the above sense.

We fix a basis $\mathbf{z}_1, \ldots, \mathbf{z}_r$ for a free module M over $A = \mathbb{F}[x_1, \ldots, x_m]$. A term of M is an element of the form $\mathbf{x}^\alpha \mathbf{z}_i$, where $\alpha \in \mathbb{N}^m$ and $1 \leq i \leq r$. Term orders on the terms of M are defined in much the same way that we define monomial orders on rings. Two types of term orders are used most of the time. A position-over-term (POT) order examines terms by first looking at the basis element. To compare terms which have identical basis elements, further discrimination is based on some monomial order on A. Conversely, in a term-over-position (TOP) order, the monomial order is applied first, and the position is the tiebreaker.

Example 1. *Suppose a* lex *order on* $\mathbb{F}[x, y]$ *with* $x < y$. *Let* $\mathbf{z}_1 = (1, 0)$ *and* $\mathbf{z}_2 = (0, 1)$, *and let* $\mathbf{z}_2 < \mathbf{z}_1$. *Under a POT order* $<_P$,

$$y^3 \mathbf{z}_2 = (0, y^3) <_P xy\mathbf{z}_1 = (xy, 0) <_P x^y \mathbf{z}_1 = (x^2 y, 0);$$

but under a TOP order $<_T$,

$$xy\mathbf{z}_1 = (xy, 0) <_T x^2 y \mathbf{z}_1 = (x^2 y, 0) <_T y^3 \mathbf{z}_2 = (0, y^3).$$

Unfortunately, most computer algebra systems allow only these two possibilities for module term orders (Maple is a notable exception). Of course, these are not the only possible term orders for modules; in fact, they represent only two extremes. By definition an order $<$ on the terms of M must meet the following requirements to qualify as a term order. An order $<$ is a term order on M if:

(i) $<$ is a total order (*i.e.*, either $\mathbf{X} < \mathbf{Y}$, $\mathbf{X} = \mathbf{Y}$ or $\mathbf{X} > \mathbf{Y}$ for any two terms $\mathbf{X}, \mathbf{Y} \in M$);

(ii) $\mathbf{X} < \mathbf{x}^\alpha \cdot \mathbf{X}$, for every term $\mathbf{X} \in M$ and every monomial $\mathbf{x}^\alpha \in A$ with $\mathbf{x}^\alpha \neq 1$;
(iii) If $\mathbf{X} < \mathbf{Y}$, then $\mathbf{x}^\alpha \cdot \mathbf{X} < \mathbf{x}^\alpha \cdot \mathbf{Y}$, for all terms $\mathbf{X}, \mathbf{Y} \in M$ and any monomial $\mathbf{x}^\alpha \in A$.

To define a weighted degree on terms of M, let $w = (w_1, \ldots, w_m, u_1, \ldots, u_r) \in \mathbb{R}^{m+r}$. Then the w-weighted degree of a term $\mathbf{x}^\alpha \cdot \mathbf{z}_j$ is defined to be

$$w \cdot (\alpha, e_j) = w_1\alpha_1 + \cdots + w_m\alpha_m + u_j,$$

where e_j is the jth unit vector $(0, \ldots, 0, 1, 0, \ldots, 0)$ and the dot product of two vectors means the inner product of vectors. An arbitrary term order can be defined by several weight vectors, represented as rows of a matrix. In general the term order matrix will be an $\ell \times (m + r)$ matrix T having block form

$$T = (W|U), \text{where } W \in \mathbb{R}^{\ell \times m} \text{and } U \in \mathbb{R}^{\ell \times r}.$$

Each column of W corresponds to the weights on one of the m variables, while each column of U corresponds to the weights on one of the r placeholder variables $\mathbf{z}_i$. For POT order $T = \begin{pmatrix} 0 & I_r \\ W & 0 \end{pmatrix}$, and for TOP order $T = \begin{pmatrix} W & 0 \\ 0 & I_r \end{pmatrix}$, where W is the monomial weight matrix.

If the rules of a legitimate term order are explicitly given, it is straightforward to construct the matrix T. Is there a characterization of T, though, that ensures that T induces a valid term order? The next theorem answers this question affirmatively for an integral matrix.

Theorem 2. *Suppose M is a free module of rank r over $\mathbb{F}[\mathbf{x}]$ with a basis $\mathbf{z}_1, \ldots, \mathbf{z}_r$ as above. A matrix $T = (W|U)$, where $W \in \mathbb{Z}^{(m+r)\times m}$ and $U \in \mathbb{Z}^{(m+r)\times r}$, defines a valid term order for the module M if and only if the following conditions hold.*

(1) *W has rank m.*
(2) *The first nonzero entry in each column of W is positive.*
(3) *Let W_1 be a left pseudo-inverse of W;* i.e., *$W_1W = (I_m, 0)^T$. Then if any two columns of W_1U agree in the final r entries, they must disagree in one of the first m entries by a nonintegral value.*

Proof: The first two requirements for T are similar to the requirements for a monomial order matrix and ensure that properties (ii) and (iii) of the definition of a term order are satisfied. The final requirement, though unusual in appearance, guarantees that T is a total order on the terms of M.

