

**RESEARCH ARTICLE**

# An algorithm for two-variable rational interpolation suitable for matrix manipulations with the evaluation-interpolation method

Katerina G. Hadjifotinou\* | Nicholas P. Karampetakis

Dept. of Mathematics, Aristotle Univ. of  
Thessaloniki, 54124 Thessaloniki, Greece

**Correspondence**

Katerina G. Hadjifotinou, Dept. of  
Mathematics, Aristotle Univ. of  
Thessaloniki, 54124 Thessaloniki, Greece.  
Email: khad@math.auth.gr

An algorithm for two-variable rational interpolation is developed. The algorithm is suitable for interpolation cases where neither the number of interpolation points to be used nor the final degrees of the rational interpolant are known a priori. Instead, a maximum degree for the interpolant's numerator and denominator is assumed, and, by testing the condition number of the interpolation system's matrix at each step, the necessary reductions are made so as to cope with non-normality and unattainability occasions. The algorithm can be used for applications of the Evaluation-Interpolation technique in matrix manipulations, such as finding the inverse of a matrix with elements rational functions in two variables. The algorithm avoids completely symbolic calculations, thus keeping the execution time very low even if the system size is large, and achieves accurate function recoveries for greater polynomial degrees than other bivariate rational interpolation methods.

**KEYWORDS:**

bivariate rational interpolation, evaluation-interpolation method

## 1 | INTRODUCTION

The problem of two-variable rational interpolation has been addressed by many authors in the recent years (see<sup>1</sup> for a thorough review). However, in all research done so far, it is assumed that the degrees of the numerator and denominator of the wanted rational interpolant are predefined - e.g. determined by the available number of interpolation points. However, there are cases where neither the number of interpolation points to be used nor the final degrees of the rational interpolant are known a priori. In these cases we have to assume a maximum degree for the interpolant's numerator and denominator, and proceed with interpolation, reducing this degree if necessary. Such an example is the application of the Evaluation-Interpolation technique in matrix manipulations, widely used in Control Theory (see e.g.<sup>2,3</sup>). If for example we want to calculate the inverse of a matrix with

elements rational functions in two variables without the use of symbolic calculations, we can arrive to the solution doing only numerical calculations, by means of interpolating at numerical “instances” of the solution. That is, we first calculate numerical instances of the matrix inverse for specific data points and then interpolate the analytic solution at these points. In order to implement this technique, the first thing to be calculated is the maximum global degree of the numerator and denominator of each element of the wanted inverse matrix. The initial number of needed interpolation points will be determined by this degree, but as the interpolation procedure develops, it may come out that fewer interpolation points are actually needed for the calculation of each symbolic element of the inverse matrix.

In the sections that follow: a) we define the problem of two-variable rational interpolation, b) we develop the Algorithm of Successive Reductions, which starts with a maximum interpolant degree and successively reduces the number of interpolation points used, in order to arrive to the correct rational interpolant for the specific problem, c) we apply this algorithm to the calculation of the inverse of a matrix with elements rational functions in two variables using the Evaluation-Interpolation technique and d) we compare the efficacy of the Algorithm of Successive Reductions with that of the recently developed Bulirsch-Stoer bivariate rational interpolation algorithm<sup>4</sup>.

## 2 | THE TWO-VARIABLE RATIONAL INTERPOLATION PROBLEM

Extending the canonical formularity (see e.g.<sup>5</sup>) of the univariate case to two dimensions, the rational interpolation problem in two-variables can be defined as follows:

Let  $f(x, y)$  be an unknown function of  $x, y$ . We need to determine a rational function of  $x, y$

$$r(x, y) = p(x, y)/q(x, y)$$

where  $p(x, y)$  and  $q(x, y)$  are polynomials of  $x, y$ , of priorly unknown global degrees  $n_p$  and  $n_q$  respectively. The rational function  $r(x, y)$  must be such, that, for a specific number  $M$  of interpolation points  $(x_i, y_i, f(x_i, y_i))$ , it holds

$$r(x_i, y_i) = f_i, \quad i = 1, \dots, M,$$

where for reasons of simplicity we denote  $f(x_i, y_i)$  with  $f_i$ . Then,  $r(x, y)$  interpolates  $f(x, y)$  at the given points. The number of interpolation points  $M$  and the global degrees  $n_p$  and  $n_q$  are directly related. In fact, if we focus on finding a rational function  $r(x, y) = p(x, y)/q(x, y)$  with  $p(x, y)$ ,  $q(x, y)$  of global degree at most  $n$  in  $x, y$ , then  $M$  can be specified as follows:

$p(x, y)$ ,  $q(x, y)$  can be presented as

$$\begin{aligned}
 p(x, y) &= \sum_{i=0}^n \sum_{j=0}^{n-i} a_{i,j} x^i y^j \\
 &= a_{0,0} + a_{0,1} y + \dots + a_{0,n-1} y^{n-1} + a_{0,n} y^n \\
 &\quad + a_{1,0} x + a_{1,1} xy + \dots + a_{1,n-1} xy^{n-1} \\
 &\quad \dots \\
 &\quad + a_{n-1,0} x^{n-1} + a_{n-1,1} x^{n-1} y \\
 &\quad + a_{n,0} x^n
 \end{aligned} \tag{1}$$

$$\begin{aligned}
 q(x, y) &= \sum_{i=0}^n \sum_{j=0}^{n-i} b_{i,j} x^i y^j \\
 &= b_{0,0} + b_{0,1} y + \dots + b_{0,n-1} y^{n-1} + b_{0,n} y^n \\
 &\quad + b_{1,0} x + b_{1,1} xy + \dots + b_{1,n-1} xy^{n-1} \\
 &\quad \dots \\
 &\quad + b_{n-1,0} x^{n-1} + b_{n-1,1} x^{n-1} y \\
 &\quad + b_{n,0} x^n.
 \end{aligned} \tag{2}$$

