INTEGRAL FUNCTION APPROXIMATIONS DERIVED FROM INHOMOGENEOUS EQUATIONS

by

A.W. McInnes

Department of Mathematics, University of Canterbury, Christchurch, New Zealand.

No. 61 December, 1990.

Abstract

The formulation of the problem of obtaining a unique integral function approximation to a real-valued locally analytic function is given. The integral function in this case is derived from an inhomogeneous, linear differential equation. A careful distinction is made between the approximation of the integral form which defines the polynomial coefficients of the differential equation which defines the integral function, and the approximation by the integral function itself. This formulation enables us to obtain results for existence, uniqueness and order of approximation in both normal and non-normal cases.

Subject classification: AMS(MOS) : 41A21

Key words and phrases: integral function approximation, Hermite-Padé approximation, order of approximation.

§ 1. Introduction.

This paper considers the problem of approximating a real-valued, locally analytic function by an integral function which is derived from a linear inhomogeneous ordinary differential equation. The approximations are determined by sufficient derivative information about the given function at the origin. The objective of the paper is to investigate these approximations in the same way that approximations by algebraic functions [9,11], and approximations by integral functions derived from homogeneous equations [10], have been previously studied. The main results are that a clear formulation of the problem leads to the existence and uniqueness of an approximating integral form which determines the polynomial coefficients of the linear inhomogeneous ordinary differential equation for the integral function, and the existence and uniqueness of a distinguished integral function which is a solution of this differential equation. In addition the order of approximation of the integral function is quantified in a variety of circumstances.

The class of integral functions of degree p consists of those functions which are solutions of a linear inhomogeneous ordinary differential equation of order p with polynomial coefficients. The name of integral approximants in this context appears to have been introduced by Hunter and Baker [7]. There is some confusion with this nomenclature since the original G^3J approximants [1], as well as [14,11] define integral function approximations in terms of linear homogeneous ordinary differential equations, while [7,3,2] define the approximations in terms of linear inhomogeneous equations. The latter option is being taken in this paper even though it may seem more appropriate to consider this case as part of a more general structure.

The class of integral functions includes polynomials, algebraic functions and the usual elementary functions. However, apart from equations with constant coefficients, the general linear equation of order p cannot be explicitly integrated in terms of elementary functions. As a general rule, the equations which arise from problems in applied mathematics and which cannot be transformed into equations with constant coefficients, have solutions which are transcendental functions. Since the linear homogeneous ordinary differential equation is a particular case of the inhomogeneous equation, the class of integral functions includes the well-known transcendental functions like the Bessel functions, but does not include functions like the Riemann-Zeta function which does not satisfy an ordinary differential equation of finite order [8].

As has been noted by Baker and Graves-Morris [1,2] and Baker and Lubinsky [3] and the references contained therein, these approximants have been applied to various calculations in mathematical physics. Approximants chosen from this class can be viewed as a particular case of the variety of generalizations of the Taylor polynomial approximation and the Padé rational approximation. These generalizations are often called Hermite-Padé approximants [3,4,6,12,14]. A number of recent papers [2,3,6,10,14] have considered a variety of theoretical aspects of integral function approximations, but few have clearly formulated the problem. This paper clarifies the formulation of integral function approximations and obtains explicit results for the order of approximation.

Since there are obvious analogies with approximation by algebraic functions (another particular case of the general class of Hermite-Padé approximants), the basic formulation follows that first set out in [9] and [11] for the algebraic function case, and in [10] for the case of integral functions defined in terms of the homogeneous equation. The important feature is that a careful distinction is made between the approximation of the *integral form* which determines the polynomial coefficients of the differential equation for the integral function, and the approximating properties of the integral function itself. The distinction between these two concepts leads to a separation of the degenerate cases arising from different sources, and consequently a clear treatment of the effects of each type of degenerate case. Since these two concepts overlap in case of Padé rational approximations, the differences have tended to be obscured in previous generalizations. As was the case in [9,10,11], this formulation leads to the definition of the surplus in the case of the integral form, and the definition of deficiency in the case of the approximation by integral functions.

In Section 2 the problem of the approximation of the integral form which determines the coefficient polynomials is considered. The existence of a unique integral form is established and the concept of the surplus in the approximation of the integral form is introduced. In Section 3 the approximating properties of the integral function are investigated. The concept of a normal integral form is introduced and the deficiency of a non-normal integral form is defined. The existence of a unique integral function is established and the order of approximation by this integral function is quantified. Section 4 contains a series of illustrative examples to demonstrate the results of the previous sections in particular cases.

§ 2. The Integral Form.

To formulate the problem, the class of integral functions from which we seek an approximation is first defined.

Definition 1 (Integral Function).

Let $\mathrm{n} = (n_{-1}, n_0, n_1, \ldots, n_p)$, where n_i, $i = -1(1)p$, $n_i \geq -1$, are integers and $p \geq 0$.

An n *integral function of degree-*p is a function $Q(x)$ which satisfies

$$P_{\mathbf{n},p}(Q,x) \equiv a_{-1}(x) + \sum_{i=0}^{p} a_i(x) Q^{(i)}(x) = 0 \tag{1}$$

where the $a_i(x)$ are algebraic polynomials with degree $a_i(x) \leq n_i$ for $i = -1(1)p$, with $a_i(x) \not\equiv 0$ for at least one value of i for $i = 0(1)p$, and $Q^{(i)}(x)$ is the ith formal derivative of $Q(x)$. □

By convention, a polynomial of degree -1 is identically zero. *Generally the subscripts* n, p *on* P *will be dropped when the context makes them obvious.*

The integral function $Q(x)$ is the solution of a linear, nonhomogeneous ordinary differential equation (1) of order p. It is well-known that in general the solutions to the differential equation (1) form a linear space of dimension p. A unique solution may be determined by given initial conditions which will be imposed when considering the approximation by an integral function. Note that the integral function will be regarded as being defined by the polynomial coefficients $a_i(x)$, in analogy with the way that the algebraic function is also regarded as being defined by its polynomial coefficients [9,11].

Now consider the problem of approximating a real-valued, locally analytic function $f(x)$ by an integral function $Q(x)$. We may suppose that the function $f(x)$ is analytic in the neighbourhood of the origin and thus that a sufficient number of derivative values of $f(x)$ at the origin are available. The given information is used to determine the integral form which is the best approximation to the equation (1) in an appropriate sense. That is the given information used to determine appropriate polynomial coefficients for equation (1).

Definition 2 (Integral Form).

Let $f(x)$ be a real-valued function, locally analytic at the origin.

Let $p \geq 0$, and $\mathbf{n} = (n_{-1}, n_0, n_1, \ldots, n_p)$ where $n_i \geq -1$ are integers for $i = -1(1)p$. The function

$$P_{\mathbf{n},p}(f,x) \equiv a_{-1}(x) + \sum_{i=0}^{p} a_i(x) f^{(i)}(x) = O\left(x^N\right) \tag{2}$$

will be called an $\mathbf{n}$ *integral form of degree* p, where the $a_i(x)$ are algebraic polynomials of degree $a_i(x) \leq n_i$ for $i = -1(1)p$, with $a_i(x) \not\equiv 0$ for at least one value of i for $i = 0(1)p$, $N + 1 = \sum_{i=-1}^{p}(n_i + 1)$, and $f^{(i)}(x)$ is the ith formal derivative of $f(x)$. □

Note that $P(f,x)$ may also be written as a formal power series, $r(x) = \sum_{i=0}^{\infty} r_i x^i$, and that conditions (2) are equivalent to the requirement that the linear functionals D^k satisfy

$$D^k(P(f,x)) = D^k r(x) = r^{(k)}(0)/k! = 0 \text{ for } k = 0(1)N-1\,.$$

As noted in [3], if $p = 0$ in this form the problem reduces to the usual Padé approximation problem. If $p = 1, n_{-1} = -1$ the problem reduces to Baker's D–log approximant [1], and if $p = 2$ the problem reduces to the generalized form of the G^3J approximants [1].

The existence of the polynomial coefficients $a_i(x)$ in (2) is essentially given in [3], but the theorem is included here for completeness.

Theorem 3 (Existence).

There always exists an $\mathbf{n}$ integral form of degree p for a given real-valued function $f(x)$ which is locally analytic at the origin.

Proof:

The integral form is defined by the coefficient polynomials $a_i(x)$. The existence of the $a_i(x)$ follows, since the application of the linear functionals D^k, $k = 0(1)N-1$, (where $D^k r(x) = r^{(k)}(0)/k!$) to equation (2) leads to a system of N homogeneous linear equations for the $N+1$ unknown coefficients of the $a_i(x)$. Hence a non-trivial solution, with the $a_i(x)$ not all identically zero, exists.

Further, if the $a_i(x) \equiv 0$ for $i = 0(1)p$, then $a_{-1}(x) \equiv 0$, and hence for a non-trivial solution $\mathbf{a}(x) = (a_0(x), a_1(x), \ldots, a_p(x)) \not\equiv 0$. □

As in the case of algebraic functions [9,11], the uniqueness of this integral form requires a more careful argument. The matrix form of the system of linear equations represented by (2) has the coefficient matrix

$$F = \left[F_{n_{-1}} \;:\; F_{n_0} \;:\; F_{n_1} \;:\; \cdots \;:\; F_{n_p}\right] \tag{3}$$

where

$$F_{n_{-1}} = \begin{bmatrix} I_{n_{-1}+1} \\ 0 \end{bmatrix}$$

and

$$F_{n_k} = \begin{bmatrix} d_k f_k & & & \\ d_{k+1} f_{k+1} & d_k f_k & & \\ \cdot & \cdot & & \\ d_{k+n_k} f_{k+n_k} & d_{k+n_k-1} f_{k+n_k-1} & \cdots & d_k f_k \\ \cdot & \cdot & & \cdot \\ d_{k+N-1} f_{k+N-1} & d_{k+N-2} f_{k+N-2} & \cdots & d_{k+N-1-n_k} f_{k+N-1-n_k} \end{bmatrix}$$

for $k = 0(1)p$, where $f(x) = \sum_{i=0}^{\infty} f_i x^i$, and the constants d_{k+j} in the submatrix F_{n_k} are given by $d_{k+j} = (k+j)!/j!$ with the convention that $0! = 1$.

Note that the coefficients f_i, $i = 0(1)N+p-1$ will be needed to define the integral form.

The matrix F has dimension $N \times (N+1)$ and hence the solution space has dimension at least 1. If the rank of the matrix F is N (and hence the solution space has dimension exactly 1), then any constant multiple of the coefficient polynomials $\{a_i(x)\}_{i=-1}^{p}$ will also be a solution. These solutions may be called *essentially unique* [3]. A unique representative of this class of solutions may be defined by choosing any suitable normalization of the non-trivial coefficients $\mathbf{a}(x)$. Thus if $\mathbf{a}$ represents the vector of coefficients of the non-trivial vector $\mathbf{a}(x)$, then $\|\mathbf{a}\| = 1$ for some convenient norm may be used. The use of the ℓ_∞ norm on this vector is in fact the normalization suggested by Baker and Lubinsky [3]. However, the implied normalization of $a_2^{(1)}(0) = 1$ used in [1] for the G^3J approximants (an integral form for $p = 2$) is not a suitable normalization in general since there is no requirement that $a_2^{(1)}(0) \neq 0$ in the general case for $p = 2$.

