ELECTRONIC STRUCTURE OF MATERIALS CENTRE

Least squares orthogonal polynomial approximation in several independent variables

R.S. Caprari

ESM.46 June 1992

Least squares orthogonal polynomial approximation in several independent variables

R.S. Caprari

Electronic Structure of Materials Centre,

School of Physical Sciences,

The Flinders University of South Australia,

GPO Box 2100, Adelaide 5001, Australia

This paper begins with an exposition of a systematic technique for generating orthonormal polynomials in two independent variables by application of the Gram-Schmidt orthogonalisation procedure of linear algebra. It is then demonstrated how a linear least squares approximation for experimental data or an arbitrary function can be generated from these polynomials. The least squares coefficients are computed without recourse to matrix arithmetic, which ensures both numerical stability and simplicity of implementation as a self contained numerical algorithm. The Gram-Schmidt procedure is then utilised to generate a complete set of orthogonal polynomials of fourth degree. A theory for the transformation of the polynomial representation from an arbitrary basis into the familiar sum of products form is presented, together with a specific implementation for fourth degree polynomials. Finally, the computational integrity of this algorithm is verified by reconstructing arbitrary fourth degree polynomials from their values at randomly chosen points in their domain.

I. INTRODUCTION

A recurrent theme in the calibration of experimental apparatus and interpretation of experimental results is the approximation of discrete experimental data by a continuous analytic function. Typically, an experiment measures a 'dependent' variable with some statistical uncertainty or systematic perturbation. It is conjectured that this dependent variable is influenced only by a finite number of 'independent' variables, each of which is known with essentially zero error. Furthermore, only a finite number of measurements are undertaken, thereby sampling the relation between independent and dependent variables at discrete locations in the vector space formed by the independent variables.

To maximise the utility of the experimental results one often wishes to know the relation between the dependent variable and independent variables at arbitrary values of the independent variables. In situations where physical intuition suggests a smoothly varying relation, one also requires the rapid variations that are artifacts of statistical uncertainty in the measurement process to be eliminated. These two criteria identify the present problem as a prime candidate for the application of the general technique of function approximation, in which a continuous function with a finite number of arbitrary parameters has its parameter values chosen to conform most closely to the experimental data, without having sufficient freedom to duplicate the erratic variations in the experimental data.

Function approximation of experimental data is most frequently applied in the context of only one independent variable, which is both conceptually and computationally the simplest case, as well as being the one of greatest algorithmic maturity. However, some experimentally measured parameters are manifestly dependent on two (or more) independent variables, thus necessitating the development of workable function approximation techniques for several independent variables.

To demonstrate the applicability of such techniques to scientific experiments, a description of the actual experimental situation that motivated the development of the algorithm that is the primary subject of this paper will be presented in Sec. II.

II. MOTIVATION

The ESM Centre at Flinders University has recently developed a coincidence electron spectrometer for studying (e,2e) reactions within condensed matter specimens. (e,2e) reactions are kinematically complete electron impact ionisation collisions, both the energies and momentum vectors of each of the incident, scattered and ejected electrons being measured [1, 2]. Specifically, one requires a measurement of the energy and angles (relative to some coordinate system) of both of the outgoing electrons. Electrons of different energy are dispersed by a static electric field that is not parallel to the original electron trajectory, therefore electrons with either different energies or emission angles follow different trajectories in space subsequent to dispersion. The Flinders spectrometer is configured to detect outgoing electrons that have definite values for their polar angles, but a broad range of azimuthal angle values.

Each outgoing electron ends its trajectory by impinging on a 2-dimensional position sensitive detector consisting of a microchannel plate electron multiplier followed by a resistive anode position encoder [3, 4]. The (x, y) position coordinates of any incident electron are obtained directly from the position encoder signals. It is required to separately map the position coordinates to an energy (E) value and an azimuthal angle (ϕ) value, that is, it is required to determine the continuous functions of two variables $E(x, y)$ and $\phi(x, y)$.

A calibration procedure, whereby electrons of known energy and azimuthal angle have their (x, y) coordinates observed provides one with discrete and 'noisy' samples of $E(x, y)$ and $\phi(x, y)$. One now requires a procedure for using these samples to

converge upon an analytic approximation for the actual continuous functions, thus motivating the theory developed in the following sections.

An often desirable property of any algorithm that is developed is that it be self-contained, without recourse to the use of (often unavailable) sophisticated mathematical program libraries (such as matrix algebra). Also, the numerical procedure should be inherently stable (numerical errors not growing exponentially as the computation proceeds), a property not necessarily ensured by an algorithm that uses matrix algebra. Consequently, the algorithm developed in this paper is devoid of matrix algebra. Such a congenial situation results from setting the analytic approximating function equal to a linear combination of mutually orthogonal functions of two variables. A convenient choice of orthogonal functions is polynomials, because the systematic generation of orthogonal polynomials in two variables is an entirely tractable problem, as will now be demonstrated.

III. GENERATION OF ORTHOGONAL POLYNOMIALS IN TWO INDEPENDENT VARIABLES

A. *Vector space concepts*

It is instructive to elucidate the content of the theory developed in this paper by the use of vector space concepts from linear algebra [5]. Let there be N samples to which the function is to be fitted, that is,

> for independent variables (x_i, y_i) there corresponds the dependent variable z_i, where $i = 1, 2, \ldots, N$.

One can consider $(z_1, z_2, \ldots, z_N)$ as a vector in the N-dimensional real vector space $\mathcal{R}^N$. The basis that is implied by this representation, which will be denoted by $\mathcal{S}_N$, is the set of vectors for which the jth basis vector is a function that has the value 0 at all points (x_i, y_i), except for (x_j, y_j) where it has the value 1.

Define the *Euclidean scalar* (or *inner*) *product* of the two members of a family of real valued functions $p_{ij}(x,y)$ and $p_{kl}(x,y)$ by

$$(p_{ij}, p_{kl}) \equiv \langle ij|kl\rangle \equiv \sum_{n=1}^{N} p_{ij}(x_n, y_n) p_{kl}(x_n, y_n) \,. \tag{1}$$

Define the *weighted scalar product* of the same two functions $p_{ij}(x,y)$ and $p_{kl}(x,y)$ with respect to the real weight function $h(x,y)$ by

$$(p_{ij}, hp_{kl}) \equiv \langle ij|h|kl\rangle \equiv \sum_{n=1}^{N} p_{ij}(x_n, y_n) h(x_n, y_n) p_{kl}(x_n, y_n) \,. \tag{2}$$

Note that

$$\langle ij|kl\rangle = \langle kl|ij\rangle \quad \text{and} \quad \langle ij|h|kl\rangle = \langle kl|h|ij\rangle \,. \tag{3}$$

An *orthonormal* set of functions p_{ij}, $i,j = 0,1,2,\ldots$ has the property

$$\langle ij|kl\rangle = \delta_{ik}\delta_{jl} \quad \forall\, i,j,k,l \;, \tag{4}$$

where δ_{mn} is the *Kronecker delta*.

Note that $\mathcal{S}_N$ is a perfectly good orthonormal basis for $\mathcal{R}^N$, but it is unsuitable for constructing continuous approximating functions, because the natural extension of the basis functions to the (x,y) continuum is *Dirac delta* functions, which are certainly not 'well behaved'. An orthonormal basis for $\mathcal{R}^N$ (or a subspace of $\mathcal{R}^N$), which extends to the (x,y) continuum in a well behaved manner must be sought. The basis vectors will be orthonormal polynomials in x and y (more precisely, samples of polynomials at the points (x_i, y_i) , $i = 1,2,\ldots,N$, at which the continuous function is to be fitted, where the orthonormality conditions of Eq. (4) hold).

