 CMC, vol.28, no.3, pp.231-259, 2012

A Multiple-Precision Study on the Modified Collocation Trefftz Method

Chia-Cheng Tsai[1] **and Po-Ho Lin**[2]

Abstract: Recently, Liu (CMES 21(2007), 53) developed the modified collocation Trefftz method (MCTM) by setting a characteristic length slightly larger than the maximum radius of the computational domain. In this study, we find that the range of admissible characteristic length can be significantly enlarged if the LU decomposition is applied for solving the resulted dense unsymmetric matrix. Furthermore, we discover a range formula for admissible characteristic length, in which the number of the T-complete functions, the shape of the computation domain, and the exponent bits of the involved floating-point arithmetic have been taken into consideration. In order to validate the prescribed formula for different exponent bits, the multiple precision floating-point reliable (MPFR) library is used. In addition, we find that the MCTM is a numerical method of exponential convergence. In other words, increasing the numbers of the T-complete functions can reduce the logarithmic error proportionally till the precision limit, which can be set up for the MPFR library. Numerical experiments are carried out to demonstrate that the proposed MCTM with the LU decomposition can solve the Laplace equation stably and accurately, even for a Cauchy problem. A multiple-precision comparison between the MCTM and the method of fundamental solution is also preformed.

Keywords: exponential convergence, modified collocation Trefftz method, multiple precision floating-point reliable library, Laplace equation

1 Introduction

In recent years, the meshless numerical methods have considered as alternative numerical schemes to the classical mesh-dependent numerical methods, such as the finite difference method (FDM), the finite element method (FEM), and the boundary element method (BEM). In general, the meshless numerical methods can be

[1] Corresponding author, Department of Marine Environmental Engineering, National Kaohsiung Marine University, Kaohsiung 811, Taiwan, E-mail: tsaichiacheng@mail.nkmu.edu.tw

[2] Department of Marine Environmental Engineering, National Kaohsiung Marine University, Kaohsiung 811, Taiwan

divided loosely into two categories: the domain-type and the boundary-type. Examples of the domain-type numerical methods include the Kansa's method [Kansa (1990b); Kansa (1990a); Huang, Lee and Cheng (2007); Huang, Yen and Cheng (2010)] and the meshless local Petrov-Galerkin method (MLPG) [Atluri and Zhu (1998); Atluri and Shen (2002); Atluri (2004)]. On the other hand, a popularly utilized method in the boundary-type category is the method of fundamental solutions (MFS) [Bogomolny (1985); Fairweather and Karageorghis (1998); Golberg and Chen (1999)]. In this paper, we study the Trefftz method [Liu (2007a); Liu (2007b); Liu (2008b); Liu (2008c); Liu (2008d); Liu (2008a); Yeih, Liu, Kuo and Atluri (2010); Fan and Chan (2011)], which is another boundary-type meshless numerical method.

The Trefftz method was first proposed by Trefftz in a conference [Trefftz (1926)]. The basic idea of the Trefftz method is to find a set of the so-called T-complete functions which satisfy the governing equation identically, and then use these functions to approximate the boundary condition. Since then, the Trefftz method was extensively studied and applied to many problems in science and engineering. Examples include the harmonic equation [Cheung, Jin and Zienkiewicz (1989)], the plane elasticity problems [Jin, Cheung and Zienkiewicz (1990)], the Helmholtz equation [Cheung, Jin and Zienkiewicz (1991); Kamiya and Wu (1994); Chang, Liu, Yeih and Kuo (2002); Chang, Liu, Kuo and Yeih (2003)], the biharmonic problems [Jin, Cheung and Zienkiewicz (1993)], the piezoelectric problems [Sheng, Sze and Cheung (2006); Dziatkiewicz and Fedeliǹski (2011)] and others. For a useful survey of literature, one can refer to the article [Kita and Kamiya (1995)].

In addition, it is noticeable that a significant superiority of the Trefftz method is its capability of dealing with singular problems, such as the Motz problem [Lu, Hu and Li (2004)], the singular biharmonic [Li, Lu and Hu (2004)] and Helmholz problems [Li, Lu, Tsai and Cheng (2006)]. An excellent review is this direction can be referred to the article [Li, Lu, Hu and Cheng (2008)].

The early applications of the Trefftz method are limited and less popular due to its ill-posed nature even for a well-posed boundary value problem [Kita and Kamiya (1995); Yeih, Liu, Chang and Kuo (2006); Li, Lu, Hu and Cheng (2008)]. In order to circumvent the ill-posed behavior of the Trefftz method, the traditional consideration is either to combine the Trefftz method with the domain decomposition method [Leitão (1997); Kita, Kamiya and Iio (1999)] or to use the Tikhonov's regularization method [Yeih, Liu, Chang and Kuo (2006)].

On the other hand, Liu modified the Trefftz method for dealing with the ill-posed behavior, in which a characteristic length was introduced to scale the basis functions [Liu (2007a); Liu (2007b)]. Later, the modified collocation Trefftz method (MCTM) was extended to deal with the potential problems in two-dimensional

doubly [Liu (2008a)] or multiply [Yeih, Liu, Kuo and Atluri (2010)] connected domains and Cauchy problems [Liu (2008c); Liu (2008d)]. The MCTM was also applied for solving Cauchy problems of the biharmonic equation [Liu (2008b)] and the nonlinear geometry boundary identification problem of heat conduction [Fan and Chan (2011)]. Basically, their results have demonstrated that the MCTM does not require any regularization technique.

However, the improvement of the characteristic length seems to be overestimated after performing a study by using a direct matrix solver instead of an iterative one for solving the resulted dense unsymmetric matrix of the MCTM. We find that there is a wide range of admissible characteristic length in which the solution are equally accurate if the LU decomposition [William and Teukolsky (1988)] is used for solving the resulted matrix. Furthermore, we discover a range formula to predict the admissible characteristic length, in which the number of the T-complete functions, the shape of the computation domain, and the exponent bits of the involved floating-point arithmetic have been taken into consideration.

All of the prescribed studies were performed upon the ANSI/IEEE 754-1985 standard of floating-point arithmetic [IEEE (1985)]. Although the IEEE 64-bit floating-point arithmetic is sufficient for most of the scientific applications, there are still certain scientific computing applications require a higher level of numeric precision. Therefore, the multiple precision floating-point reliable (MPFR) library [Hanrot, Lefevre, Pelissier and Zimmermann (2005)] was innovated. The MPFR provides correct rounding for all the operations and mathematical functions it implements. Therefore, the MPFR library is a multiple-precision extension of the IEEE 754 standard of floating-point arithmetic. For a comprehensive review of the MPFR, one can refer to the article [Fousse, Hanrot, Lefevre, Pelissier and Zimmermann (2007)].

The only multiple-precision study of the Trefftz method, to the best knowledge of the authors, was performed by the symbolic software Mathematica for solving the Motz problem [Li and Lu (2000)]. However, this implementation cannot be easily communicated with other existing scientific programs written in high-level programming languages. In this paper, in order to understand the behavior of the MCTM beyond the limit of IEEE 754 standard, the MPFR is used for its implementation. We find that the MCTM is a highly accurate numerical method of exponential convergence. In other words, increasing the numbers of the T-complete functions can reduce the logarithmic error proportionally till the machine epsilon of the working floating-point arithmetic, which can be set up for the MPFR library. On the other hand, we also validate our range formula of admissible characteristic length for different bit numbers of exponent via the MPFR library.

The organization of the paper is given as follows: the MCTM is reviewed in Section

2. Then, the range formula of admissible characteristic length is derived in Section 3. Some numerical results are presented in Section 4 and conclusions are drawn in Section 5.

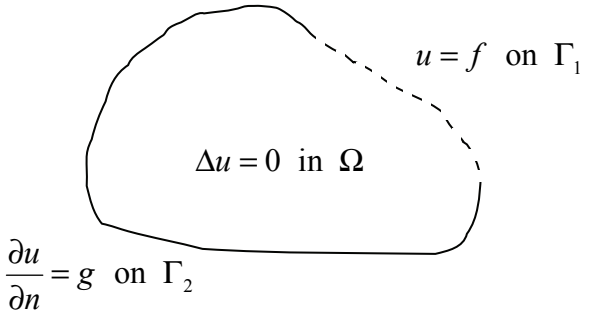


Figure 1: Schematic diagram of the problem of heat conduction.

