INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN ENGINEERING
Int. J. Numer. Meth. Engng 2011; **88**:157–179
Published online 1 March 2011 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/nme.3168

Perturbation and stability analysis of strong form collocation with reproducing kernel approximation

Hsin-Yun Hu[1], Jiun-Shyan Chen[2,*,†] and Sheng-Wei Chi[2]

[1]*Department of Mathematics, Tunghai University, Taichung 407, Taiwan*
[2]*Department of Civil and Environmental Engineering, University of California, Los Angeles (UCLA), CA 90095, U.S.A.*

SUMMARY

Solving partial differential equations using strong form collocation with nonlocal approximation functions such as orthogonal polynomials and radial basis functions offers an exponential convergence, but with the cost of a dense and ill-conditioned linear system. In this work, the local approximation functions based on reproducing kernel approximation are introduced for strong form collocation method, called the reproducing kernel collocation method (RKCM). We perform the perturbation and stability analysis of RKCM, and estimate the condition numbers of the discrete equation. Our stability analyses, validated with numerical tests, show that this approach yields a well-conditioned and stable linear system similar to that in the finite element method. We also introduce an effective condition number where the properties of both matrix and right-hand side vector of a linear system are taken into consideration in the measure of conditioning. We first derive the effective condition number of the linear systems resulting from RKCM, and show that using the effective condition number offers a tighter estimation of stability of a linear system. The mathematical analysis also suggests that the effective condition number of RKPM does not grow with model refinement. The numerical results are also presented to validate the mathematical analysis. Copyright © 2011 John Wiley & Sons, Ltd.

Received 1 September 2010; Revised 24 January 2011; Accepted 26 January 2011

KEY WORDS: reproducing kernel approximation; strong form collocation; stability; perturbation; condition number; effective condition number

1. INTRODUCTION

In the past 15 years, meshfree methods [1, 2] have emerged into a new class of computational methods that have been applied to many engineering and scientific problems. Meshfree methods all share a common feature: the approximation of unknown in the partial differential equation is constructed based on scattered points without mesh connectivity. While no mesh is needed in the construction of approximation in meshfree methods, domain integration presents some difficulties if the discrete equation is formulated based on weak formulation [3, 4]. Alternatively, collocation on strong forms has been introduced in meshfree method, such as the radial basis collocation methods (RBCM) [5–7] or the reproducing kernel collocation method (RKCM) [8–10]. These methods employ approximation functions with higher order continuities and thus allow calculation of higher order derivatives in the strong forms.

*Correspondence to: Jiun-Shyan Chen, Department of Civil and Environmental Engineering, University of California, Los Angeles (UCLA), CA 90095, U.S.A.
†E-mail: jschen@seas.ucla.edu

The approximation functions commonly used in the weak form-based meshfree methods are the moving least-squares (MLS) [11, 12] and reproducing kernel (RK) [13–15] approximations, while the radial basis functions (RBFs) [5, 6, 16, 17] are usually employed in the strong form collocation method. Since taking derivatives of RBF is relatively straightforward than taking derivatives of MLS/RK, using RBFs in the strong form approaches justifies its simplicity. From convergence standpoint, the compactly supported MLS/RK approximations with monomial reproducibility render an algebraic convergence, while the nonlocal RBFs with certain regularity offer exponential convergence. Nevertheless, the linear system of RBFs with collocation method is typically more ill-conditioned compared with those based on the compactly supported MLS/RK approximations with weak formulation. The RKCM [10, 18], where MLS/RK approximation is used in conjunction with strong form collocation, is a compromise for enhanced conditioning with an algebraic rate of convergence.

The stability of a linear system is strongly related to the conditioning of the matrix, which is measured by the condition number. The traditional condition number is used to measure the solution errors resulting from the round-off perturbations in the matrix [19–21], whereas the new effective condition number takes into account round-off perturbation in both matrix and right-hand side vector of a linear system [22] and offers a better measure of conditioning than the traditional condition numbers. Christiansen and Hansen [23] proposed the effective condition number for boundary collocation method, and Li *et al.* [24] applied the approach to finite difference method. Problems with singularity are examples where the linear system is not as ill-conditioned as what the traditional condition number indicates, and this has been demonstrated by the use of effective condition number [25].

In this work, we investigate the stability of RKCM by introducing perturbation analysis of the RKCM, estimating the bound of solution perturbation due to the perturbation of the linear system, and analyzing the conditioning of the linear system using the traditional condition number and the effective condition number. We show that the traditional condition number of RKCM is of the order $O(h^{-2})$, where h is the maximal nodal distance. However, we show that this measure of conditioning sometimes overexaggerates the ill-conditioning of a linear system and leads to a loose bound of stability estimation. The effective condition number of RKCM is shown to be insensitive to model refinement, and it offers a better conditioning measure and yields a tighter bound of stability estimation when validated with numerical tests. In this paper, we call the approach using *strong form collocation with reproducing kernel approximation* the RKCM, and one term is used over the other depending on the condition of the sentence.

This paper is organized as follows. Section 2 introduces reproducing kernel approximation, its approximation properties, as well as its inverse inequalities. In Section 3, we discuss perturbation and stability analyses, as well as condition numbers, for linear systems resulting from discretization of function approximation and differential equations based on strong form collocation using reproducing kernel approximation. In Section 4, we introduce the effective condition number, and apply it for the stability estimation of RKCM linear system. The concluding remarks are given in Section 5.

2. REPRODUCING KERNEL APPROXIMATION

2.1. Basic equations

We first describe the RK approximation in one dimension. The multi-dimensional formation can be easily obtained with a similar construction [13, 14]. Let a function $f(x)$ be approximated by

$$f^h(x)=\sum_{I=1}^{Np}\psi_I(x)d_I, \quad x\in\Omega\subset\mathbb{R} \tag{1}$$

where $\psi_I(x)$ are the RK shape functions centered at x_I, and d_I are the coefficients to be sought. The shape functions are constructed based on the locations of a set of nodal points S with