A *polynomial* of *degree* K in two independent variables x and y will be defined as follows,

$$p(x,y) \equiv \sum_{j=0}^{K} \sum_{i=0}^{j} d_{(j-i)i} x^{(j-i)} y^{i} \;,\; d_{ij} = \text{constant}. \tag{5}$$

The *Gram-Schmidt* orthogonalisation procedure for vector spaces [5] will be applied to generate a set of orthonormal basis polynomials in (x, y) for a subspace of $\mathcal{R}^N$, following the general technique elucidated by Courant and Hilbert [6, p. 50–51] and elaborated upon by Cadwell and Williams [7] and Hayes [8].

B. *Dimension of the subspace spanned by the orthonormal basis*

Suppose that one decides to seek a set of L ($\leq N$) orthonormal polynomials to form a basis, denoted by $\mathcal{O}_L$, which spans a subspace of $\mathcal{R}^N$ of anticipated dimension L, to be denoted $\mathcal{P}_L$. Assume that there exists a 'conventional' basis for $\mathcal{P}_L$, denoted by $\mathcal{C}_L$, that consists of L product terms $(x^i y^j)$ taken individually. These two sets of L vectors span the same subspace, and if it can be shown that one set of vectors is linearly dependent, it follows that the subspace has dimension less than L, which implies that the other set of vectors is linearly dependent.

One can determine whether $\mathcal{C}_L$ is a linearly independent set of vectors by (at least) two methods. The first method consists of forming the $L \times N$ matrix whose rows are simply the individual members of $\mathcal{C}_L$ evaluated at the N sample points (x_i, y_i), $i = 1, 2, \ldots, N$. The *rank* [5, sec. 4.6] of this matrix is the dimension of the subspace spanned by $\mathcal{C}_L$ (this subspace being $\mathcal{P}_L$); only if the rank of this matrix is L is $\mathcal{C}_L$ a linearly independent set of vectors. Alternatively, one can compute the *Gram determinant* [6. p. 34–36] of the members of $\mathcal{C}_L$; if $\mathcal{C}_L$ is a linearly independent set of vectors then the Gram determinant is definitely positive, a zero value indicating linear dependence. The Gram determinant does not directly yield the dimension of $\mathcal{P}_L$.

If $\mathcal{C}_L$ is linearly independent then all L members of $\mathcal{O}_L$ can be constructed by the Gram-Schmidt procedure to be non-zero vectors, reflecting the fact that $\mathcal{P}_L$ has dimension L. If $\mathcal{C}_L$ is linearly dependent then the Gram-Schmidt procedure will

yield zero vectors for one or more members of $\mathcal{O}_L$, reflecting the fact that $\mathcal{P}_L$ has dimension less than L. Although these members of $\mathcal{O}_L$ are not identically zero when one progresses to the continuum in (x,y), it will be revealed in Sec. IV that the coefficient of a basis function in the total approximating function is a scalar product of the corresponding vector in $\mathcal{R}^N$ (zero vector in this case) with another vector (see Eq. (20)), which certainly is zero. Therefore, the members of $\mathcal{O}_L$ that are zero vectors in $\mathcal{R}^N$ will be absent from the approximating function in the continuum. Since $\mathcal{P}_L$ has dimension less than L, it is expected that there will be a larger approximation error than if the sample points had been chosen differently (different positions or more of them) to give $\mathcal{P}_L$ dimension L. This situation is most unlikely to occur in practice and in any case is not catastrophic, therefore the test for whether $\mathcal{P}_L$ has dimension L (and if not, proceeding to choose more or different sample points to obtain a subspace $\mathcal{P}_L$ with dimension L) can be omitted with considerable practical justification.

C. *Gram-Schmidt orthogonalisation procedure*

The essence of the Gram-Schmidt orthogonalisation procedure is as follows. One constructs orthonormal polynomials one at a time. Each time a new orthonormal polynomial is required, it is constructed so that it is orthogonal to all previously constructed polynomials (and is normalised). In this way, at all stages, each polynomial is orthogonal to every other, as well as being normalised.

Members of the orthonormal set of polynomials are denoted $p_{ij}(x,y)$ where the subscripts indicate that $p_{ij}(x,y)$ is the first polynomial to be constructed that includes the product term $x^i y^j$. The order in which the $p_{ij}(x,y)$ are constructed is indicated by the sequence

$$\{\ \{p_{(n-j)j}\}_{j=0,1,2,\ldots,n-1,n}\ \}_{n=0,1,2,\ldots,K-1,K}\ . \tag{6}$$

Any polynomial of degree K, as defined by Eq. (5), can be expressed as a linear

combination of only these $p_{ij}(x,y)$.

Instead of choosing the non-orthonormal basis polynomials that are the 'input' to the Gram-Schmidt procedure to be the members of C_L, one extends the method developed by Forsythe [9] for polynomials of one variable to polynomials of two variables. In this method the non-orthonormal basis, which still spans the same subspace $\mathcal{P}_L$, is constructed as the Gram-Schmidt procedure progresses by multiplying a judiciously chosen, previously determined member of $\mathcal{O}_L$ by either x or y. Specifically, the Gram-Schmidt procedure constructs a new orthonormal polynomial from previous orthonormal polynomials in one of two ways, using either of the two *Gram-Schmidt equations*,

$$\lambda_{ij}p_{ij}(x,y) = xp_{(i-1)j}(x,y) - \sum_{k=0}^{i+j-1}\sum_{l=0}^{k} {}^{ij}e_{(k-l)l}p_{(k-l)l}(x,y) - \sum_{l=0}^{j-1} {}^{ij}e_{(i+j-l)l}p_{(i+j-l)l}(x,y) \quad , \text{for } i \neq 0 , \tag{7}$$

$$\lambda_{0j}p_{0j}(x,y) = yp_{0(j-1)}(x,y) - \sum_{k=0}^{j-1}\sum_{l=0}^{k} {}^{0j}e_{(k-l)l}p_{(k-l)l}(x,y) - \sum_{l=0}^{j-1} {}^{0j}e_{(j-l)l}p_{(j-l)l}(x,y) \quad , \text{for } i = 0 , \tag{8}$$

where λ_{ij} and ${}^{ij}e_{mn}$ are constant coefficients to be determined from the orthonormality conditions (4) by taking the scalar product of Eq. (7) or (8) with $n_{mn}(x,y)$, one of the orthonormal polynomials that have been generated prior to the current one, immediately obtaining the relations

$${}^{ij}e_{mn} = \langle mn|x|(i-1)j\rangle \quad , \text{for } i \neq 0 , \tag{9}$$

$${}^{0j}e_{mn} = \langle mn|y|0(j-1)\rangle \quad , \text{for } i = 0 , \tag{10}$$

$$\lambda_{ij} = (\lambda_{ij}p_{ij}, \lambda_{ij}p_{ij})^{1/2} \quad , \text{for all } i . \tag{11}$$

With the calculation of all the *orthogonalisation coefficients* (${}^{ij}e_{mn}$) and *normal-*

isation coefficients (λ_{ij}) the generation of orthonormal polynomials is complete. In Sec. V a basis of orthonormal polynomials will be explicitly constructed, but prior to that, a derivation of the optimal approximating function that can be expressed in terms of this basis will be undertaken in Sec. IV.