A more serious problem however, is the case when the solution space is multi-dimensional. This may occur when the matrix has rank $< N$. If the rank of F is $N + 1 - k$, then the solution space of the integral form has dimension $k > 1$. The corresponding case for algebraic forms was first set out in [9,4,11].

Definition 4 (Order of the integral form).

The *order* of the $\mathbf{n}$ integral form of degree p is defined to be

$$R = \text{Ord}(P(f,x)) \text{ if } P(f,x) = O\left(x^R\right), \neq O\left(x^{R+1}\right) \text{ as } x \to 0\,.$$

If the solution space of equations (2) is multi-dimensional, then the order of the integral form is to be interpreted as the *minimum* of the orders of the multiple solutions. □

Remarks:

(i) Note that by the nature of the order relation it makes sense to define the order of a set of solutions to be the minimum of the orders of the elements of that set.

(ii) If $P(f,x) = 0$ in Definition 4, then $f(x)$ is in fact an integral function and we may regard the order $R = \infty$ in this case.

(iii) It may be observed that if $P_{\mathbf{n},p}(f,x) = 0$ then for any non-negative integer k it follows that $D^k P_{\mathbf{n},p}(f,x) = 0$, which may also be expressed as $P_{\mathbf{m},p+k}(f,x) = 0$ for a suitable

$\mathbf{m}$ (with $m_{p+k} = n_p$). Hence if $f(x)$ is an integral function, the most appropriate integral form $P_{\mathbf{n},p}(f,x)$ may be the minimum value of p for which this form vanishes. But in general, for a fixed degree p, we seek the order R as large as possible since this will give a better approximation.

If the dimension of the solution space is $k > 1$, then to obtain a unique representative we seek a one-dimensional subspace whose elements satisfy $P(f,x) = O(x^R)$ where the order R is maximal over the space of $\mathbf{n}$ integral forms of degree p. For example let $P(f,x)$ have order R and $\bar{\mathbf{a}}^{(i)}(x) = (a_{-1}^{(i)}(x), a_0^{(i)}(x), a_1^{(i)}(x), \cdots, a_p^{(i)}(x))$, $i = 1,2$, be two linearly independent solutions to equations (2). (Note that the superscript (i) merely represents an index in this context and not a derivative as previously). Then by taking a suitable linear combination, $c_1\bar{\mathbf{a}}^{(1)}(x) + c_2\bar{\mathbf{a}}^{(2)}(x)$, of these solutions, the term $(c_1 r_R^{(1)} + c_2 r_R^{(2)})x^R$ may be eliminated and an $\mathbf{n}$ integral form of order at least $R+1$ is obtained, with the dimension of the solution space decreased by 1.

Theorem 5 (Uniqueness).

There always exists an essentially unique $\mathbf{n}$ integral form of degree p, which is of maximal order $R \geq N$, and which may be chosen uniquely by a suitable normalization of the vector of coefficients of the non-trivial vector of coefficient polynomials $\mathbf{a}(x)$. This unique representative will be denoted by $P^*(f,x)$.

Proof:

If the coefficient matrix F (equation (3)), has rank N then the solution space has dimension 1 and the result is trivial.

If the matrix F has rank $N+1-k$ for $k > 1$, then the solution space has dimension k. Suppose $P(f,x)$ has order R such that $N \leq R$, and let $\mathbf{a}^{(i)}(x)$, $i = 1(1)k$ represent a basis for this solution space.

Then $\sum_{i=1}^{k} c_i \mathbf{a}^{(i)}(x)$ represents an $\mathbf{n}$ integral form of order at least $R+k-1$, where the constants c_i are defined by the linear system

$$\sum_{i=1}^{k} c_i r_{R+j}^{(i)} = 0\,, \quad j = 0(1)k-2\,, \tag{4}$$

and where the r_{R+j} are the coefficients of x^{R+j} of the formal power series associated with $P(f,x) = r(x)$, i.e. $P(f,x) = r(x) = \sum_{i=0}^{\infty} r_i x^i$.

If the matrix of the linear system (4) has rank $k-1$, then there exists an essentially unique solution. However, if this matrix has rank $< k-1$, then we must iterate this process since there are still linearly independent solutions. Since the rank reduces by at least one at each step, the number of iterations must be finite, noting of course, that if we obtain a zero matrix or a zero value for the integral form then $R = \infty$ and the exact integral form, and hence the exact integral function, has been obtained.

A unique representative of this essentially unique solution lying in a one-dimensional space, may be obtained by a suitable normalization of the coefficients. That is, if $\mathbf{a}$ represents the vector of coefficients of the non-trivial vector $\mathbf{a}(x)$, then $\|\mathbf{a}\| = 1$ for some convenient norm may be used. $\square$

Note that if the orders of the linearly independent solutions are all different, then by taking zero multiples of all solutions except that of maximum order, we identify the unique solution of maximal order. However it is not sufficient to merely identify that particular solution of maximal order in the set as is shown by the example below.

Example 6.

Let $p = 1$ and $f(x) = \exp(x^6)$. Thus the $(1,1,1)$ integral form has $N = 5$ and the following four linearly independent solutions of equations (2) are all integral forms for $f(x)$.

$$\begin{aligned}
P^{(1)}(f,x) &\equiv -1 + f(x) &&= O(x^5) = x^6 + O(x^{12}) \\
P^{(2)}(f,x) &\equiv -x + xf(x) &&= O(x^5) = x^7 + O(x^{13}) \\
P^{(3)}(f,x) &\equiv f^{(1)}(x) &&= O(x^5) = 6x^5 + O(x^{11}) \\
P^{(4)}(f,x) &\equiv xf^{(1)}(x) &&= O(x^5) = 6x^6 + O(x^{12}) .
\end{aligned}$$

Note that the particular solution of maximal order is $P^{(2)}(f,x) \equiv -x + xf(x) = O(x^7)$, but that if the algorithm of Theorem 5 is applied then the linear system (4) is

$$\begin{bmatrix} 0 & 0 & 6 & 0 \\ 1 & 0 & 0 & 6 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{bmatrix} = 0 .$$

These equations have the solution

$$c_1 = -6,\ c_2 = 0,\ c_3 = 0,\ c_4 = 1 ,$$

and hence

$$P^*(f,x) \equiv 6 - 6f(x) + xf^{(1)}(x) = O(x^{12})$$

is the $(1,1,1)$ integral form of maximal order. A unique representative of this essentially unique integral form may be obtained by a suitable normalization of the coefficients. □

Example 7.

For a less trivial example consider $f(x) = \exp(\cos(x) - 1)$ with $p = 2$. The $(2,2,-1,0)$ integral form has $N = 6$ and the following two linearly independent solutions of equations (2) are both integral forms of this type for $f(x)$.

$$\begin{aligned}
P^{(1)}(f,x) &\equiv -6 + x^2 + (6 + 2x^2)f(x) &&= O(x^6) = \frac{3}{40}x^6 + O(x^8) \\
P^{(2)}(f,x) &\equiv -54 + 15x^2 + 62f(x) + 8f^{(2)}(x) &&= O(x^6) = \frac{37}{24}x^6 + O(x^8) .
\end{aligned}$$

The algorithm of Theorem 5 then implies that

$$\begin{aligned}
P^*(f,x) &= \tfrac{1}{2}\Big(185P^{(1)}(f,x) - 9P^{(2)}(f,x)\Big) \\
&\equiv 312 - 25x^2 - (276 + 185x^2)f(x) + 36f^{(2)}(x) = O(x^8)
\end{aligned}$$

is the $(2,2,-1,0)$ integral form of maximal order. □

Remarks:

(i) Theorem 5 establishes a unique integral form in all cases. This should be compared to Stahl [14] who remarks that integral approximants are not unique, and Baker and Lubinsky [3] who obtain a unique integral form only in special cases.

(ii) The concept of the minimal integral form is introduced by Baker and Graves-Morris [2]. This concept identifies that solution of equations (2) which has minimal degree of the coefficient polynomial $a_i(x)$ where i is the largest index of a non-vanishing coefficient polynomial. Thus in both Example 6 and Example 7, $P^{(1)}(f, x)$ would be the minimal solution. Although the concept enables the authors to identify an essentially unique integral form, there seems to be little other advantage since the essentially unique integral form of maximal order is clearly a better approximation to the differential equation defining the integral function and, as will be shown later, defines an integral function which also has a better order of approximation.

(iii) Example 6 is in contrast to the assertion by Baker and Graves-Morris [2] that there are at most $p+1$ linearly independent solutions. An examination of the coefficient matrix F (equation (3)) shows that F must have rank $n_{-1}+1$, even if $f_i = 0$ for $i = 0(1)N+p-1$ and hence $F_{n_k} = [0]$ for $k = 0(1)p$. Hence the solution space of n integral forms of degree p has dimension at most $N+1-(n_{-1}+1) = N-n_{-1}$. If $n_k = 0$ for $k = 0(1)p$, or more generally, if $\sum_{k=0}^{p} n_k = 0$, then $N - n_{-1} = p+1$, but if $\sum_{k=0}^{p} n_k > 0$ then $N - n_{-1} > p+1$.

Corollary 8.

(i) There are at most $N - n_{-1}$ linearly independent n integral forms of degree p.

(ii) Further, if $f^{(k)}(0) = k! f_k \neq 0$ for some particular value of k in the range $[0, p]$, then there are at most $N - \max(n_{-1}, n_k)$ linearly independent n integral forms of degree p.

Proof:

(i) The submatrix $F_{n_{-1}}$ of the coefficient matrix F (3) contains the identity matrix of order $n_{-1}+1$. Hence F must have at least rank $n_{-1}+1$, even if $F_{n_k} = [0]$ for $k = 0(1)p$. Hence the solution space of n integral forms of degree p has dimension at most $N+1-(n_{-1}-1) = N - n_{-1}$.

(ii) Again considering the coefficient matrix F (3), we note that if $f_k \neq 0$ for some k then the submatrix F_{n_k} has the leading principal diagonal of n_k+1 non-zero elements. Hence there exist at least $\max(n_{-1}, n_k)+1$ linearly independent columns of F and hence rank (F) is at least $\max(n_{-1}, n_k)+1$. As before, the solution space of n integral forms of degree p thus has dimension at most

$$N+1-\{\max(n_{-1}, n_k)+1\} = N - \max(n_{-1}, n_k).$$

That is there are at most $N - \max(n_{-1}, n_k)$ linearly independent integral forms of degree p. $\square$

An example of the situation when the process in the proof of Theorem 5 is necessary was given in [9,11] for the case of an algebraic form. It was noted that if $f(x)$ is an even function, the n_k are all even, and $p = 2, 4$, then an examination of the coefficient matrix F reveals that the matrix has rank of at most $N-1$, and hence the solution space has dimension of at least 2.

In fact the situation is more general and applies to both algebraic and integral forms since the basic structure of the coefficient matrix F is similar under certain circumstances. If $f(x)$ is an even function then odd derivatives of $f(x)$ are odd, but even derivatives of $f(x)$ are even. Hence by considering integral forms $P(f,x)$ in which the "basis" functions are even derivatives of $f(x)$, the coefficient matrix F for the integral form of an even function has a similar structure to that of the algebraic form for $f(x)$ an even function. Although more general theorems have been obtained, the following theorem is illustrative of the possibility of multiple solutions.