DOI: 10.1002/nme

maximal nodal distance h

$$S=\{x_I\}_{I=1}^{Np}=\{x_1, x_2, \ldots, x_{Np}\} \tag{2}$$

where Np is the number of nodal points. Based on MLS or reproducing kernel approximation, the shape function is given as follows [13, 14]:

$$\psi_I(x)=\mathbf{h}^{\mathrm{T}}(0)\mathbf{M}^{-1}(x)\mathbf{h}(x-x_I)\phi_a(x-x_I) \tag{3}$$

where

$$\mathbf{M}(x)=\sum_{I=1}^{Np}\mathbf{h}(x-x_I)\mathbf{h}^{\mathrm{T}}(x-x_I)\phi_a(x-x_I) \tag{4}$$

$$\mathbf{h}^{\mathrm{T}}(x-x_I)=[1, x-x_I, (x-x_I)^2, \ldots, (x-x_I)^n] \tag{5}$$

$$\mathbf{h}^{\mathrm{T}}(0)=[1, 0, \ldots, 0] \tag{6}$$

The vectors $\mathbf{h}(x-x_I)$ have dimension $n+1$, and $\mathbf{M}(x)$ is a moment matrix with dimension $(n+1)\times(n+1)$. The function $\phi_a(x-x_I)$ is called the kernel function, for example, the B-spline function

$$\phi_a(z)=\begin{cases}\dfrac{2}{3}-4z^2+4z^3, & 0\leqslant z<\dfrac{1}{2}\\ \dfrac{4}{3}-4z+4z^2-\dfrac{4}{3}z^3, & \dfrac{1}{2}\leqslant z<1\\ 0, & z\geqslant 1\end{cases} \tag{7}$$

where $z=|x-x_I|/a$ and a is called the support size. The support sizes are allowed to vary in space and be dependent on I. Another choice of kernel function is the quintic B-spline

$$\phi_a(z)=\begin{cases}\dfrac{11}{20}-\dfrac{9z^2}{2}+\dfrac{81z^4}{4}-\dfrac{81z^5}{4}, & 0\leqslant z<\dfrac{1}{3}\\ \dfrac{17}{40}+\dfrac{15z}{8}-\dfrac{63z^2}{4}+\dfrac{135z^3}{4}-\dfrac{243z^4}{8}+\dfrac{81z^5}{8}, & \dfrac{1}{3}\leqslant z<\dfrac{2}{3}\\ \dfrac{81}{40}-\dfrac{81z}{8}+\dfrac{81z^2}{4}-\dfrac{81z^3}{4}+\dfrac{81z^4}{8}-\dfrac{81z^5}{40}, & \dfrac{2}{3}\leqslant z<1\\ 0, & z\geqslant 1\end{cases} \tag{8}$$

The RK shape functions are so constructed such that they satisfy the following reproducing conditions when the complete nth-order polynomials are used as the bases in (5):

$$\sum_{I=1}^{Np}\psi_I(x)x_I^i=x^i, \quad i=0, 1, \ldots, n \tag{9}$$

where n is the order of polynomial introduced in (5), and is called the reproducing order. The reproducing conditions are also applied to multi-dimensions in the construction of multi-dimensional RK approximation functions [13, 14].

2.2. *Properties of RK approximation*

The RK approximation described in the previous section will be used for solving boundary value problem (BVP) under a strong form collocation framework. For a set of the RK shape function satisfying reproduction conditions in (9), they have the following derivative reproducing properties:

$$\sum_{I=1}^{Np}\psi_I^{(\ell)}(x)x_I^i=(x^i)^{(\ell)}, \quad i=0, 1, \ldots, n \tag{10}$$

DOI: 10.1002/nme

where $(\ )^{(\ell)} = \mathrm{d}^{\ell}(\ )/\mathrm{d}x^{\ell}$ is a differential operator. Same properties applied to general multi-dimensional case. Consider an one-dimensional domain $\Omega = \{x | 0 < x < 1\}$ with discretization by Np equally or non-equally spaced points in $\bar{\Omega}$, we have the bounds [15]

$$|\psi_I(x)|_{\infty} < C_1 \tag{11}$$

$$|\psi_I^{(\ell)}(x)|_{\infty} \leqslant C_2 a^{-\ell}, \quad \ell = 1, 2, \ldots \tag{12}$$

where C_1 and C_2 are generic constants.

Let function $v = v(x)$ be approximated by the linear combination of RK shape functions

$$v(x) = \sum_{I=1}^{Np} \psi_I(x) d_I \tag{13}$$

Define a finite-dimensional space V as:

$$V = \mathrm{span}\{\psi_1(x), \psi_2(x), \ldots, \psi_{Np}(x)\} \subset H^2(\Omega) \tag{14}$$

The inverse inequalities given below [10] for the high-order derivatives of function $\forall v \in V$ are needed to show the convergence and stability in the remainder of this paper.

Lemma 2.1

Let the set of nodal points be quasi-uniformly distributed. For $\forall v \in V$, there exist the following inequalities:

$$\|v\|_{\ell,\Omega} \leqslant C_1 \kappa^{1/2} a^{-\ell} n^{2\ell} \|v\|_{0,\Omega}, \quad \ell = 1, 2, 3, \ldots \tag{15}$$

$$\|v\|_{\ell,\Gamma} \leqslant C_2 \kappa^{1/2} a^{-\ell} n^{2\ell} \|v\|_{1,\Omega}, \quad \ell = 1, 2, 3, \ldots \tag{16}$$

$$\|v_{,x}\|_{\ell,\Gamma} \leqslant C_3 \kappa^{1/2} a^{-(\ell+1)} n^{2(\ell+1)} \|v\|_{1,\Omega}, \quad \ell = 1, 2, 3, \ldots \tag{17}$$

where κ is the maximal overlapping number of RK discretization in the domain, a is the maximal support size of kernel functions, $\Gamma = \partial\Omega$ is the boundary of Ω, and C_i are generic constants.

3. PERTURBATION AND STABILITY ANALYSIS OF REPRODUCING KERNEL COLLOCATION METHOD

In this section, we present the perturbation analysis and stability estimation of linear systems resulting from (i) function approximation and (ii) discretization of partial differential equations by strong form collocation with RK approximation.

3.1. Stability of RKCM for function approximation

Here, we study the stability of linear systems resulting from function approximation by collocation method using RK approximation. Consider RK approximation of $f(\mathbf{x})$ in (1) by a set of RK shape functions $\{\psi_I(\mathbf{x})\}_{I=1}^{Np}$ centered at $\{\mathbf{x}_I\}_{I=1}^{Np}$. We define another set of evaluation points, called the collocation points, denoted as

$$E = \{\boldsymbol{\xi}_I\}_{I=1}^{Nc} \tag{18}$$

The set of collocation points may or may not be the same as the set of nodal points. In the collocation method, the residuals of the approximation are enforced to be zero at the collocation points:

$$f^h(\xi_I) = f(\xi_I) \quad \forall\ I = 1, 2, \ldots, Nc \tag{19}$$

where Nc is the number of collocation points. This gives a linear system

$$\mathbf{A}\mathbf{x} = \mathbf{b} \tag{20}$$

DOI: 10.1002/nme

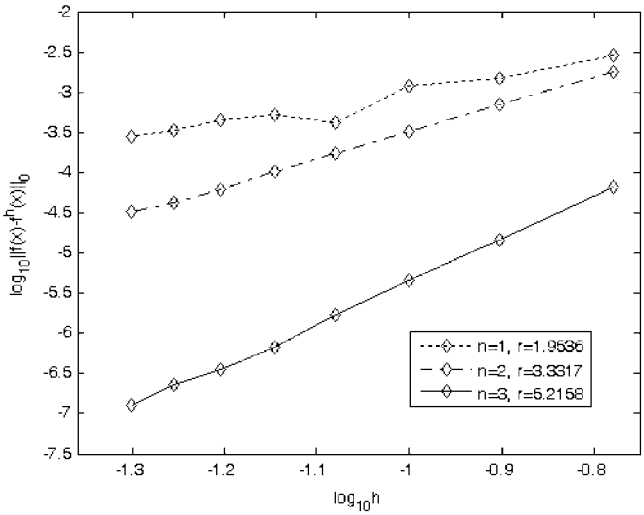

Figure 1. The approximation errors of RK approximation by collocation method with various reproducing degrees.

where

$$A_{IJ}=\psi_J(\xi_I),\quad b_I=f(\xi_I),\quad x_I=d_I\ \ \forall\ I=1,2,\ldots,Nc,\ \ J=1,2,\ldots,Np \tag{21}$$

In the collocation method, typically $Nc\geqslant Np$. If $Nc>Np$, we have an overdetermined system, and the least-squares method, QR decomposition or singular value decomposition (SVD) can be used to solve the system.

First, we discuss the convergence behavior for the system (20). Let n be the reproducing degree in RK approximation, and if $f^{(n+1)}(\mathbf{x})\in L_2$, where the superscript $(n+1)$ denotes the order of derivative, there exists the following convergence property [10, 18]:

$$\|f(\mathbf{x})-f^h(\mathbf{x})\|_{\ell,\Omega}\leqslant ca^{n+1-\ell}|f(\mathbf{x})|_{n+1,\Omega} \tag{22}$$

where $\ell\geqslant 0$, and $a=(n+1)h$ is the support size with h the nodal distance. Taking $\ell=0$, we have the L_2 error bound:

$$\|f(\mathbf{x})-f^h(\mathbf{x})\|_{0,\Omega}\leqslant ca^{n+1}|f(\mathbf{x})|_{n+1,\Omega} \tag{23}$$

The L_2 error norms of RK approximation of a function $f(x)=\sin(\pi x)$, $0\leqslant x\leqslant 1$, by collocation method are shown in Figure 1, in which the number of collocation points is taken to be the same as the number of nodal points, the support sizes vary with the reproducing degree n, i.e., $a=(n+1)h$, and $\mathbf{r}$ denotes the rate of convergence. The results in Figure 1 agree well with (23) for $n=1$ and 2, and show a superconvergence for $n=3$.

To study the stability of the system (20), we begin with a perturbation analysis. For function approximation, we consider $Nc=Np$. Let square matrix $\mathbf{A}$ be positive definite with full rank. The diagonal canonical form of $\mathbf{A}$ is

$$\mathbf{U}^{\mathrm{T}}\mathbf{A}\mathbf{U}=\mathbf{D}=\mathrm{daig}(\lambda_1,\lambda_2,\ldots,\lambda_{Np}) \tag{24}$$

where λ_i are the eigenvalues of $\mathbf{A}$, with the order $\lambda_{\max}=\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_{Np}=\lambda_{\min}>0$. The columns of $\mathbf{U}$ are the corresponding orthonormal eigenvectors $\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_{Np}$, $\mathbf{u}_i^{\mathrm{T}}\mathbf{u}_i=1$, satisfying

$$\mathbf{A}\mathbf{u}_i=\lambda_i\mathbf{u}_i,\quad i=1,2,\ldots,Np \tag{25}$$

Two perturbed problems and their properties for system (20) are summarized as follows. Detail derivations are given in [19, 20].

DOI: 10.1002/nme

Table I. The condition numbers of matrix **A** associated with RK approximation of a function $f(x)=\sin(\pi x)$, $0\leqslant x\leqslant 1$ by collocation.

$n=1$		$n=2$		$n=3$	
Np	Cond(**A**)	Np	Cond(**A**)	Np	Cond(**A**)
6	1.53	6	1.45	6	2.68
11	1.58	11	1.58	11	6.54
21	1.60	21	1.62	21	8.77

Table II. Stability of a linear system $\mathbf{Ax}=\mathbf{b}$ associated with RK approximation of a function $f(x)=\sin(\pi x)$, $0\leqslant x\leqslant 1$ by collocation.

	$n=2,\ Np=11$	$n=3,\ Np=11$
Cond(**A**)	1.58	6.54
$\|\Delta\mathbf{A}\|$	1×10^{-13}	1×10^{-13}
$\|\mathbf{A}\|$	1.0000	1.0000
$\|\Delta\mathbf{b}\|$	1×10^{-13}	1×10^{-13}
$\|\mathbf{b}\|$	2.23607	2.23607
$\|\Delta\mathbf{x}\|$	1.2247×10^{-13}	3.2529×10^{-13}
$\|\mathbf{x}\|$	2.23634	2.23767

Case I: The vector **b** on the right side of the system (20) is perturbed:

$$\mathbf{A}\hat{\mathbf{x}}=\mathbf{b}+\Delta\mathbf{b} \tag{26}$$

Let the perturbed solution of (26) be expressed as $\hat{\mathbf{x}}=\mathbf{x}+\Delta\mathbf{x}$, where $\Delta\mathbf{x}$ is the perturbation of solution due to the perturbation of vector $\Delta\mathbf{b}$. We have the following perturbation property (see [19, 20] for derivation):

$$\frac{\|\hat{\mathbf{x}}-\mathbf{x}\|}{\|\mathbf{x}\|}=\frac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|}\leqslant\text{Cond}(\mathbf{A})\frac{\|\Delta\mathbf{b}\|}{\|\mathbf{b}\|} \tag{27}$$

where $\text{Cond}(\mathbf{A})=\|\mathbf{A}\|\|\mathbf{A}^{-1}\|$ is the condition number of matrix **A**, and $\|\|$ is the matrix norm. Note that $\text{Cond}(\mathbf{A})=\lambda_{\max}/\lambda_{\min}$ when matrix 2-norm is used.