In particular, two distinct terms $\mathbf{x}^{\alpha_1} \cdot \mathbf{z}_i$ and $\mathbf{x}^{\alpha_2} \cdot \mathbf{z}_j$ are distinguishable if and only if $W(\alpha_1 - \alpha_2)^T \neq U(e_j - e_i)^T$. Hence $\mathbf{x}^{\alpha_1} \cdot \mathbf{z}_i$ and $\mathbf{x}^{\alpha_2} \cdot \mathbf{z}_j$ are distinguishable if and only if

$$\binom{I_m}{0} (\alpha_1 - \alpha_2)^T \neq W_1U(e_j - e_i)^T. \tag{5}$$

The righthand side of (5) is simply the difference of the ith and jth columns of W_1U. Hence, if any two columns of W_1U differ by integers on the first m entries and are identical on the final r, then an α_1 and α_2 may be found so that the two sides in (5) are equal and the terms $\mathbf{x}^{\alpha_1} \cdot \mathbf{z}_i$ and $\mathbf{x}^{\alpha_2} \cdot \mathbf{z}_j$ are indistinguishable. □

We now describe the role of term orders in applying Gröbner bases to Padé approximation. Define M to be the free $\mathbb{F}[\mathbf{x}]$-module of rank two; that is,

$$M = \mathbb{F}[\mathbf{x}] + \mathbf{z} \cdot \mathbb{F}[\mathbf{x}] \cong \mathbb{F}[\mathbf{x}]^2.$$

If we fix a polynomial $f \in \mathbb{F}[\mathbf{x}]$ and an ideal $I \subseteq \mathbb{F}[\mathbf{x}]$, then the module M_f defined earlier can also be written as

$$M_f = \{\mathbf{z} \cdot b - a | f \cdot b - a \in I\}. \tag{6}$$

Fitzpatrick and Flynn introduce a term-over-position (TOP) order for M. The monomial order on $\mathbb{F}[\mathbf{x}]$ is a weighted order built around a given initial order combined with a special order; this special order must cooperate with the desired (unknown) solution to satisfy a condition which they call a "weak term order" condition. This algorithm, while laying the necessary groundwork for using Gröbner bases in rational approximation, is quite restrictive.

Little *et al.* [16] were able to extend the use of Gröbner bases to a wider class of rational approximation problems by defining a new term order, denoted $\prec_\tau$. This term order can be thought of as a term-position-term order since the total degree (*tdeg*) of the term is considered first, then the position as a module element, and, finally, the size of the term with respect to a given monomial order. Specifically, if W is the $m \times m$ weight matrix for the given monomial order, then the $(m+2) \times (m+2)$ weight matrix for $\prec_\tau$ is

$$\begin{pmatrix} 1 & 1 & \dots & 1 & 0 & 0 \\ 0 & 0 & \dots & 0 & 1 & 0 \\ & & & & 0 & 0 \\ & & W & & \vdots & \vdots \\ & & & & 0 & 0 \end{pmatrix}.$$

Note that it is possible for one of the columns corresponding to a placeholder variable to be zero. In such a case, the zero column may be omitted as long as the remaining placeholder columns are clearly labeled.

Theorem 3 (Little *et al.*, 2003 [16])**.** *Assume that* $N = \{\mathbf{x}^\alpha : |\alpha| \leq t_1\}$, $D = \{\mathbf{x}^\beta : |\beta| \leq t_2\}$ *and* $|N| + |D| \geq |\mathcal{B}(I)| + 1$. *Then a Gröbner basis for* M_f *with respect to the* $\prec_\tau$ *order contains an element* (a, b) *such that* $\deg(a) < \deg(b) \leq t_2$, *assuming such a solution exists in* M_f.

As Little *et al.* [16] point out, these hypotheses are still very limiting, although they do allow for some important cases which the weak term order method does not. We highlight two of these limitations. First, N and D are limited to a triangular shape because of the dependence of $\prec_\tau$ on the *tdeg* order. The second drawback is that the Gröbner basis is only guaranteed to have a solution if that solution has a particular form. Of course solutions will not always be of this form. Remedies to this problem are suggested in [16]; however, they are either inconvenient or, in some cases, inadequate, particularly if N and D are to be of some shape besides triangular.

We address both of these shortcomings by introducing a new term order. Unlike the two previous orders, this new order intertwines the given monomial order with the module position at each step. Our approach allows N and D to take on a much

wider variety of shapes than has been studied previously. Included in these shapes are the types most frequently cited in the literature: the triangular ones mentioned previously (see also [14]) and rectangular shapes (see [17]). Many other shapes are possible by customizing I or the monomial order; see the examples in Section 4.