2 The MCTM formulation

In Fig. 1, we consider the following heat conduction problem

$$\begin{cases} \Delta u = u_{rr} + \frac{u_r}{r} + \frac{u_{\theta\theta}}{r^2} = 0 & \text{in } \Omega \\ u = f(\mathbf{x}) & \text{on } \Gamma_1 \\ \frac{\partial u}{\partial n} = g(\mathbf{x}) & \text{on } \Gamma_2 \end{cases} \tag{1}$$

where Δ is the Laplace operator, $u = u(\mathbf{x})$ is the desired temperature field with $\mathbf{x} = (r, \theta)$ being the polar coordinate, Ω is the domain occupied by the medium being conducted and $\Gamma = \Gamma_1 + \Gamma_2$ is the boundary of Ω which is described by

$$r = \rho(\theta) \tag{2}$$

In addition, f is a given temperature on Γ_1, and $\frac{\partial u}{\partial n}$ is the outward normal derivative with g being a given heat flux on Γ_2. Here, the outward normal derivative is defined as

$$\frac{\partial u}{\partial n}(\mathbf{x}) = \nabla u \cdot \mathbf{n} \tag{3}$$

with

$$\nabla u = \frac{\partial u}{\partial r}\mathbf{r} + \frac{1}{r}\frac{\partial u}{\partial \theta}\boldsymbol{\theta} \tag{4}$$

and $\mathbf{n}$ is the unit outward normal vector of Γ which can be obtained from Eq. (2) as

$$\mathbf{n} = \frac{\rho(\theta)\mathbf{r} + \rho'(\theta)\boldsymbol{\theta}}{\sqrt{\rho(\theta)^2 + \rho'(\theta)^2}} \tag{5}$$

In Cartesian coordinate, the unit vectors $\mathbf{r}$ and θ are given as

$$\mathbf{r} = (\cos\theta, \sin\theta) \tag{6}$$

$$\boldsymbol{\theta} = (-\sin\theta, \cos\theta) \tag{7}$$

In the MCTM, the solution u is approximated by $\tilde{u}(\mathbf{x};\mathbf{c},\mathbf{d})$ as follows:

$$u(\mathbf{x}) \cong \tilde{u}(\mathbf{x};\mathbf{c},\mathbf{d}) = c_0 + \sum_{i=1}^{N}\left[c_i\left(\frac{r}{R_0}\right)^n \cos n\theta + d_i\left(\frac{r}{R_0}\right)^n \sin n\theta\right] \tag{8}$$

where R_0 is a characteristic length for scaling T-complete functions [Liu (2007b)], $2N+1$ is the number of the T-complete functions, $\mathbf{c} = (c_0, c_1, ..., c_N)$ and $\mathbf{d} = (d_1, d_2, ..., d_N)$ are the unknown coefficients to be determined. Here, we have assumed that the Trefftz origin is inside the computational domain Ω. When the number of the T-complete functions is sufficient large, the prescribed formula (8) can approximate a harmonic function arbitrarily well. However the resulted system of linear equations is highly ill-conditioned.

In order to apply the MCTM, the unknown coefficients $\mathbf{c}$ and $\mathbf{d}$ should be determined so that the boundary conditions are satisfied at $2N+1$ boundary points, $(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{2N+1})$, which result in

$$\begin{cases} f(\mathbf{x}_j) = c_0 + \sum\limits_{i=1}^{N}\left[c_i\left(\frac{r}{R_0}\right)^n \cos n\theta + d_i\left(\frac{r}{R_0}\right)^n \sin n\theta\right]\Big|_{\mathbf{x}=\mathbf{x}_j} & \text{for } \mathbf{x}_j \text{ on } \Gamma_1 \\ g(\mathbf{x}_j) = \sum\limits_{i=0}^{N} c_i \left.\frac{\partial\left(\left(\frac{r}{R_0}\right)^n \cos n\theta\right)}{\partial n(\mathbf{x})}\right|_{\mathbf{x}=\mathbf{x}_j} + \sum\limits_{i=1}^{N} d_i \left.\frac{\partial\left(\left(\frac{r}{R_0}\right)^n \sin n\theta\right)}{\partial n(\mathbf{x})}\right|_{\mathbf{x}=\mathbf{x}_j} & \text{for } \mathbf{x}_j \text{ on } \Gamma_2 \end{cases} \tag{9}$$

Eq. (9) is a system of $2N+1$ linear equations with $2N+1$ unknowns $\mathbf{c}$ and $\mathbf{d}$ which can also be rewritten in matrix form as

$$\mathbf{A}\left\{\begin{array}{c}\mathbf{c}^T\\ \mathbf{d}^T\end{array}\right\} = \left\{\begin{array}{c}\mathbf{f}\\ \mathbf{g}\end{array}\right\} \tag{10}$$

where $\mathbf{f}$ and $\mathbf{g}$ are the column vector constructed from the left-hand side of the equation and $\mathbf{A}$ is an unsymmetric matrix formed from the kernel functions in Eq. (9).

In the original development of the MCTM, Eq. (10) was solved by the conjugate gradient method (CGM) [Liu (2007b)] through the following symmetric form:

$$\mathbf{A}^T\mathbf{A}\left\{\begin{array}{c}\mathbf{c}^T\\ \mathbf{d}^T\end{array}\right\}=\mathbf{A}^T\left\{\begin{array}{c}\mathbf{f}\\ \mathbf{g}\end{array}\right\} \tag{11}$$

Alternatively, one can apply the bi-conjugate gradient method (CGM) or the LU decomposition [William and Teukolsky (1988)] to solve the unsymmetric matrix in Eq. (10). We will demonstrate that the LU decomposition is much more stable compared to the iterative matrix solvers.

After **c** and **d** are solved, the temperature field u can be determined by Eq. (8). This has finished the MCTM formulation.

3 Floating-point arithmetic and admissible characteristic length

In order to study the range of the admissible characteristic length, it is required to understand the behavior of implemented floating-point arithmetic. The ANSI/IEEE 754-1985 standard for floating-point arithmetic [IEEE (1985)] is the common standard for most of the programming languages, including C++, Fortran and Java. Therefore, every program using the formats and operations specified by IEEE 754 standard has exactly the same behavior on every configuration. Considering the 64-bits double precision of IEEE 754, it consists of three fields: one bit for sign, 11 bits of exponent and 52 bits of mantissa. Therefore, it has precision or machine epsilon equal to $2^{-53}\cong 10^{-16}$, which is defined as the minimum difference between two successive mantissa representations and gives a lower bound on the relative error due to rounding in a floating-point computing.

Then, we should introduce the range of the 64-bits double precision. First, the furthest positive and negative numbers from zero are

$$\pm\left(1-\left(\frac{1}{2}\right)^{53}\right)2^{1024} \tag{12}$$

And, the positive and negative normalized numbers closest to zero are

$$\pm 2^{-1022} \tag{13}$$

In the above equation, the normalized number means that the implicit leading binary digit is one. To reduce the loss of precision when an underflow occurs, the IEEE 754 standard includes the ability to represent a number smaller than the possible normalized representation, by making the implicit leading digit being zero.

Such numbers are called denormal. They don't include as many significant digits as a normalized number, but they enable a gradual loss of precision when the result of an arithmetic operation is not exactly zero but is too close to zero to be represented by a normalized number. For more details, one can refer to the article [IEEE (2008)].

While a multiple-precision computation is required, one can consider the MPFR library which is an extension of the IEEE 754 standard. In its application, the bit numbers of the exponent and mantissa can be set up. Therefore, the machine epsilon becomes

$$2^{-(p+1)} \tag{14}$$

with p being the mantissa bits. And the furthest positive and negative numbers from zero are given as

$$\pm\left(1-\left(\frac{1}{2}\right)^{p+1}\right)2^{2^{e-1}} \tag{15}$$

with e being the exponent bits. And, the positive and negative normalized numbers closest to zero are

$$\pm 2^{-2^{e-1}+2} \tag{16}$$

To derive the admissible range of characteristic length when applying the MCTM for a Dirichlet problem, we assume that the maximum radius of the boundary Γ is $R_{\max} = 2^{r_{\max}}$ and the minimum radius is equal to $R_{\min} = 2^{r_{\min}}$. By observing Eq. (9), it is clear that one can prevent the overflow by considering

$$\left(\frac{R_{\max}}{R_0}\right)^N \leq \left(1-\left(\frac{1}{2}\right)^{p+1}\right)2^{2^{e-1}} \tag{17}$$

or equivalently

$$N\left(r_{\max}-r_0\right) \leq \log_2\left(1-\left(\frac{1}{2}\right)^{p+1}\right)+2^{e-1} \tag{18}$$

where $r_0 = \log_2 R_0$. On the other hand, avoiding the underflow of normalized numbers requires

$$2^{-2^{e-1}+2} \leq \left(\frac{R_{\min}}{R_0}\right)^N \tag{19}$$

or equivalently

$$-2^{e-1}+2 \leq N\left(r_{\min}-r_0\right) \tag{20}$$

Eqs. (18) and (20) can be combined to give the admissible range as

$$r_{\max}-\frac{\log_2\left(1-\left(\frac{1}{2}\right)^{p+1}\right)+2^{e-1}}{N} \leq r_0 \leq r_{\min}+\frac{2^{e-1}-2}{N} \tag{21}$$

In applying Eq. (21), it should be noticed that the overflow will make the MCTM immediately crashed when r_0 is smaller than the lower bound. On the other hand, the underflow only results in a denormalized number when r_0 is larger than the upper bound. In other words, the characteristic length can still get a little larger beyond the upper bound.

4 Numerical results

Now, we are in a position to study the stability and accuracy of the MCTM. In the following studies, the absolute maximum errors are adopted and the IEEE results are obtained upon the IEEE double precision of 64 bits.

4.1 Review case one

First, we consider an epitrochoid boundary shape defined by

$$\rho\left(\theta\right)=\sqrt{26-10\cos\left(4\theta\right)} \tag{22}$$

Dirichlet boundary condition is set up according to the following analytical solution

$$u\left(\mathbf{x}\right)=\exp x\cos y \tag{23}$$

where

$$\begin{cases} x=r\cos\theta \\ y=r\sin\theta \end{cases} \tag{24}$$

This problem has been solved by the MCTM with the CGM [Liu (2007b)]. In the numerical computations, we have set $R_0=R_{\max}=6$ and $N=25$. The numerical errors along a circle with radius equal to 3 for the solutions obtained by the MCTM with the CGM, the BiCGM and the LU decomposition are plotted in Fig. 3(a). In the figure, we have exactly reproduced the result of the CGM [Liu (2007b)] and

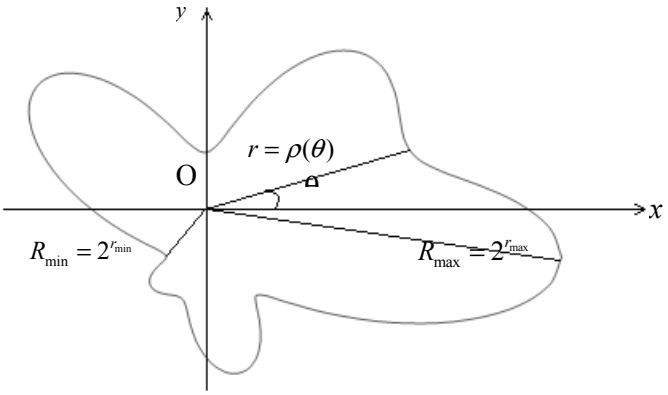


Figure 2: Definition of R_{max} and R_{min}.