Therefore, the number of coefficients to be specified is  $(n+1)(n+2)/2$  for the numerator  $p(x, y)$  and the same for the denominator  $q(x, y)$ , so the total number of coefficients is:

$$N = (n+1)(n+2).$$

Since for the interpolation points  $(x_i, y_i, f_i)$   $i = 1, \dots, M$  it must be

$$r(x_i, y_i) = f_i, \quad i = 1, \dots, M,$$

this leads to the relation

$$p(x_i, y_i) - f_i q(x_i, y_i) = 0, \quad i = 1, \dots, M. \tag{3}$$

Substituting (1) and (2) to (3) we get

$$a_{0,0} + a_{0,1} y_i + \dots + a_{0,n} y_i^n + \dots + a_{n,0} x_i^n - f_i b_{0,0} - f_i b_{0,1} y_i - \dots - f_i b_{0,n} y_i^n - \dots - f_i b_{n,0} x_i^n = 0, \quad i = 1, \dots, M. \tag{4}$$

The system (4) is a homogeneous system of  $M$  equations with  $N$  unknowns. Let us denote the system's matrix as

$$B = \begin{bmatrix} 1 & \cdots & y_1^n & \cdots & x_1 y_1^{n-1} & \cdots & x_1^n & -f_1 & -f_1 y_1 & \cdots & -f_1 x_1^n \\ 1 & \cdots & y_2^n & \cdots & x_2 y_2^{n-1} & \cdots & x_2^n & -f_2 & -f_2 y_2 & \cdots & -f_2 x_2^n \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \cdots & y_M^n & \cdots & x_M y_M^{n-1} & \cdots & x_M^n & -f_M & -f_M y_M & \cdots & -f_M x_M^n \end{bmatrix} \quad (5)$$

In order to be able to obtain a non-trivial solution of (4) for any set of interpolation points, it must be  $M < N$ . If we use for example  $M = N - 1$  interpolation points, then we can normalize the rational function  $r(x, y)$  by assuming  $b_{0,0} = 1$  and thus we obtain the  $(N - 1) \times (N - 1)$  non-homogeneous system

$$\begin{bmatrix} 1 & \cdots & y_1^n & \cdots & x_1 y_1^{n-1} & \cdots & x_1^n & -f_1 y_1 & \cdots & -f_1 x_1^n \\ 1 & \cdots & y_2^n & \cdots & x_2 y_2^{n-1} & \cdots & x_2^n & -f_2 y_2 & \cdots & -f_2 x_2^n \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \cdots & y_{N-1}^n & \cdots & x_{N-1} y_{N-1}^{n-1} & \cdots & x_{N-1}^n & -f_{N-1} y_{N-1} & \cdots & -f_{N-1} x_{N-1}^n \end{bmatrix} \cdot \begin{bmatrix} a_{0,0} \\ a_{0,1} \\ \vdots \\ a_{n,0} \\ b_{0,1} \\ \vdots \\ b_{n,0} \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_{N-1} \end{bmatrix} \quad (6)$$

However, system (6) has a non-zero determinant and therefore a unique solution only if  $b_{0,0}$  is indeed non-zero. Consider for example the case when the function we want to interpolate is

$$f(x, y) = \frac{x^2 - 1}{x + y}.$$

In this case the interpolant  $r(x, y)$  must have  $b_{0,0} = 0$ .

Nevertheless, there is still solution to the problem. All we need to do, is normalize  $r(x, y)$  by assuming  $a_{0,0} = 1$  instead of  $b_{0,0} = 1$ .

In this case the  $(N - 1) \times (N - 1)$  system would be

$$\begin{bmatrix} y_1 & \cdots & y_1^n & \cdots & x_1 y_1^{n-1} & \cdots & x_1^n & -f_1 & \cdots & -f_1 x_1^n \\ y_2 & \cdots & y_2^n & \cdots & x_2 y_2^{n-1} & \cdots & x_2^n & -f_2 & \cdots & -f_2 x_2^n \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ y_{N-1} & \cdots & y_{N-1}^n & \cdots & x_{N-1} y_{N-1}^{n-1} & \cdots & x_{N-1}^n & -f_{N-1} & \cdots & -f_{N-1} x_{N-1}^n \end{bmatrix} \cdot \begin{bmatrix} a_{0,1} \\ \vdots \\ a_{n,0} \\ b_{0,0} \\ b_{0,1} \\ \vdots \\ b_{n,0} \end{bmatrix} = \begin{bmatrix} -1 \\ \vdots \\ -1 \end{bmatrix} \quad (7)$$

whilst its solution would contain  $b_{0,0} = 0$ .

Of course, there still remains the case when both  $a_{0,0}$  and  $b_{0,0}$  are zero. Let us consider e.g.

$$f(x, y) = \frac{x^2}{x + y}.$$

This case can still be solved if we reduce the size  $N$  of our problem by 2 and repeat the process. Then the number of interpolation points needed, would be  $N - 3$ . In this way we can overpass the problems of unattainability and non-normality of the canonical representation of the system, stated by Salazar Celis<sup>6</sup> as well as other researchers. The procedure described above can be standardized into the following Algorithm of Successive Reductions (ASR).