IV. LEAST SQUARES APPROXIMATION

The optimal analytic approximating function that will be sought will be a linear combination of the vectors of $\mathcal{O}_L$ generalised to the continuum in (x, y), in which case they emerge as the polynomials $p_{ij}(x, y)$ of Sec. III C. Without loss of generality, choose an appropriate set of p_{ij} to form a *complete set* of functions for polynomials in two variables of degree K as defined by Eq. (5); this is precisely the set of p_{ij} that is identified in Eq. (6). Consequently, the analytic approximating function $\hat{p}(x, y)$ is

$$\hat{p}(x, y) = \sum_{n=0}^{K} \sum_{j=0}^{n} c_{(n-j)j} p_{(n-j)j}(x, y) \,, \tag{12}$$

where the c_{ij} are variation parameters to be chosen to maximise the fidelity of the analytic approximation.

The *linear least squares* method [6, p. 52] of optimising the approximating function will be adopted in this paper. Define the *mean square error* in the approximation of the sample values z_i by the analytic function values $\hat{p}(x_i, y_i)$ for $i = 1, 2, \ldots, N$ by

$$\epsilon^2 \equiv \sum_{i=1}^{N} (z_i - \hat{p}(x_i, y_i))^2 \,. \tag{13}$$

According to Eq. (1) (generalised beyond functions of only a specific family) the mean square error can be expressed as the Euclidean scalar product of a vector in $\mathcal{R}^N$ with itself (ie. the square of the *norm* of the vector),

$$\epsilon^2 = ((z - \hat{p}), (z - \hat{p})) \,. \tag{14}$$

Using a property of the scalar product, Eq. (14) can be expanded to give

$$\epsilon^2 = (z,z) + (\hat{p},\hat{p}) - 2(z,\hat{p}) \,. \tag{15}$$

By making reference to the scalar product definition (Eq. (1)), the orthonormality of the p_{ij} (Eq. (4)) and the definition of $\hat{p}$ (Eq. (12)), one can easily derive the following identities,

$$(z,z) = \sum_{i=1}^{N} z_i^2 \,, \tag{16}$$

$$(\hat{p},\hat{p}) = \sum_{n=0}^{K}\sum_{j=0}^{n} c_{(n-j)j}^2 \,, \tag{17}$$

$$(z,\hat{p}) = \sum_{n=0}^{K}\sum_{j=0}^{n} c_{(n-j)j}(z, p_{(n-j)j}) \,, \tag{18}$$

and substituting these identities into Eq. (15) gives (on adding and subtracting extra terms)

$$\epsilon^2 = \sum_{i=1}^{N} z_i^2 + \sum_{n=0}^{K}\sum_{j=0}^{n}(c_{(n-j)j} - (z, p_{(n-j)j}))^2 - \sum_{n=0}^{K}\sum_{j=0}^{n}(z, p_{(n-j)j})^2 \,. \tag{19}$$

In Eq. (19) the first and third terms on the right side are constant with respect to variations of the approximating function. Only the second term on the right depends on the approximating function through the appearance of the variation parameters of $\hat{p}$, that is c_{ij}. Since the second term makes a non-negative contribution to the mean square error, the mean square error is minimised when this term is zero, a condition that is satisfied only if the variation parameters have the specific values

$$c_{ij} = (z, p_{ij}) \quad \forall\, i,j \,. \tag{20}$$

Equation (12) with the specific variation parameter values given by Eq. (20) is the least squares approximating polynomial of degree K, in which case the c_{ij} are denoted as *least squares coefficients*. Substituting Eq. (20) into Eq. (19) yields the *minimum mean square error* corresponding to the least squares approximation,

$$\widehat{\epsilon^2} = \sum_{i=1}^{N} z_i^2 - \sum_{n=0}^{K}\sum_{j=0}^{n} c_{(n-j)j}^2 \,. \tag{21}$$

The *root mean square error* of the least squares approximant is best defined as

$$\widehat{\epsilon_{\text{RMS}}} = \sqrt{\widehat{\epsilon^2}/N} \,. \tag{22}$$

Anton [5, p. 268-270] reveals an intuitive geometrical interpretation of the least squares solution, which in the current context can be stated as follows :

> The least squares approximation to the vector $z = (z_1, z_2, \ldots, z_{(N-1)}, z_N)$ in $\mathcal{R}^N$ by the subspace $\mathcal{P}_L$ is the *orthogonal projection of z on* $\mathcal{P}_L$, given by the vector $((z, p_{00}), (z, p_{10}), \ldots, (z, p_{1(K-1)}), (z, p_{0K}))$ with respect to the basis $\mathcal{O}_L$, and the residual mean square error is the square of the norm of the *component of z orthogonal to* $\mathcal{P}_L$.

A theory for constructing a basis of functions that can be linearly combined to give an approximation to samples of an arbitrary function has been developed in Sec. III, and a theory for choosing the coefficients of the linear combination to maximise the fidelity of the approximation has been developed in the present section. Within the next section these results will be applied to explicitly generate a finite basis of $p_{ij}(x, y)$, in the process fully exploiting all redundancies to provide a general numerical algorithm that involves the minimum amount of computation.

V. BASIS OF ORTHONORMAL POLYNOMIALS OF DEGREE 0, 1, 2, 3 AND 4

The theoretical foundations developed in the two previous sections will now be applied to the case of the approximation function being a polynomial in two variables of degree 4 (ie. $K = 4$). In this case Eq. (6) identifies 15 orthonormal polynomials p_{ij} that are to be constructed if possible (see Sec. III B for reasons why possibly only less than 15 such polynomials are significant). Subspace $\mathcal{P}_L$ will therefore have dimension 15 (possibly less).

The following orthonormal polynomials are generated by the Gram-Schmidt orthogonalisation procedure of Sec. III C; in particular, a new orthonormal polynomial is expressed in terms of all previously constructed orthonormal polynomials by either of the Gram-Schmidt equations (Eqs. (7) and (8)), and the arbitrary parameters in this equation are evaluated (Eqs. (9) and (10) for orthogonalisation coefficients (${}^{ij}e_{mn}$) and Eq. (11) for normalisation coefficients (λ_{ij})). Least squares coefficients (c_{ij}) are as derived in Sec. IV, and expressed in Eq. (20), and can be evaluated progressively as each orthonormal polynomial is constructed, to completely determine the least squares approximation polynomial of degree 4.

The Gram-Schmidt equations defining the orthonormal polynomials are stated in this section in a form that fully exploits redundancies in the orthogonalisation and normalisation coefficients, by not explicitly including coefficients that are identically zero and by calculating degenerate coefficients only once. Degenerate coefficients are subject to the following policy. On the first occasion that a memeber of a set of degenerate coefficients arises in the Gram-Schmidt procedure (this coefficient will be referred to as the 'initial' member of the set), the coefficient is stated in the equation in which it arose without modification. On subsequent occasions that members of the same set of degenerate coefficients arise ('subsequent' members), the coefficients are replaced by the initial memeber of the same set in the statement of the Gram-Schmidt equations in which the subsequent members occur. In summary, this section states the Gram-Schmidt equations in a form which omits coefficients that are identically zero, and all the non-zero coefficients that are present are either non-degenerate (in which case they appear only once), or initial members of a degenerate set (in which case they appear a number of times equal to the multiplicity of the degenerate set).

Perusing the following equations indicates that degeneracies between orthogonalisation and normalisation coefficients can occur. Polynomials and coefficients must

be evaluated in the indicated order, since succeeding polynomials make reference to preceding polynomials and coefficients.