Case II: Both the matrix **A** and vector **b** are perturbed

$$(\mathbf{A}+\Delta\mathbf{A})\tilde{\mathbf{x}}=\mathbf{b}+\Delta\mathbf{b} \tag{28}$$

Let the solution of this perturbed system be expressed as $\tilde{\mathbf{x}}=\mathbf{x}+\Delta\mathbf{x}$, where $\Delta\mathbf{x}$ is the perturbation of solution due to perturbation of matrix $\Delta\mathbf{A}$ and vector $\Delta\mathbf{b}$. We have the following perturbation property (see [19, 20] for derivation):

$$\frac{\|\tilde{\mathbf{x}}-\mathbf{x}\|}{\|\mathbf{x}\|}=\frac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|}\leqslant\frac{\text{Cond}(\mathbf{A})}{1-\text{Cond}(\mathbf{A})\frac{\|\Delta\mathbf{A}\|}{\|\mathbf{A}\|}}\left\{\frac{\|\Delta\mathbf{A}\|}{\|\mathbf{A}\|}+\frac{\|\Delta\mathbf{b}\|}{\|\mathbf{b}\|}\right\} \tag{29}$$

Example 3.1

This example studies the conditioning of matrix associated with RK approximation of a function $f(x)=\sin(\pi x)$, $0\leqslant x\leqslant 1$ by collocation, and validates the dependency of the solution perturbation to the condition number of matrix **A** as demonstrated in Table I. We consider the number of collocation points to be the same as the number of nodal points in function approximation. The kernel function $\phi_a(x-x_I)$ is chosen as the quintic B-spline defined in (8). The results in Table I show that the condition number of the matrix associated with RK approximation by collocation is fairly insensitive to the reproducing degree n and the number of nodal points. The small condition number in the linear system $\mathbf{Ax}=\mathbf{b}$ of RKCM for function approximation suggests its good stability. This stable property is reflected in Table II, where a small perturbation in **A** and **b** results in small

DOI: 10.1002/nme

perturbation in $\mathbf{x}$ due to the small condition number of $\mathbf{A}$. The matrix 2-norm is used in Table II. The results in Table II also validate the perturbation property given in (A3).

3.2. Stability of RKCM for solving boundary value problem

In this subsection, solving BVP by strong form collocation with RK approximation is introduced. For demonstration purposes, consider the following Poisson problem:

$$-\Delta u = f \quad \text{in } \Omega \tag{30}$$

$$u_n = q_1 \quad \text{on } \Gamma_N \tag{31}$$

$$u_n + \beta u = q_2 \quad \text{on } \Gamma_R \tag{32}$$

where $\partial\Omega = \Gamma_N \cup \Gamma_R$, $\Gamma_N \cap \Gamma_R = \varnothing$, Γ_N is the Neumann boundary, Γ_R is the Robin boundary, $\beta > 0$, $u_n = \nabla u \cdot \mathbf{n}$, and $\mathbf{n}$ is the outward normal. Let v be the RK approximation of u by

$$u(\mathbf{x}) \approx v(\mathbf{x}) = \sum_{I=1}^{Np} \psi_I(\mathbf{x}) a_I \tag{33}$$

Introducing approximation of u into (30)–(32), and enforce the residual to be zero at the Nc collocation points $\Xi = \{\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_{Nc}\}$, we have

$$-\sqrt{\alpha_J} \sum_{I=1}^{Np} \Delta\psi_I(\xi_J) a_I = \sqrt{\alpha_J} f(\xi_J) \quad \forall\ \xi_J \in \Omega \tag{34}$$

$$\sqrt{\alpha_J^N} \sum_{I=1}^{Np} \psi_{I,n}(\xi_J) a_I = \sqrt{\alpha_J^N} q_1(\xi_J) \quad \forall\ \xi_J \in \Gamma_N \tag{35}$$

$$\sqrt{\alpha_J^R} \sum_{I=1}^{Np} \{\psi_{I,n} + \beta\psi_I\}(\xi_J) a_I = \sqrt{\alpha_J^R} q_2(\xi_J) \quad \forall\ \xi_J \in \Gamma_R \tag{36}$$

The parameters α_J, α_J^N, α_J^R are the weights on domain and boundaries, respectively, for balanced errors between domain and boundaries, see [26] for details. Equations (34)–(36) can be rewritten as

$$\mathbf{F}\mathbf{y} = \mathbf{r} \tag{37}$$

where $\mathbf{F}$ is an $Nc \times Np$ matrix, $Nc \geqslant Np$, and $\mathbf{r}$ is an $Nc \times 1$ vector. For optimal solution accuracy, the number of collocation points Nc should be greater than the number of nodal points Np [17]. Using RK approximation in the strong form collocation in (37) has the following convergence [10]:

$$\|u - v\|_H \leqslant C a^{n-1} |u|_{n+1,\Omega} \tag{38}$$

where

$$\|v\|_H = \{\|v\|_{1,\Omega}^2 + \|\Delta v\|_{0,\Omega}^2 + \|v_n\|_{0,\Gamma_N}^2 + \|v_n + \beta v\|_{0,\Gamma_R}^2\}^{\frac{1}{2}} \tag{39}$$

According to above results, the solution convergence of RKCM requires at least quadratic bases in the RK approximation, that is $n \geqslant 2$. For the solution error in L_2 norm, we have the following bound:

$$\|u - v\|_{L_2} \leqslant C a^{n+1} |u|_{n+1,\Omega}, \quad n \geqslant 2 \tag{40}$$

The convergence properties of RKCM are verified in this numerical example. Consider the following BVP:

$$\frac{\mathrm{d}^2 u}{\mathrm{d}x^2} = -\pi^2 \sin \pi x, \quad 0 < x < 1 \tag{41}$$

$$u(0) = 0 \tag{42}$$

$$u(1) = 0 \tag{43}$$

DOI: 10.1002/nme

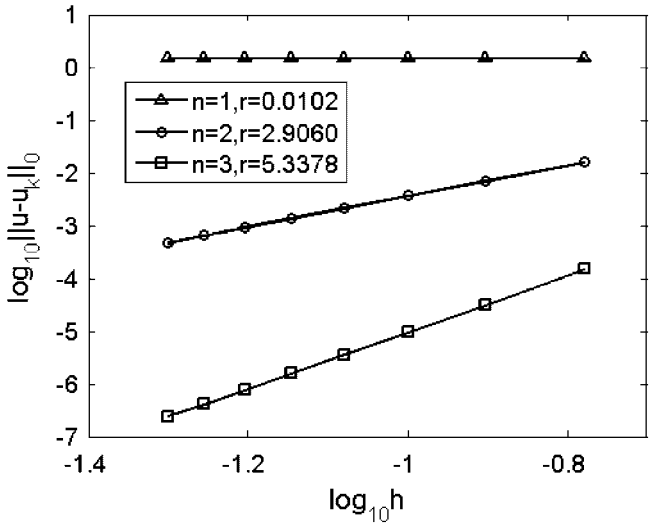

Figure 2. Solution convergence of BVP in (41)–(43) by RKCM.