### 3 | ALGORITHM OF SUCCESSIVE REDUCTIONS (ASR)

#### Step 1

Determine the maximum global degree  $maxn$  of the numerator and the denominator of the wanted interpolant.

#### Step 2

Calculate the initial size of the problem:  $N = (maxn + 1)(maxn + 2)$ .

#### Step 3

Choose  $N - 1$  interpolation points  $(x_i, y_i, f_i)$ . In order to avoid system singularities, not only due to the poles of  $f(x, y)$  but also due to a possible production of two identical columns in the system's matrix, a fairly safe strategy is a random choice of  $x_i, y_i \in (0, 1)$  instead of using e.g. an orthogonal distribution of  $(x_i, y_i)$ .

#### Step 4

Calculate homogeneous system matrix  $B$  as of (5).

#### Step 5

Calculate the reciprocal condition numbers (in 1-norm, see e.g the Matlab function *rcond*)  $c_1$  and  $c_2$  of submatrices  $B_1 = B - \{\text{column } N/2 + 1, \text{ that is, the column of } b_{0,0}\}$  and  $B_2 = B - \{\text{column } 1, \text{ that is, the column of } a_{0,0}\}$  respectively. A matrix is well-conditioned if it's reciprocal condition number is close to 1, while it is badly conditioned if it's reciprocal condition number is close to zero - that is, close to machine accuracy  $\epsilon$ .

If the maximum of  $c_1$  and  $c_2$  is not close to zero, assume the corresponding coefficient ( $b_{0,0}$  or  $a_{0,0}$ ) to be 1 and solve the thus derived non-singular non-homogeneous system (6) or (7) respectively.

Else, reduce system size by 2 removing the first and the  $(N/2 + 1)$ -th columns of matrix  $B$ , together with it's last two rows (since now the number of needed interpolation points is also reduced by 2) and repeat Step 5. In the most extreme scenario, after successive executions of Step 5, the size of the system will be reduced to  $N = 2$  and the interpolant will be a constant.

Step 5 of the above algorithm can be further enriched by an extra procedure that acts on the final non-singular system to be solved: Since we are dealing with polynomials in two variables, it is highly probable that most of the coefficients of the solution are zero, especially in cases of global degrees greater than three. It is well known that a zero element in the solution vector of a non-singular linear system corresponds to linear dependence of the vector of system constants with the rest columns of the system's matrix (that is, if we substitute the system's column that corresponds to the zero element with the vector of system constants, the determinant of the corresponding matrix is zeroed). This property can be used as an extra check, column by column, to further reduce the system size and finally solve the much smaller system of non-zero solution elements. This can be better illustrated in the following example:

If wanting to recover the function

$$f(x, y) = \frac{x}{2y^2}$$

with the only knowledge that the maximum global degree of it's numerator and denominator is  $\max n = 3$ , then the initial size of the problem will be  $N = 20$ . That is, we will have to calculate 20 unknown coefficients and use  $N - 1 = 19$  interpolation points, thus creating a matrix  $B$  of the form (5) with dimensions  $19 \times 20$ . Each column of  $B$  corresponds to an unknown coefficient, the coefficients being ordered as follows:

$$\begin{array}{cccccccccccccccccccccccc} 1 & y & y^2 & y^3 & x & xy & xy^2 & x^2 & x^2y & x^3 & 1 & y & y^2 & y^3 & x & xy & xy^2 & x^2 & x^2y & x^3 \\ a_{0,0} & a_{0,1} & a_{0,2} & a_{0,3} & a_{1,0} & a_{1,1} & a_{1,2} & a_{2,0} & a_{2,1} & a_{3,0} & b_{0,0} & b_{0,1} & b_{0,2} & b_{0,3} & b_{1,0} & b_{1,1} & b_{1,2} & b_{2,0} & b_{2,1} & b_{3,0} \end{array}$$

However, since in our example both coefficients  $a_{0,0}$  and  $b_{0,0}$  are zero, both reciprocal condition numbers  $c_1$  and  $c_2$  of the submatrices  $B_1$  and  $B_2$  described in Step 5 of ASR are zero. As a result, the columns 1 and 11 of  $B$  have to be omitted. The last two lines of  $B$  need also be discarded, since now the number of unknowns, and together the needed interpolation points, is reduced by 2.

Step 5 of ASR is then repeated, examining the reciprocal condition numbers that correspond to the coefficients  $a_{0,1}$  and  $b_{0,1}$ . These are also zero, and Step 5 is repeated again, now with the dimensions of  $B$  being  $15 \times 16$  and it's columns corresponding to:

$$\begin{array}{cccccccccccccccccccc} y^2 & y^3 & x & xy & xy^2 & x^2 & x^2y & x^3 & y^2 & y^3 & x & xy & xy^2 & x^2 & x^2y & x^3 \\ a_{0,2} & a_{0,3} & a_{1,0} & a_{1,1} & a_{1,2} & a_{2,0} & a_{2,1} & a_{3,0} & b_{0,2} & b_{0,3} & b_{1,0} & b_{1,1} & b_{1,2} & b_{2,0} & b_{2,1} & b_{3,0} \end{array}$$

However now  $c_1$  is non-zero, since it corresponds to the non-zero denominator's coefficient  $b_{0,2}$  of  $y^2$ . Therefore we can normalize by assuming that  $b_{0,2} = 1$  and solve the corresponding non-homogeneous uniquely solvable  $15 \times 15$  system.