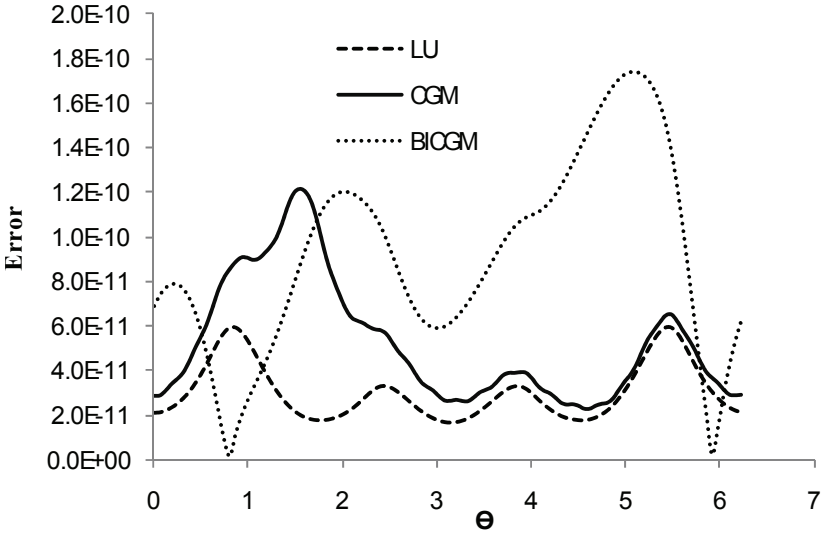
their accuracy are similar. Then we solve the same problem by setting $R_0 = 2$, which is a circle smaller than the computational domain. In Fig. 3(b), the solutions obtained by iterative solvers are not accurate as mentioned by Liu. However, the solutions obtained by the LU decomposition are still highly accurate.

To further investigate the topic, we perform computations by different R_0 and plot the maximum errors against $r_0 = \log_2 R_0$ as depicted in Fig. 4. The results have clearly indicated that the solution obtained by the LU decomposition is much more stable with respect to the characteristic length compared with those by iterative solvers. Then, we plot the maximum errors against r_0 for different numbers of T-complete functions. Figs. 5(a) and 5(b) give the plots for the solutions obtained by the CGM and the LU decomposition, respectively. The superiority of the LU decomposition over the iterative methods for solving the resulted linear equation system of the MCTM is also clear in these figures. Also, it can be found that the optimal accuracy in Fig. 5 is around 10^{-14}, which is very close to the machine epsilon of IEEE 754 double precision. The gap between the computed optimal accuracy and the machine epsilon is produced by the accumulated round-off errors.

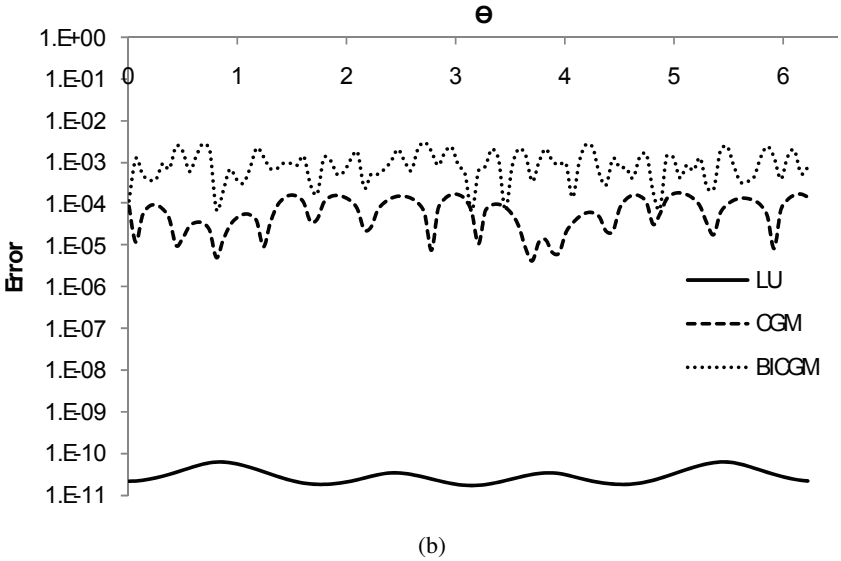
4.2 Review case two

Then we consider a circular domain with a radius equal to 2. To illustrate the accuracy and stability of the MCTM with the LU decomposition, we consider the following analytical solution [Jin (2004); Liu (2007b)]:

$$u(\mathbf{x}) = \cos x \cosh y + \sin x \sinh y \tag{25}$$



(a)



(b)

Figure 3: Numerical errors for the review case one solved by the MCTM with (a) $R_0 = 6$ and (b) $R_0 = 2$.

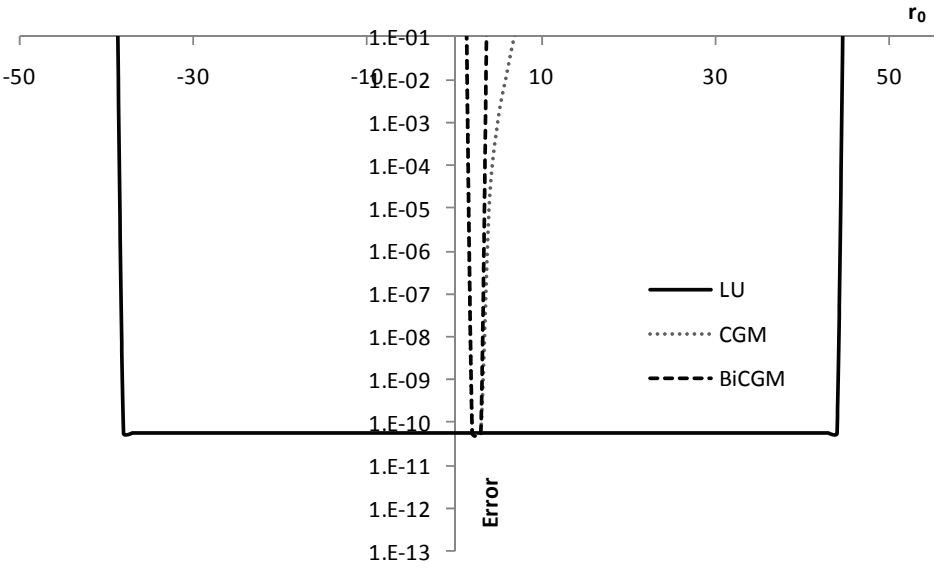


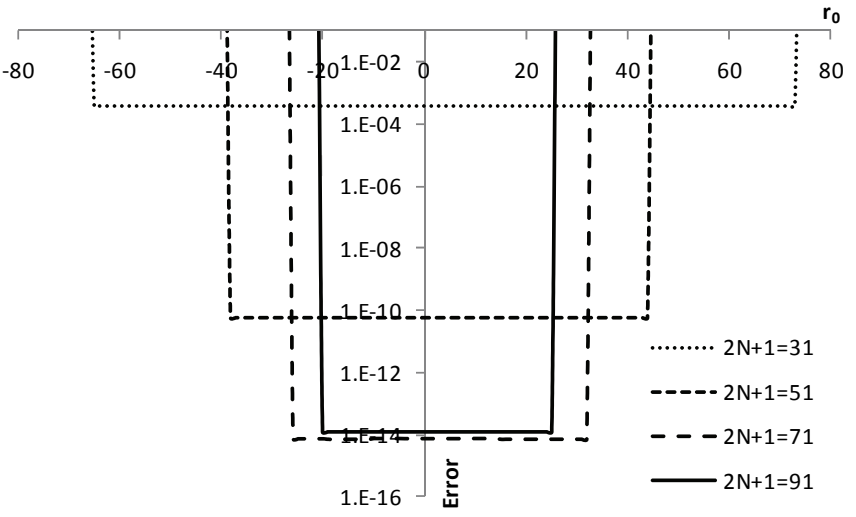
Figure 4: Maximum error against r_0 for the review case one solved by the MCTM with CGM, BiCGM and LU decomposition.

The exact boundary data are derived by inserting $x = 2\cos\theta$ and $y = 2\sin\theta$ into the above equation. In the numerical computations, we have set $N = 50$. The numerical errors along a circle with radius $r = 1$ for the solutions obtained by the CGM, the BiCGM and the LU decomposition are plotted in Fig. 6(a) and 6(b) for $R_0 = 2.5$ and $R_0 = 1.55$, respectively. In the figure, it can be observed that the modification by setting the characteristic length slightly larger than the maximum radius of the computational domain is significant for the MCTM with the iterative solvers as mentioned in the literature [Liu (2007b)]. However, a further observation has shown that the solution obtained by the LU decomposition is much more stable with respect to the characteristic length as shown in Fig. 7.