The solution of this problem is $u(x)=\sin \pi x$. The following numerical parameters are employed: reproducing degrees $n=1,2,3$ numbers of points $Np=6$–21, and the number of collocations $Nc=4Np$. Equally spaced collocation points and nodal points are used, and the kernel function $\phi_a(x-x_I)$ is chosen as the quintic B-spline defined in (8), and support size a is selected as $a=(n+1)h$, where h is the nodal distance. The weights for the boundary collocation equations are selected according to Hu *et al.* [26] for balanced domain error and boundary error. The errors of solution in L_2-norm with various levels of refinement and reproducing degrees n are shown in Figure 2. The results agree with the convergence behavior of RKCM given in (38) for $n=1$ and 2, and a superconvergence behavior is observed for $n=3$. Further, as was suggested in the error analysis in (38), no convergence is achieved when linear basis is used in RKCM.

Next, we study the stability of the overdetermined discrete system constructed by RKCM of a BVP in (37). Let matrix $\mathbf{F}$ be full rank and have an SVD as follows:

$$\mathbf{U}^{\mathrm{T}}\mathbf{F}\mathbf{V}=\sum=\mathrm{daig}(\sigma_1,\sigma_2,\ldots,\sigma_{Np}) \tag{44}$$

where σ_i are the singular values of $\mathbf{F}$, $\sigma_{\max}=\sigma_1\geqslant\sigma_2\geqslant\cdots\geqslant\sigma_{Np}=\sigma_{\min}>0$, the matrices $\sum$ and $\mathbf{F}$ are with the dimension $Nc\times Np$, $\mathbf{U}=[\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_{Nc}]$, $\mathbf{U}^{\mathrm{T}}\mathbf{U}=\mathbf{I}$, $\mathbf{V}=[\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_{Np}]$, $\mathbf{V}^{\mathrm{T}}\mathbf{V}=\mathbf{I}$, $\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_{Nc}$ are the left singular vectors, and $\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_{Np}$ are the right singular vectors. It follows from (44) that

$$\mathbf{F}\mathbf{v}_i=\sigma_i\mathbf{u}_i,\quad i=1,2,\ldots,Np \tag{45}$$

Since $\mathbf{F}=\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathrm{T}}$, we have $\mathbf{F}^{\mathrm{T}}\mathbf{F}=\mathbf{V}\boldsymbol{\Sigma}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{V}^{\mathrm{T}}$. The pseudoinverse of $\mathbf{F}$, denoted as $\mathbf{F}^{+}$, is defined as

$$\mathbf{F}^{+}=(\mathbf{F}^{\mathrm{T}}\mathbf{F})^{-1}\mathbf{F}^{\mathrm{T}}=\mathbf{V}(\boldsymbol{\Sigma}^{\mathrm{T}}\boldsymbol{\Sigma})^{-1}\mathbf{V}^{\mathrm{T}}\mathbf{V}\boldsymbol{\Sigma}^{\mathrm{T}}\mathbf{U}^{\mathrm{T}}=\mathbf{V}(\boldsymbol{\Sigma}^{\mathrm{T}}\boldsymbol{\Sigma})^{-1}\boldsymbol{\Sigma}^{\mathrm{T}}\mathbf{U}^{\mathrm{T}}=:\mathbf{V}\boldsymbol{\Sigma}^{+}\mathbf{U}^{\mathrm{T}} \tag{46}$$

We further denote an orthogonal projection $\mathbf{P}$ onto $R(\mathbf{F})$ as

$$\mathbf{P}=\mathbf{F}\mathbf{F}^{+} \tag{47}$$

where $R(\mathbf{F})$ is the range of $\mathbf{F}$ spanned by all column vectors of matrix $\mathbf{F}$. This projection satisfies $\mathbf{P}^2=\mathbf{P}$. The matrix $\mathbf{I}-\mathbf{P}$ is orthogonal to the pseudoinverse $\mathbf{F}^{+}$, i.e.

$$\mathbf{F}^{+}(\mathbf{I}-\mathbf{P})=\mathbf{F}^{+}-\mathbf{F}^{+}\mathbf{P}=\mathbf{F}^{+}-\mathbf{F}^{+}\mathbf{F}\mathbf{F}^{+}=0 \tag{48}$$

The optimal solution of the overdetermined system (37) is

$$\mathbf{y}=\mathbf{F}^{+}\mathbf{r}=\mathbf{F}^{+}(\mathbf{P}+\mathbf{I}-\mathbf{P})\mathbf{r}=\mathbf{F}^{+}\mathbf{P}\mathbf{r} \tag{49}$$

DOI: 10.1002/nme

We discuss the perturbation properties of the system (37) as follows. The detail derivations are given in Appendix A.

Case I: The vector $\mathbf{r}$ on the right-hand side of the linear system (37) is perturbed

$$\mathbf{F}\hat{\mathbf{y}}=\mathbf{r}+\Delta\mathbf{r} \tag{50}$$

Let the solution of the perturbed system (50) be expressed as $\hat{\mathbf{y}}=\mathbf{y}+\Delta\mathbf{y}$. The perturbed solution due to the perturbation of right-hand side vector has the following perturbation property (see Appendix A for derivation):

$$\frac{\|\hat{\mathbf{y}}-\mathbf{y}\|}{\|\mathbf{y}\|}=\frac{\|\Delta\mathbf{y}\|}{\|\mathbf{y}\|}\leqslant \mathrm{Cond}(\mathbf{F})\frac{\|\mathbf{P}\Delta\mathbf{r}\|}{\|\mathbf{P}\mathbf{r}\|} \tag{51}$$

where $\mathrm{Cond}(\mathbf{F})=\|\mathbf{F}\|\|\mathbf{F}^{+}\|$ is the condition number of matrix $\mathbf{F}$, and $\|\|$ is the matrix norm. Note that $\mathrm{Cond}(\mathbf{F})=\sigma_{\max}/\sigma_{\min}$ when matrix 2-norm is adopted.