Lemma 4. *Suppose we have an arbitrary monomial order on $\mathbb{F}[\mathbf{x}]$ defined by a matrix $W \in \mathbb{R}^{\ell \times m}$. For any two terms $\mathbf{x}^{\alpha_1}$ and $\mathbf{x}^{\alpha_2} \cdot \mathbf{z}$ in $M = \mathbb{F}[\mathbf{x}] + \mathbf{z} \cdot \mathbb{F}[\mathbf{x}]$, we can define a term order on M so that $\mathbf{x}^{\alpha_1} < \mathbf{x}^{\alpha_2} \cdot \mathbf{z}$ and so that they are consecutive, that is, there is no other term in M that lies between them.*

Proof: Let w_i be the ith row of W and $c_i = w_i \cdot \alpha_1 - w_i \cdot \alpha_2$, $1 \le i \le \ell$. Define a term order matrix of the form

$$T = \begin{pmatrix} & & & c_1 \\ & W & & \vdots \\ & & & c_\ell \\ 0 & \dots & 0 & 1 \end{pmatrix},$$

where the last column is for $\mathbf{z}_2 = \mathbf{z}$. The column for $\mathbf{z}_1 = 1$ is the zero column, so omitted from T. By this term order matrix, we do have $\mathbf{x}^{\alpha_1} < \mathbf{x}^{\alpha_2} \cdot \mathbf{z}$ as they have the same weighted degree for the first ℓ rows of T but the last row distinguishes them. We need to show that there is no term in M that lies between $\mathbf{x}^{\alpha_1}$ and $\mathbf{x}^{\alpha_2} \cdot \mathbf{z}$. Let $\mathbf{x}^\beta$ be an arbitrary monomial in $\mathbb{F}[\mathbf{x}]$. It suffices to show that if $\mathbf{x}^{\alpha_1} < \mathbf{x}^\beta$ then $\mathbf{x}^{\alpha_2} \cdot \mathbf{z} < \mathbf{x}^\beta$ and if $\mathbf{x}^\beta \cdot \mathbf{z} < \mathbf{x}^{\alpha_2} \cdot \mathbf{z}$ then $\mathbf{x}^\beta \cdot \mathbf{z} < \mathbf{x}^{\alpha_1}$. We show the latter, as the former is similar. Note that $\mathbf{x}^\beta \cdot \mathbf{z} < \mathbf{x}^{\alpha_2} \cdot \mathbf{z}$ implies that $\mathbf{x}^\beta < \mathbf{x}^{\alpha_2}$, so there is an index k such $w_i \cdot \beta = w_i \cdot \alpha_2$ for $1 \le i < k$, but $w_k \cdot \beta < w_k \cdot \alpha_2$. Hence, $w_i \cdot \beta + c_i = w_i \cdot \alpha_2 + c_i = w_i \cdot \alpha_1$ for $1 \le i < k$, but $w_k \cdot \beta + c_k < w_k \cdot \alpha_2 + c_k = w_k \cdot \alpha_1$. Thus, $\mathbf{x}^\beta \cdot \mathbf{z} < \mathbf{x}^{\alpha_1}$ as claimed. □

3. Proof of Main Result

We begin with the following simple but useful lemma. Under a fixed term order on M, for any $G \subseteq M$, $B(G)$ denotes the set of terms in M that are not divisible by any leading terms of elements in G. Also, if S is a finite set of elements in M, then $Span(S)$ denotes the set of all linear combinations of elements in S with coefficients in $\mathbb{F}$; *i.e.*,

$$Span(S) = \left\{ \sum_{h \in S} a_h h : a_h \in \mathbb{F} \right\}.$$

Lemma 5. *Let $I \subseteq \mathbb{F}[\mathbf{x}]$ be a zero-dimensional ideal with $\dim(\mathbb{F}[\mathbf{x}]/I) = t$, and fix any term order on the module $M = \mathbb{F}[\mathbf{x}] + \mathbf{z} \cdot \mathbb{F}[\mathbf{x}] \cong \mathbb{F}[\mathbf{x}]^2$. Then, for any $f \in \mathbb{F}[\mathbf{x}]$ and M_f as defined by (6),*

(i) *the quotient module M/M_f has dimension t as a vector space over $\mathbb{F}$;*
(ii) *$|\mathcal{B}(u_1, \dots, u_s)| \ge t$, for any $u_1, \dots, u_s \in M_f$;*
(iii) *$\{u_1, \dots, u_s\}$ is a Gröbner basis for M_f if and only if $|\mathcal{B}(u_1, \dots, u_s)| = t$.*

Proof: (i) Since $\dim_{\mathbb{F}}(M/M_f)$ does not depend on the term order, we can assume a POT order on M with $\mathbf{z} >$ all x_i and any fixed monomial order on the x_i.

We immediately notice that $I \subset M_f$ (where polynomials in $\mathbb{F}[\mathbf{x}]$ are now viewed as module elements) and that $\mathbf{z} - f \in M_f$.