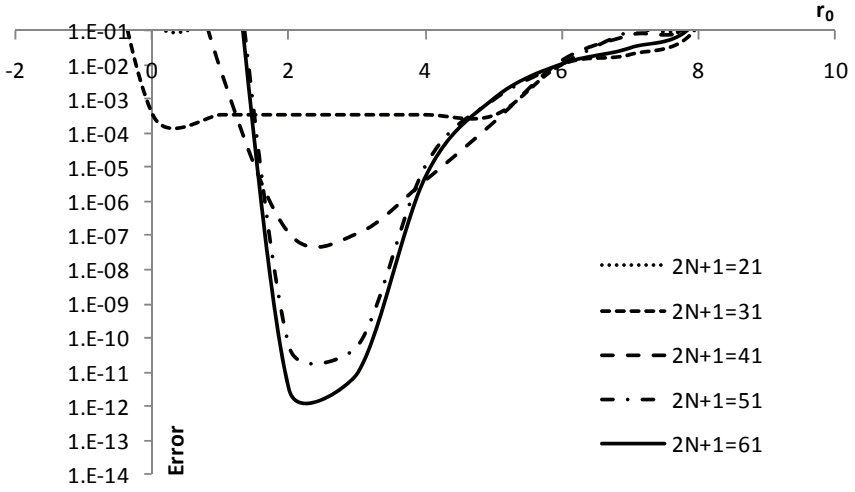
4.3 The admissible characteristic length

After revisiting the two cases in the literature, we should study the derived range formula for admissible characteristic length. We consider Dirichlet problems in a circular domain with radius of 2 and a 2-by-16 ellipse. In the computations, the analytical solutions are set up according to Eq. (23).

Fig. 8 gives the maximum errors against r_0 for the solutions in the prescribed two domains. For all of the cases, it can be observed that the solutions are very accurate and stable in certain wide admissible ranges. And the optimal accuracy is around

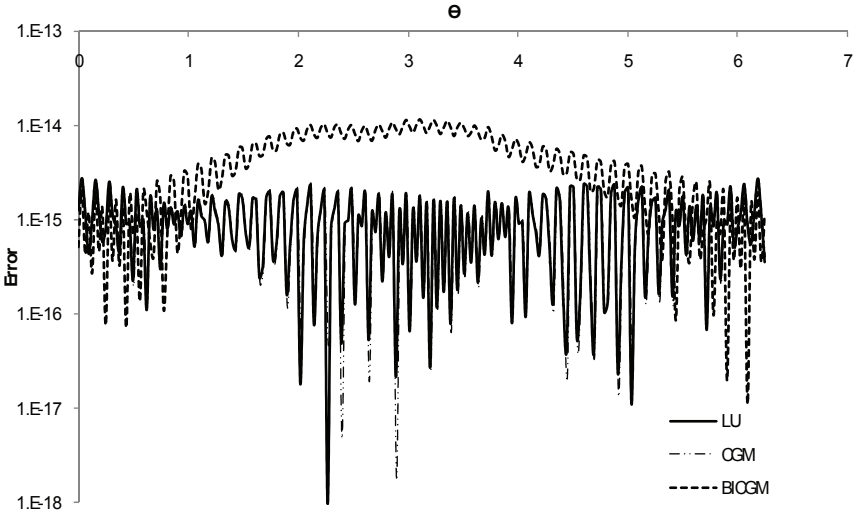


(a)

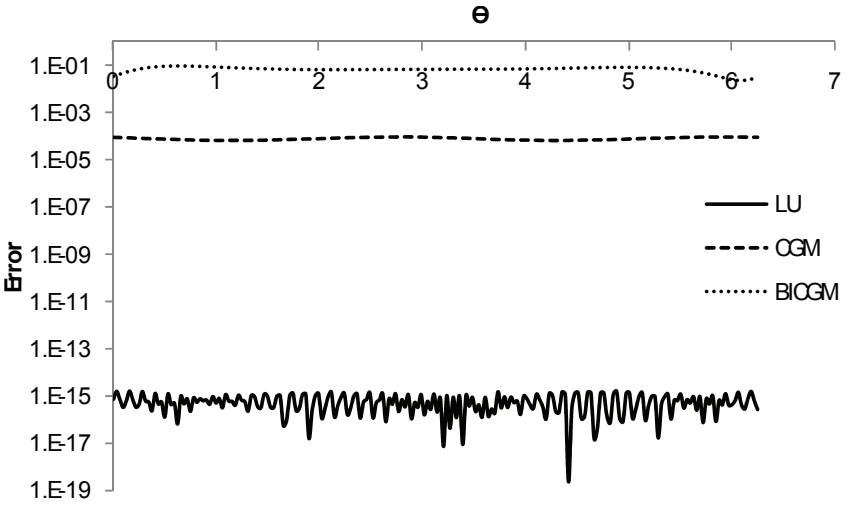


(b)

Figure 5: Maximum error against r_0 for different numbers of T-complete functions for the review case one solved by the MCTM with (a) LU decomposition and (b) CGM.



(a)



(b)

Figure 6: Numerical errors for the review case two solved by the MCTM with (a) $R_0 = 2.5$ and (b) $R_0 = 1.55$

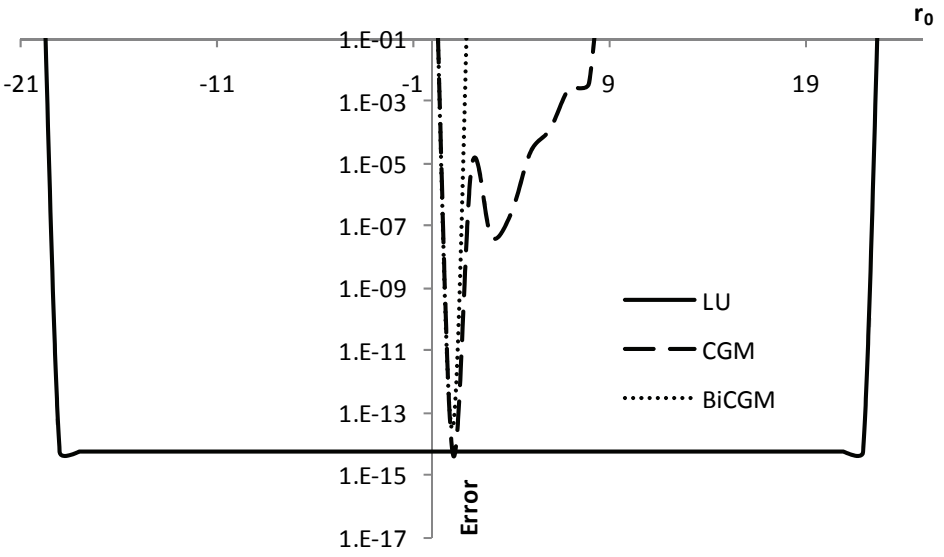


Figure 7: Maximum error against r_0 for the review case two solved by the MCTM with CGM, BiCGM and LU decomposition.

10^{-14} and 10^{-8} respectively for the circular and elliptic problems. The differences between the computed optimal accuracy and the machine epsilon are caused by the round-off errors. In addition, the round-off errors are more significant for the problem in the more slender computational domain.

Table 1: Admissible range of r_0 for the circular problem by the IEEE double precision.

2N+1	computed lower bound	predicted lower bound	computed upper bound	predicted upper bound
21	-101	-101.4	104	103.2
31	-67	-67.3	70	69.1
41	-49	-50.2	54	52.1
61	-33	-33.1	36	35.1
81	-24	-24.6	27	26.6
101	-19	-19.5	22	21.4

Then, the computed lower and upper bounds of r_0 and the corresponding predicted values by Eq. (21) are tabulated in Tables 1 and 2 for the circular and elliptic problems, respectively. In the tables, it is clear that the predicted lower bound of

Table 2: Admissible range of r_0 for the 2-by-16 elliptic problem by the IEEE double precision.

2N+1	computed lower bound	predicted lower bound	computed upper bound	predicted upper bound
41	-47	-47.2	54	52.1
51	-36	-37.0	44	41.9
61	-30	-30.1	37	35.1
71	-25	-25.3	33	30.2
81	-21	-21.6	29	26.6
101	-16	-16.5	24	21.4

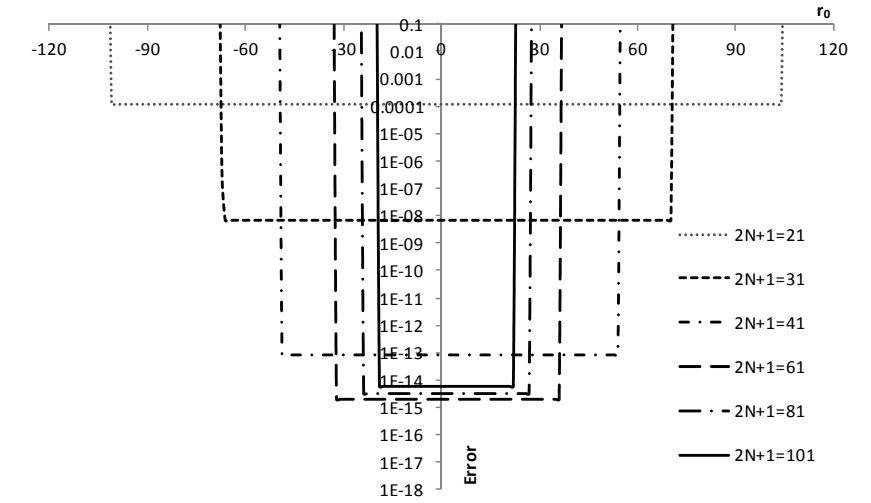
r_0 is exact since the overflow has immediately made the program crashed. On the other hand, the computed upper bound of r_0 are slightly larger than the predicted value which is caused by the denormal mechanism of the underflow in the IEEE 754 floating-point standard. In the computation, the values of r_0 are limited to integers for simplicity.