At this point, we can further simplify our system by successively substituting each column of the system's  $15 \times 15$  matrix with the vector of system constants and examining the reciprocal condition number of the thus derived matrix. In fact, at our example, the 15-element solution vector contains only one non-zero element, that concerning the coefficient  $a_{1,0}$ . Therefore, if we perform this procedure for our example, we will find only one non-zero reciprocal condition number, the one that corresponds to the third column of the reduced  $15 \times 15$  matrix  $B$ . This practically means that only one interpolation point is finally needed in order to reproduce the interpolant.

If, for example, the first of our interpolation points was  $(x_1, y_1) = (2, 1)$  with  $f_1 = 1$ , our final system, after deducting the 14 unnecessary columns of  $B$ , would degenerate into one trivial equation:

$$a_{1,0}x_1 - f_1y_1^2 = 0 \Leftrightarrow 2a_{1,0} = 1$$

which yields  $a_{1,0} = 0.5$ . This coefficient, together with the assumed  $b_{0,2} = 1$  are adequate for reproducing the wanted interpolant as

$$f(x, y) = \frac{0.5x}{y^2}.$$

A complete numerical illustration of all stages of the ASR procedure is presented in the next section, where the algorithm is connected to an application of the Evaluation-Interpolation technique.

#### 4 | APPLICATION IN THE EVALUATION-INTERPOLATION TECHNIQUE: CALCULATION OF THE INVERSE OF A MATRIX WITH ELEMENTS RATIONAL FUNCTIONS IN TWO VARIABLES

Let  $A$  be a square invertible matrix with elements rational functions of  $x, y$ , whose inverse we want to calculate without the use of symbolic calculations. The Evaluation-Interpolation technique (see e.g.<sup>2,3</sup>) uses numerical computations to compute the exact solution by means of interpolating at numerical "instances" of the solution. In our case, we first calculate numerical instances of  $A^{-1}$  for specific data points and then interpolate the analytic solution  $A^{-1}$  at these points. In fact, each element of  $A^{-1}$  will also be a rational function of  $x, y$ . Assigning specific values of  $x_i, y_i$  on  $A$ , we can numerically calculate  $A^{-1}(x_i, y_i)$ . Each element  $A_{r,c}^{-1}$ ,  $r, c = 1, \dots, \text{size}(A)$  of  $A^{-1}(x_i, y_i)$  is actually a value  $f_{i,r,c}$  of the wanted function  $f_{r,c}$  at the point  $(x_i, y_i)$  ( $i = 1, \dots, M$ ).

As a result, ASR must be run as many times as are the elements of  $A$ . In order to implement the algorithm, the first thing to be considered is the maximum global degree of the numerator and denominator of each element of the wanted  $A^{-1}$ . This can be easily done as follows: It is easy to show (e.g.<sup>3</sup>, p. 227) that the maximum degree of the numerator and denominator of each element of  $A^{-1}$  is the corresponding degree of the numerator of  $\det(A)$ . The latter can be calculated as follows:

Let  $Deg_1$  be the matrix of global degrees of all numerators of the elements of  $A$  and  $Deg_2$  be the matrix of global degrees of all denominators.  $\det(A)$  is also a rational function of  $x, y$ . The maximum global degree  $D_2$  of its denominator is the sum of all denominator degrees (that is, the sum of all elements of  $Deg_2$ ). The maximum global degree  $D_1$  of its numerator is  $D_2 + D_m$ ,  $D_m$  being the minimum between the sum of the maximum global degrees of all rows and the sum of the maximum global degrees of all columns of  $Deg_1$  (see e.g.<sup>3</sup>, p. 199). As a result, for each element of  $A^{-1}$  we can take as  $maxn$  (that is, maximum global degrees of its numerator and denominator) the value  $D_1$ . Now we can proceed to Step 2 of the algorithm. Again the initial size of the problem is  $N = (maxn + 1)(maxn + 2)$  and is the same for each element of  $A^{-1}$ . The choice of the  $N - 1$  interpolation points  $(x_i, y_i)$  at step 3 can also be the same for each element. However, the values  $f_i$ ,  $i = 1, \dots, N - 1$  are different for each element as we have explained above.

This is why we need to calculate matrix  $B$  of algorithm's Step 4 from the beginning, for each element of  $A^{-1}$ . Step 5 must also be run for each element separately.

### An example

An example of the complete procedure follows. Let

$$A(x, y) = \begin{bmatrix} \frac{1}{x^2} & \frac{y+3}{x} \\ 1 & 2x \end{bmatrix}$$

be the matrix, with elements rational functions of  $x, y$ , whose inverse we want to calculate. It is easy to find analytically that the determinant of  $A$  is

$$\det(A) = -\frac{y+1}{x}$$

while the wanted inverse is

$$A^{-1}(x, y) = \begin{bmatrix} \frac{-2x^2}{y+1} & \frac{y+3}{y+1} \\ \frac{x}{y+1} & \frac{-1}{xy+x} \end{bmatrix}. \quad (8)$$

Application of ASR for each element of  $A$ :