Next, we claim that if $\{w_1, \ldots, w_s\}$ is a Gröbner basis for I, then $\{w_1, \ldots, w_s, \mathbf{z} - f\}$ is a Gröbner basis for M_f. Certainly, for any element $u \in M_f$, either $\mathrm{LT}(u) = \mathbf{z} \cdot \mathbf{x}^\alpha$ so that $\mathrm{LT}(u)$ is divisible by $\mathrm{LT}(\mathbf{z} - f) = \mathbf{z}$ or $u = 0 \cdot \mathbf{z} + a$ in which case $a \in I$ and $\mathrm{LT}(u) = \mathrm{LT}(a)$ is divisible by some $\mathrm{LT}(w_i)$, $1 \leq i \leq s$. Finally, $\mathcal{B}(w_1, \ldots, w_s, \mathbf{z} - f) = \mathcal{B}(w_1, \ldots, w_s)$, implying $\dim_{\mathbb{F}}(M/M_f) = t$.

(ii) Suppose $G = \{u_1, \ldots, u_s\} \subseteq M_f$. Then each $w \in M$ can be reduced by G to a polynomial in $Span\{B(G)\}$. So $Span\{B(G)\}$ contains a basis for M/M_f; consequently, $|\mathcal{B}(G)| \geq \dim(M/M_f) = t$.

(iii) If G is a Gröbner basis for M_f, then each $w \in M$ has a unique reduction with respect to G. So $|\mathcal{B}(G)| = t$. Conversely, if $|\mathcal{B}(G)| = t$, then, since the terms in $\mathcal{B}(G)$ are linearly independent, they must form a basis for M/M_f. So, for any $w \in M_f$, as $w \equiv 0$ in M/M_f, we see that w must be reduced to zero by G. Hence G is a Gröbner basis. □

Proof of Theorem 1: Notice that in a and b, we have $t_1 + t_2 = t + 1$ coefficients to determine. Further, notice that we have t linear equations implied by the congruence $f \cdot b - a \equiv 0 \mod I$. Since the system is homogeneous, it can not be inconsistent, and, since the system has more unknowns than equations, a nontrivial solution is guaranteed. So, there is at least one element, $b\mathbf{z} - a$ satisfying (2), in the module M_f with $a, b \in Span\{\mathbf{x}^{\alpha_1}, \ldots, \mathbf{x}^{\alpha_t}\}$ not both zero.

Suppose the monomial order on $\mathbb{F}[x_1, \ldots, x_m]$ is defined by an $\ell \times m$ matrix W. Using Lemma 4, we define a term order on $M = \mathbb{F}[\mathbf{x}] + \mathbb{F}[\mathbf{x}] \cdot \mathbf{z}$ so that $\mathbf{x}^{\alpha_{t_1}} < \mathbf{x}^{\alpha_{t_2}} \cdot z$ and so that they are consecutive. Now suppose the last part of the theorem is false. Thus, for any nonzero solution $b\mathbf{z} - a \in M_f$, we would have either $\mathrm{LT}(a) > \mathbf{x}^{\alpha_{t_1}}$ or $\mathrm{LT}(b) > \mathbf{x}^{\alpha_{t_2}}$. Let G be any Gröbner basis for M_f. Then none of the terms $\mathbf{x}^{\alpha_1}, \ldots, \mathbf{x}^{\alpha_{t_1}}, \mathbf{z} \cdot \mathbf{x}^{\alpha_1}, \ldots, \mathbf{z} \cdot \mathbf{x}^{\alpha_{t_2}}$ is divisible by a leading term of G. Hence, $\{\mathbf{x}^{\alpha_1}, \ldots, \mathbf{x}^{\alpha_{t_1}}, \mathbf{z} \cdot \mathbf{x}^{\alpha_1}, \ldots, \mathbf{z} \cdot \mathbf{x}^{\alpha_{t_2}}\} \subseteq \mathcal{B}(G)$, implying that $|\mathcal{B}(G)| \geq t + 1$. This contradicts Lemma 5, and the theorem is proved. □

Theorem 1 proves that at least one solution exists in the Gröbner basis, but uniqueness of the solution is not guaranteed even if $\gcd(a, b) = 1$ is required, as Example 4 below shows. Additionally, to find a solution, one can simply find a Gröbner basis for M_f. How the Gröbner basis is computed depends on how I is given. Specifically, if we are already given a Gröbner basis for I, then the Gröbner basis for M_f can be computed by a Gröbner basis conversion method such as Gröbner walk or FGLM. On the other hand, if I is given by a vanishing set of points (distinct or with multiplicities), then the algorithm in [1, 4, 8, 18] may be used.

We note that the guaranteed solution is minimal with respect to the degree of b. Finally, if $t_2 = 1$ then we are back to the case of polynomial interpolation.

4. Examples

We now illustrate the power and flexibility of Theorem 1. Specifically, the following examples show the impact of size, shape and monomial order on the Padé approximant.

Except in Examples 3 and 5, we work over the finite field with two elements, $\mathbb{F}_2$. Computations are done with the computer algebra package Magma using the algorithm in [8].

Example 2. We first examine a triangle-shaped Karlsson-Wallin approximant discussed by Little *et al.* in [16]. Unlike their approach, though, our method has the flexibility to allow $D \subset N$, $N \subset D$ or $N = D$. Additionally, we are able to choose N or D to be "incomplete" triangles, *e.g.*, $\{1, x, y, x^2, xy\}$ rather than $\{\mathbf{x}^\alpha : |\alpha| \leq 2\}$, although Karlsson-Wallin approximants assume complete triangles.