The Gram-Schmidt equations for the generation of orthonormal polynomials of degree 0, 1, 2, 3 and 4 are as follows,

$$\lambda_{00}p_{00} = 1 \ , \tag{23}$$

$$\lambda_{10}p_{10} = xp_{00} - {}^{10}e_{00}p_{00} \ , \tag{24}$$

$$\lambda_{01}p_{01} = yp_{00} - {}^{01}e_{00}p_{00} - {}^{01}e_{10}p_{10} \ , \tag{25}$$

$$\lambda_{20}p_{20} = xp_{10} - \lambda_{10}p_{00} - {}^{20}e_{10}p_{10} - {}^{20}e_{01}p_{01} \ , \tag{26}$$

$$\lambda_{11}p_{11} = xp_{01} - {}^{20}e_{01}p_{10} - {}^{11}e_{01}p_{01} - {}^{11}e_{20}p_{20} \ , \tag{27}$$

$$\lambda_{02}p_{02} = yp_{01} - \lambda_{01}p_{00} - {}^{02}e_{10}p_{10} - {}^{02}e_{01}p_{01} - {}^{02}e_{20}p_{20} - {}^{02}e_{11}p_{11} \ , \tag{28}$$

$$\lambda_{30}p_{30} = xp_{20} - \lambda_{20}p_{10} - {}^{11}e_{20}p_{01} - {}^{30}e_{20}p_{20} - {}^{30}e_{11}p_{11} - {}^{30}e_{02}p_{02} \ , \tag{29}$$

$$\lambda_{21}p_{21} = xp_{11} - \lambda_{11}p_{01} - {}^{30}e_{11}p_{20} - {}^{21}e_{11}p_{11} - {}^{21}e_{02}p_{02} - {}^{21}e_{30}p_{30} \ , \tag{30}$$

$$\lambda_{12}p_{12} = xp_{02} - {}^{30}e_{02}p_{20} - {}^{21}e_{02}p_{11} - {}^{12}e_{02}p_{02} - {}^{12}e_{30}p_{30} - {}^{12}e_{21}p_{21} \ , \tag{31}$$

$$\lambda_{03}p_{03} = yp_{02} - {}^{03}e_{10}p_{10} - \lambda_{02}p_{01} - {}^{03}e_{20}p_{20} - {}^{03}e_{11}p_{11} - {}^{03}e_{02}p_{02} - {}^{03}e_{30}p_{30} - {}^{03}e_{21}p_{21} - {}^{03}e_{12}p_{12} \ , \tag{32}$$

$$\lambda_{40}p_{40} = xp_{30} - \lambda_{30}p_{20} - {}^{21}e_{30}p_{11} - {}^{12}e_{30}p_{02} - {}^{40}e_{30}p_{30} - {}^{40}e_{21}p_{21} - {}^{40}e_{12}p_{12} - {}^{40}e_{03}p_{03} \ , \tag{33}$$

$$\lambda_{31}p_{31} = xp_{21} - \lambda_{21}p_{11} - {}^{12}e_{21}p_{02} - {}^{40}e_{21}p_{30} - {}^{31}e_{21}p_{21} - {}^{31}e_{12}p_{12} - {}^{31}e_{03}p_{03} - {}^{31}e_{40}p_{40} \ , \tag{34}$$

$$\lambda_{22}p_{22} = xp_{12} - \lambda_{12}p_{02} - {}^{40}e_{12}p_{30} - {}^{31}e_{12}p_{21} - {}^{22}e_{12}p_{12} - {}^{22}e_{03}p_{03} - {}^{22}e_{40}p_{40} - {}^{22}e_{31}p_{31} \ , \tag{35}$$

$$\lambda_{13}p_{13} = xp_{03} - {}^{40}e_{03}p_{30} - {}^{31}e_{03}p_{21} - {}^{22}e_{03}p_{12} - {}^{13}e_{03}p_{03} - {}^{13}e_{40}p_{40} - {}^{13}e_{31}p_{31} - {}^{13}e_{22}p_{22} \ , \tag{36}$$

$$\lambda_{04}p_{04} = yp_{03} - {}^{04}e_{10}p_{10} - {}^{04}e_{20}p_{20} - {}^{04}e_{11}p_{11} - \lambda_{03}p_{02} - {}^{04}e_{30}p_{30} -$$
$${}^{04}e_{21}p_{21} - {}^{04}e_{12}p_{12} - {}^{04}e_{03}p_{03} - {}^{04}e_{40}p_{40} - {}^{04}e_{31}p_{31} - {}^{04}e_{22}p_{22} -$$
$${}^{04}e_{13}p_{13} \ . \qquad (37)$$

The results of Secs. III C and IV are amenable to automatic and progressive computation of orthogonalisation, normalisation and least squares coefficients, for an arbitrarily large basis of orthonormal polynomials. Although this approach is simple and flexible from a programming viewpoint, direct application of this technique does not exploit the considerable redundancy in the orthogonalisation coefficients (vanishing and degenerate coefficients) that becomes apparent only under explicit analysis such as that elucidated in this section. Therefore, this 'mechanical' approach is significantly more computationally intensive than the analysis and reduction of the general equations arising from the Gram-Schmidt procedure, as has been implemented in this section for a complete set of orthonormal polynomials of degree 4. Nevertheless, the mechanical approach has been used in similar contexts by other researchers [10, and references therein].

In this section, the least squares approximating polynomial of degree 4 has been derived as a linear combination of orthonormal polynomials. The least squares polynomial is expressed in a more familiar form that is evaluated with much less computational effort in the next section.

VI. TRANSFORMATION FROM THE ORTHONORMAL BASIS $\mathcal{O}_L$ TO THE CONVENTIONAL BASIS $\mathcal{C}_L$

A. *Preliminary considerations*

The least squares approximating polynomial $\hat{p}(x,y)$ (Eq. (12)) has been derived as a superposition of orthonormal polynomials $p_{ij}(x,y)$ with coefficients given by

Eq. (20). In principle the $p_{ij}(x,y)$ can be expanded as a sum of *monomials* (terms of the form $d_{ij}x^iy^j$, where d_{ij} is a constant), thereby allowing $\hat{p}(x,y)$ to be expressed as a sum of monomials (that is, *sum of products*), as in Eq. (5).

This transformation is equivalent to a change of basis for the subspace $\mathcal{P}_L$ from the orthonormal basis $\mathcal{O}_L$ (basis vectors being p_{ij}) to the conventional basis $\mathcal{C}_L$ (basis vectors being x^iy^j) (see Sec. III B). Although the conventional basis is neither normalised nor orthogonal, there are distinct benefits in the transition to the (x,y) continuum associated with the use of $\mathcal{C}_L$ rather than $\mathcal{O}_L$. These benefits include :

- much lower computational effort required to evaluate a typical x^iy^j term than a typical p_{ij} ;
- algebraic simplicity, compactness and familiarity in the expression of the least squares polynomial as a sum of products, as opposed to a sum of orthonormal polynomials which themselves have to be defined ;
- algebraic invariance in the (x,y) continuum of x^iy^j, but not p_{ij}, with respect to different vector spaces $\mathcal{R}^N$ established by different sets of (x_i,y_i) samples, since different sample values lead to different orthonormal polynomials, whereas there is no freedom to alter the x-y dependence of a monomial.

The change of basis problem in subspace $\mathcal{P}_L$, from the old basis $\mathcal{O}_L$ to the new basis $\mathcal{C}_L$, has a formal solution [5, sec. 4.10]. If $\mathcal{P}_L$ has dimension L, then the *transition matrix* from $\mathcal{O}_L$ to $\mathcal{C}_L$ has as its elements the coefficients of the p_{ij} vectors in $\mathcal{O}_L$ expressed as linear combinations of the x^iy^j vectors in $\mathcal{C}_L$. Elements of the transition matrix are too multitudinous (L^2 in number), and individually too cumbersome (being complicated products, quotients, sums and differences of the λ_{ij} and ${}^{ij}e_{mn}$ coefficients), to even contemplate symbolically expressing the basis transformation equations, however the transition matrix elements are amenable to algorithmic

evaluation by a computer program, if desired. If $\mathcal{P}_L$ has dimension less than L, then both $\mathcal{O}_L$ and $\mathcal{C}_L$ contain less than L vectors, and it is necessary to determine a linearly independent set of monomials that span $\mathcal{P}_L$ (some monomials will be absent from this set) to be the basis $\mathcal{C}_L$. The comments made above for $\mathcal{P}_L$ having dimension L will then be directly applicable to $\mathcal{P}_L$ having dimension less than L .