#### Step 1

The matrix  $Deg_1$  of global degrees of all numerators of  $A$  is:

$$Deg_1 = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$$

while the matrix  $Deg_2$  of global degrees of all denominators of  $A$  is:

$$Deg_2 = \begin{bmatrix} 2 & 1 \\ 0 & 0 \end{bmatrix}$$



The sum of all elements of  $Deg_2$  is

$$D_2 = 3.$$

The minimum between the sum of the maximum global degrees of all rows and the sum of the maximum global degrees of all columns of  $Deg_1$  is

$$D_m = \min\{1 + 1, 0 + 1\} = 1.$$

As a result, the maximum global degree  $D_1$  of the numerator of  $\det(A)$  is

$$D_1 = D_2 + D_m = 4$$

and therefore we take

$$maxn = 4.$$

## Step 2

The initial size of the problem is

$$N = (maxn + 1)(maxn + 2) = 30.$$

## Step 3

We must randomly choose  $N - 1 = 29$  interpolation points  $x_i, y_i \in (0, 1)$ , substitute each of these points in  $A$  and calculate numerically  $f_i = A^{-1}(x_i, y_i)$ . The list of points of an execution of the algorithm is shown in Table 1.

## Step 4

The homogeneous system matrix  $B$  whose size is  $29 \times 30$ , has again to be calculated 4 times, one for each element of  $A^{-1}$ . Indicatively in Table 2 we present matrix  $B_{1,1}$  (that corresponds to the  $(1,1)$ -element of wanted matrix  $A^{-1}$ ) rounded at 1 decimal digit. At the first line of the table, the  $x, y$ -monomials that correspond to each column are shown.

## Step 5

Repeat for each of the four matrices  $B_{i,j}, i, j = 1, 2$ :

Calculate the reciprocal condition numbers  $c_1$  and  $c_2$  of submatrices  $B_{i,j} - \{\text{column } 16\}$  and  $B_{i,j} - \{\text{column } 1\}$  respectively. If the maximum of  $c_1$  and  $c_2$  is not close to zero, assume the corresponding coefficient to be 1 and solve the thus derived non-singular non-homogeneous system of the form (6) or (7). Else, reduce system size by 2, remove the two columns of matrix  $B_{i,j}$  that were tested with the above process to be linearly dependent to the rest, remove also the last two rows of  $B_{i,j}$  and repeat Step 5.

E.g. for  $B_{1,1}$  the complete process of it's reductions is shown in Table 3, where for the calculated reciprocal condition numbers, their order of magnitude is presented.

From Table 3 we notice that, in order to find a reciprocal condition number that is not close to zero, we need to reduce the system size as much as  $N = 12$ . That is, we need to omit columns 1-9 and 16-24 and rows 12-29 of  $B_{1,1}$ . The matrix after these reductions is shown in Table 4 (rounded at two decimal digits).

Since, as shown in Table 3, for  $N = 12$   $c_1$  is non-zero, the coefficient of column 7 of Table 4 can be considered as  $b_{2,0} = 1$ .

As a result, from matrix  $B_{1,1}$  of Table 4 we can obtain the following uniquely solvable  $11 \times 11$  system:

$$\begin{bmatrix} 0.11 & 0.03 & 0.01 & 0.04 & 0.01 & 0.01 & 0.00 & 0.00 & 0.01 & 0.00 & 0.00 \\ 0.10 & 0.10 & 0.10 & 0.03 & 0.03 & 0.01 & 0.01 & 0.01 & 0.00 & 0.00 & 0.00 \\ 0.30 & 0.23 & 0.17 & 0.16 & 0.12 & 0.09 & 0.08 & 0.06 & 0.06 & 0.04 & 0.03 \\ 0.71 & 0.12 & 0.02 & 0.60 & 0.10 & 0.50 & 0.14 & 0.02 & 0.72 & 0.12 & 0.61 \\ 0.82 & 0.09 & 0.01 & 0.74 & 0.08 & 0.67 & 0.13 & 0.01 & 1.09 & 0.11 & 0.98 \\ 0.56 & 0.40 & 0.30 & 0.41 & 0.30 & 0.31 & 0.26 & 0.19 & 0.27 & 0.19 & 0.20 \\ 0.51 & 0.07 & 0.01 & 0.37 & 0.05 & 0.26 & 0.06 & 0.01 & 0.34 & 0.04 & 0.24 \\ 0.20 & 0.10 & 0.05 & 0.09 & 0.05 & 0.04 & 0.03 & 0.01 & 0.02 & 0.01 & 0.01 \\ 0.28 & 0.24 & 0.21 & 0.15 & 0.13 & 0.08 & 0.07 & 0.06 & 0.05 & 0.04 & 0.02 \\ 0.46 & 0.37 & 0.30 & 0.31 & 0.25 & 0.21 & 0.19 & 0.15 & 0.16 & 0.13 & 0.11 \\ 0.28 & 0.27 & 0.26 & 0.15 & 0.14 & 0.08 & 0.08 & 0.07 & 0.04 & 0.04 & 0.02 \end{bmatrix} \cdot \begin{bmatrix} a_{2,0} \\ a_{2,1} \\ a_{2,2} \\ a_{3,0} \\ a_{3,1} \\ a_{4,0} \\ b_{2,1} \\ b_{2,2} \\ b_{3,0} \\ b_{3,1} \\ b_{4,0} \end{bmatrix} = \begin{bmatrix} -0.02 \\ -0.01 \\ -0.10 \\ -0.86 \\ -1.20 \\ -0.36 \\ -0.47 \\ -0.05 \\ -0.09 \\ -0.23 \\ -0.08 \end{bmatrix} \quad (9)$$

while we have taken  $b_{2,0} = 1$  and all the rest of the coefficients that do not appear in (9) are zero.