Let $I = \langle xy^4, y^5, x^6, x^5y, x^4y^2, x^3y^3\rangle \subset \mathbb{F}_2[x, y]$, and let $N = D = \{\mathbf{x}^\alpha : |\alpha| \leq 3\}$. So $|\mathcal{B}(I)| = 19$, and $|N| = |D| = 10$. If we assume a graded lexicographical order (*glex*) with $y > x$, then the conditions for Theorem 1 are satisfied.

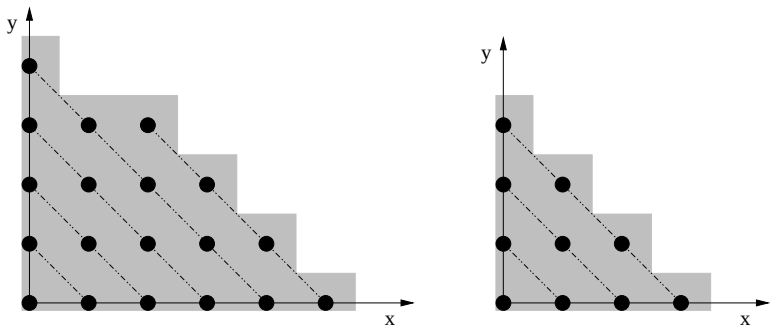

Figure 1. The shape of $\mathcal{B}(I)$ (left) and N and D in Example 2

Let $f = 1 + x + y + y^2 + xy + x^3 + xy^2 + x^4 + x^2y^2 + xy^3 + x^4y + x^2y^3$. We find that the Gröbner basis for M_f contains the desired solution

$$(a, b) = (y^3 + xy^2 + x^2y + y^2 + y + x, xy^2 + x^2y + x^3 + x^2 + y + x).$$

It is often the case that we wish $b(0) \neq 0$. Little *et al.* point out that we can assume this is true for sufficiently general f. This assumption is accurate when working over fields of characteristic zero; however, $b(0) = 0$ with probability q^{-1} over the field $\mathbb{F}_q$. Unfortunately, for small fields such as $\mathbb{F}_2$, b will not have the desired property a significant portion of the time.

Keeping I and the monomial order on $\mathbb{F}_2[x, y]$ the same, we can vary the sizes of N and D as long as we ensure that $|N| + |D| = |\mathcal{B}(I)| + 1$. In the extreme case $N = \mathcal{B}(I)$, $D = \{1\}$, the problem reduces to the multivariate polynomial interpolation problem, and the module order is the simple POT order.

Example 3 (Coding theory)**.** In this example we work over the finite field with nine elements generated by a primitive eighth root of unity ω; *i.e.*, if ω satisfies the primitive polynomial $p(x) = x^2 - x - 1 \in \mathbb{F}_3[x]$, then $\mathbb{F}_9 = \{0, \omega^0 = 1, \omega, \omega^2, \ldots, \omega^7\}$. We use a weighted degree (*wdeg*) order of $(3, 4)$ on x and y, ties broken by $y > x$. Set $I = \langle x^9 - x, y^3 - y - x^4 \rangle$ and $|N| = 22$ and $|D| = 6$. Here, I is radical; *i.e.*, each point in $\mathbf{V}(I)$ is distinct. This corresponds to the problem of decoding the five-error-correcting $[27, 14, 11]_9$ Hermitian code—see [9] for details. The only difference is that in the coding problem the function f, corresponding to the received vector, is given implicitly.

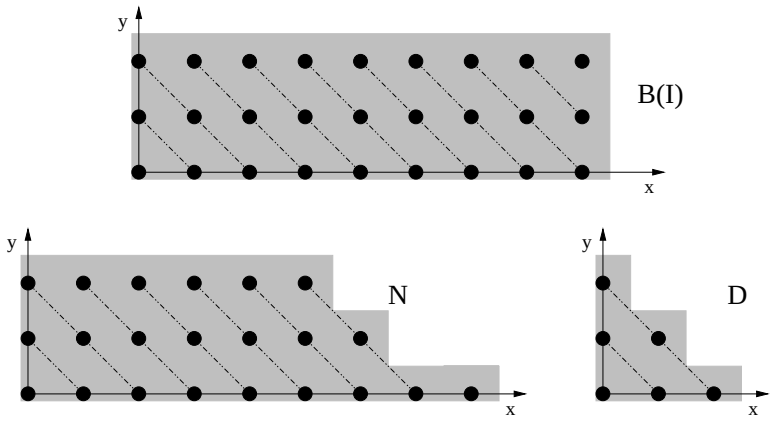

FIGURE 2. The shape of $\mathcal{B}(I)$, N and D in Example 3

We let

$$\begin{aligned} f &= \omega^6 x^8 y^2 + \omega^5 x^8 y + 2x^6 y^2 + x^7 y + \omega^3 x^8 + x^6 y + x^7 + \omega^6 x^4 y^2 + \omega^2 x^5 y \\ &\quad + \omega^5 x^6 + \omega^3 x^4 y + x^5 + 2x^2 y^2 + \omega^7 x^4 + \omega x y^2 + \omega^6 x^2 y + x^3 + \omega^2 x y + \omega^2 x^2 \\ &\quad + \omega y + \omega^3 x + \omega^6. \end{aligned}$$