Case II: The matrix $\mathbf{F}$ and the vector $\mathbf{r}$ of (37) are perturbed

$$(\mathbf{F}+\Delta\mathbf{F})\tilde{\mathbf{y}}=\mathbf{r}+\Delta\mathbf{r} \tag{52}$$

Let the perturbed solution of the perturbed system (52) be expressed as $\tilde{\mathbf{y}}=\mathbf{y}+\Delta\mathbf{y}$. The perturbed solution due to the perturbation of matrix and right-hand side vector has the following perturbation property (see Appendix A for derivation):

$$\frac{\|\tilde{\mathbf{y}}-\mathbf{y}\|}{\|\mathbf{y}\|}=\frac{\|\Delta\mathbf{y}\|}{\|\mathbf{y}\|}\leqslant\frac{\left(\dfrac{\|\mathbf{r}\|}{\|\mathbf{P}\mathbf{r}\|}+\dfrac{\|\Delta\mathbf{r}\|}{\|\mathbf{P}\mathbf{r}\|}\right)\dfrac{\|\Delta\mathbf{F}\|}{\|F\|}+\mathrm{Cond}(\mathbf{F})\dfrac{\|\mathbf{P}\Delta\mathbf{r}\|}{\|\mathbf{P}\mathbf{r}\|}}{1-\left(\dfrac{\|\Delta\mathbf{F}\|^{2}}{\|\mathbf{F}\|^{2}}+(1+\mathrm{Cond}(\mathbf{F}))\dfrac{\|\Delta\mathbf{F}\|}{\|\mathbf{F}\|}\right)} \tag{53}$$

When $\Delta\mathbf{F}=\mathbf{0}$, the error bound in (53) reduces to (51). Note that these results are in alternative forms of those results given in [21], where the orthogonal projection $\mathbf{P}$ onto $R(\mathbf{F})$ was not considered and thus the higher order terms in $\Delta\mathbf{F}$ and $\Delta\mathbf{r}$ were not included in [21] compared with our results in (53).

3.3. Estimation of condition number of RKCM linear system

In this section, we provide the upper bounds for the condition number of RKCM linear system. As a first step, it is crucial to realize that the collocation method in (34)–(36) solved by least-squares method is equivalent to the minimization of discrete least-squares functional with quadrature [17] as shown below

$$\hat{E}(u^{R})=\min_{v\in V}\hat{E}(v) \tag{54}$$

where V is the finite-dimensional space (14), $\hat{E}(\cdot)$ is a quadrature version of least-squares functional $E(\cdot)$ defined as

$$\hat{E}(v)=\tfrac{1}{2}\{\hat{\int}_{\Omega}(\Delta v+f)^{2}\mathrm{d}\Omega+\hat{\int}_{\Gamma_N}(v_n-q_1)^{2}\mathrm{d}\ell+\hat{\int}_{\Gamma_R}(v_n+\beta v-q_2)^{2}\mathrm{d}\ell\} \tag{55}$$

Here, $\hat{\int}$ denotes the quadrature version of $\int$. Minimization of least-squares functional in (55) yields

$$\hat{B}(u^{R},v)=\hat{F}(v) \tag{56}$$

where

$$\hat{B}(u,v)=\hat{\int}_{\Omega}\Delta u\Delta v\mathrm{d}\Omega+\hat{\int}_{\Gamma_N}u_n v_n\mathrm{d}\ell+\hat{\int}_{\Gamma_R}(u_n+\beta v)(v_n+\beta v)\mathrm{d}\ell \tag{57}$$

$$\hat{F}(v)=-\hat{\int}_{\Omega}f\Delta v\mathrm{d}\Omega+\hat{\int}_{\Gamma_N}q_1 v_n\mathrm{d}\ell+\hat{\int}_{\Gamma_R}q_2(v_n+\beta v)\mathrm{d}\ell \tag{58}$$

DOI: 10.1002/nme

We define two norms for stability analysis

$$\|v\|_E = B(v,v)^{1/2}, \quad \widehat{\|v\|}_E = \hat{B}(v,v)^{1/2} \tag{59}$$

where $B(v,v)$ is the continuous version of $\hat{B}(v,v)$. The discrete norm in (B7) can be expressed as a quadric form

$$\widehat{\|v\|}_E^2 = \mathbf{y}^{\mathrm{T}}\mathbf{F}^{\mathrm{T}}\mathbf{F}\mathbf{y} =: \mathbf{y}^{\mathrm{T}}\mathbf{G}\mathbf{y} \tag{60}$$

where matrix $\mathbf{F}$ and vector $\mathbf{y}$ are given in (34)–(37), and matrix $\mathbf{G}=\mathbf{F}^{\mathrm{T}}\mathbf{F}$ is symmetric and positive definite. There exist the following relationships:

$$c_1\|v\|_E \leqslant \widehat{\|v\|}_E \leqslant c_2\|v\|_E \tag{61}$$

$$c_3\|v\|_{0,\Omega} \leqslant \|v\|_E \leqslant c_4\|v\|_{2,\Omega} \tag{62}$$

where c_i are generic constants and $\|\cdot\|_{2,\Omega}$ is the Sobolev H^2-norm. We define a discrete zero norm of $\|v\|_{0,\Omega}=(v,v)^{1/2}$, denoted as $\widehat{\|v\|}_{0,\Omega}$

$$\widehat{\|v\|}_{0,\Omega}^2 = \mathbf{y}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{y} =: \mathbf{y}^{\mathrm{T}}\mathbf{N}\mathbf{y} \tag{63}$$

where $A_{IJ}=\psi_J(\xi_I)$is the collocation matrix of RK shape functions with $\mathrm{Cond}(\mathbf{N})\approx O(1)$. Using the Rayleigh–Ritz theorem and the inequalities as well as (61) and (62), we obtain the bounds for maximal and minimal eigenvalues of matrix $\mathbf{G}$ as follows:

$$\lambda\min(\mathbf{G}) = \min\frac{\mathbf{y}^{\mathrm{T}}\mathbf{G}\mathbf{y}}{\mathbf{y}^{\mathrm{T}}\mathbf{y}} \geqslant \min\frac{\mathbf{y}^{\mathrm{T}}\mathbf{G}\mathbf{y}}{\mathbf{y}^{\mathrm{T}}\mathbf{N}\mathbf{y}}\cdot\min\frac{\mathbf{y}^{\mathrm{T}}\mathbf{N}\mathbf{y}}{\mathbf{y}^{\mathrm{T}}\mathbf{y}} \geqslant \min\frac{c_1c_3\|v\|_{0,\Omega}^2}{\|v\|_{0,\Omega}^2}\cdot\lambda\min(\mathbf{N}) \tag{64}$$

$$\lambda\max(\mathbf{G}) = \max\frac{\mathbf{y}^{\mathrm{T}}\mathbf{G}\mathbf{y}}{\mathbf{y}^{\mathrm{T}}\mathbf{y}} \leqslant \max\frac{\mathbf{y}^{\mathrm{T}}\mathbf{G}\mathbf{y}}{\mathbf{y}^{\mathrm{T}}\mathbf{N}\mathbf{y}}\cdot\max\frac{\mathbf{y}^{\mathrm{T}}\mathbf{N}\mathbf{y}}{\mathbf{y}^{\mathrm{T}}\mathbf{y}} \leqslant \max\frac{c_2c_4\|v\|_{2,\Omega}^2}{\|v\|_{0,\Omega}^2}\cdot\lambda\max(\mathbf{N}) \tag{65}$$