Before proceeding to finding the unique solution of (9), we can make further reductions to the system by eliminating the columns that correspond to zero elements of the solution vector. In this way we avoid the accumulation of rounding errors that can arise when trying to solve a large system. In fact, we substitute successively each column of the matrix of (9) with the vector of system constants and calculate the reciprocal condition number of the new matrix. The exact process for system (9) is illustrated in Table 5, where the order of magnitude of reciprocal condition numbers is presented for each corresponding column.

Examining Table 5, we can deduce that only coefficients  $a_{4,0}$  and  $b_{2,1}$  of the solution vector of (9) are non-zero. Therefore, after omitting the last 9 rows of data of (9) and the columns 1-5 and 8-11 of the system matrix, we obtain the  $2 \times 2$  solvable system:

$$\begin{bmatrix} 0.01 & 0 \\ 0.01 & 0.01 \end{bmatrix} \cdot \begin{bmatrix} a_{4,0} \\ b_{2,1} \end{bmatrix} = \begin{bmatrix} -0.02 \\ -0.01 \end{bmatrix}$$

whose solution is

$$\begin{bmatrix} -2 \\ 1 \end{bmatrix}.$$

The coefficients  $a_{4,0} = -2$  and  $b_{2,1} = 1$ , together with  $b_{2,0} = 1$  are the only non-zero elements of the 30-element solution vector of the original system and lead to the rational function

$$r(x, y) = \frac{-2x^4}{x^2 + x^2y} = \frac{-2x^2}{y + 1}$$

which is indeed the element (1,1) of  $A^{-1}(x, y)$  (see (8)).

Working the same way, we can find that, for the element (1,2) of  $A^{-1}(x, y)$  the reduced system  $B_{1,2}$  is of size  $N = 6$ , and, after considering  $b_{3,0}$  to be 1, the non-singular non-homogeneous 5x5 system to be solved is (rounded in 4 digits):

$$\begin{bmatrix} 0.0386 & 0.0091 & 0.0131 & -0.0239 & -0.0342 \\ 0.0322 & 0.0316 & 0.0102 & -0.0635 & -0.0205 \\ 0.1646 & 0.1233 & 0.0902 & -0.2642 & -0.1933 \\ 0.5969 & 0.0997 & 0.5026 & -0.2705 & -1.3640 \\ 0.7363 & 0.0773 & 0.6649 & -0.2172 & -1.8683 \end{bmatrix} \cdot \begin{bmatrix} a_{3,0} \\ a_{3,1} \\ a_{4,0} \\ b_{3,1} \\ b_{4,0} \end{bmatrix} = \begin{bmatrix} 0.1011 \\ 0.0646 \\ 0.3528 \\ 1.6200 \\ 2.0690 \end{bmatrix} \quad (10)$$

By testing the reciprocal condition numbers of (10) as done before for system (9), we notice that only the coefficients  $a_{3,0}$ ,  $a_{3,1}$  and  $b_{3,1}$  are non-zero. Thus, system (10) can be reduced to

$$\begin{bmatrix} 0.0386 & 0.0091 & -0.0239 \\ 0.0322 & 0.0316 & -0.0635 \\ 0.1646 & 0.1233 & -0.2642 \end{bmatrix} \cdot \begin{bmatrix} a_{3,0} \\ a_{3,1} \\ b_{3,1} \end{bmatrix} = \begin{bmatrix} 0.1011 \\ 0.0646 \\ 0.3528 \end{bmatrix}$$

which yields the solution

$$\begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}.$$

Therefore, the non-zero elements of the solution vector are  $a_{3,0} = 3$ ,  $a_{3,1} = 1$ ,  $b_{3,1} = 1$  and  $b_{3,0} = 1$ . This means that the wanted rational function is

$$r(x, y) = \frac{3x^3 + x^3y}{x^3 + x^3y} = \frac{y + 3}{y + 1}$$

which is the element (1,2) of  $A^{-1}(x, y)$  (see (8)).

A similar situation holds for the element (2,1) of  $A^{-1}(x, y)$ . The reduced system  $B_{2,1}$  is again of size  $N = 6$ , while the further reduced system of non-zero solution elements (rounded in 4 digits) is:

$$\begin{bmatrix} 0.0131 & -0.0025 \\ 0.0102 & -0.0051 \end{bmatrix} \cdot \begin{bmatrix} a_{4,0} \\ b_{3,1} \end{bmatrix} = \begin{bmatrix} 0.0106 \\ 0.0052 \end{bmatrix} \quad (11)$$

while we have again considered  $b_{3,0} = 1$ .

The solution of system (11) is

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

and the wanted rational function is

$$r(x, y) = \frac{x^4}{x^3 + x^3 y} = \frac{x}{y + 1}$$

which is the element (2,1) of  $A^{-1}(x, y)$  (see (8)).