Theorem 9 (Multiple solutions).

Let $f(x)$ be an even function and let the (even) n integral form of degree $2p$ (with a "basis" of even derivatives) have n_k all even for the non-null "basis" functions (i.e. n_k even for $k = 0(2)2p$).

Then the dimension of the solution space and the rank of the coefficient matrix F for this (even) integral form satisfy the following conditions:

(i) If p is even then rank(F) is at most $N - p/2$ and hence the solution space has dimension at least $p/2 + 1$.

(ii) If p is odd then there are two cases:

  (a) If n_{-1} is even then rank(F) is at most $N - (p+1)/2$ and hence the solution space has dimension at least $(p+1)/2 + 1$.

  (b) If n_{-1} is odd then rank(F) is at most $N - (p-1)/2$ and hence the solution space has dimension at least $(p-1)/2 + 1$.

Proof:

The coefficient matrix F has its $N+1$ columns partitioned into $p+2$ non-null blocks as given by (3) since F_{n_k}, $k = 1(2)2p-1$, are all null.

Since the first block, $F_{n_{-1}}$, has the identity matrix in the first $n_{-1}+1$ rows, it is sufficient to consider the submatrix $\bar{F}$ of F which excludes the first $n_{-1}+1$ rows and first $n_{-1}+1$ columns. Subsequent references to the block F_{n_k} will mean that part of the block F_{n_k} lying in this submatrix. In this connection, the number of rows in the submatrix $\bar{F}$ is $\bar{N} = N - (n_{-1}+1)$.

Let the columns of $\bar{F}$ be rearranged into blocks of up to $p+1$ columns, in the following way. The first block consists of the first columns from each of the non-null blocks F_{n_k}, $k = 0(2)2p$. The second block of up to $p+1$ columns consists of the third columns from each of the non-null blocks F_{n_k}, $k = 0(2)2p$. If any block F_{n_k} does not have a third column then this null column is ignored and the block in the rearranged matrix will have less than $p+1$ columns. Continuing in this manner, when all the odd numbered columns of the non-null blocks F_{n_k} have been dealt with, the procedure is repeated for the even numbered columns of the non-null blocks F_{n_k}.

The $(1,1)$ element of the transformed matrix $\bar{F}$ will be zero if n_{-1} is even and can be non-zero if n_{-1} is odd. Now if p is even, the rows of the transformed matrix $\bar{F}$ are now interchanged so that, since $\bar{N}$ is even in this case:

(a) If n_{-1} is even, the even numbered rows are collected into the first $\bar{N}/2$ rows and the odd numbered rows are collected into the final $\bar{N}/2$ rows.

(b) If n_{-1} is odd, the odd numbered rows are collected into the first $\bar{N}/2$ rows and the even numbered rows are collected into the final $\bar{N}/2$ rows.

If p is odd (and hence $\bar{N}$ is odd), the same procedure is carried out except in this case there are $(\bar{N}-1)/2$ even numbered rows and $(\bar{N}+1)/2$ odd numbered rows in each case.

Thus the matrix $\bar{F}$ is now transformed into the form

$$\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} \tag{5}$$

where

(i) if p is even then A, B both have $\bar{N}/2$ rows, and
(ii) if p is odd, then if n_{-1} is even A has $(\bar{N}-1)/2$ rows and B has $(\bar{N}+1)/2$ rows, while if n_{-1} is odd A has $(\bar{N}+1)/2$ rows and B has $(\bar{N}-1)/2$ rows.

The one exception to this arrangement is the case where $n_k = 0,\ k = 0(2)2p$, when the matrix B is null. In this case $\bar{N} = p$, and if p is even then A has $p/2$ rows of zeros, while if p is odd then A has $(p+1)/2$ or $(p-1)/2$ rows of zeros according to whether n_{-1} is even or odd respectively. Hence the matrix (5) has rank at most $p/2 = \bar{N} - p/2$ if p is even, and if p is odd has rank at most $\bar{N} - (p+1)/2$ or $\bar{N} - (p-1)/2$ according as to whether n_{-1} is even or odd respectively. Thus the conclusion of the theorem follows in this case.

In the more general case if $n_k > 0$ for at least one value of k then the corresponding block F_{n_k} in (3) has one more odd numbered column than even numbered column, since n_k is even and F_{n_k} has $n_k + 1$ columns. Hence A has $p+1$ more columns than B. By considering a Laplacian expansion of the determinant of an arbitrary $\bar{N} \times \bar{N}$ submatrix of (5), in terms of the first $\bar{N}/2$ rows, in the case p is even, it is clear that each cofactor has at least $p/2$ columns of zeros. Hence the rank of (5) is at most $\bar{N} - p/2$, and hence that of F is at most $N - p/2$, in the case p is even. In the case p is odd and n_{-1} is even, we consider a Laplacian expansion in terms of the first $(\bar{N}-1)/2$ rows. Since the sum of the columns of A and B is $\bar{N}+1$ and A has $p+1$ more columns than B, then each cofactor in the expansion has at least $(p+1)/2$ columns of zeros. Hence the rank of (5) is at most $\bar{N} - (p+1)/2$, and hence the rank of F is at most $N - (p+1)/2$. In the case p is odd and n_{-1} is odd, a similar Laplacian expansion in terms of the first $(\bar{N}+1)/2$ rows has cofactors with at least $(p-1)/2$ columns of zeros and hence leads to the result that the rank of F is at most $N - (p-1)/2$.

The rank of F may in fact be even less than these values if the ranks of A or B are not full. □

This theorem is illustrated by Example 7 where the even function has an integral form of degree $p = 2$ with a basis of even derivatives (i.e. the coefficient of $f^{(1)}(x)$ is required to be null). This corresponds to $p = 1$ in the notation of Theorem 9 and since $n_{-1} = 2$ is even, case (ii)(a) applies, which states that the solution space has dimension at least $(p+1)/2 + 1 = (1+1)/2 + 1 = 2$, and two linearly independent solutions were obtained as expected.

The theorem is further illustrated by the following example.

Example 10.

Let $f(x) = \exp(\cos(x) - 1)$ and consider the (even) integral form of degree 6 of the type $(0, 0, -1, 0, -1, 0, -1, 0)$. For this example $N = 4$.

This example satisfies Theorem 9 with $p = 3$. Case (ii)(a) applies and hence the solution space has dimension $(p+1)/2 + 1 = (3+1)/2 + 1 = 3$. Hence 3 linearly independent solutions are expected, which are

$$\begin{array}{llll} P^{(1)}(f,x) \equiv & -3+4f(x)+f^{(2)}(x) & & = O(x^4) \\ P^{(2)}(f,x) \equiv & 27-31f(x)+ & f^{(4)}(x) & = O(x^4) \\ P^{(3)}(f,x) \equiv & -348+379f(x)+ & \qquad f^{(6)}(x) & = O(x^4) \,. \end{array}$$

The algorithm of Theorem 5 then implies that the linear system (4) is

$$\begin{bmatrix} -\frac{5}{8} & \frac{85}{8} & -210 \\ \frac{17}{48} & -\frac{373}{48} & \frac{385}{2} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = 0 \,.$$

These equations have the solution

$$c_3 = 1, \quad c_2 = 42, \quad c_1 = 378,$$

and hence

$$P^*(f,x) \equiv -348 + 589f(x) + 378f^{(2)}(x) + 42f^{(4)}(x) + f^{(6)}(x) = O(x^8)$$

is the $(0,0,-1,0,-1,0,-1,0)$ integral form of maximal order. □

§ 2.1 Degeneracies in the integral form

In the integral form defined by equation (2) it is possible that some of the coefficient polynomials are not of full degree, i.e. $\deg(a_i(x)) < n_i$ for some i. Except when *all* the coefficient polynomials are not of full degree, this is of no particular consequence. It may be compared to the occurrence in the algebraic function case when the Taylor polynomial approximation of degree 1 to the function $f(x) = 1 + x^2$ is given by $p_1(x) = 1$, i.e. there is a zero coefficient of the term in x.

Of greater importance is the fact that a particular set of coefficient polynomials which solves (2), may in fact eliminate more of the coefficients of $r(x)$ than just the first N. A simple example for $p = 1$ is the $(0,0,0)$ type approximation to $f(x) = (x+2)^3 - 3(x+1)^2 + 2$. In this case $N = 2$, but the integral form is

$$1 - 1.f(x) + 1.f^{(1)}(x) = O(x^3) \,.$$

In [9,11] the term "surplus" was introduced to describe this phenomenon for approximation by algebraic functions.

Definition 11 (Surplus of integral form).

The *surplus*, $S(\mathbf{n})$, of the $\mathbf{n}$ integral form of degree p, $P^*(f,x)$ is defined by

$$S(\mathbf{n}) = \mathrm{Ord}(P^*(f,x)) - N$$

where the order of the integral form is defined in Definition 4. □

The surplus, $S(\mathbf{n}) = S \geq 0$, is the amount of extra matching obtained from $P^*(f,x)$. This may be achieved by serendipity for a particular function (as in the simple example above), or by the process of obtaining a unique solution as outlined in the proof of Theorem 5. It is clear that in general we would like the surplus to be as large as possible since if $S = \infty$ then the integral form represents an integral function exactly.

In [9,11], the surplus was used to define an S–table of the algebraic forms. It was indicated that in the rational case ($p = 1$), this table illustrates the block structure of the Padé table in a somewhat easier fashion than the traditional C–table [1]. In the case of integral forms of degree p, the S–table of the integral forms would be a $(p+1)$–dimensional table. The structure of this table could be expected to lead to the structure of the table for the $\mathbf{n}$ integral forms of degree p in an analogous way. An analogous theorem for basic block structure can also be given for the integral forms.

Theorem 12 (Basic block structure).