To further validate the range formula of admissible characteristic length, we solve the 2-by-16 elliptic problem by the MPFR library with two different configurations ($\begin{cases} e=12 \\ p=100 \end{cases}$ and $\begin{cases} e=13 \\ p=150 \end{cases}$). Fig. 9 gives the maximum errors against r_0 for the solutions. In the figure, it can also be observed that the solutions are very accurate and stable in the wide admissible ranges, and the optimal accuracy is about $2^{-100} \cong 8\times 10^{-31}$ and $2^{-150} \cong 7\times 10^{-46}$ for the two configurations, respectively. Furthermore, Table 3 and 4 give the computed and predicted ranges of the characteristic length. An excellent agreement can also be observed.

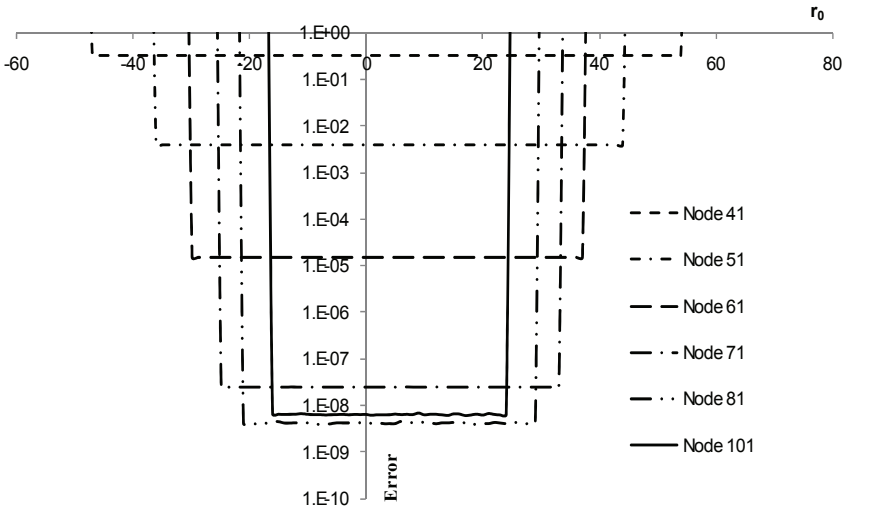
Table 3: Admissible range of r_0 for the 2-by-16 elliptic problem by the MPFR library ($e=12$ and $p=100$).

2N+1	computed lower bound	predicted lower bound	computed upper bound	predicted upper bound
61	-64	-64.3	71	69.2
81	-47	-47.2	54	52.2
101	-36	-37.0	44	41.9
121	-30	-30.1	39	35.1
141	-25	-25.3	31	30.2

For a very slender ellipse, reasonable solutions cannot be obtained by the IEEE

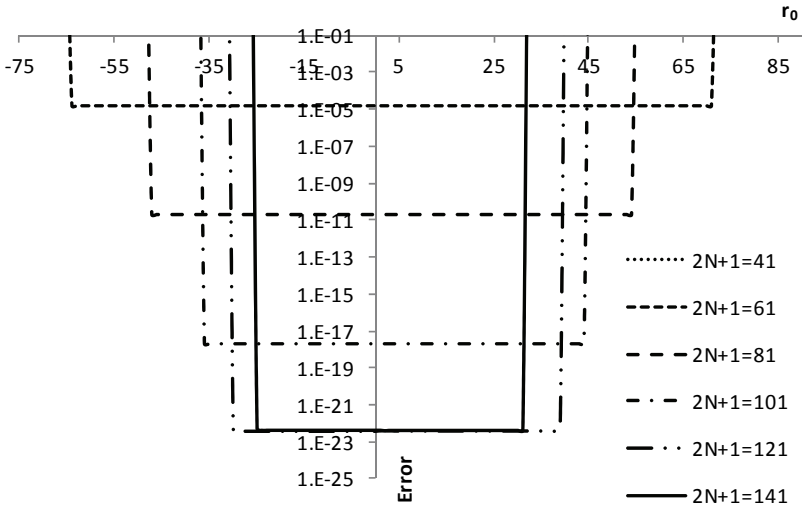


(a)

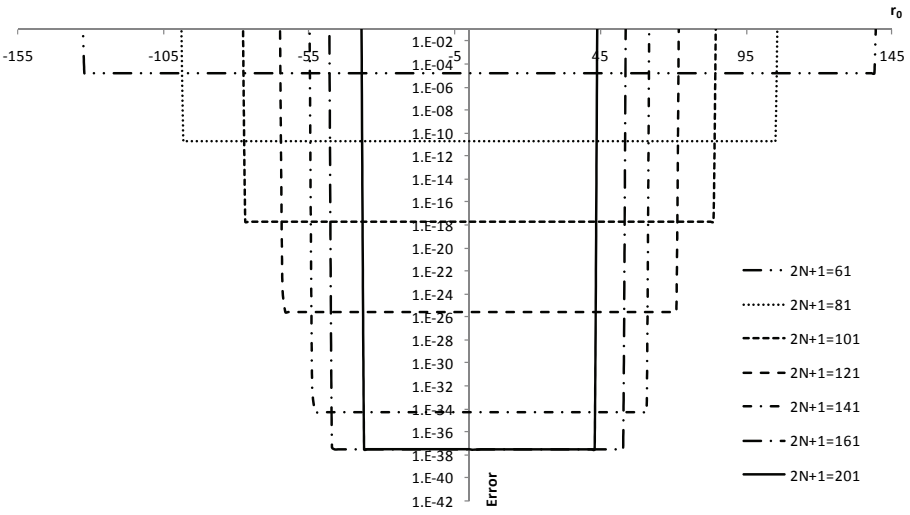


(b)

Figure 8: Maximum error against r_0 for different numbers of T-complete functions solved by the IEEE double precision for (a) circular and (2) 2-by-16 elliptic problems.



(a)



(b)

Figure 9: Maximum error against r_0 for different numbers of T-complete functions of the 2-by-16 elliptic problem solved by the MPFR library with (a)$e = 12\&p = 100$ and (b)$e = 13\&p = 150$.

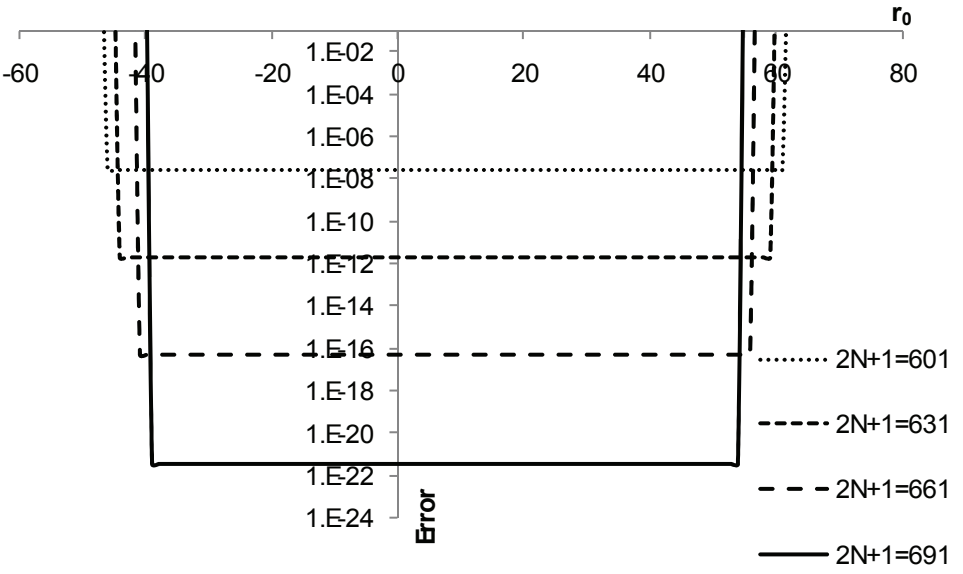


Figure 10: Maximum error against r_0 for different numbers of T-complete functions of the 2-by-256 elliptic problem solved by the MPFR library with $e = 15$ and $p = 600$.

Table 4: Admissible range of r_0 for the 2-by-16 elliptic problem by the MPFR library ($e = 13$ and $p = 150$).

2N+1	computed lower bound	predicted lower bound	computed upper bound	predicted upper bound
61	-132	-132.5	139	137.5
81	-98	-98.4	105	103.4
101	-77	-77.9	84	82.9
121	-64	-64.3	71	69.2
141	-54	-54.5	61	59.5
161	-47	-47.2	53	52.2
201	-36	-37.0	43	41.9

754 arithmetic. Therefore, we use the MPFR library with $e = 15$ and $p = 600$ for solving a problem in a 2-by-256 ellipse. Fig. 10 and Table 5 gives the maximum errors against r_0 and the range of admissible characteristic length, which behave reasonably and similarly as in the previous study.

Table 5: Admissible range of r_0 for the 2-by-256 elliptic problem by the MPFR library ($e = 15$ and $p = 600$).

2N+1	computed lower bound	predicted lower bound	computed upper bound	predicted upper bound
601	-46	-46.6	61	55.6
631	-44	-44.0	59	53.0
661	-41	-41.6	56	50.6
691	-39	-39.5	54	48.5

Table 6: Admissible range of r_0 for the Amoeba problem by the MPFR library ($e = 13$ and $p = 300$).