B. *General theory of the transformation*

The transformation proposed in the previous section encompassed a change of basis for subspace $\mathcal{P}_L$ of $\mathcal{R}^N$ from $\mathcal{O}_L$ to $\mathcal{C}_L$, followed by the extension of $\mathcal{C}_L$ to the (x,y) continuum. These two operations may be transposed; first the basis $\mathcal{O}_L$ of subspace $\mathcal{P}_L$ is extended to the (x,y) continuum, then the specific $\hat{p}(x,y)$ is expressed as a linear combination of $x^i y^j$ terms, as required. This transformation is not as general as the determination of the transition matrix for the change of basis, but is endowed with the advantages of complete expressibility of the transformation equations concisely in symbolic form, and absence of the need for special consideration of subspaces $\mathcal{P}_L$ with dimension less than L (as explained in the previous section). The theoretical foundation of the latter transformation of the representation of $\hat{p}(x,y)$ will be developed in this section.

The least squares polynomial $\hat{p}(x,y)$ of degree K will be defined to have the following specific sum of products form,

$$\hat{p}(x,y) = \sum_{j=0}^{K} \sum_{i=0}^{j} \hat{d}_{(j-i)i} x^{(j-i)} y^i \quad , \tag{38}$$

where the monomial coefficients $\hat{d}_{(j-i)i}$ are to be determined.

Polynomials of degree K in the (x,y) continuum span a subspace of infinite dimensional *Hilbert space* [6, p. 55] that has dimension L given by

$$L = \sum_{j=0}^{K} (j+1) = \tfrac{1}{2}(K+1)(K+2) \quad . \tag{39}$$

The first equality in Eq. (39) follows directly from enumerating the linearly independent terms in either Eq. (5) or Eq. (6), while the second equality can be proven by mathematical induction. This L dimensional subspace of Hilbert space will be denoted by $\mathcal{K}_L$.

Introduce another L dimensional subspace of Hilbert space as the subspace spanned by the L Dirac delta functions,

$$\delta(x - X_l, y - Y_l) \, , \; l = 1, \ldots, L \, , \tag{40}$$

for distinct, arbitrary points (X_l, Y_l) in the (x, y) continuum. This subspace is denoted by $\mathcal{D}_L$. Note that the L Dirac delta functions of Eq. (40) constitute an orthonormal basis for subspace $\mathcal{D}_L$.

Subspace $\mathcal{K}_L$ is invariant, being precisely that subspace that is spanned by all polynomials of degree K, but flexibility in the choice of subspace $\mathcal{D}_L$ may be exercised by choosing the points (X_l, Y_l). Assume that $\mathcal{D}_L$ is chosen such that the orthogonal projection of subspace $\mathcal{K}_L$ onto subspace $\mathcal{D}_L$ spans $\mathcal{D}_L$. Equivalently, the orthogonal projection of any basis of $\mathcal{K}_L$ onto $\mathcal{D}_L$ produces a basis of $\mathcal{D}_L$. Denote the orthogonal projection of an arbitrary vector v in Hilbert space (that is an arbitrary function $v(x, y)$) onto subspace $\mathcal{D}_L$ by "$\mathrm{Proj}_{\mathcal{D}_L}(v)$".

$\hat{p}(x, y)$ is a specific vector in subspace $\mathcal{K}_L$, by virtue of the fact that it is a polynomial of degree K. That the orthogonal projection of $\hat{p}$ onto subspace $\mathcal{D}_L$ has the coordinates

$$(\, \mathrm{Proj}_{\mathcal{D}_L}(\hat{p}) \,)_l \; = \; \hat{p}(X_l, Y_l) \, , \; l = 1, \ldots, L \, , \tag{41}$$

relative to the Dirac delta function basis of $\mathcal{D}_L$ (Eq. (40)) follows directly from the following property of Dirac delta functions [11, p. 58–61],

$$\hat{p}(X_l, Y_l) \; = \; \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \hat{p}(x, y) \delta(x - X_l, y - Y_l) \, dx \, dy \, . \tag{42}$$

Similarly, monomial functions $x^i y^j$ considered as vectors in Hilbert space, have the following coordinates relative to the Dirac delta function basis of $\mathcal{D}_L$ for their orthogonal projections onto subspace $\mathcal{D}_L$,

$$(\,\mathrm{Proj}_{\mathcal{D}_L}(x^i y^j))_l \;=\; X_l^i Y_l^j \;,\; l = 1, \ldots, L \,. \tag{43}$$

Note that the monomials appearing in Eq. (38), that is $x^i y^j$, $i + j \leq K$, constitute a basis for $\mathcal{K}_L$, and as a consequence of the assumption of the previous paragraph, the orthogonal projections onto subspace $\mathcal{D}_L$ of Eq. (43) for $i + j \leq K$, constitute another basis for $\mathcal{D}_L$ (as distinct from the Dirac delta function basis of Eq. (40)).

Taking the orthogonal projection of Eq. (38) onto subspace $\mathcal{D}_L$, and substituting Eqs. (41) and (43) for the orthogonal projections, gives the following equations for the coordinates relative to the Dirac delta function basis of $\mathcal{D}_L$,

$$\hat{p}(X_l, Y_l) = \sum_{j=0}^{K} \sum_{i=0}^{j} \hat{d}_{(j-i)i} X_l^{(j-i)} Y_l^i \;,\; l = 1, \ldots, L \,. \tag{44}$$

These equations, expressed in matrix form, can be interpreted as a change of basis equation for subspace $\mathcal{D}_L$. The 'old' basis is the set of vectors $\mathrm{Proj}_{\mathcal{D}_L}(x^i y^j)$, $i+j \leq K$, relative to which the coordinates of the vector $\mathrm{Proj}_{\mathcal{D}_L}(\hat{p}(x,y))$ are $\hat{d}_{ij}$, $i + j \leq K$; these coordinates are collected into the $L \times 1$ *coordinate matrix* $\hat{\mathbf{d}}$. The 'new' basis is the Dirac delta function basis of Eq. (40), relative to which the coordinates of the vector $\mathrm{Proj}_{\mathcal{D}_L}(\hat{p}(x,y))$ are $\hat{p}(X_l, Y_l)$, $l = 1, \ldots, L$ (Eq. (41)); these coordinates are collected into the $L \times 1$ coordinate matrix $\hat{\mathbf{p}}$. The transition matrix transforming the coordinate matrix from the old basis to the new basis is the $L \times L$ matrix, to be denoted $\mathbf{M}$, whose columns are the coordinate matrices of the old basis vectors relative to the new basis, that is, the coordinate matrices of $\mathrm{Proj}_{\mathcal{D}_L}(x^i y^j)$, $i + j \leq K$ relative to the Dirac delta function basis, as in Eq. (43).