If the $\mathbf{n}$ integral form of degree p, $P_{\mathbf{n}}^*(f,x)$, has surplus $S(\mathbf{n}) = S > 0$, then

$$x^r P_{\mathbf{n}}^*(f,x)\,, \qquad r = 0(1)(S/(p+1))$$

is an $\mathbf{m}$ integral form with a surplus of $S(\mathbf{m}) \geq S(\mathbf{n}) - (p+1)r - \sum_{k=-1}^{p} i_k$, where

$$\begin{aligned}
&\mathbf{m} = (m_{-1}, m_0, m_1, \ldots, m_p) \quad \text{with} \\
&m_k = n_k + r + i_k\,, \quad i_k \geq 0 \quad \text{for } k = -1(1)p\,, \\
&\text{satisfies} \quad \sum_{k=-1}^{p} i_k \leq S - (p+1)r\,.
\end{aligned}$$

Proof:

Since $\quad \mathrm{Ord}(P_{\mathbf{n}}^*(f,x)) = N + S$, then

$$\mathrm{Ord}(x^r P_{\mathbf{n}}^*(f,x)) = N + S + r\,.$$

Hence $x^r P_{\mathbf{n}}^*(f,x)$ will be an $\mathbf{n} + r\mathbf{o}$ (where $\mathbf{o} = (1,1,\ldots,1)$) integral form of degree p provided $(p+1)r \leq S$. The surplus of this integral form, $P_{\mathbf{n}+r\mathbf{o}}(f,x)$, is $S-(p+1)r$. Further, this integral form is also an integral form of the type $\mathbf{m}$ where $\mathbf{m} = (m_{-1}, m_0, m_1, \ldots, m_p)$ and $m_k = n_k + r + i_k$, $i_k \geq 0$, for $k = -1(1)p$, satisfies $\sum_{k=-1}^{p} i_k \leq S - (p+1)r$. This follows since $P_{\mathbf{m}}(f,x) = O(x^{M+S(\mathbf{m})})$ with $M + 1 = \sum_{k=-1}^{p}(m_k + 1)$. Hence $\mathrm{Ord}(P_{\mathbf{m}}(f,x)) \geq M + S(\mathbf{m}) = N + S + r$ (from above). Substituting for M and N, and using the relation between m_k and n_k gives

$$\sum_{k=-1}^{p} m_k + p + 1 + S(\mathbf{m}) = \sum_{k=-1}^{p} n_k + p + 1 + S + r\,.$$

Hence $\sum_{k=-1}^{p} i_k + (p+1)r + S(\mathbf{m}) = S$, and $S(\mathbf{m}) \geq 0$ gives the required relation. $\square$

This type of basic structure was first identified in the case of algebraic functions in [9,4]. The same structure was noted for integral functions in [10]. Some remarks along these lines about the structure of the set of solutions have also been made in [2]. The full determination of the structure is complicated by the occurence of overlapping structures and will be deferred to future reports. However a few basic observations on the result of Theorem 12 can be made.

Remarks:

(i) It should be emphasized that $P_{\mathbf{m}}(f,x)$ in Theorem 12 is *an* integral form of type $\mathbf{m}$, but is not necessarily the $\mathbf{m}$ integral form of maximal order. This is particularly true in the obvious case if $P_{\mathbf{n}}^{*}(f,x)$ has $a_i(x) \equiv 0$ for some value of i, $i = -1(1)p$.

(ii) If there is another structure of this type which overlaps the structure of Theorem 12, then for some of these values of $\mathbf{m}$ there will be additionally linearly independent solutions. Theorem 5 will need to be applied to find $P_{\mathbf{m}}^{*}(f,x)$, which will come from the set of additional solutions since these have greater order. However, if there is no overlapping structure, then for those $\mathbf{m}$ such that $i_k = 0$ for at least one value of k, $P_{\mathbf{m}}(f,x)$ has maximal order. This follows since the polynomial coefficient corresponding to this value of k has full degree and hence no additional factors of x may be multiplied through in equation (2) to raise the nominal order. Thus, if there is no overlapping structure and $\mathbf{m}$ is such that $i_k = 0$ for at least one value of k then $P_{\mathbf{m}}(f,x) = P_{\mathbf{m}}^{*}(f,x)$.

(iii) Note also that it may be assumed that the $\mathbf{n}$ integral form $P_{\mathbf{n}}^{*}(f,x)$ does not have polynomial coefficients with a common polynomial factor $b(x)$, $b(0) \neq 0$. This follows since if $P_{\mathbf{n}}(f,x)$ (with order R) has polynomial coefficients with such a common factor then the solution space is multi-dimensional since there exists an integral form of the same type $\mathbf{n}$ with a higher order. Hence $P_{\mathbf{n}}^{*}(f,x)$ (with maximal order) has the form $x^q P_{\mathbf{n}-q\mathbf{o}}(f,x)$ (where $\mathbf{o} = (1,1\ldots,1)$) with order $R+q$. In this argument it is assumed further that $\mathbf{n}$ is a type which does not belong to an overlapping structure, so that it may be asserted that this $P_{\mathbf{n}}(f,x)$ has maximal order (compare Remark (ii) above). If $\mathbf{n}$ is a type that does belong to an overlapping structure then the argument is unchanged but $P_{\mathbf{n}}^{*}(f,x)$ may have an even higher order (compare Remark (ii)). These remarks will be amplified in a forthcoming report.

Some simple examples of the basic structure can be given.

Example 13.

Let $p = 1$ and $f(x) = (x+2)^3 - 3(x+1)^2 + 2$. The $(0,0,0)$ integral form is

$$1 - f(x) + f^{(1)}(x) = O(x^3)\,.$$

Since $N = 2$, this integral form has surplus $S = 1$. Since $p = 1$, Theorem 12 implies that this is also an integral form of the types $(1,0,0),(0,1,0),(0,0,1)$ with $S = 0$. □

Example 14.

Consider again Example 6. The $(1,1,1)$ integral form of degree 1 for $f(x) = \exp(x^6)$ is

$$P^{*}(f,x) = 6 - 6f(x) + xf^{(1)}(x) = O(x^{12})\,.$$

Since $N = 5$, this integral form has surplus $S = 7$.

Using Theorem 12 this is also an integral form of types $(2,1,1),(1,2,1),(1,1,2)$ with S=6 and of types $(2,2,1),(2,1,2),(1,2,2)$ with $S = 5$. Also by Theorem 12, the $(2,2,2)$ integral form is $xP^{*}(f,x)$ with $N = 8$ and $S = 5$. Also the $(3,3,3)$ integral form is $x^2P^{*}(f,x)$ with $N = 11$ and $S = 3$, and the $(4,4,4)$ integral form is $x^3P^{*}(f,x)$ with

$N = 14$ and $S = 1$. The intermediate members of the "lattice" may be fitted in in the same was as before.

Finally we observe that $x^3 P^*(f,x)$ is also an integral form of types $(5,4,4), (4,5,4)$ and $(4,4,5)$ with $N = 15$ and $S = 0$. However the integral form $x^4 P^*(f,x)$ is also a $(4,4,5)$ integral form of order 16 and will thus be the form of maximal order for the type $(4,4,5)$. This is an example of the overlapping structure noted above.

Also there is another solution for the form of type $(4,5,4)$. This is the integral form

$$-6x^5 f(x) + f^{(1)}(x) = 0\,,$$

which has $S = \infty$ and implies that $f(x)$ is an integral function of this type as expected. □

§ 3. The Integral Approximation

Once the unique n integral form of degree p, $P^*(f,x)$, satisfying equation (2) has been obtained, it is clear that we can define an n integral function of degree p, $Q(x)$, which satisfies the equation

$$P^*(Q,x) \equiv a_{-1}(x) + \sum_{i=0}^{p} a_i(x) Q^{(i)}(x) = 0\,. \tag{6}$$

Since the coefficients $a_i(x)$ are determined by the given function $f(x)$, this function, $Q(x)$, represents an approximation to $f(x)$. Recall that $f(x)$ is assumed given in the sense that a sufficient number of derivative values at the origin are known.

It is important to note that if we are considering approximations to $f(x)$ then the differential equation (6) *must* be that derived from equations (2). Thus although an arbitrary differential equation may be constructed to show anomalous behaviour, this is irrelevant unless the differential equation is one corresponding to an integral form derived from a locally analytic function $f(x)$.

From the general theory of linear ordinary differential equations [8], it is known that equation (6) *normally* has p linearly independent analytic solutions at the origin. To keep the notation consistent with that introduced for algebraic functions in [9,11], we will call this normal behaviour. This normal behaviour follows provided that $a_p(0) \neq 0$ [8]. In fact some authors [13] call the equation (6) "normal on I" if $a_p(x)$ is never zero on the interval I. However, if $a_p(0) = 0$ then the origin is a singular point of the differential equation and possible solutions will have a different character.

Definition 15 (Normal form).

Given a real-valued, locally analytic function $f(x)$, the corresponding n integral form of degree p, $P^*(f,x)$, is called *normal* if

$$\partial P^*(f,x)/\partial f^{(p)}|_{x=0} \neq 0 \quad [\text{or } a_p(0) \neq 0]\,.$$

□

If $p = 1$ this condition becomes $a_1(0) \neq 0$ which corresponds to the requirement for the D–log approximation in this case. The condition also corresponds to the condition in [3] that the approximation to a meromorphic function be "pole-matching", i.e., that $a_p(x)$ vanishes only at the poles of $f(z)$ in $|z| < R$. However this definition of normality does not correspond to that used in [12] where the given condition corresponds more closely to the condition that the surplus $S = 0$. This confusion probably arises from the fact that the notions of surplus and normality tend to overlap in the case $p = 1$, as noted previously [9,11].

Since the solution space of equation (6) has dimension p it will be necessary to impose p linearly independent conditions in order to distinguish a unique solution $Q^*(x)$ in this space. The usual conditions used to distinguish a unique solution are

$$Q^{*(i)}(0) = f^{(i)}(0), \quad i = 0(1)p-1\,.$$

We may consider the approximation properties of this unique solution as an approximation to $f(x)$.

Theorem 16 (Order of approximation).

Let $P^*(f,x)$ be the *normal* n integral form of degree p defined by equation (2), and satisfying $P^*(f,x) = O(x^{N+S})$ where $S \geq 0$ is the surplus.

The unique n integral function of degree p, $Q^*(x)$, defined by the equation

$$P^*(Q^*,x) = 0,$$

subject to the initial conditions

$$Q^{*(i)}(0) = f^{(i)}(0), \quad i = 0(1)p, \tag{7}$$

is an approximation to $f(x)$ satisfying

$$Q^*(x) = f(x) + O\left(x^{N+S+p}\right).$$

Proof:

Let $N + S = R$. Since we have

$$P^*(f,x) = O\left(x^R\right), \quad P^*(Q^*,x) = 0,$$

then

$$\frac{d^i}{dx^i}P^*(f,x)|_{x=0} = 0 = \frac{d^i}{dx^i}P^*(Q^*,x)|_{x=0}, \quad i = 0(1)R-1\,. \tag{8}$$

For $i = 0$ in (8)

$$P^*(f,x)|_{x=0} = P^*(Q^*,x)|_{x=0}\,.$$

The initial conditions (7) imply this equation becomes

$$a_p(0)f^{(p)}(0) = a_p(0)Q^{*(p)}(0)$$

and some P^* is normal, $a_p(0) \neq 0$, and hence

$$Q^{*(p)}(0) = f^{(p)}(0)\,. \tag{9}$$

For $i = 1$ in (8)

$$\left[P^*\left(f^{(1)}(x), x\right) + \partial P^*(f, x)/\partial x\right] \Big|_{x=0} = \left[P^*\left(Q^{*(1)}(x), x\right) + \partial P^*(Q^*, x)/\partial x\right] \Big|_{x=0} .$$