Finally, for the element (2,2) of  $A^{-1}(x, y)$  it comes out that the reduced system  $B_{2,2}$  is of size  $N = 12$ , while the further reduced system of non-zero solution elements (rounded in 4 digits) is:

$$\begin{bmatrix} 0.0924 & 0.0218 \\ 0.0510 & 0.0502 \end{bmatrix} \cdot \begin{bmatrix} b_{3,0} \\ b_{3,1} \end{bmatrix} = \begin{bmatrix} -0.1142 \\ -0.1011 \end{bmatrix}. \quad (12)$$

This time,  $a_{2,0}$  is considered 1. The solution of system (12) is

$$\begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

and the wanted rational function is

$$r(x, y) = \frac{x^2}{-x^3 - x^3 y} = \frac{-1}{x + xy}$$

which is the element (2,2) of  $A^{-1}(x, y)$  (see (8)).

## 5 | COMPARISON OF ASR WITH THE BULIRSCH-STOER BIVARIATE RATIONAL INTERPOLATION METHOD

Xia *et.al*<sup>4</sup> have recently developped a Bulirsch-Stoer multivariate rational interpolation algorithm. They compared their method to other methods for recovering rational functions such as Thiele-Thiele contined fraction method<sup>7</sup> and the two-variable Löwner matrix method<sup>8</sup> and have found their method significantly superior in terms of execution time. Since the Bulirsch-Stoer method is also suitable for using symbolic calculations and therefore could be used for problems such as finding the inverse of a matrix with elements rational functions in two variables, we have compared the bivariate iterative Bulirsch-Stoer method (BS2v) to the ASR algorithm, using simple bivariate rational functions as the test functions shown in Table 6. In all tested cases, the *maxn* of ASR was the minimum necessary (as was also the case with BS2v). Both algorithms were implemented in Matlab R2016a and run on an Intel Core i5-3470, 3.60GHz CPU, 8GB RAM, Windows 10 Pro x64 computer.

In Table 6 we denote with ‘-’ the cases where the algorithm gave the wrong results. Examining the execution times we can see that, as the maximum global polynomial degree  $n$  increases, the execution time increases significantly at BS2v while this is not the case with ASR. Furthermore, BS2v stops giving accurate results at about  $n = 4$  while the same holds for ASR for  $n = 7$ . Finally, neither the BS2v execution times nor the accuracy did improve when we also tested the LC-version of Bulirsch-Stoer of Xia *et.al*<sup>4</sup>.

## 6 | DISCUSSION

The ASR technique can numerically compute any rational functions, without the need of symbolic arithmetic and without limitations concerning the difference in degrees of numerator and denominator. However, depending on the programming environment used, limitations may arise when the maximum global polynomial degree  $n$  becomes large enough so that all reciprocal condition numbers are computed to be near zero, without the numerical ability to distinguish between real linear dependence cases. This is mainly due to the fact that, as shown by Gautschi and Inglese<sup>9</sup>, the condition number of Vandermonde-like matrices grows exponentially with  $n$ . In practice, for ASR, problems arise around  $n = 7$ , while other methods may become inaccurate even for smaller values of  $n$ . Concerning ASR, in order to handle the ill-conditioning problems and increase the values of  $n$  for which accurate results are obtained, we can make use of the pseudoinverse (e.g. Matlab function *pinv*) for the solution of systems of the form (6) or (7). In this way, the method’s production of accurate results can in most cases be extended up to  $n = 10$ .

## References

1. Cuyt A, Verdonk B. Different techniques for the construction of multivariate rational interpolants. In: Cuyt A (ed.); *Nonlinear Numerical Methods and Rational Approximation*. Dordrecht: Springer; 1988. 167-190.
2. Varsamis DN, Karampetakis NP. On the Newton multivariate polynomial interpolation with applications. In: *7th International Workshop on Multidimensional (nD) Systems (nDs11)*. France: Poitiers; 2011, 1-8.
3. Varsamis DN. *Development of computational methods for solving problems of Control Theory*, Ph.D. Thesis. Aristotle Univ. of Thessaloniki; 2012.
4. Xia P, Dong T, Zhang S, Lei N. The Bulirsch-Stoer algorithm for multivariate rational interpolation. *Math Meth Appl Sci*. 2018; 41(9): 7698-7710. DOI: 10.1002/mma.5233
5. Berrut JP, Mittelmann HD. Matrices for the direct determination of the barycentric weights of rational interpolation. *J. Comput. Appl. Math*. 1997; 78:355-370.

6. Salazar Celis O. *Practical Rational Interpolation of Exact and Inexact Data - Theory and Algorithms*, Ph.D. Thesis. Univ. of Antwerpen; 2008.
7. Tan JQ. *Continued Fractions Theory and Applications*. Beijing: Science; 2007.
8. Antoulas A, Ionita A, Lefteriu S. On two-variable rational interpolation. *Linear Algebra Appl.* 2012; 436(8):2889-2915.
9. Gautschi W, Inglese G. Lower Bounds for the Condition Number of Vandermonde Matrices. *Numer Math.* 1988; 52:241-250.