2N+1	computed lower bound	predicted lower bound	computed upper bound	predicted upper bound
81	-98	-98.4	104	103.4
101	-77	-77.9	84	82.9
121	-64	-64.3	70	69.2
141	-54	-54.5	60	59.5
161	-47	-47.2	53	52.2
181	-41	-41.5	47	46.5
201	-36	-37.0	43	41.9
221	-33	-33.2	38	38.2

4.4 The exponential convergence

It is well-known that the MCTM is a numerical method of exponential convergence [Schaback (2008)]. In other words, increasing the numbers of the T-complete functions can reduce the logarithmic error proportionally till the precision limit, which can be set up by the MPFR library. To demonstrate this numerical phenomenon, the previous Laplace problems are reconsidered. The logarithmic errors against the numbers of T-complete functions are plotted in Figs. 11, 12 and 13 respectively for the solutions of the Laplace problems in a circular domain of radius 2 as well as 2-by-16 and 2-by-256 ellipses. In the figures, the exponential convergence is significant and the improvement on the optimal accuracy by increasing the precisions via the MPFR library is also clear, especially for the problem in a 2-by-256 ellipse which cannot be solved by an IEEE computing. Furthermore, the gaps between the optimal accuracy and the machine epsilon of a given precision are more relevant for the problems in more sender domains. In addition, the optimal accuracy obtained by the MPFR library for the problem in the 2-by-256 ellipse is 3.69×10^{-69}, which

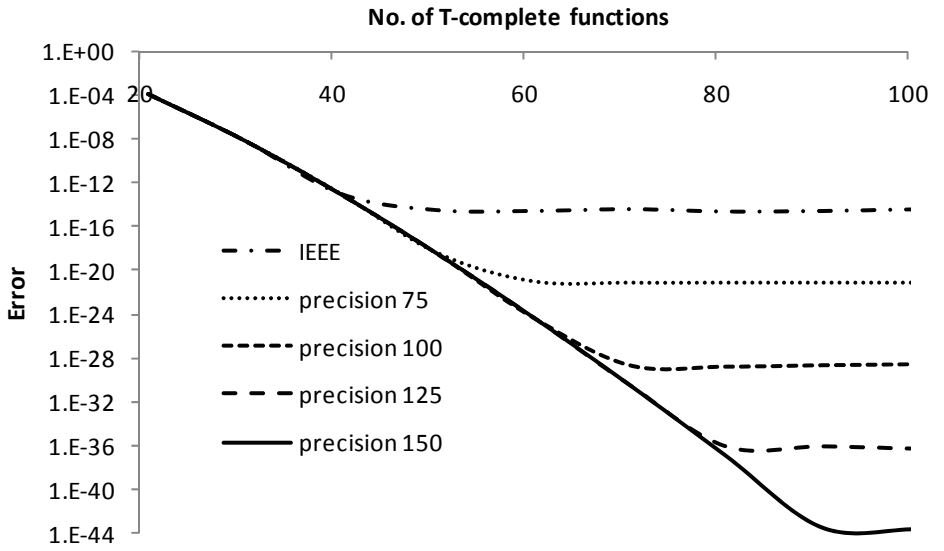


Figure 11: The exponential convergence for the circular problem.

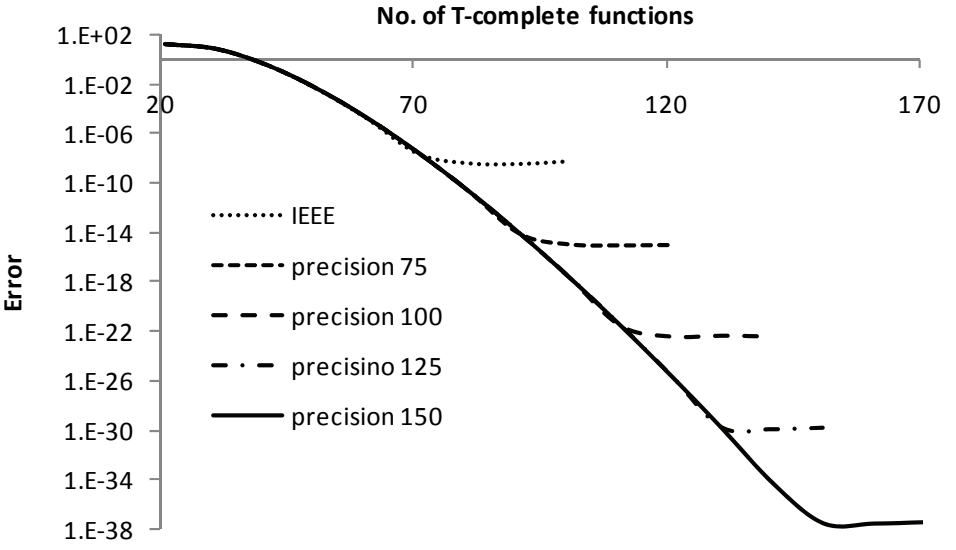


Figure 12: The exponential convergence for the 2-by-16 elliptic problem.

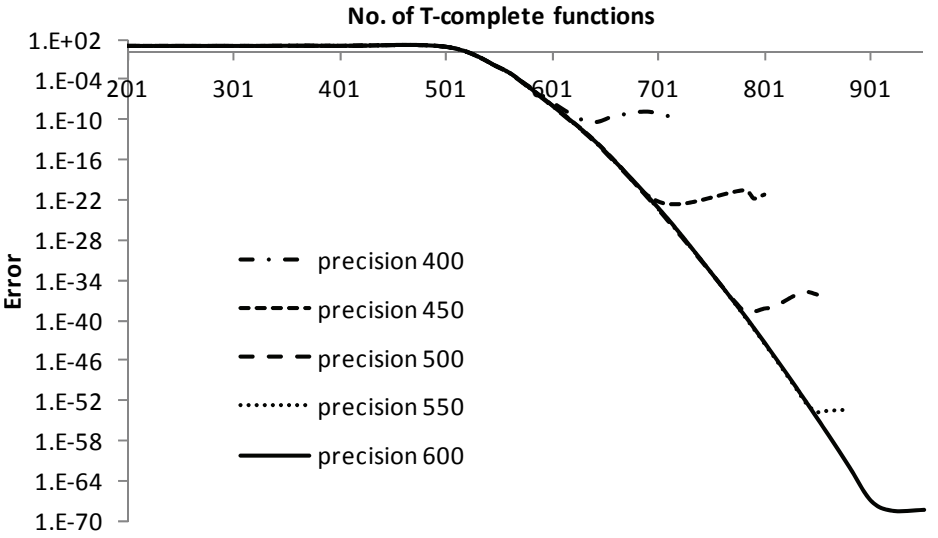


Figure 13: The exponential convergence for the 2-by-256 elliptic problem.

is obtained in 462 seconds by a CPU of Intel(R) core i7 2.67GHz.

4.5 Amoeba-like domain

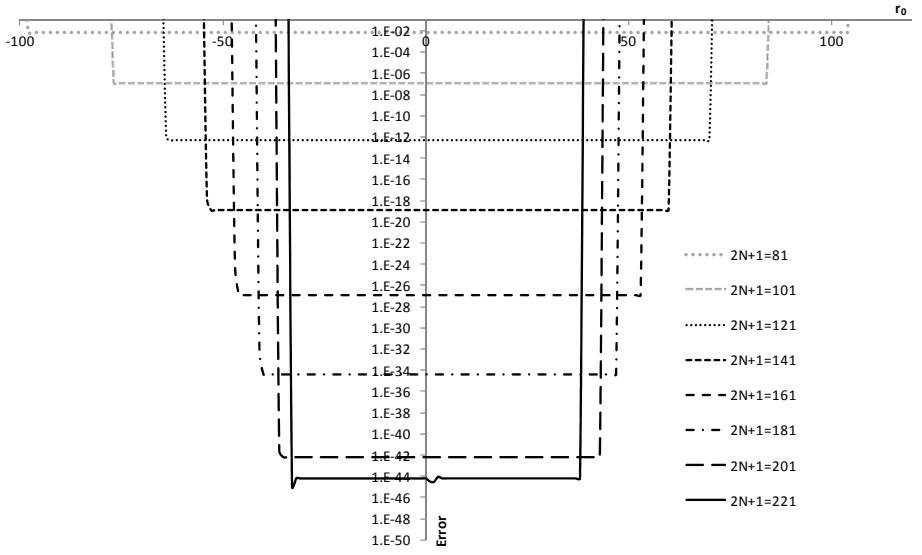
The application of the proposed MCTM to an irregular domain is straightforward. In this example, the computer domain is defined as

$$\begin{cases} x = a\rho(\theta)\cos\theta \\ y = b\rho(\theta)\sin\theta \end{cases} \tag{26}$$