Again using the initial conditions (7) and also (9), this equation reduces to

$$a_p(0) f^{(p+1)}(0) = a_p(0) Q^{*(p+1)}(0),$$

which implies since $a_p(0) \neq 0$,

$$Q^{*(p+1)}(0) = f^{(p+1)}(0) .$$

In general

$$\left[P^*\left(f^{(i)}(x), x\right) + F_i\right] \Big|_{x=0} = \left[P^*\left(Q^{*(i)}(x), x\right) + G_i\right] \Big|_{x=0}, \quad i = 0(1)R-1 \tag{10}$$

where

$$\begin{aligned} F_0 &= 0 = G_0 , \\ F_1 &= \partial P^*(f, x)/\partial x = F_1\left(x, f(x), f^{(1)}(x), \ldots, f^{(p)}(x)\right) , \\ G_1 &= \partial P^*(Q^*, x)/\partial x = F_1\left(x, Q^*(x), Q^{*(1)}(x), \ldots, Q^{*(p)}(x)\right) , \end{aligned}$$

and

$$\begin{aligned} F_i &= \partial P^*\left(f^{(i-1)}(x), x\right)/\partial x + dF_{i-1}/dx = F_i\left(x, f(x), f^{(1)}(x), \ldots, f^{(p+i-1)}(x)\right) , \\ G_i &= \partial P^*\left(Q^{*(i-1)}(x), x\right)/\partial x + dF_{i-1}/dx = F_i\left(x, Q^*(x), Q^{*(1)}(x), \ldots, Q^{*(p+i-1)}(x)\right) , \\ &\qquad \text{for } i = 1(1)R-1 . \end{aligned}$$

The induction step implies that since

$$Q^{*(j)}(0) = f^{(j)}(0) , \qquad j = 0(1)p+i-1 ,$$

then $F_i|_{x=0} = G_i|_{x=0}$ and hence, since P^* is normal, equation (10) implies that

$$Q^{*(p+i)}(0) = f^{(p+i)}(0) .$$

It follows that

$$Q^{*(i)}(0) = f^{(i)}(0) , \qquad i = 0(1)R+p-1 ,$$

and hence

$$Q^*(x) = f(x) + O\left(x^{R+p}\right) . \qquad \square$$

Hence, in the case that $P^*(f, x)$ is a normal form, a unique integral function of degree p is obtained from the unique corresponding integral form of order $R = N + S$, and this unique function, $Q^*(x)$, approximates $f(x)$ in the neighbourhood of the origin with order of approximation $N + S + p$. This improves the result in [2, eqn. (5.10)].

§ 3.1 The Non-Normal Case.

Differential equations in which the corresponding integral form is not normal are not uncommon. A simple example for the special case of homogeneous differential equations is Bessel's equation, whose solutions are the well-known Bessel functions. A further example occurs when the integral form has a factor of $x^r, r > 0$, as was obtained in Theorem 12. This situation leads to an approximation whose order is less than that expected in the normal case. This concept has been previously identified (but not formalized) in the case of quadratic algebraic functions in [5].

In [9,10,11] a deficiency index was defined to measure the amount by which the approximation falls short of the expected order. An analogous concept is appropriate in this case for integral functions.

Definition 17 (Deficiency).

The *deficiency*, $D(\mathbf{n}) = D \geq 0$, of the $\mathbf{n}$ integral function of degree p, $Q^*(x)$, which satisfies $P^*(Q, x) = 0$, is defined in terms of the corresponding integral form by

$$\frac{\partial}{\partial f^{(p)}}\left[\frac{d^k P^*(f,x)}{dx^k}\right] = O\left(x^{D-k}\right) \qquad \text{for } k = 0(1)D, \text{ as } x \to 0,$$
$$\text{and} \quad \neq O\left(x^{D-k+1}\right) \quad \text{for at least one value of } k\,.$$

□

This condition can also be expressed alternatively in terms of the coefficients of $P^*(f, x)$. Thus for $k = 0$ we obtain $a_p^*(x) = O(x^D)$. For $k = 1$ we obtain $a_{p-1}^*(x) + a_p^{*(1)}(x) = O(x^{D-1})$, which implies $a_{p-1}^*(x) = O(x^{D-1})$. For $k = 2$, $a_{p-2}^*(x) + 2a_{p-1}^{*(1)}(x) + a_p^{*(2)}(x) = O(x^{D-2})$, which may be written as $a_{p-2}^*(x) = O(x^{D-2})$ using the previous results.

Thus, in general, the conditions for the deficiency, D, become

$$\begin{aligned} a_{p-k}^*(x) &= O\left(x^{D-k}\right) \qquad \text{for } k = 0(1)D, \text{ as } x \to 0 \\ &\neq O\left(x^{D-k+1}\right) \quad \text{for at least one value of } k\,. \end{aligned}$$

Note that if $D = 0$, the integral form is normal. It will be shown that the deficiency, D, is the amount by which the order of approximation falls short of the expected order.

Firstly we consider the case where the integral form has a factor of $x^r, r > 0$ an integer.

Lemma 18 (Common factor).

If $P_{\mathbf{n}}^*(f, x) = x^r P_{\mathbf{m}}^*(f, x)$, where $P_{\mathbf{m}}^*(f, x)$ has polynomial coefficients with no common factor of x (i.e. $\sum_{i=-1}^{p} |a_i^*(0)| \neq 0$), then

$$D = D(\mathbf{n}) = D(\mathbf{m}) + r\,.$$

Proof:

$$\frac{d^k P_{\mathbf{n}}^*(f,x)}{dx^k} = \sum_{j=0}^{k} \binom{k}{j} \frac{\partial^j \left(P_{\mathbf{n}}^*\left(f^{(k-j)}, x\right)\right)}{\partial x^j}\,. \tag{11}$$

If the polynomial coefficients of $P^*_{\mathbf{m}}(f,x)$ are $a^*_i(x)$, $i = -1(1)p$, then the coefficients of $P^*_{\mathbf{n}}(f,x)$ are $x^r a^*_i(x)$. Applying the operator $\partial/\partial f^{(p)}$ to equation (11), the right side becomes

$$\sum_{j=0}^{k} \binom{k}{j} \frac{d^j}{dx^j} \left[x^r a^*_{p-k+j}(x)\right] .$$

If $k < r$ then this expression will still have a factor of x, and if $k = r$ then the term $k! a^*_p(x)$ is the only term without an explicit factor of x. Provided $a^*_p(0) \neq 0$, then by Definition 17, the integral form $P^*_{\mathbf{n}}(f,x)$ has deficiency r. However if $a^*_p(0) = 0$, there will be a further factor of x. In fact if $a^{*(1)}_p(0) = 0 = a^*_{p-1}(0)$ there will be another additional factor of x. Continuing in this way, we are simply restating the conditions for the deficiency of $P^*_{\mathbf{m}}(f,x)$. That is the total deficiency of $P^*_{\mathbf{n}}(f,x)$ will be r plus the deficiency of $P^*_{\mathbf{m}}(f,x)$. □

Theorem 19.

If $P^*_{\mathbf{n}}(f,x) = x^r P^*_{\mathbf{m}}(f,x)$, where $P^*_{\mathbf{m}}(f,x)$ is the *normal* integral form satisfying

$$P^*_{\mathbf{m}}(f,x) = O\left(x^{M+S(\mathbf{m})}\right)$$

where $S(\mathbf{m})$ is the surplus and $M + 1 = \sum_{i=-1}^{p}(m_i + 1)$, then the $\mathbf{n}$ integral function, $Q^*(x)$, defined by the equation

$$P^*_{\mathbf{n}}(Q^*, x) = 0$$

subject to the initial conditions

$$Q^{*(i)}(0) = f^{(i)}(0) , \qquad i = 0(1)p-1 ,$$

is an approximation to $f(x)$ satisfying

$$Q^*(x) = f(x) + O\left(x^{N+S+p-D}\right) ,$$

where $D = D(\mathbf{n}) \geq 0$ is the deficiency of $P^*_{\mathbf{n}}(f,x)$,
and $S = S(\mathbf{n}) \leq 0$ is the surplus of $P^*_{\mathbf{n}}(f,x)$.

Proof:

By Lemma 18, $D = D(\mathbf{n}) = D(\mathbf{m}) + r$. Since $P^*_{\mathbf{m}}(f,x)$ is normal, $D(\mathbf{m}) = 0$ and hence $D = r$.

Now $P^*_{\mathbf{n}}(Q,x) = x^r P^*_{\mathbf{m}}(Q,x) = 0$, and hence the $\mathbf{n}$ integral function of degree p defined by $P^*_{\mathbf{n}}(Q,x) = 0$ is in fact the same as the $\mathbf{m}$ algebraic function of degree p defined by $P^*_{\mathbf{m}}(Q,x) = 0$. Hence by Theorem 16 we have

$$Q^*(x) = f(x) + O\left(x^{M+S(\mathbf{m})+p}\right) .$$

By Theorem 12, we have $N = M + (p+2)r$ and $S(\mathbf{n}) = S(\mathbf{m}) - (p+1)r$. Hence $M + S(\mathbf{m}) + p = N + S(\mathbf{n}) + p - r$. □

As with the case of algebraic functions [9,11], additional complications can arise beyond the basic definition of the deficiency. As in those previous papers, we illustrate the basic process involved and avoid the additional complications by making the further assumption that in fact *each* $a^*_{p-k}(x)$ does not have further zeros than that needed for the deficiency. That is we assume that

$$\operatorname{Ord}\left(\frac{\partial P^*_{\mathrm{n}}(f,x)}{\partial f^{(p-k)}}\right) = \operatorname{Ord}(a^*_{p-k}(x)) = D-k \quad \text{for } k = 0(1)D\,.$$

If the integral form $P^*_{\mathrm{n}}(f,x)$ has deficiency $D > 0$ then $a^*_{p-k}(0) = 0$ for $k = 0(1)D-1$. Hence for $i = 0$ in equation (8) in Theorem 16, we need the initial conditions $Q^{*(i)}(0) = f^{(i)}(0)$ only for $i = 0(1)p-1-D$ in order to deduce $Q^{*(p-D)}(0) = f^{(p-D)}(0)$. (If $p-1-D < 0$ then the equation is automatically satisfied). However these initial conditions still leave a D–dimensional solution space for the differential equation. Thus we need another D conditions in order to identify a unique solution. These auxiliary conditions are provided by the D roots of the auxiliary equation $\sum_{j=0}^{D}\binom{i}{j}a^*_{p-D+j}(0) = 0$, which is a polynomial of degree D in the variable i. Since we seek a solution of the differential equation which is locally analytic at the origin, non-analytic solutions to the differential equation will automatically be eliminated by the requirement that we consider only integral solutions to the auxiliary equation which lie in the interval $[1, R-1]$.

Suppose now that $P^*_{\mathrm{n}}(f,x)$ is an integral form whose polynomial coefficients have no common factor of x and that $Q(x)$ defined by $P^*_{\mathrm{n}}(Q,x) = 0$ has deficiency $D > 0$. (The subscript n will be dropped as being understood for the remainder of this section).

Theorem 20 (Order of approximation).

Let S be the surplus of the n integral form of degree p, $P^*(f,x)$, and let D be the deficiency of the corresponding n integral function defined by $P^*(Q,x) = 0$.