**TABLE 1** The interpolation points used for finding the four interpolants - elements of matrix  $A^{-1}$ .

$i$	$x_i$	$y_i$	$f_{i,1,1}$	$f_{i,1,2}$	$f_{i,2,1}$	$f_{i,2,2}$
1	0.338	0.236	-0.185	2.618	0.273	-2.395
2	0.318	0.984	-0.102	2.008	0.160	-1.586
3	0.548	0.749	-0.344	2.143	0.313	-1.043
4	0.842	0.167	-1.215	2.714	0.721	-1.018
5	0.903	0.105	-1.476	2.810	0.817	-1.002
6	0.745	0.729	-0.642	2.156	0.431	-0.776
7	0.717	0.133	-0.908	2.765	0.633	-1.230
8	0.446	0.509	-0.263	2.326	0.295	-1.487
9	0.530	0.860	-0.303	2.075	0.285	-1.014
10	0.678	0.806	-0.509	2.108	0.375	-0.817
11	0.531	0.956	-0.289	2.023	0.272	-0.962
12	0.067	0.542	-0.006	2.297	0.043	-9.729
13	0.282	0.481	-0.107	2.351	0.190	-2.397
14	0.685	0.208	-0.776	2.655	0.567	-1.208
15	0.608	0.326	-0.558	2.508	0.459	-1.240
16	0.881	0.133	-1.369	2.765	0.777	-1.002
17	0.102	0.959	-0.011	2.021	0.052	-4.984
18	0.153	0.153	-0.041	2.735	0.133	-5.675
19	0.156	0.090	-0.044	2.836	0.143	-5.900
20	0.454	0.669	-0.247	2.198	0.272	-1.319
21	0.831	0.790	-0.772	2.117	0.464	-0.672
22	0.713	0.473	-0.690	2.358	0.484	-0.953
23	0.709	0.958	-0.513	2.021	0.362	-0.721
24	0.506	0.305	-0.392	2.533	0.388	-1.515
25	0.790	0.236	-1.009	2.618	0.639	-1.024
26	0.234	0.465	-0.075	2.365	0.160	-2.914
27	0.619	0.615	-0.475	2.238	0.383	-0.999
28	0.123	0.124	-0.027	2.780	0.109	-7.257
29	0.284	0.736	-0.093	2.152	0.164	-2.025

**How to cite this article:** Hadjifotinou, K.G. and Karampetakis, N.P. (2020), An algorithm for two-variable rational interpolation suitable for matrix manipulations with the evaluation-interpolation method, *Math Meth Appl Sci.*, 2020;–.





**TABLE 3** The reciprocal condition numbers of the submatrices of  $B_{1,1}$  of size  $N$ .

$N$	$c_1$	$c_2$
30	$10^{-21}$	$10^{-19}$
28	$10^{-19}$	$10^{-18}$
26	$10^{-18}$	$10^{-18}$
24	$10^{-19}$	$10^{-18}$
22	$10^{-19}$	$10^{-19}$
20	$10^{-19}$	$10^{-19}$
18	$10^{-18}$	$10^{-18}$
16	$10^{-18}$	$10^{-19}$
14	$10^{-18}$	$10^{-18}$
12	$10^{-05}$	$10^{-18}$

**TABLE 4** Matrix  $B_{1,1}$  after eliminating the linearly dependent columns

$x^2$	$x^2y$	$x^2y^2$	$x^3$	$x^3y$	$x^4$	$x^2$	$x^2y$	$x^2y^2$	$x^3$	$x^3y$	$x^4$
0.11	0.03	0.01	0.04	0.01	0.01	0.02	0.00	0.00	0.01	0.00	0.00
0.10	0.10	0.10	0.03	0.03	0.01	0.01	0.01	0.01	0.00	0.00	0.00
0.30	0.23	0.17	0.16	0.12	0.09	0.10	0.08	0.06	0.06	0.04	0.03
0.71	0.12	0.02	0.60	0.10	0.50	0.86	0.14	0.02	0.72	0.12	0.61
0.82	0.09	0.01	0.74	0.08	0.67	1.20	0.13	0.01	1.09	0.11	0.98
0.56	0.40	0.30	0.41	0.30	0.31	0.36	0.26	0.19	0.27	0.19	0.20
0.51	0.07	0.01	0.37	0.05	0.26	0.47	0.06	0.01	0.34	0.04	0.24
0.20	0.10	0.05	0.09	0.05	0.04	0.05	0.03	0.01	0.02	0.01	0.01
0.28	0.24	0.21	0.15	0.13	0.08	0.09	0.07	0.06	0.05	0.04	0.02
0.46	0.37	0.30	0.31	0.25	0.21	0.23	0.19	0.15	0.16	0.13	0.11
0.28	0.27	0.26	0.15	0.14	0.08	0.08	0.08	0.07	0.04	0.04	0.02

**TABLE 5** Finding the non-zero elements of the solution of system (9)

Column	Corresponding reciprocal condition number
1	$10^{-18}$
2	$10^{-19}$
3	$10^{-18}$
4	$10^{-17}$
5	$10^{-18}$
6	$10^{-05}$
7	$10^{-05}$
8	$10^{-17}$
9	$10^{-17}$
10	$10^{-18}$
11	$10^{-17}$

**TABLE 6** Comparison of algorithms BS2v and ASR

test function	BS2v time(secs)	ASR time(secs)
$(7x + 3y - 2)/(5x - 4y - 1)$	0.4	0.07
$(x^2 + 5xy - 4y^2 - 7x + 3y - 2)/(xy - 5x - 4y - 1)$	2.8	0.08
$(x^3 - 2)/(y - 1)$	6.0	0.09
$(x^4 - 2)/(y - 1)$	19.9	0.10
$(x^4 - 2)/(y^2x - 1)$	24.0	0.10
$(32y^4 - 28y^3x + 17yx - 27)/(x^4 - 3xy - 25)$	-	0.10
$(x - 2)/(y^5 - 1)$	-	0.10
$y^5/x^5$	-	0.10
$y^6/x^6$	-	0.15
$y^7/x^7$	-	-