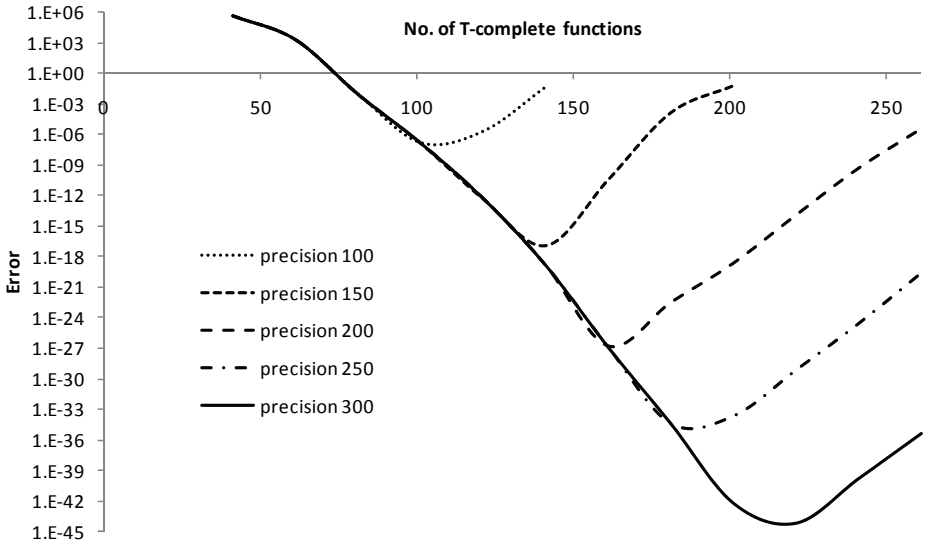
with

$$\rho(\theta) = \exp(\sin\theta)\sin^2 2\theta + \exp(\cos\theta)\sin^2 2\theta \tag{27}$$

and a and b are constants to stretching the computational domain. A typical amoeba-like domain is shown in Fig. 2. In order to make $R_{\max} = 2^4$ and $R_{\min} = 2$, $a = 5.333$ and $b = 7.831$ are typically set up. The Dirichlet boundary condition is set up according to the analytical solution in Eq. (23). Then, the computation is performed by using the MPFR library with $e = 13$ and $p = 300$, the maximum errors against r_0 for different node numbers are given in Fig. 14(a), in which stable and accuracy solutions can be found in a wide range of admissible characteristic length. In addition, the computed and predicted ranges of the characteristic length are tabulated in



(a)



(b)

Figure 14: Maximum error against (a) r_0 and (b) numbers of T-complete functions for the problem in amoeba-like domain.

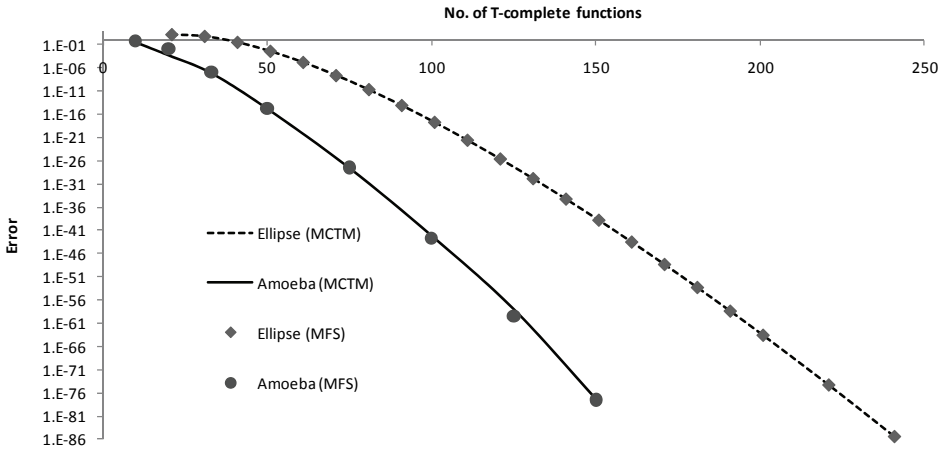


Figure 15: The exponential convergence of the MCTM and the MFS and their comparison.

Table 6. A good agreement between the computed and predicted values can also be found. Then, the exponential convergence of the present problem is demonstrated in Fig. 14(b), which behaves similarly as in the previous cases.

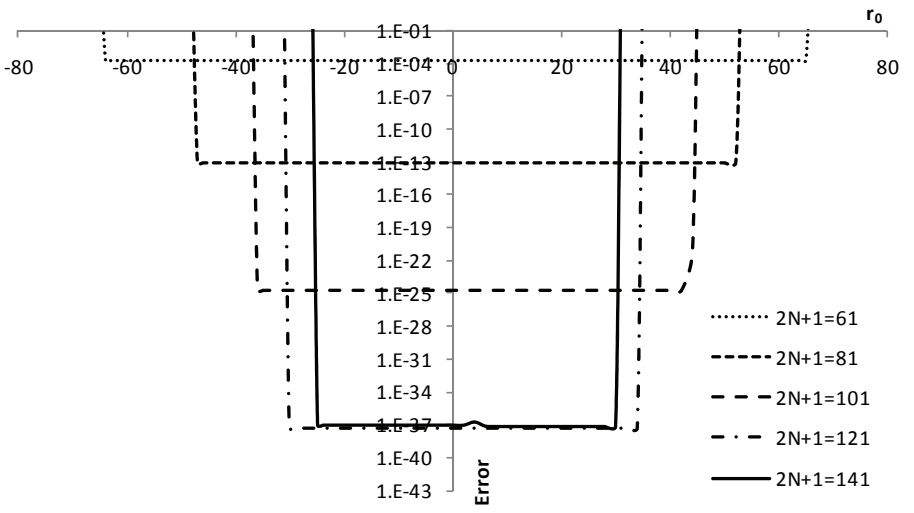
4.6 Comparison with the method of fundamental solutions

Then, we also compare the exponential convergence of the MCTM and the MFS for the solutions of the Laplace equation in a 2-by-16 ellipse and an amoeba-like domain. The Dirichlet boundary conditions are set up according to the analytical solution in Eq. (23). For the amoeba-like domain, its boundary is described by Eq. (26) with $a = b = 1$. The numerical results are obtained by the MPFR library with infinite precision. And the sources of the MFS are located as far as possible.

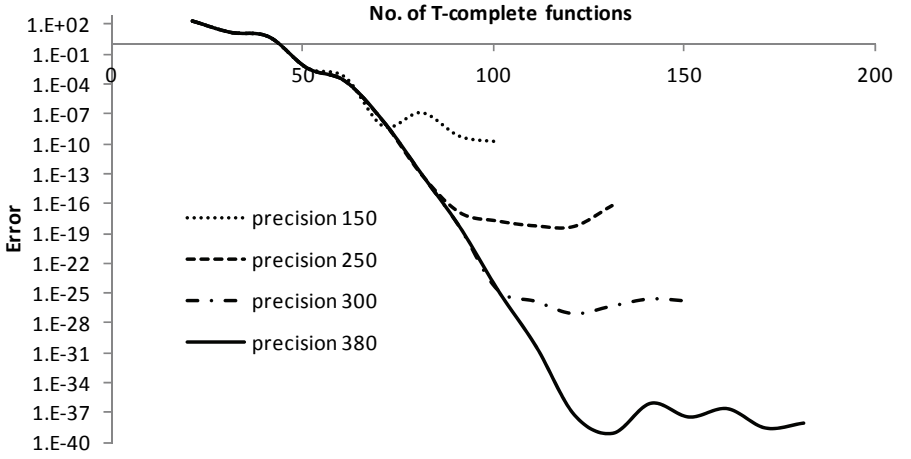
Fig. 15 gives the logarithmic errors against the numbers of the T-complete functions or the fundamental solutions. In the figure, it can be observed that the numerical solutions obtained by the MCTM and the MFS are equally accurate. In the literature, it has been indicated that the MFS for far–away source points is asymptotically nothing else than a fit of the boundary data by specific harmonic polynomials [Schaback (2008)] or alternatively the MCTM and the MFS are equivalent [Chen, Wu, Lee and Chen (2007)]. Our numerical solutions seem to support their comments.

4.7 Cauchy problem

The application of the proposed MCTM with the LU decomposition for a Cauchy problem is also of ease. In this example, we consider a Cauchy problem in a 2-by-



(a)



(b)

Figure 16: Maximum error against (a) r_0 and (b) numbers of T-complete functions for the Cauchy problem.

16 ellipse. Both Dirichlet and Neumann boundary conditions are set up according to the analytical solution in Eq. (23) on the right boundary of the ellipse while no boundary conditions are set up on the left boundary of the ellipse. Fig. 16(a) and 16(b) gives the maximum errors against r_0 for different numbers of the T-complete functions and the logarithmic errors against the numbers of the T-complete functions, respectively. The results behave similarly as those of the Dirichlet problems.

5 Conclusion

In this study, the numerical solution of the modified collocation Trefftz method (MCTM) was reviewed. The LU decomposition was used for solving the resulted unsymmetric dense matrix. A range formula for admissible characteristic lengths was derived by considering the number of the T-complete functions, the shape of the computation domain, and the exponent bits of the involved floating-point arithmetic. In order to change the exponent bits, the multiple precision floating-point reliable (MPFR) library was used. Numerical solutions were in general very accuracy and stable in the predicted range of admissible characteristic length. These results have suggested that iterative matrix solvers should be replaced by the LU decomposition or other direct matrix solvers in the application of the MCTM.

Furthermore, the exponential convergence of the MCTM was demonstrated. Numerical results have indicated that increasing the numbers of the T-complete functions can reduce the logarithmic error proportionally till the precision limit, which can be set up for the MPFR library. For a Dirichlet problem in a 2-by-256 ellipse, the optimal accuracy of the obtained solution was 3.69×10^{-69} in 462 seconds of computing time by Intel(R) core i7 2.67GHz. And the applicability of the proposed MCTM with the LU decomposition was also demonstrated for a Cauchy problem. Furthermore, numerical solutions also demonstrated that using the MCTM and the MFS for approximating harmonic boundary were about equally accurate.

Acknowledgement: The National Science Council of Taiwan under NSC 100-2221-E-022-006 is gratefully acknowledged for providing financial support to carry out the present work.

References

IEEE (1985): IEEE Standard for Binary Floating-Point Arithmetic. *ANSI/IEEE Std 754-1985*, pp. 0-1.