If $P^*(f,x)$ satisfies

(i) $\left[\sum_{i=-1}^{p} |a^*_i(x)|\right]\Big|_{x=0} \neq 0,$

(ii) $\operatorname{Ord}\left(\dfrac{\partial P^*(f,x)}{\partial f^{(p-k)}}\right) = D-k, \quad \text{for } k = 0(1)D,$

then the associated n integral function of degree p, $Q^*(x)$, defined by the equation

$$P^*(Q^*,x) = 0$$

subject to the initial conditions

$$Q^{*(i)}(0) = f^{(i)}(0)\,, \qquad i = 0(1)(p-1-D)\,, \tag{12}$$

and the auxiliary conditions

$$Q^{*(p+i-D)}(0) = f^{(p+i-D)}(0)\,, \tag{13}$$

where $i \in \{1,2,\ldots,R-1\}$ satisfies the auxiliary equation

$$\sum_{j=0}^{D}\binom{i}{j}a^*_{p-D+j}(0) = 0\,, \tag{14}$$

is an approximation to $f(x)$ satisfying

$$Q^*(x) = f(x) + O\left(x^{N+S+p-D}\right).$$

Proof:

Firstly note that the auxiliary equation (14) is a polynomial of degree D in the variable i, with all the constant coefficients being non-zero on account of condition (ii) on the integral form $P^*(f,x)$. We are concerned only with integer solutions in the interval $[1, R-1]$ and it is clear that there can be at most D such solutions, and hence at most D auxiliary conditions (13).

Let $N+S=R$. Following a similar argument to the proof of Theorem 16 we have

$$P^*(f,x) = O\left(x^R\right), \qquad P^*(Q^*,x) = 0, \tag{15}$$

and

$$\frac{\partial P^*(f,x)}{\partial f^{(p-k)}} = O\left(x^{D-k}\right), \quad \frac{\partial P^*(Q^*,x)}{\partial Q^{*(p-k)}} = O\left(x^{D-k}\right), \quad k = 0(1)D. \tag{16}$$

From equations (15)

$$\frac{d^i}{dx^i} P^*(f,x)|_{x=0} = 0 = \frac{d^i}{dx^i} P^*(Q^*,x)|_{x=0}, \quad i = 0(1)R-1,$$

and hence

$$\sum_{j=0}^{i} \binom{i}{j} \frac{\partial^j P^*\left(f^{(i-j)}(x), x\right)}{\partial x^j}\Bigg|_{x=0} = \sum_{j=0}^{i} \binom{i}{j} \frac{\partial^j P^*\left(Q^{*(i-j)}(x), x\right)}{\partial x^j}\Bigg|_{x=0}, \quad i = 0(1)R-1. \tag{17}$$

From equations (16)

$$\frac{d^j}{dx^j}\left(\frac{\partial P^*(f,x)}{\partial f^{(p-k)}}\right)\Bigg|_{x=0} = 0 = \frac{d^j}{dx^j}\left(\frac{\partial P^*(Q^*,x)}{\partial Q^{*(p-k)}}\right)\Bigg|_{x=0}, \quad k = 0(1)D, \quad j = 0(1)D-k-1 \tag{18}$$

and

$$\frac{d^{D-k}}{dx^{D-k}}\left(\frac{\partial P^*(f,x)}{\partial f^{(p-k)}}\right)\Bigg|_{x=0} = \frac{d^{D-k}}{dx^{D-k}}\left(\frac{\partial P^*(Q^*,x)}{\partial Q^{*(p-k)}}\right)\Bigg|_{x=0} \neq 0, \quad k = 0(1)D. \tag{19}$$

(This notation is a little redundant, but has been written this way to be consistent with that in [9,11])

(a) Consider the case $D = 1$.
For $i = 0$ in equation (17)

$$P^*(f,x)|_{x=0} = P^*(Q^*,x)|_{x=0}.$$

Using the initial conditions (12) and equation (18) this reduces to

$$a^*_{p-1}(0)\left(f^{(p-1)}(0) - Q^{*(p-1)}(0)\right) = 0,$$

which implies, using (19),

$$Q^{*(p-1)}(0) = f^{(p-1)}(0).$$

For $i = 1$ in equation (17)

$$\left[P^*\left(f^{(1)}(x), x\right) + \frac{\partial P^*(f,x)}{\partial x}\right]\Big|_{x=0} = \left[P^*\left(Q^{*(1)}(x), x\right) + \frac{\partial P^*(Q,x)}{\partial Q}\right]\Big|_{x=0}.$$

Using the initial conditions (12), the previous result from $i = 0$, and equation (18), this reduces to

$$\left(a^*_{p-1}(0) + a^{*(1)}_p(0)\right)\left(f^{(p)}(0) - Q^{*(p)}(0)\right) = 0,$$

which implies, using (19) or the auxiliary condition (13) if $i = 1$ satisfies the auxiliary equation (14),

$$Q^{*(p)}(0) = f^{(p)}(0).$$

Proceeding now as in the proof of Theorem 16, we obtain in general from equation (17)

$$\begin{aligned}&\left[P^*\left(f^{(i)}(x), x\right) + \binom{i}{1}\frac{\partial P^*\left(f^{(i-1)}(x), x\right)}{\partial x} + F_i\right]\Bigg|_{x=0}\\ &= \left[P^*, \left(Q^{*(i)}(x), x\right) + \binom{i}{1}\frac{\partial P^*\left(Q^{*(i-1)}(x), x\right)}{\partial x} + G_i\right]\Bigg|_{x=0}, \quad i = 2(1)R-1\end{aligned} \tag{20}$$

where $F_i = F_i(x, f(x), f^{(1)}(x), \ldots, f^{(p+i-2)}(x))$
$G_i = F_i(x, Q^*(x), Q^{*(1)}(x), \ldots, Q^{*(p+i-2)}(x))$ for $i = 2(1)R-1$.
As before, the term $a^*_p(x)f^{(p+i)}(x)$ in $P^*(f^{(i)}, x)$ vanishes at $x = 0$ by (18), and the induction step implies that since

$$Q^{*(j)}(0) = f^{(j)}(0), \quad j = 0(1)(p+i-2),$$

equation (20) reduces to

$$\left(a^*_{p-1}(0) + ia^{*(1)}_p(0)\right)\left(f^{(p+i-1)}(0) - Q^{*(p+i-1)}(0)\right) = 0,$$

which implies, using (19) or the auxiliary condition (13) if this value of i satisfies the auxiliary equation (14),

$$Q^{*(p+i-1)}(0) = f^{(p+i-1)}(0) \quad \text{for } i = 1(1)R-1.$$

It follows that

$$Q^{*(j)}(0) = f^{(j)}(0) \quad , \ j = 0(1)R+p-2,$$

and hence

$$Q^*(x) = f(x) + O\left(x^{R+p-1}\right) = f(x) + O\left(x^{R+p-D}\right).$$

(b) Let $D = 2$.
For $i = 0$ in equation (17)

$$P^*(f,x)|_{x=0} = P^*(Q^*,x)|_{x=0}\,.$$

Using the initial conditions (12) and equations (18), this reduces to

$$a^*_{p-2}(0)\left(f^{(p-2)}(0) - Q^{*(p-2)}(0)\right) = 0\,,$$

which implies, using (19),

$$Q^{*(p-2)}(0) = f^{(p-2)}(0)\,.$$

For $i = 1$ in equation (17), using this result, the initial conditions (12) and equations (18), we obtain

$$\left(a^*_{p-2}(0) + a^{*(1)}_{p-1}(0)\right)\left(f^{(p-1)}(0) - Q^{*(p-1)}(0)\right) = 0\,.$$

By (19), $a^*_{p-2}(0) \neq 0$, $a^*_{p-1}(0) \neq 0$, and so the first term in this product is non-zero unless $a^*_{p-2}(0) + a^{*(1)}_{p-1}(0) = 0$. But this latter possibility corresponds to the solution $i = 1$ in the auxiliary equation (14). Hence, either from the above equation, or from the auxiliary condition (13) we have

$$Q^{*(p-1)}(0) = f^{(p-1)}(0)\,.$$

In general we obtain from equation (17)

$$\left[P^*\left(f^{(i)}(x),x\right) + \binom{i}{1}\frac{\partial P^*\left(f^{(i-1)}(x),x\right)}{\partial x} + \binom{i}{2}\frac{\partial^2 P^*\left(f^{(i-2)}(x),x\right)}{\partial x^2} + F_i\right]\Bigg|_{x=0} = 0$$

and

$$\left[P^*\left(Q^{*(i)}(x),x\right) + \binom{i}{1}\frac{\partial P^*\left(Q^{*(i-1)}(x),x\right)}{\partial x} + \binom{i}{2}\frac{\partial^2 P^*\left(Q^{*(i-2)}(x),x\right)}{\partial x^2} + G_i\right]\Bigg|_{x=0} = 0$$

for $i = 2(1)R-1$,

where $F_i = F_i(x, f(x), f^{(1)}(x), \ldots, f^{(p+i-3)}(x))$,
and $G_i = F_i(x, Q^*(x), Q^{*(1)}(x), \ldots, Q^{*(p+i-3)}(x))$.
The coefficients of $f^{(p+i)}(x)$ and $f^{(p+i-1)}(x)$ vanish by (18), and hence, since the induction step implies that since

$$Q^{*(j)}(0) = f^{(j)}(0) \qquad \text{for } j = 0(1)(p+i-3)\,,$$

these equations imply

$$\left(a_{p-2}^{*}(0)+ia_{p-1}^{*(1)}(0)+\frac{i(i-1)}{2}\,a_{p}^{*(2)}(0)\right)\left(f^{(p+i-2)}(0)-Q^{k(p+i-2)}(0)\right)=0\,,$$

$$\text{for } i=2(1)R-1\,.$$

Either this value of i satisfies the auxiliary equation (14) (in which case we use the auxiliary condition (13)) or else the expression in the first backet is non-zero. In either case we deduce

$$Q^{*(p+i-2)}(0)=f^{(p+i-2)}(0)\qquad,\ i=2(1)R-1\,.$$

It follows that

$$Q^{*(j)}(0)=f^{(j)}(0)\qquad,\ j=0(1)R+p-3\,,$$

and hence

$$Q^{*}(x)=f(x)+O\left(x^{R+p-2}\right)=f(x)+O\left(x^{R+p-D}\right).$$

(c) The general case follows in a similar fashion.
For $i=0$ in equation (17)

$$P^{*}(f,x)|_{x=0}=P^{*}(Q^{*},x)|_{x=0}\,.$$

Using the initial conditions (12) and equations (18), this reduces to

$$a_{p-D}^{*}(0)\left(f^{(p-D)}(0)-Q^{*(p-D)}(0)\right)=0\,,$$

which implies, using (19)

$$Q^{*(p-D)}(0)=f^{(p-D)}(0)\,.$$

Also from equation (17)

$$\left[\sum_{j=0}^{i}\binom{i}{j}\frac{\partial^{j}P^{*}\left(f^{(i-j)}(x),x\right)}{\partial x^{j}}\right]\Bigg|_{x=0}=\left[\sum_{j=0}^{i}\binom{i}{j}\frac{\partial^{j}P^{*}\left(Q^{*(i-j)}(x),x\right)}{\partial x^{j}}\right]\Bigg|_{x=0}$$

$$\text{for } i=1(1)D-1\,.$$

Using the result for $i=0$, the initial conditions (12) and equations (18), this reduces to

$$\left[\sum_{j=0}^{i}\binom{i}{j}a_{p-D+i}^{*(j)}(0)\right]\left(f^{(p+i-D)}(0)-Q^{*(p+i-D)}(0)\right)=0,\quad i=1(1)D-1\,.$$