Furthermore, by means of the inverse inequality (16) in Lemma 1, the condition number for matrix $\mathbf{G}$ is given by

$$\mathrm{Cond}(\mathbf{G}) = \frac{\lambda\max(\mathbf{G})}{\lambda\min(\mathbf{G})} \leqslant C_1\frac{\|v\|_{2,\Omega}^2}{\|v\|_{0,\Omega}^2}\cdot\mathrm{Cond}(\mathbf{N}) \leqslant C_2\frac{\|v\|_{2,\Omega}^2}{\|v\|_{0,\Omega}^2} \leqslant C_3\kappa a^{-4}n^8 \tag{66}$$

where κ is the overlapping constant of reproducing kernel and n is the reproducing degree. Therefore, the condition number of matrix F has the following bound:

$$\mathrm{Cond}(\mathbf{F}) = \{\mathrm{Cond}(\mathbf{G})\}^{1/2} \leqslant \tilde{C}k^{1/2}a^{-2}n^4 \tag{67}$$

According to the bound in (67), we see that the condition number of $\mathbf{F}$ is closely related to the support size of the RK shape function. As the support size is proportional to the nodal distance, for example, $a=(n+1)\cdot h$, and n is typically small, we have

$$\mathrm{Cond}(\mathbf{F}) \leqslant C\frac{n^4}{(n+1)^2}\cdot h^{-2} \approx O(h^{-2}) \tag{68}$$

This result shows that the condition number increases with the order of $O(h^{-2})$, similar to that of FEM. It also indicates that increasing the reproducing degree n yields a marginal increase in the condition number for small n.

Example 3.2

The condition numbers of a linear system resulting from RKCM discretization of BVP (41)–(43) with different reproducing degrees $n=2$, 3 ($n=1$ does not converge) are demonstrated in Table III. The results show that the condition numbers of matrix $\mathbf{F}$ in RKCM are insensitive

DOI: 10.1002/nme

Table III. The condition numbers of matrix **F** associated with the BVP in (41)–(43) solved by RKCM.

$n=2$		$n=3$	
Np	Cond(**F**)	Np	Cond(**F**)
6	60.42	6	67.99
11	184.69	11	191.45
21	549.76	21	565.44

Table IV. Stability of RKCM linear system associated with the BVP in (41)–(43) and the corresponding traditional condition number.

	$n=2,\ Np=21$	$n=3,\ Np=21$
Cond(**F**)	549.76	565.44
$\|\Delta\mathbf{F}\|$	1×10^{-13}	1×10^{-13}
$\|\mathbf{F}\|$	9453.6	6599.1
$\|\mathbf{P}\Delta\mathbf{r}\|$	0.15×10^{-13}	0.17×10^{-13}
$\|\mathbf{Pr}\|$	62.4	62.4
$\|\Delta\mathbf{y}\|$	8.62×10^{-15}	2.54×10^{-15}
$\|\mathbf{y}\|$	3.16	3.16

to the reproducing degree n, while they increase significantly as the number of nodal points Np increases. By examining the numerical results in Table III the condition numbers increase by five times when the number of nodal points Np is doubled. In addition, the condition numbers slightly increase as the reproducing degree n increases from 2 to 3, consistent with analytical result in (68). The results in (38) and (68) suggest that increasing reproducing order in RK approximation is an effective way of achieving higher convergent rate in solving PDE by RKCM without sacrificing stability.

To identify stability, we then examine how a small perturbation in **F** and **r** affects the perturbation in **x** as shown in Table IV. Interestingly, even with large condition numbers in matrix **F** in this case, the perturbations in the solution are still at the order of the small perturbations in matrix **F** and vector **r**. This indicates that using the tradition condition number overexaggerates the ill-conditioning of a linear system, and consequently the bounds in the error estimates in (51) and (53) become too loose. As shown in Table IV, the measure of conditioning using maximum and minimum eigenvalues in the traditional condition number over-amplifies the ill-conditioning of a matrix, and thus yields misleading instability estimate of the linear system according to the numerical test in Table IV and prediction in (41)–(43). This has also been observed in [27–29]. The stability of the collocation methods for the BVP may not be as ill-conditioned as what the traditional condition number indicates. The effective condition number will be introduced in the following section which provides a more precious conditioning measure of a linear system.

4. STABILITY ESTIMATION BY AN EFFECTIVE CONDITION NUMBER

The perturbation and stability analysis in Section 3 shows that condition number plays a critical role in the stability of linear systems. However, the numerical results in Table IV suggest the need for a more precious conditioning measure to provide tighter bounds in stability estimation in (41)–(43). Here, we discuss an alternative measure of matrix conditioning and its implication to the stability of a linear system. We start with a matrix 2-norm used in the traditional definition of condition number.

DOI: 10.1002/nme

(i) For a square linear system $\mathbf{Ax}=\mathbf{b}$, where $\mathbf{A}\in\mathbb{R}^{Np\times Np}$, $\mathbf{b}\in\mathbb{R}^{Np}$

$$\mathrm{Cond}(\mathbf{A})=\frac{\lambda_{\max}}{\lambda_{\min}} \tag{69}$$

where $\lambda_{\max}$ and $\lambda_{\min}$ are the maximum and minimum eigenvalues of $\mathbf{A}$.