IEEE (2008): IEEE Standard for Floating-Point Arithmetic. *IEEE Std 754-2008*, pp. 1-58.

Atluri, S. N. (2004): *The meshless method (MLPG) for domain & BIE discretizations*, Tech Science Press.

Atluri, S. N.; Shen, S. (2002): The meshless local Petrov-Galerkin (MLPG) method: A simple & less-costly alternative to the finite element and boundary element methods. *CMES: Computer Modeling in Engineering & Sciences*, vol. 3, pp. 11-52.

Atluri, S. N.; Zhu, T. (1998): A new meshless local Petrov-Galerkin (MLPG) approach in computational mechanics. *Computational Mechanics*, vol. 22, pp. 117-127.

Bogomolny, A. (1985): Fundamental solutions method for elliptic boundary value problems. *SIAM Journal on Numerical Analysis*, vol. 22, pp. 644-669.

Chang, J. R.; Liu, R. F.; Kuo, S. R.; Yeih, W. (2003): Application of symmetric indirect Trefftz method to free vibration problems in 2D. *International Journal for Numerical Methods in Engineering*, vol. 56, pp. 1175-1192.

Chang, J. R.; Liu, R. F.; Yeih, W.; Kuo, S. R. (2002): Applications of the direct Trefftz boundary element method to the free-vibration problem of a membrane. *The Journal of the Acoustical Society of America*, vol. 112, pp. 518.

Chen, J. T.; Wu, C. S.; Lee, Y. T.; Chen, K. H. (2007): On the equivalence of the Trefftz method and method of fundamental solutions for Laplace and biharmonic equations. *Computers & Mathematics with Applications*, vol. 53, pp. 851-879.

Cheung, Y. K.; Jin, W. G.; Zienkiewicz, O. C. (1989): Direct solution procedure for solution of harmonic problems using complete, non-singular, Trefftz functions. *Communications in Applied Numerical Methods*, vol. 5, pp. 159-169.

Cheung, Y. K.; Jin, W. G.; Zienkiewicz, O. C. (1991): Solution of Helmholtz equation by Trefftz method. *International Journal for Numerical Methods in Engineering*, vol. 32, pp. 63-78.

Dziatkiewicz, G.; Fedelìnski, P. (2011): Indirect Trefftz Method for Solving Cauchy Problem of Linear Piezoelectricity. Computational Modelling and Advanced Simulations. Murìn, J., Kompiš, V. and Kutiš, V., Springer Netherlands. 24: 49-65.

Fairweather, G.; Karageorghis, A. (1998): The method of fundamental solutions for elliptic boundary value problems. *Advances in Computational Mathematics*, vol. 9, pp. 69-95.

Fan, C.-M.; Chan, H.-F. (2011): Modified Collocation Trefftz Method for the Geometry Boundary Identification Problem of Heat Conduction. *Numerical Heat Transfer, Part B: Fundamentals*, vol. 59, pp. 58-75.

Fousse, L.; Hanrot, G.; Lefevre, V.; Pelissier, P.; Zimmermann, P. (2007): MPFR: A multiple-precision binary floating-point library with correct rounding.

ACM Trans. Math. Softw., vol. 33, pp. 13.

Golberg, M. A.; Chen, C. S., Eds. (1999): *The method of fundamental solutions for potential, Helmholtz and diffusion problems*. Boundary Integral Methods: Numerical and Mathematical Aspects. Southampton, Computational Mechanics Publications.

Hanrot, G.; Lefevre, V.; Pelissier, P.; Zimmermann, P. (2005): "The GNU MPFR library." from http://www.mpfr.org/.

Huang, C. S.; Lee, C. F.; Cheng, A. H. D. (2007): Error estimate, optimal shape factor, and high precision computation of multiquadric collocation method. *Engineering Analysis with Boundary Elements*, vol. 31, pp. 614-623.

Huang, C. S.; Yen, H. D.; Cheng, A. H. D. (2010): On the increasingly flat radial basis function and optimal shape parameter for the solution of elliptic PDEs. *Engineering Analysis with Boundary Elements*, vol. 34, pp. 802-809.

Jin, B. (2004): A meshless method for the Laplace and biharmonic equations subjected to noisy boundary data. *CMES: Computer Modeling in Engineering and Sciences*, vol. 6, pp. 253-262.

Jin, W. G.; Cheung, Y. K.; Zienkiewicz, O. C. (1990): Application of the Trefftz method in plane elasticity problems. *International Journal for Numerical Methods in Engineering*, vol. 30, pp. 1147-1161.

Jin, W. G.; Cheung, Y. K.; Zienkiewicz, O. C. (1993): Trefftz method for Kirchhoff plate bending problems. *International Journal for Numerical Methods in Engineering*, vol. 36, pp. 765-781.

Kamiya, N.; Wu, S. T. (1994): Generalized eigenvalue formulation of the Helmholtz equation by the Trefftz method. *Engineering Computations*, vol. 11, pp. 177-186.

Kansa, E. J. (1990a): Multiquadrics - A scattered data approximation scheme with applications to computational fluid-dynamics–I surface approximations and partial derivative estimates. *Computers & Mathematics with Applications*, vol. 19, pp. 127-145.

Kansa, E. J. (1990b): Multiquadrics - A scattered data approximation scheme with applications to computational fluid-dynamics–II solutions to parabolic, hyperbolic and elliptic partial differential equations. *Computers & Mathematics with Applications*, vol. 19, pp. 147-161.

Kita, E.; Kamiya, N. (1995): Trefftz method: an overview. *Advances in Engineering Software*, vol. 24, pp. 3-12.

Kita, E.; Kamiya, N.; Iio, T. (1999): Application of a direct Trefftz method with domain decomposition to 2D potential problems. *Engineering Analysis with Boundary Elements*, vol. 23, pp. 539-548.

Leitão, V. M. A. (1997): On the implementation of a multi-region Trefftz-collocation formulation for 2-D potential problems. *Engineering Analysis with Boundary Elements*, vol. 20, pp. 51-61.

Li, Z.-C.; Lu, T.-T.; Tsai, H.-S.; Cheng, A. H. D. (2006): The Trefftz method for solving eigenvalue problems. *Engineering Analysis with Boundary Elements*, vol. 30, pp. 292-308.

Li, Z. C.; Lu, T. T. (2000): Singularities and treatments of elliptic boundary value problems. *Mathematical and Computer Modelling*, vol. 31, pp. 97-145.

Li, Z. C.; Lu, T. T.; Hu, H. Y. (2004): The collocation Trefftz method for biharmonic equations with crack singularities. *Engineering Analysis with Boundary Elements*, vol. 28, pp. 79-96.

Li, Z. C.; Lu, T. T.; Hu, H. Y.; Cheng, A. H. D. (2008): *Trefftz and collocation methods*, WIT Press.

Liu, C. S. (2007a): An effectively modified direct Trefftz method for 2D potential problems considering the domain's characteristic length. *Engineering Analysis with Boundary Elements*, vol. 31, pp. 983-993.

Liu, C. S. (2007b): A modified Trefftz method for two-dimensional Laplace equation considering the domain's characteristic length. *CMES: Computer Modeling in Engineering & Sciences*, vol. 21, pp. 53-65.

Liu, C. S. (2008a): A highly accurate collocation Trefftz method for solving the Laplace equation in the doubly connected domains. *Numerical Methods for Partial Differential Equations*, vol. 24, pp. 179-192.

Liu, C. S. (2008b): A highly accurate MCTM for direct and inverse problems of biharmonic equation in arbitrary plane domains. *CMES: Computer Modeling in Engineering & Sciences*, vol. 30, pp. 65-75.

Liu, C. S. (2008c): A highly accurate MCTM for inverse Cauchy problems of Laplace equation in arbitrary plane domains. *CMES: Computer Modeling in Engineering & Sciences*, vol. 35, pp. 91-111.

Liu, C. S. (2008d): A modified collocation Trefftz method for the inverse Cauchy problem of Laplace equation. *Engineering Analysis with Boundary Elements*, vol. 32, pp. 778-785.

Lu, T. T.; Hu, H. Y.; Li, Z. C. (2004): Highly accurate solutions of Motz's and the cracked beam problems. *Engineering Analysis with Boundary Elements*, vol. 28, pp. 1387-1403.

Schaback, R., Ed. (2008): *Adaptive numerical solution of MFS systems*. The Method of Fundamental Solutions - A Meshless Method, Dynamic Publishers.

Sheng, N.; Sze, K. Y.; Cheung, Y. K. (2006): Trefftz solutions for piezoelec-

tricity by Lekhnitskii's formalism and boundary-collocation method. *International Journal for Numerical Methods in Engineering*, vol. 65, pp. 2113-2138.

Trefftz, E. (1926): *Ein gegenst uck zum ritzschen verfahren*.

William, H.; Teukolsky, S. A. (1988): *Numerical Recipes in C: The art of scientific computing*, Cambridge university press.

Yeih, W.; Liu, C. S.; Kuo, C. L.; Atluri, S. N. (2010): On Solving the Direct/Inverse Cauchy Problems of Laplace Equation in a Multiply Connected Domain, Using the Generalized Multiple-Source-Point Boundary-Collocation Trefftz Method & Characteristic Lengths. *CMC: Computers Materials & Continua*, vol. 17, pp. 275-302.

Yeih, W.; Liu, R. F.; Chang, J. R.; Kuo, S. R. (2006): Numerical instability of the direct Trefftz method for Laplace problems for a 2D finite domain. *International Journal of Applied Mathematics and Mechanics*, vol. 2, pp. 41-66.