If the first term of this product is zero for say $i=k$, then this corresponds to a solution $i=k$ in the auxiliary equation (14). Hence, either because the first term is non-zero, or by using the auxiliary condition for $i=k$, we obtain

$$Q^{*(p+i-D)}(0)=f^{(p+i-D)}(0)\,,\quad \text{for } i=0(1)D-1\,.$$

In general, we obtain from equation (17)

$$\left[\sum_{j=0}^{i}\binom{i}{j}\frac{\partial^j P^*\left(f^{(i-j)}(x),x\right)}{\partial x^j}+F_i\right]\Bigg|_{x=0}=\left[\sum_{j=0}^{i}\binom{i}{j}\frac{\partial^j P^*\left(Q^{*(i-j)}(x),x\right)}{\partial x^j}+G_i\right]\Bigg|_{x=0}$$
$$\text{for } i=D(1)R-1\,,$$

where $F_i = F_i(x, f(x), f^{(1)}(x), \ldots, f^{(p+i-D-1)}(x))$,
and $G_i = F_i(x, Q^*(x), Q^{*(1)}(x), \ldots, Q^{*(p+i-D-1)}(x))$.
The coefficients of $f^{(p+k)}(x)$, $k = (i - D + 1)(1)i$, vanish by (18) and hence the induction step implies that since

$$Q^{*(j)}(0) = f^{(j)}(0) \quad , \quad j = 0(1)(p + i - D - 1)\,,$$

these equations imply

$$\left[\sum_{j=0}^{D}\binom{i}{j}a^{*(j)}_{p-D+j}(0)\right]\left(f^{(p+i-D)}(0)-Q^{*(p+i-D)}(0)\right)=0\,,\quad i=D(1)R-1\,.$$

Either this value of i satisfies the auxiliary equation (14) (in which case we use the auxiliary condition (13)), or else the expression in the first term of this product is non-zero. In either case we deduce

$$Q^{*(p+i-D)}(0) = f^{(p+i-D)}(0)\,,\quad i = D(1)R-1\,.$$

It follows that

$$Q^{*(j)}(0) = f^{(j)}(0)\,,\quad j = 0(1)(R-p-D-1)\,,$$

and hence

$$Q^*(x) = f(x) + O\left(x^{R+p-D}\right).$$

□

A similar result was achieved by an entirely different method in [2, eqn. (5.14)] even though equations (5.12) and (5.13) appear to be in error.

Corollary 21.

For an integral form satisfying the conditions of Theorem 20, the deficiency, D, satisfies $D \le p$, and the order of approximation by an integral function will always be at least R. That is, at worst, $Q^*(x)$ satisfies

$$Q^*(x) = f(x) + O\left(x^R\right).$$

Proof:

The integral form $P^*(f,x)$ satisfies $[\sum_{i=-1}^{p}|a_i(x)|\,]|_{x=0} \neq 0$. The hypotheses of the theorem include

$$\mathrm{Ord}\left(\frac{\partial P^*(f,x)}{\partial f^{(p-k)}}\right) = D-k\,,\quad k = 0(1)D\,.$$

For $D = p$ this means that

$$\operatorname{Ord}\left(\frac{\partial P^*(f,x)}{\partial f}\right) = 0 \quad \text{for} \quad k = D\,.$$

That is $a_0(x) = O(1)$ and $a_0(x) \neq O(x)$. Thus $a_0(0) \neq 0$.

If $D > p$ then $a_0(x)$ would have a factor of x (and hence $a_{-1}(0) = 0$ since $P^*(f,x)|_{x=0} = 0$), which contradicts the condition above that not all coefficients $a_k(x)$ vanish at the origin. If there were a common factor of $x^r, r > 0$, for all coefficients, then Theorem 19 would apply.

The techniques used in the above proofs have been modelled after the approach in [9,11] for algebraic functions. However, viewed as a method for the solution of differential equations, the approach used is similar to the method of Frobenius [8] for the series solution of linear homogeneous differential equations. Thus the integral form is normal if the origin is an ordinary point of the associated integral function. The deficiency of the solution corresponds to the origin being a regular singular point. However it is relevant to note that we are not concerned with finding all solutions of the differential equation, but only with choosing from the solution space that solution which approximates the given locally analytic function. Hence we seek only analytic solutions even though other solutions may have singularities at the origin. The auxiliary equation (14) is clearly closely related to the indicial equation [8,13].

There may be other singular points away from the origin as was the case for algebraic functions. For quadratic algebraic functions, a preliminary analysis of the behaviour at such points has been undertaken and more details will be given in forthcoming reports.

The fundamental process for the order of approximation of a non-normal integral function may be summarized by the following theorem.

Theorem 22.

Let S be the surplus and D be the deficiency of the $\mathbf{n}$ integral form of degree p, $P^*_{\mathbf{n}}(f,x)$.

Let $P^*_{\mathbf{n}}(f,x) = x^r P^*_{\mathbf{m}}(f,x)$, r a non-negative integer, where $P^*_{\mathbf{m}}(f,x)$ has polynomial coefficients $a^*_i(x)$ which satisfy

(i) $\left[\sum_{i=-1}^{p} |a^*_i(x)|\right]\Big|_{x=0} \neq 0\,,$

(ii) $\operatorname{Ord}\left(\dfrac{\partial P^*_{\mathbf{n}}(f,x)}{\partial f^{(p-k)}}\right) = D_1 - k \quad \text{for} \quad k = 0(1)D_1\,,$ where $D(\mathbf{m}) = D_1$.

Then the associated $\mathbf{n}$ integral function of degree p, $Q^*(x)$, defined by the equation

$$P^*_{\mathbf{n}}(Q^*, x) = 0$$

subject to the initial conditions

$$Q^{*(i)}(0) = f^{(i)}(0)\,, \quad i = 0(1)(p-1-D_1)\,,$$

and the auxiliary conditions

$$Q^{*(p+i-D_1)}(0) = f^{(p+i-D_1)}(0)\,,$$

where $i \in \{1,2,\ldots,R-1\}$ satisfies the auxiliary equation

$$\sum_{j=0}^{D_1} \binom{i}{j} a^{*(j)} \quad (0) = 0 .$$

is an approximation to $f(x)$ satisfying

$$Q^*(x) = f(x) + O\left(x^{N+S+p-D}\right) .$$

Proof:

By Lemma 18, $D = D(\mathrm{m}) + r = D_1 + r$. By combining the results of Theorems 19 and 20 the result is obtained. □

This section has developed theorems for the basic behaviour of the order of approximation by an n integral function of degree p, determined by collocation at the single node $x = 0$. A simple change of variable to $x - x_0$ will generalize these results to collocation at an arbitrary node $\{x_0\}$. It is interesting to observe the close analogy of these results with the case of approximation by algebraic functions [9,11]. For algebraic functions it was found that the order of approximation was determined by the expression $N + S - D$, while for the case of integral functions of degree p it has been shown that the order of approximation is determined by the expression $N + S + p - D$. It is conjectured that these basic results will also hold true for the cases of more general deficiencies than the basic case that was considered in this paper. These details will be considered in a future report.

§ 4. Examples

This section contains some examples which illustrate the results of the previous sections.

Example 23.

This is the same as Example 13.

The $(0,0,0)$ integral form for $f(x) = (x+2)^3 - 3(x+1)^2 + 2$ is

$$1 - f(x) + f^{(1)}(x) = O(x^3) .$$

This form has $p = 1, N = 2, S = 1$. The integral form is normal since $\partial P^*(f,x)/\partial f^{(1)}|_{x=0} = 1 \neq 0$.

The corresponding $(0,0,0)$ integral function $Q^*(x)$ is defined by

$$1 - Q^*(x) + Q^{*(1)}(x) = 0 ,$$

subject to the initial condition

$$Q^*(0) = f(0) = 7 .$$

The solution to this initial value problem is

$$\begin{aligned} Q^*(x) = 1 + 6e^x &= 7 + 6x + 3x^2 + x^3 + O(x^4) \\ &= f(x) + O(x^4) \\ &= f(x) + O\left(x^{N+S+p}\right) , \end{aligned}$$

which illustrates Theorem 16.

Note that $Q^*(x)$ is also the integral function approximation of types $(1,0,0),(0,1,0),(0,0,1)$ since N increases by 1 and S decreases by 1 in these three cases.

Example 24.

This is the same as Example 14 in Section 2.

(a) The $(1,1,1)$ integral form of degree 1 for $f(x) = \exp(x^6)$ is

$$P^*(f,x) \equiv 6 - 6f(x) + xf^{(1)}(x) = O(x^{12}) \,.$$

$N = 5$ for this form and hence $S = 7$. In this case $\partial P^*(f,x)/\partial f^{(1)} = x$ and $\partial P^*(f,x)/\partial f = -6$, so that the hypotheses of Theorem 20 are satisfied with $D = 1$. The auxiliary equation (14) is

$$-6 + i = 0 \,,$$

which has the integer solution in [1,11] of $i = 6$. Thus we have the auxiliary condition (13) for $i = 6$, and since $p = D = 1$ there are no initial conditions in this case.

The $(1,1,1)$ integral function satisfies

$$6 - 6Q^*(x) + xQ^{*(1)}(x) = 0$$

subject to the auxiliary condition

$$Q^{*(6)}(0) = f^{(6)}(0) = 6!$$

Thus $$\begin{aligned} Q^*(x) &= 1 + x^6 \\ &= f(x) + O(x^{12}) \\ &= f(x) + O(x^{N+S+p-D}) \end{aligned}$$

which illustrates Theorem 20.

(b) As in Example 14, the $(2,2,2)$ integral form of degree $p = 1$ is

$$6x - 6xf(x) + 6x^2 f^{(1)}(x) = O(x^{13}) \,,$$

with $N = 8$ and $S = 5$. In this case Theorem 22 applies with $r = 1$ and $D_1 = D(\mathbf{m}) = 1$. The auxiliary equation is again

$$-6 + i = 0$$

and the $(2,2,2)$ integral function $Q^*(x)$ satisfies

$$6 - 6Q^*(x) + xQ^{*(1)}(x) = 0$$

subject to the auxiliary condition

$$Q^{*(6)}(0) = f^{(6)}(0) = 6!$$

Thus $$\begin{aligned} Q^*(x) &= 1 + x^6 \\ &= f(x) + O(x^{12}) \\ &= f(x) + O(x^{N+S+p-D}) \end{aligned}$$

since $D = D_1 + r = 1 + 1 = 2$.

The results for the other types mentioned in Example 14 follow similarly.

Example 25. [2, p361]

Let $p = 1$ and $f(x) = (2/x)(1 - \sqrt{1-x}) + x^5$. The $(0, 1, 2)$ integral form is

$$2 + (-2 + x)f(x) + (-2x + 2x^2)f^{(1)}(x) = O(x^5).$$

Thus $N = 5$ and $S = 0$ for this integral form.

Now $\partial P^*(f, x)/\partial f^{(1)}|_{x=0} = 0$ so the form is not normal. The coefficient polynomials do not have a common factor of x and since

$$\begin{aligned} &a_1^*(0) = 0 \quad , \ a_1^{*(1)}(0) = -2, \\ &a_0^*(0) = -2, \end{aligned}$$

the hypotheses of Theorem 20 are satisfied with $D = 1$. The auxiliary equation (14) is thus

$$-2 - 2i = 0$$

which has the root $i = -1 \notin [1, 4]$.