Introducing these matrices, whose elements are defined as follows (where the pair of indices (ij) are to be interpreted as a single 'vector' index for the row or column),

$$\hat{\mathbf{d}}_{(ij)1} \equiv \hat{d}_{ij} \ , \ i+j \leq K \ , \tag{45}$$

$$\mathbf{M}_{l(ij)} \equiv X_l^i Y_l^j \ , \ l = 1, \ldots, L \ , \ i+j \leq K \ , \tag{46}$$

$$\hat{\mathbf{p}}_{l1} \equiv \hat{p}(X_l, Y_l) \ , \ l = 1, \ldots, L \ , \tag{47}$$

Eq. (44) is conveniently expressed as the matrix equation

$$\hat{\mathbf{p}} = \mathbf{M}\hat{\mathbf{d}} \ . \tag{48}$$

$\mathbf{M}$, being a transition matrix, must be invertible [5, p. 195]. If in practice one makes an unfortunate choice of subspace $\mathcal{D}_L$, so that, contrary to the previous assumption, the orthogonal projection of subspace $\mathcal{K}_L$ on subspace $\mathcal{D}_L$ does not span $\mathcal{D}_L$, then a new subspace $\mathcal{D}_L$ must be sought by seeking a new set of points (X_l, Y_l), $l = 1, \ldots, L$ in the (x, y) continuum, that do indeed convert to an invertible $\mathbf{M}$. A systematic strategy for choosing the subspace $\mathcal{D}_L$ is as follows. $\mathbf{M}$ is progressively constructed one row at a time by choosing one point (X_l, Y_l) at a time, each time forming a new row of $\mathbf{M}$. This results in an $l \times L$ intermediate matrix, with $1 \leq l \leq L$ at each stage of the procedure. If the rank of the intermediate matrix is l, then the chosen (X_l, Y_l) is a suitable point for generating subspace $\mathcal{D}_L$, and one can proceed to choose the next $((l+1)\text{th})$ point. If the rank of the intermediate matrix is $(l-1)$, then any resulting matrix $\mathbf{M}$ wil not be invertible, so the chosen (X_l, Y_l) is unsuitable for generating subspace $\mathcal{D}_L$ (when considered in conjunction with all previously chosen points), therefore a new choice must be made for (X_l, Y_l) and the procedure repeated.

$\mathbf{M}^{-1}$, the inverse of $\mathbf{M}$, is actually the transition matrix from the Dirac delta function basis of $\mathcal{D}_L$ (Eq. (40)) to the basis of $\mathcal{D}_L$ given by $\text{Proj}_{\mathcal{D}_L}(x^i y^j)$, $i+j \leq K$ (Eq. (43)), and Eq. (48) can be rewritten as

$$\hat{\mathbf{d}} = \mathbf{M}^{-1}\hat{\mathbf{p}} \ . \tag{49}$$

This matrix equation yields the required monomial coefficients $\hat{d}_{ij}$ as the elements of $\hat{\mathbf{d}}$. Note that the coordinate matrix $\hat{\mathbf{p}}$ and the transition matrix $\mathbf{M}$ (or $\mathbf{M}^{-1}$) both depend on the particular subspace $\mathcal{D}_L$ in such a way that the other coordinate matrix $\hat{\mathbf{d}}$ is invariant with respect to changes in the subspace $\mathcal{D}_L$. Note also that the basis that forms the original representation of $\hat{p}(x,y)$, which in this case is the set of $p_{ij}(x,y)$, is irrelevant; this technique is entirely suitable for transforming from an arbitrary basis to the basis defined by the set of product terms $x^i y^j$.

The theory for the basis transformation that was noted in Sec. VI A is conceptually simpler and more direct than the theory that is derived in this section, however it is explained that symbolic expression of the coefficients of the transformation equations is precluded by their severe complexity. These coefficients, whose values are dependant on the particular basis of orthogonal polynomials, would have to be computed by a numerical algorithm on each occasion that orthogonal polynomials are generated. These factors support the contention that the theory that is elucidated in this section, although being more convoluted when stated in abstract form, emerges with greater clarity, conciseness and efficiency of implementation when the abstraction is elaborated upon to yield explicit transformation equations. This is the justification for dismissing the 'obvious' theory of Sec. VI A in favour of the formally more complicated theory developed in this section.

C. Specific transformation equations

The result of a specific application of the theory of Sec. VI B for the case of degree 4 polynomials ($K = 4$, $L = 15$) is stated below. The set of points (X_l, Y_l) , $l = 1, \ldots, L$, that define the subspace $\mathcal{D}_L$ are chosen to be ordered pairs of integers, as is apparent from the elements of $\hat{\mathbf{p}}$. Consequently, $\mathbf{M}$ has integral elements and $\mathbf{M}^{-1}$ has elements that are rational numbers. Inversion of $\mathbf{M}$ using

exact algebra was accomplished by using the symbolic mathematics computer program MATHEMATICA [12]. Expanding the matrices of Eq. (49) into their individual elements, gives one set of possible transformation equations as

$$\begin{bmatrix} \hat{d}_{00} \\ \hat{d}_{10} \\ \hat{d}_{01} \\ \hat{d}_{20} \\ \hat{d}_{11} \\ \hat{d}_{02} \\ \hat{d}_{30} \\ \hat{d}_{21} \\ \hat{d}_{12} \\ \hat{d}_{03} \\ \hat{d}_{40} \\ \hat{d}_{31} \\ \hat{d}_{22} \\ \hat{d}_{13} \\ \hat{d}_{04} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{2}{3} & 0 & -\frac{2}{3} & 0 & -\frac{1}{12} & 0 & \frac{1}{12} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{2}{3} & 0 & -\frac{2}{3} & 0 & -\frac{1}{12} & 0 & \frac{1}{12} & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{5}{4} & \frac{2}{3} & 0 & \frac{2}{3} & 0 & -\frac{1}{24} & 0 & -\frac{1}{24} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & -1 & -\frac{1}{6} & -\frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & 0 & \frac{5}{4} & -\frac{1}{12} & \frac{1}{4} & -\frac{1}{12} & -\frac{1}{6} & -\frac{1}{6} \\ -\frac{5}{4} & 0 & \frac{2}{3} & 0 & \frac{2}{3} & 0 & -\frac{1}{24} & 0 & -\frac{1}{24} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\frac{1}{6} & 0 & \frac{1}{6} & 0 & \frac{1}{12} & 0 & -\frac{1}{12} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & 0 & 0 \\ 0 & -\frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & 0 & 0 \\ 0 & 0 & -\frac{1}{6} & 0 & \frac{1}{6} & 0 & \frac{1}{12} & 0 & -\frac{1}{12} & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & -\frac{1}{6} & 0 & -\frac{1}{6} & 0 & \frac{1}{24} & 0 & \frac{1}{24} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{6} & 0 & -\frac{1}{6} & 0 & 0 & 0 & -\frac{1}{2} & 0 & 0 & -\frac{1}{6} & 0 & \frac{1}{6} \\ 1 & -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 \\ -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} & 0 & \frac{1}{6} & 0 & -\frac{1}{6} & 0 & 0 & -\frac{1}{2} & -\frac{1}{6} & 0 & 0 & \frac{1}{6} & 0 \\ \frac{1}{4} & 0 & -\frac{1}{6} & 0 & -\frac{1}{6} & 0 & \frac{1}{24} & 0 & \frac{1}{24} & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{p}(0,0) \\ \hat{p}(1,0) \\ \hat{p}(0,1) \\ \hat{p}(-1,0) \\ \hat{p}(0,-1) \\ \hat{p}(2,0) \\ \hat{p}(0,2) \\ \hat{p}(-2,0) \\ \hat{p}(0,-2) \\ \hat{p}(1,1) \\ \hat{p}(1,-1) \\ \hat{p}(-1,-1) \\ \hat{p}(-1,1) \\ \hat{p}(1,2) \\ \hat{p}(2,1) \end{bmatrix} \tag{50}$$

Expanding Eq. (50) into its components yields the individual equations for the desired monomial coefficients $\hat{d}_{ij}$ in the representation of the least squares solution $\hat{p}(x,y)$ as a superposition of monomials. The $\hat{d}_{ij}$ coefficients are calculated exactly if the $\hat{p}(x,y)$ values can be calculated exactly from Eq. (12). An assessment of the severity with which numerical errors in $\hat{p}(x,y)$ propagate through to numerical errors in $\hat{d}_{ij}$, and how these latter errors can be minimised, is given in the next section.