The Gröbner basis contains the desired entry

$$\begin{aligned} (a, b) &= (\omega y^2 x^5 + \omega y x^6 + \omega^2 x^7 + \omega^6 y^2 x^4 + \omega^7 y x^5 + x^6 + \omega^3 y^2 x^3 + y x^4 + y x^3 \\ &\quad + \omega^5 x^4 + \omega^5 y^2 x + \omega^5 y x^2 + \omega x^3 + \omega^6 y^2 + \omega^7 y x + 2x^2 + \omega^6 y + x + \omega^5, \\ &\quad x y + 2x^2 + \omega^5 y + \omega^3 x + \omega^7). \end{aligned}$$

Here,

$$\begin{aligned} b \cdot f - a &= \omega^6 y^3 x^9 + \omega^2 y^2 x^{10} + \omega^3 y^3 x^8 + \omega y x^{10} + 2y^3 x^7 + \omega^3 y x^9 + \omega y^3 x^6 \\ &\quad + \omega^7 x^{10} + \omega^3 y x^8 + \omega^6 y^3 x^5 + \omega^3 x^9 + 2y x^7 + \omega^3 y^3 x^4 + \omega^7 x^8 + \omega y x^6 \\ &\quad + 2y^3 x^3 + x^7 + \omega^6 y x^5 + \omega^5 y^3 x^2 + \omega x^6 + \omega^3 y x^4 + \omega^6 y^3 x + \omega^6 x^5 \\ &\quad + \omega^6 y^2 x^2 + 2y x^3 + \omega^7 x^4 + \omega y x^2 + x^3 + \omega^5 y x + \omega^2 x^2 + \omega^5 x. \end{aligned}$$

Since I is not a monomial ideal, it is not immediately obvious that $bf - a$ is zero modulo I. However, $bf - a$ can indeed be reduced to zero by $x^9 - x$ and $y^3 - y - x^4$, the two polynomials in the Gröbner basis for I.

In the coding theory setting, b is an error-locator polynomial, and a "contains" the original message polynomial. We remark that the method of [16] is able to handle the *wdeg* order in this case since it happens to be equivalent to a total degree order on $\mathcal{B}(I)$; however, the method still is unable to handle the incomplete row in N.

Example 4. Next, we compare the effects of using three different pairs of N and D with a single I; in each case, though, $|N| = |D| = 11$. We again choose a rectangular shape for $\mathcal{B}(I)$ by letting $I = \{y^3, x^7\}$. Our function to approximate is

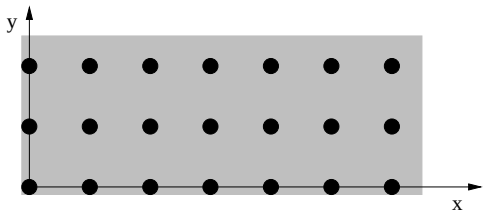

FIGURE 3. The shape of $\mathcal{B}(I)$ in Example 4

$$f = 1 + x + y + x^2 + xy + xy^2 + x^4 + x^2y^2 + x^5 + x^4y + x^6 + x^4y^2 + x^6y + x^6y^2.$$

First, we impose a *lex* order on $\mathbb{F}_2[x, y]$ with $x > y$. N_1 and D_1 have x^3y as the leading term and have the shape indicated in Figure 4. The resulting approximant is

$$(a_1, b_1) = (x^3y + x^2y^2 + xy + y^2, x^3y + x^2y + x^3 + xy + y^2).$$

Compare this output with that of choosing N_2 and D_2 by allowing a *lex* order on $\mathbb{F}_2[x, y]$ with $y > x$. Again, the leading term is x^3y, but Figure 4 shows that the shape of N_2 and D_2 is different than that of N_1 and D_1. The approximant appearing in the Gröbner basis is also different:

$$(a_2, b_2) = (x^2y + y + x^6 + x^5 + x^3 + x^2 + x + 1, xy + x^6 + x^5 + x^3 + 1).$$

Thus, the shape of N and D, influenced by the monomial order and by the chosen cardinality, results in very different approximants even though the leading term of the sets is the same.