The $(0, 1, 2)$ integral function which is analytic at the origin satisfies

$$2 + (-2 + x)Q^*(x) + (-2x + 2x^2)Q^{*(1)}(x) = 0$$

which has the series solution

$$\begin{aligned} Q^*(x) &= 1 + \frac{1}{4}x + \frac{1}{8}x^2 + \frac{5}{64}x^3 + \frac{7}{128}x^4 + \frac{21}{512}x^5 + O(x^6) \\ &= f(x) + O(x^5) \\ &= f(x) + O\left(x^{N+S+p-D}\right), \end{aligned}$$

which agrees with Theorem 20. The fact that another solution to the differential equation has a singularity at the origin is irrelevant to our objective of seeking an approximation to a function which is locally analytic at the origin. □

Example 26.

This continues Example 7 in Section 2.

The $(2, 2, -1, 0)$ integral form for $f(x) = \exp(\cos(x) - 1)$ is

$$312 - 25x^2 - (276 + 185x^2)f(x) + 36f^{(2)}(x) = O(x^8)$$

with $N = 6, p = 2$ and $S = 2$.

Since $\partial P^*(f, x)/\partial f^{(2)}|_{x=0} = 36 \neq 0$, the integral form is normal and $D = 0$.

The $(2, 2, -1, 0)$ integral function satisfies

$$312 - 25x^2 - (276 + 185x^2)Q^*(x) + 36Q^{*(2)}(x) = 0$$

subject to the initial conditions

$$\begin{aligned} Q^*(0) &= f(0) = 1, \\ Q^{*(1)}(0) &= f^{(1)}(0) = 0. \end{aligned}$$

Thus the unique $(2,2,-1,0)$ integral function is

$$\begin{aligned}Q^*(x) &= 1 - \frac{1}{2}x^2 + \frac{1}{6}x^4 - \frac{31}{720}x^6 + \frac{379}{40320}x^8 - \frac{54139}{32659200}x^{10} + O(x^{12}) \\ &= f(x) + O(x^{10}) \\ &= f(x) + O\left(x^{N+S+p-D}\right)\end{aligned}$$

in agreement with Theorem 16.

This is also the $(3,2,-1,0), (2,3,-1,0), (2,2,0,0)$ and $(2,2,-1,1)$ integral function approximation to $f(x)$ since N increases by 1 and S decreases by 1 in these cases. According to Theorem 12, it is also the integral function approximation of the types $(3,3,-1,0), (3,2,0,0), (3,2,-1,1), (2,3,0,0), (2,3,-1,1)$ and $(2,2,0,1)$ since $N = 8, S = 0$ for these cases. However, as noted in the remarks after Theorem 12, the integral form of the type $(4,2,-1,0)$ will have an overlapping structure and hence another linearly independent solution to (2). Consequently there will be a different integral function approximation of higher order for this type. □

Example 27.

This continues Example 10 in Section 2.

The $(0,0,-1,0,-1,0,-1,0)$ integral form for $f(x) = \exp(\cos(x) - 1)$ is

$$-348 + 589f(x) + 378f^{(2)}(x) + 42f^{(4)}(x) + f^{(6)}(x) = O(x^8),$$

with $N = 4, p = 6$ and $S = 4$.

Since $\partial P^*(f,x)/\partial f^{(6)}|_{x=0} = 1 \neq 0$, this integral form is normal and $D = 0$.

The $(0,0,-1,0,-1,0,-1,0)$ integral function satisfies

$$-348 + 589Q^*(x) + 378Q^{*(2)}(x) + 42Q^{*(4)}(x) + Q^{*(6)}(x) = 0$$

subject to the initial conditions

$$\begin{aligned}Q^*(0) &= f(0) = 1, \\ Q^{*(1)}(0) &= f^{(1)}(0) = 0, \\ Q^{*(2)}(0) &= f^{(2)}(0) = -1, \\ Q^{*(3)}(0) &= f^{(3)}(0) = 0, \\ Q^{*(4)}(0) &= f^{(4)}(0) = 4, \\ Q^{*(5)}(0) &= f^{(5)}(0) = 0.\end{aligned}$$

Thus the unique $(0,0,-1,0,-1,0,-1,0)$ integral function is

$$\begin{aligned}Q^*(x) &= 1 - \frac{1}{2}x^2 + \frac{1}{6}x^4 - \frac{31}{720}x^6 + \frac{379}{40320}x^8 - \frac{1639}{907200}x^{10} \\ &\qquad + \frac{150349}{479001600}x^{12} - \frac{4059721}{87178291200}x^{14} + \cdots \\ &= f(x) + O(x^{14}) \\ &= f(x) + O\left(x^{N+S+p-D}\right)\end{aligned}$$

in agreement with Theorem 16.

Following Theorem 12, this function is also the integral function of degree 6 approximation of a large number of additional types. Examples of these types are $(1,1,-1,1,-1,1,-1,0), (0,0,0,0,0,0,0\ 1)$, and $(0,0,-1,0,0,1,0\ 1)$. However there are also a number of overlapping structures in this example and additional analysis would be needed for these cases. □

Example 28. [2, p368]

In the example in [2] there was no function $f(x)$ given to approximate. Let

$$f(x) = 1 - \frac{1}{16}x^2 - \frac{1}{384}x^4 - \frac{1}{9216}x^6 - \frac{1}{147456}x^8 + x^{10} + \cdots$$

Let $p = 2$ and consider the $(1,1,0,1)$ integral form. Since $N = 6$ and the coefficients of f up to $f_{N+p-1} = f_7$ are required to determine the integral form, the portion of the series for $f(x)$ given above will be sufficient to determine an integral form of this type.

$$P^*(f,x) \equiv -2x + xf(x) - 9f^{(1)}(x) + xf^{(2)}(x) = O(x^9)\,.$$

Hence $S = 3$. In this case $\partial P^*(f,x)/\partial f^{(2)} = x$ and $\partial P^*(f,x)/\partial f^{(1)} = -9$ so that the hypotheses of Theorem 20 are satisfied with $D = 1$. The auxiliary equation (14) is

$$-9 + i = 0$$

which has the solution $i = 9 \notin [1,8]$. Hence there are no auxiliary conditions and since $p = 2, D = 1$, there is just one initial condition.

Thus the $(1,1,0,1)$ integral function satisfies

$$-2x + xQ^*(x) - 9Q^{*(1)}(x) + xQ^{*(2)}(x) = 0$$

subject to the initial condition

$$Q^*(0) = f(0) = 1\,.$$

The series solution is

$$\begin{aligned} Q^*(x) &= 1 - \frac{1}{16}x^2 - \frac{1}{384}x^4 - \frac{1}{9216}x^6 - \frac{1}{147456}x^8 + O(x^{10}) \\ &= f(x) + O(x^{10}) \\ &= f(x) + O\left(x^{N+S+p-D}\right), \end{aligned}$$

in agreement with Theorem 20. Note that the other solution to the differential equation has a logarithmic singularity and so we are not concerned with that solution when seeking an analytic approximation. □

§ 5. Conclusion

This paper has considered the problem of approximating a real-valued, locally analytic function, $f(x)$, by an integral function, $Q(x)$, which is derived from an inhomogeneous differential equation. Following the analogous procedure in [9,11] for algebraic functions, a careful distinction was made between the approximating properties of the integral form, by which the polynomial coefficients of the integral form are defined, and the approximating properties of the integral function itself.

By using an elimination procedure when the space of solutions for the integral form has dimension greater than 1, a unique integral form of maximal order was identified in all cases. This form may well have an order greater than N and the concept of the surplus was defined to measure this additional order. A basic block structure within the table of integral forms was identified, but the global nature of such a table is complicated by the possibility of overlapping structures.

A normal integral form of degree p (which corresponds to the differential equation having the origin as an ordinary point) has deficiency $D = 0$, and by imposing the appropriate initial conditions a unique element in the solution space of the corresponding inhomogeneous linear differential equation may be identified with specified approximating properties. For a non-normal integral form of degree p (with deficiency $D > 0$) it was shown that a unique element in the solution space of the differential equation may still be determined with order of approximation given by the expression $N + S + p - D$. This may be compared to the case of algebraic functions [9,11] where it was shown that the order of approximation is determined by the expression $N + S - D$.

Although these results have been obtained for a function locally analytic at the origin, it is clear that a simple change of variable to $x - x_0$ will generalize the results for a function which is locally analytic at an arbitrary point x_0.

The important feature of this paper is that a clear formulation of the problem in this way leads to the existence and uniqueness of an approximating integral form of maximal order, and to the existence and uniqueness of a distinguished integral function (solution of the inhomogeneous differential equation) which is an approximation to the given function with a specified order.

The results obtained have been illustrated by some simple examples in Section 4. Further examples in the case of integral functions derived from a homogeneous differential equation are given in [10].

Acknowledgement

The author would like to acknowledge the hospitality of the Center for Applied Mathematics, Purdue University, where this research was initiated.

References

1. G.A. Baker Jr. and P.R. Graves-Morris, "Padé Approximants, Part I: Basic Theory, Part II: Extensions and Applications", Addison-Wesley, Reading, MA, 1981.
2. G.A. Baker Jr. and P.R. Graves-Morris, Definition and uniqueness of integral approximants, *J. Comput. Appl. Math.* **31** (1990), 357–372.
3. G.A. Baker Jr. and D.S. Lubinsky, Convergence theorems of rows of differential and algebraic Hermite-Padé approximations, *J. Comput. Appl. Math.* **18** (1987), 29–52.
4. R.G. Brookes and A.W. McInnes, Structure in the Hermite-Padé table, Research Report 49, Department of Mathematics, University of Canterbury (May, 1989).
5. R.G. Brookes and A.W. McInnes, The existence and local behaviour of the quadratic function approximation, *J. Approx. Theory* **62** (1990), 383–395.
6. J. Della Dora and C. di Crescenzo, Approximations de Padé-Hermite, *Numer. Math.* **43** (1984), 23–57.
7. D.L. Hunter and G.A. Baker Jr., Methods of series analysis III. Integral approximant methods, *Phys. Rev. B.* **19** (1979), 3808–3821.
8. E.L. Ince, "Ordinary Differential Equations", Dover, New York, 1956.
9. A.W. McInnes, The existence of approximations by algebraic functions, Technical Report 96, Center for Applied Mathematics, Purdue University (April, 1989).
10. A.W. McInnes, Existence and uniqueness of integral function approximations, Research Report 52, Department of Mathematics, University of Canterbury (October, 1989); submitted to *Numer. Math.*
11. A.W. McInnes, Existence and uniqueness of algebraic function approximations, *Constr. Approx.* (to appear).
12. J. Nuttall, Asymptotics of diagonal Hermite-Padé polynomials, *J. Approx. Theory* **42** (1984), 299–386.
13. E.D. Rainville and P.E. Bedient, "Elementary Differential Equations", 7th ed., Macmillan, New York, 1989.
14. H. Stahl, Asymptotics of Hermite-Padé polynomials and related convergence results – a summary of results, in "Nonlinear Numerical Mathods and Rational Approximation" (A. Cuyt, Ed.), pp. 25–53, Reidel, Dordrecht, 1988.