D. *Domain scaling*

In evaluating the $\hat{d}_{ij}$ by Eq. (50), reference is made to the values of $\hat{p}(x,y)$ at selected points in the domain $(x,y) \in ([-2,2],[-2,2])$. However, suppose that $\hat{p}(x,y)$ has been determined to be the optimal analytic approximation for samples distributed over a substantial region of the domain $(x,y) \in ([-A,A],[-B,B])$, for constants A and B.

Assume that $A \gg 2$ and $B \gg 2$. In this case it is (almost) certain that $\hat{p}(x,y)$ has an exceedingly small variation over the interval $(x,y) \in ([-2,2],[-2,2])$. Consequently, as can be verified from Eq. (50), all of the $\hat{d}_{ij}$ with the exception of $\hat{d}_{00}$ are very small, that is, $|\hat{d}_{ij}| \ll |\hat{p}(0,0)|$ for $\{ij\} \neq \{00\}$. However, absolute errors in the evaluation of $\hat{p}(x,y)$ at the individual selected points are not generally similar, so when linearly combined to give the absolute errors in the $\hat{d}_{ij}$, the absolute errors do not cancel in the same way as the $\hat{p}(x,y)$ cancel in yielding the $\hat{d}_{ij}$. Summarising these observations, one concludes that the $\hat{d}_{ij}$ have the same absolute errors as the $\hat{p}(x,y)$ (in terms of order of magnitude), but the values of the $\hat{d}_{ij}$ are much smaller than the $\hat{p}(x,y)$ (generally). Therefore, the relative errors in the $\hat{d}_{ij}$ are much greater than the relative errors in the $\hat{p}(x,y)$.

This problem can be completely rectified by utilising a more appropriate domain for the selected values of $\hat{p}(x,y)$ appearing in Eq. (50), as will be immediately demonstrated. Eq. (38) can be rewritten as

$$\hat{p}(A\left(\frac{x}{A}\right), B\left(\frac{y}{B}\right)) = \sum_{j=0}^{K} \sum_{i=0}^{j} A^{(j-i)} B^{i} \hat{d}_{(j-i)i} \left(\frac{x}{A}\right)^{(j-i)} \left(\frac{y}{B}\right)^{i} . \tag{51}$$

Define new coefficients $\hat{d}'_{ij} \equiv A^{i} B^{j} \hat{d}_{ij}$. (52)

Define new independent variables $x' \equiv \frac{x}{A}$, $y' \equiv \frac{y}{B}$. (53)

Define a new polynomial $\hat{p}'(x',y') \equiv \hat{p}(Ax', By')$. (54)

Substituting these definitions into Eq. (51) derives an equation analogous to Eq. (38),

$$\hat{p}'(x',y') = \sum_{j=0}^{K} \sum_{i=0}^{j} \hat{d}'_{(j-i)i} x'^{(j-i)} y'^{i} \quad . \tag{55}$$

Eq. (55) is essentially Eq. (38) with the domain scaled so that the (experimental) samples are confined to the interval $(x',y') \in ([-1,1],[-1,1])$. Eq. (50) is still applicable with the $\hat{p}$ values replaced by corresponding $\hat{p}'$ values evaluated at the same points (eg. $\hat{p}(x=1,y=2)$ is replaced by $\hat{p}'(x'=1,y'=2)$), and the $\hat{d}_{ij}$ replaced by $\hat{d}'_{ij}$. The domain within which $\hat{p}'(x',y')$ is evaluated at selected points remains $(x',y') \in ([-2,2].[-2,2])$, so there is now a much closer correspondence between the region of experimental samples and the region of selected points. In particular, $\hat{p}'(x',y')$ will in general exhibit considerable variation over the latter region, hence the previously identified numerical problem has been overcome.

Finally, the required unscaled monomial coefficients $\hat{d}_{ij}$ are obtained from the scaled coefficients $\hat{d}'_{ij}$ by inverting Eq. (52) to obtain

$$\hat{d}_{ij} = \frac{\hat{d}'_{ij}}{A^i B^j} \quad \forall\, i,j \; . \tag{56}$$

Adhering to this domain scaling procedure produces $\hat{d}_{ij}$ coefficients with relative errors of the same order of magnitude as the relative errors present in the evaluation of $\hat{p}(x,y)$, which implies that there is no degradation in precision introduced by the transformation from the $p_{ij}(x,y)$ representation to the $x^i y^j$ representation. There is, however, a significant reduction in computational complexity associated with this transformation. A technique for reducing computational complexity even further, with no detrimental effect on numerical accuracy, is indicated in the next section.

E. *Nested multiplication*

To evaluate the least squares polynomial $\hat{p}(x,y)$ using Eq. (38) for the case $K=4$ (ie. fourth degree polynomial), without retaining intermediate partial products, re-

quires 40 multiplications and 14 additions. However, any polynomial of two variables can be expressed with a *nested multiplication* structure separately in the x and y coordinates [13]. Using nested multiplication, the least squares polynomial is expressed as

$$\begin{aligned}\hat{p}(x,y) = {} & (_1(_2(_3\hat{d}_{04}y + \\ & (_4\hat{d}_{13}x + \hat{d}_{03})_4)_3y + \\ & (_3(_4\hat{d}_{22}x + \hat{d}_{12})_4x + \hat{d}_{02})_3)_2y + \\ & (_2(_3(_4\hat{d}_{31}x + \hat{d}_{21})_4x + \hat{d}_{11})_3x + \hat{d}_{01})_2)_1y + \\ & (_0(_1(_2(_3\hat{d}_{40}x + \hat{d}_{30})_3x + \hat{d}_{20})_2x + \hat{d}_{10})_1x + \hat{d}_{00})_0 \quad . \end{aligned} \tag{57}$$

Using this nested multiplication scheme requires 14 multiplications and 14 additions for every evaluation of the least squares polynomial $\hat{p}(x,y)$, which constitutes a worthwhile reduction in computational effort compared with the direct evaluation.

This completes the exposition of an algorithm for the least squares approximation of experimental samples (or discrete samples of an arbitrary function) by a polynomial in two independent variables of degree 4, and it is noteworthy that the stated intention of arriving at an algorithm devoid of matrix arithmetic has been satisfied. An assessment of the validity of the algorithm, by numerically executing it on several sets of actual samples, is undertaken in the next section.

VII. NUMERICAL VALIDATION OF THE ALGORITHM

The following procedure will be implemented to assess the validity of the algorithm that has been developed in previous sections. A 'test polynomial' of degree 4 in two variables is defined by choosing numerical values for its d_{ij} coefficients in Eq. (5). Values of this polynomial at randomly located positions are used as the samples upon which the least squares orthogonal polynomial approximation algorithm

operates. Upon execution of the algorithm, one identifies a 'reconstructed polynomial' by computing its d_{ij} coefficients. Comparison between the two polynomials is made by comparing corresponding coefficients. Since the least squares algorithm that has been explicitly stated in this paper spans a complete set of degree 4 polynomials, it is expected that the algorithm will faithfully reproduce the test polynomial as its reconstructed polynomial. Corresponding coefficients in the test and reconstructed polynomials are expected to be identical, to within floating point arithmetic errors.