Consider the set N_3 and D_3 using a *wdeg* order $(1, 3)$ on x and y with ties broken by $y > x$. Notice that N_3 and D_3 are exactly the shape of N_2 and D_2; however, the sets are ordered differently, indicated by the dotted lines. In this case the Gröbner basis yields two solutions

$$\begin{aligned}(a_3, b_3) &= (x^6 + x^2y + x^5 + y + x^3 + x^2 + x + 1, x^6 + x^5 + xy + x^3 + 1)\\ (a_4, b_4) &= (x^3y + x^2y + x^5 + xy + x^4 + y + 1, x^2y + x^5 + xy + x^4 + x^3 + x + 1).\end{aligned}$$

Observe that (a_3, b_3) is actually equal to (a_2, b_2). We might have anticipated this since N_3 and D_3 have the same shape as N_2 and D_2. Where did the additional solution come from? Closer inspection shows that the leading term of $(a_4, b_4) = (x^3y, 0)$ is divisible by $\mathrm{LT}(a_2, b_2) = (x^2y, 0)$, but not by $\mathrm{LT}(a_3, b_3) = (x^6, 0)$. Hence, even among N and D of the same size and shape, different solutions may arise from various monomial orders.

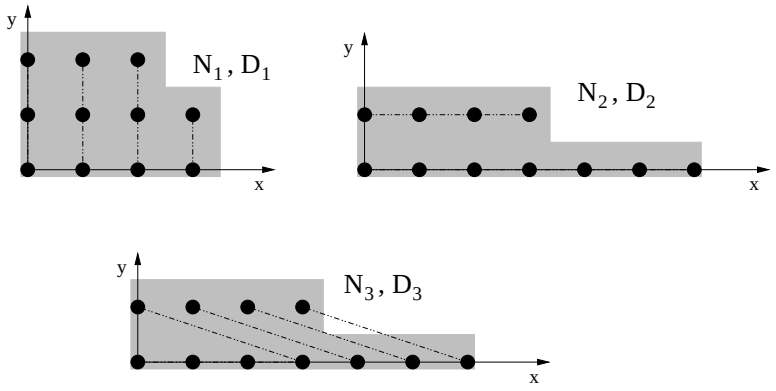

FIGURE 4. The shape of N and D in Example 4

Example 5. As was stated in Section 1, the generalized Padé approximation problem typically assumes that the ideal I is a monomial ideal so that the Taylor series expansion of a solution about a single point has zero coefficients for any term in $\mathcal{B}(I)$. In this example we consider an example of an ideal that is defined by *two* points, both with nontrivial multiplicities.

Specifically, consider the following simple example with two points in ${\mathbb{F}_3}^2$. Let $P_1 = (0,0)$, $\Delta_1 = \{(0,0),(1,0),(0,1),(2,0),(1,1),(0,2),(3,0)\}$ and $P_2 = (1,2)$, $\Delta_2 = \{(0,0),(1,0),(0,1)\}$, and assume a *glex* order on $\mathbb{F}_3[x,y]$ with $y > x$. These two points define an ideal, called a vanishing ideal, by

$$I = \left\{ f = \sum_{\alpha \in \mathbb{N}^2} f_\alpha \mathbf{x}^\alpha \in \mathbb{F}_3[x,y] : \right.$$
$$\left. \textit{the Taylor series expansion of } f \textit{ about } P_i \textit{ has } f_\alpha = 0 \textit{ for } \alpha \in \Delta_i \right\}.$$

For details on computing both a Gröbner basis for such an ideal and a Taylor series expansion for a multivariate polynomial, see [8].

It is easy to verify by hand (using Lemma 5) that $G = \{y^3 - xy^2 + x^2y, x^3y + x^4, x^2y^2 + xy^2 - x^2y, x^5 - x^4 - xy^2 - x^2y\}$ is a Gröbner basis for I in our example. The monomial basis, then, is $\mathcal{B}(I) = \{1, x, y, x^2, xy, y^2, x^3, x^2y, xy^2, x^4\}$, and $|\mathcal{B}(I)| = 10$, implying that $|N| + |D|$ should be 11. Suppose we would like the numerator to have seven terms and the deonominator to have four. That is, $N = \{1, x, y, x^2, xy, y^2, x^3\}$ and $D = \{1, x, y, x^2\}$. Then $\prec_w$ is defined by

$$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Notice that all the work that we have done so far is only precomputing. Once this information has been computed for an ideal, it can be used repeatedly. We now choose a polynomial f and compute a Gröbner basis for M_f. In this example we pick

$$f = 1 + y + x^2 + xy + y^2 + x^2y + x^4.$$

The Gröbner basis for M_f has only one element that lies within N and D, namely

$$(a, b) = (-x^3 - y^2 + xy - y - x - 1, x^2 - x - 1).$$

We can verify that this is indeed a solution by showing that $bf - a \in I$; that is, $bf - a$ has no terms $x^i y^j$, $(i, j) \in \Delta_1$, and $T_b \cdot T_f - T_a$ has no terms $x^i y^j$, $(i, j) \in \Delta_2$, where T_a, T_b, and T_f are the Taylor series expansions of a, b and f, respectively, about $P_2 = (1, 2)$. First, $bf - a = x^6 + x^4 y - x^5 + x^2 y^2 - xy^2$ satisfies the needed conditions for terms in Δ_1. Also, $T_b \cdot T_f - T_a = (x^2 + x - 1) \cdot (1 + y + y^2 + x^3 + x^2 y + x^4) - (-x^3 - y^2 + xy - y + x - 1) = x^6 + x^4 y - x^5 + x^2 y^2 + x^3 y + xy^2 + x^2$ has zero coefficients for the three terms in Δ_2.

5. Final Remarks

We have presented a general framework for multivariate Padé approximation that can be viewed as an analogue of the univariate theory, though the approximant is not unique in general. Also, the Padé approximation can be computed by Gröbner basis technique, which is a generalization of the extended Euclidean algorithm for univariate polynomials. Our Gröbner basis approach is more efficient than the linear algebra approach based on Gauss elimination when the number of variables is small relative to the degree of approximation.

Our method works for an arbitrary field $\mathbb{F}$ and we have assumed exact arithmetic in $\mathbb{F}$ throughout the paper. This is fine for finite fields which are important for coding theory and cryptography applications. For the field of real numbers, there are other important issues, *e.g.*, numerical stability, convergence of approximants when the degree of approximation goes to infinity, *etc.* These issues certainly deserve further investigation.

References

[1] J. Abbott, A. Bigatti, M. Kreuzer and L. Robbiano, Computing ideals of points, *J. Symbolic Comput.* **30** (2000), 341-356.

[2] William W. Adams and Philippe Loustaunau, *An introduction to Gröbner bases*, Graduate Studies in Mathematics, 3, American Mathematical Society, Providence, RI, 1994.

[3] E.R. Berlekamp and L.R. Welch, Error correction for algebraic block codes, U.S. Patent No. 4,633,470, issued December 30, 1986.

[4] B. Buchberger and H. M. Möller, The construction of multivariate polynomials with preassigned zeros. *Computer algebra, EUROCAM '82*, pp. 24-31, Lecture Notes in Comput. Sci., vol. 144, Springer, Berlin-New York, 1982.

[5] David Cox, John Little and Donal O'Shea, *Ideals, varieties, and algorithms*, 2nd ed., Undergraduate Texts in Mathematics, Springer-Verlag, New York, 1997.

[6] David Cox, John Little and Donal O'Shea, *Using algebraic geometry*, Graduate Texts in Mathematics, 185, Springer-Verlag, New York, 1998.

[7] Annie Cuyt, How well can the concept of Padé approximant be generalized to the multivariate case?, Continued fractions and geometric function theory (CONFUN) (Trondheim, 1997), *J. Comput. Appl. Math.* **105** (1999), no. 1-2, 25-50.

[8] Jeffrey B. Farr and Shuhong Gao, Computing Gröbner bases for vanishing ideals of finite sets of points, *preprint*. (Available at `http://www.math.clemson.edu/~sgao/pub.html`)

[9] Jeffrey B. Farr, Shuhong Gao and Daniel L. Noneaker, Construction and decoding performance of random linear codes, *in preparation*.

[10] Patrick Fitzpatrick, On the key equation, *IEEE Transactions on Information Theory*, **41** (1995), no. 5, 1290-1302.
[11] Patrick Fitzpatrick and John Flynn, A Gröbner basis technique for Padé approximation, *J. Symbolic Comput.* **13** (1992), 133-138.
[12] Mariano Gasca and Thomas Sauer, Polynomial interpolation in several variables, in Multivariate polynomial interpolation, *Adv. Comput. Math.* **12** (2000), no. 4, 377-410.
[13] Venkatesan Guruswami and Madhu Sudan, Improved decoding of Reed-Solomon and algebraic-geometry codes, *IEEE Trans. Inform. Theory* **45** (1999), no. 6, 1757-1767.
[14] J. Karlsson and H. Wallin, Rational approximation by an interpolation procedure in several variables, *Padé and rational approximation* (Proc. Internat. Sympos., Univ. South Florida, Tampa, 1976), pp. 83-100, Academic Press, New York, 1977.
[15] Martin Kreuzer and Lorenzo Robbiano, *Computational Commutative Algebra 1*, Springer-Verlag, Berlin, 2000.
[16] John B. Little, David Ortiz, Ricardo Ortiz-Rosado, Rebecca Pablo and Karen Ríos-Soto, Some remarks on Fitzpatrick and Flynn's Gröbner basis technique for Padé approximation, *J. Symbolic Comput.* **35** (2003), 451-461.
[17] C.H. Lutterodt, A two-dimensional analogue of Padé approximant theory, *J. Phys. A* **7** (1974), 1027-1037.
[18] M. G. Marinari, H.M. Möller and T. Mora, Gröbner bases of ideals defined by functionals with an application to ideals of projective points, *Appl. Algebra Engrg. Comm. Comput.* **4** (1993), no. 2, 103-145.
[19] Lorenzo Robbiano, On the theory of graded structures, *J. Symbolic Comput.* **2** (1986), no. 2, 139-170.

Centre for Experimental and Constructive Mathematics (CECM) and Department of Mathematics, Simon Fraser University, Burnaby, BC, Canada V5A 1S6 *E-mail address*: jfarr@cecm.sfu.ca

Department of Mathematical Sciences, Clemson University, Clemson, SC, USA 29634-0975 *E-mail address*: sgao@ces.clemson.edu