Results of executing this test on two sets of samples are tabulated in Table I. Floating point arithmetic is conducted entirely in double precision (8 byte numbers, approximately 15 digit accuracy). The minimum mean square error ($\widehat{\varepsilon^2}$) is computed according to Eq. (21). Although $\widehat{\varepsilon^2} \geq 0$ by definition, in both of the examples quoted in Table I, $\widehat{\varepsilon^2}$ is calculated to be a negative number as a consequence of numerical errors. On comparing corresponding coefficients in the tables, it is eminently reasonable to conclude that the reconstructed polynomial reproduces the test polynomial to within numerical errors, thus confirming the validity of the algorithm (or more precisely, the implementation of the algorithm by this particular computer program).

VIII. CONCLUSIONS AND GENERALISATIONS

The content of this paper arose out of a genuine need to establish a simple and robust technique of analysing the data arising from physics experiments being conducted by the ESM Centre, although analogous data analysis scenarios abound both within and beyond physics (Secs. I and II).

A technique of generating orthonormal polynomials in two independent variables was stated in Sec. III. In this exposition, sufficient polynomials were constructed to form a complete set for polynomials of a certain degree; no more and no less. However, this condition is not an intrinsic restriction of the theory, and an arbitrary

number of orthonormal polynomials can be constructed. Also, the condition of two independent variables is not an intrinsic limitation of the theory. The theory that is espoused in this paper can be extended to an arbitrary number of independent variables, although there will be a commensurate increase in the complexity of the statement of the resulting equations.

Within Sec. IV the orthonormal polynomials are utilised to optimally approximate empirically or analytically determined samples. A derivation of the optimal coefficient values in the superposition of the orthonormal polynomials is undertaken.

The preceding theory is applied to explicitly construct a complete set of orthonormal polynomials of degree 4 (Sec. V). Although the orthogonal polynomial generation algorithm is conducive to direct implementation as a computational procedure, the explicit analytical approach adopted here allows the identification and omission of the numerous redundancies present in the coefficients, that would otherwise be laboriously computed by the direct naïve approach.

A general theory for the transformation of a polynomial from a representation in terms of orthonormal polynomials to the conventional representation as a sum of products is developed in Sec. VI. A specific set of transformation equations, derived according to the principles of the theory, is stated for the case of degree 4 polynomials, although the same technique can be extended to polynomials of arbitrary number of monomial terms. A more direct theory for accomplishing this transformation, which is noted together with its detrimental aspects, is conducive to implementation as a general computational procedure that is executed as part of the braoader polynomial approximation algorithm.

A practical perspective to the preceding theory is provided in Sec. VII, where the algorithm that has been developed for polynomials of degree 4 is implemented as a computer program. Using this computer program, several tests of the algorithm are

executed, the results of which manifestly confirm the practical validity and utility of the algorithm.

IX. COMPUTER PROGRAM

The computer program that was used to produce Table I is written in ANSI C, apart from the use of a few non-critical intrinsic functions. It is modular and fully documented, and a substantial portion of it can be extracted intact to be used as the foundation of an implementation of the algorithm that is discussed in this paper.

The source code for this program is available by contacting the author by electronic mail (**phrsc@cc.flinders.edu.au**), facsimile (**(618/08) 201 2905**), or post.

ACKNOWLEDGEMENTS

The subject of this paper forms part of research conducted under the direction of Prof. E. Weigold, who also offered guidance in the preparation of this paper, and to whom the author is most appreciative for his assistance. Sponsorship of the author for this research by the Defence Science and Technology Organisation (Optoelectronics Division) is gratefully acknowledged.

REFERENCES

[1] I.E. McCarthy and E. Weigold, Phys. Rep. (Sec. C of Phys. Lett.) **27**, 275 (1976)

[2] I.E. McCarthy and E. Weigold, Rep. Prog. Phys. **51**, 299 (1988)

[3] J.L. Wiza, Nucl. Instr. and Meth. **162**, 587 (1979)

[4] M. Lampton, *Scientific American*, November 1981, p.46

[5] H. Anton, *Elementary Linear Algebra* (2nd ed.) (John Wiley, 1977)

[6] R. Courant and D. Hilbert, *Methods of Mathematical Physics—Volume 1* (Interscience, 1953)

[7] J.H. Cadwell and D.E. Williams, The Computer Journal **4**, 260 (1961)

[8] J.G. Hayes, *Numerical Approximation to Functions and Data* (J.G. Hayes (ed.)), chap. 7 (1970)

[9] G.E. Forsythe, J. Soc. for Indust. and Appl. Math. **5**, 74 (1957)

[10] D.D. Sarma and J.B. Selvaraj, Computers & Geosciences **16**, 897 (1990)

[11] P.A.M. Dirac, *The Principles of Quantum Mechanics* (4th ed.) (Oxford University Press, 1958)

[12] S. Wolfram, *Mathematica: A System for Doing Mathematics by Computer* (Addison-Wesley, 1988)

[13] Ž. Jeričević, D.M. Benson, J. Bryan and L.C. Smith, J. of Microscopy **149**, 233 (1988)

TABLES

TABLE I. Results of two applications of the least squares orthogonal polynomial approximation algorithm. A 'test' polynomial is defined and used to calculate N sample values within the domain $(x, y) \in ([-A, A], [-B, B])$. The algorithm computes the 'reconstructed' polynomial with a minimum mean square error $\widehat{\epsilon^2}$.

	$N = 20$, $A = 1000.0$, $B = 1000.0$, $\widehat{\epsilon^2} = -5.8208 \times 10^{-10}$ (as calculated).		$N = 100$, $A = 5000.0$, $B = 5000.0$, $\widehat{\epsilon^2} = -2.4414 \times 10^{-3}$ (as calculated).	
coeff.	test	reconstructed	test	reconstructed
d_{00}	-8.100×10^{-4}	$-8.1000000 \times 10^{-4}$	4.820×10^{2}	4.8200000×10^{2}
d_{10}	1.170×10^{-1}	1.1700000×10^{-1}	-1.380×10^{-1}	$-1.3800000 \times 10^{-1}$
d_{01}	0.000	$-1.1368684 \times 10^{-15}$	-3.700×10^{-8}	$-3.6999985 \times 10^{-8}$
d_{20}	-9.400×10^{-5}	$-9.4000000 \times 10^{-5}$	0.000	$-4.8894435 \times 10^{-17}$
d_{11}	2.800×10^{-5}	2.8000000×10^{-5}	8.470×10^{-4}	8.4700000×10^{-4}
d_{02}	3.500×10^{-11}	$3.4999998 \times 10^{-11}$	0.000	$-6.2118488 \times 10^{-18}$
d_{30}	0.000	$9.4146912 \times 10^{-23}$	-7.100×10^{-13}	$-7.1000000 \times 10^{-13}$
d_{21}	-1.900×10^{-8}	$-1.9000000 \times 10^{-8}$	1.329×10^{-6}	1.3290000×10^{-6}
d_{12}	1.840×10^{-7}	1.8400000×10^{-7}	-4.500×10^{-13}	$-4.5000000 \times 10^{-13}$
d_{03}	0.000	$1.1368684 \times 10^{-21}$	1.100×10^{-8}	1.1000000×10^{-8}
d_{40}	3.100×10^{-11}	$3.1000000 \times 10^{-11}$	-8.280×10^{-10}	$-8.2800000 \times 10^{-10}$
d_{31}	-9.800×10^{-16}	$-9.8000000 \times 10^{-16}$	0.000	$-7.4505806 \times 10^{-25}$
d_{22}	-2.540×10^{-10}	$-2.5400000 \times 10^{-10}$	0.000	$2.7939677 \times 10^{-25}$
d_{13}	0.000	$2.7284841 \times 10^{-24}$	5.040×10^{-10}	$5.0400000 \times 10^{-10}$
d_{04}	9.060×10^{-10}	$9.0600000 \times 10^{-10}$	-8.100×10^{-16}	$-8.1000000 \times 10^{-16}$