(ii) For an overdetermined linear system $\mathbf{Fy}=\mathbf{r}$, where $\mathbf{F}\in\mathbb{R}^{Nc\times Np}$, $Nc>Np$, $\mathbf{r}\in\mathbb{R}^{Nc}$

$$\mathrm{Cond}(\mathbf{F})=\frac{\sigma_{\max}}{\sigma_{\min}} \tag{70}$$

where $\sigma_{\max}$ and $\sigma_{\min}$ are the maximum and minimum singular values of $\mathbf{F}$

Various measures of condition numbers have been proposed [22–24]. An effective condition number has been introduced by considering the perturbation properties of matrix and right-hand side vector in a linear system based on eigenvector expansion. The detailed derivations are given in Appendix B, and here we summarize the important results. The main difference in perturbation analysis using the traditional condition number and the improved effective condition number is on the estimation of lower bound for norm $\|\mathbf{x}\|$. This is reflected in (B12) in Appendix B, where a rigorous derivation is given. Similarly, for the overdetermined system, the estimation of lower bounds for norm $\|\mathbf{y}\|$ given in (A7) of Appendix A for the traditional condition number is more rigorously estimated in (B25) of Appendix B for the effective condition number. The comparison of the effective condition number and the corresponding perturbation analysis in Appendix B are first summarized below, and the numerical tests are used to validate the analytical stability estimation. According to Christiansen and Hansen [23] and Huang and Li [25], as described in Appendix B, the effective condition number has the following form:

$$\mathrm{Cond}_{\mathrm{Eff}}(\mathbf{A})=\frac{\|\mathbf{b}\|}{\sqrt{\dfrac{\|\mathbf{b}\|^2-\beta_{\min}^2}{\mathrm{Cond}(\mathbf{A})^2}+\beta_{\min}^2}} \tag{71}$$

where $\beta_{\min}=\beta_{Np}$ is an inner product of the Npth eigenvector of matrix $\mathbf{A}$ and the right-hand side vector $\mathbf{b}$, i.e.

$$\beta_{\min}=\mathbf{u}_{Np}^{\mathrm{T}}\mathbf{b} \tag{72}$$

The effective condition number defined in (71) can be rewritten as

$$\mathrm{Cond}_{\mathrm{Eff}}(\mathbf{A})=\mathrm{Cond}(\mathbf{A})\times\frac{\sqrt{\sum_{i=1}^{Np}\beta_i^2}}{\sqrt{\sum_{i=1}^{Np-1}\beta_i^2+\mathrm{Cond}(\mathbf{A})^2\cdot\beta_{Np}^2}} \tag{73}$$

As shown in (73), the effective condition number is always smaller than the traditional condition number if the value $\mathrm{Cond}(\mathbf{A})^2$ in the denominator of (73) is greater than 1. This gives a tighter stability estimation by perturbation analysis of the solution $\mathbf{x}$ due to the perturbation of the vector $\mathbf{b}$ (see Appendix B for details):

$$\frac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|}\leqslant\mathrm{Cond}_{\mathrm{Eff}}(\mathbf{A})\frac{\|\Delta\mathbf{b}\|}{\|\mathbf{b}\|} \tag{74}$$

Further, the perturbation of solution $\mathbf{x}$ due to perturbation of matrix $\mathbf{A}$ and vector $\mathbf{b}$ can be derived following the same procedures in Appendices A and B:

$$\frac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|}\leqslant\frac{1}{1-\|\mathbf{A}^{-1}\|\|\Delta\mathbf{A}\|}\left\{\mathrm{Cond}(\mathbf{A})\frac{\|\Delta\mathbf{A}\|}{\|\mathbf{A}\|}+\mathrm{Cond}_{\mathrm{Eff}}(\mathbf{A})\frac{\|\Delta\mathbf{b}\|}{\|\mathbf{b}\|}\right\} \tag{75}$$

Note that when $\beta_{\min}=0$, the effective condition number in (71) reduces to the traditional condition number. In general, $\beta_{\min}\neq 0$, and from (73) we have:

$$\mathrm{Cond}_{\mathrm{Eff}}(\mathbf{A})<\mathrm{Cond}(\mathbf{A}) \tag{76}$$

DOI: 10.1002/nme

In fact, the value of effective condition number is much smaller than the traditional condition number in practice. For an overdetermined linear system, the effective condition number of matrix **F** based on eigenvector expansion has the following form (see Appendix B for details):

$$\mathrm{Cond}_{\mathrm{Eff}}(\mathbf{F}) = \frac{\|\mathbf{Pr}\|}{\sqrt{\dfrac{\|\mathbf{Pr}\|^2 - \gamma_{\min}^2}{\mathrm{Cond}(\mathbf{F})^2} + \gamma_{\min}^2}} \tag{77}$$

where $\gamma_{\min}$ is an inner product of the Npth left singular vector of matrix **F** and projection vector **Pr**, i.e.

$$\gamma_{\min} = \mathbf{u}_{Np}^{\mathrm{T}} \mathbf{Pr} \tag{78}$$

Note that this result is a correction of [24] in which the orthogonal projection **P** onto $R(\mathbf{F})$ was not considered. There exist tighter perturbation bounds for perturbed problems with perturbation in vector

$$\frac{\|\Delta \mathbf{y}\|}{\|\mathbf{y}\|} \leqslant \mathrm{Cond}_{\mathrm{Eff}}(\mathbf{F}) \frac{\|\mathbf{P}\Delta \mathbf{r}\|}{\|\mathbf{Pr}\|} \tag{79}$$

and perturbations in vector and matrix

$$\frac{\|\Delta \mathbf{y}\|}{\|\mathbf{y}\|} \leqslant \frac{1}{1 - \|\mathbf{F}^{+}\| \|\Delta \mathbf{F}\|} \left\{ \mathrm{Cond}(\mathbf{F}) \frac{\|\Delta \mathbf{F}\|}{\|\mathbf{F}\|} + \mathrm{Cond}_{\mathrm{Eff}}(\mathbf{F}) \frac{\|\mathbf{P}\Delta \mathbf{r}\|}{\|\mathbf{Pr}\|} \right\}. \tag{80}$$

We observe that the $\mathrm{Cond}_{\mathrm{Eff}}(\mathbf{F})$ reduces to $\mathrm{Cond}(\mathbf{F})$ when $\gamma_{\min} = 0$. Similarly, the effective condition number is much smaller than the traditional one, which will be numerically validated in the numerical examples

$$\mathrm{Cond}_{\mathrm{Eff}}(\mathbf{F}) < \mathrm{Cond}(\mathbf{F}) \tag{81}$$

We further give an estimate of the bound of $\mathrm{Cond}_{\mathrm{Eff}}(\mathbf{F})$. From (77), one has

$$\mathrm{Cond}_{\mathrm{Eff}}(\mathbf{F}) \leqslant \frac{\|\mathbf{Pr}\|}{|\gamma_{\min}|} \tag{82}$$

Recall BVP in (30)–(32) with the Dirichlet boundaries and their weighted collocation in (34)–(36), we have the following bound for s-dimension:

$$\|\mathbf{Pr}\| \leqslant \sqrt{\alpha_J} \hbar^{-s} \|f\|_{0,\Omega} + \sqrt{\alpha_J^R} \hbar^{-1} \|q_2\|_{0,\Gamma_R} \tag{83}$$

where $\hbar$ denotes the spacing of collocation points given in Section 3.2, and $\hbar = O(h) = O(Np^{1/s})$. For balanced domain and boundary errors derived in [26], we have introduced the following weights for the weighted collocation (34)–(36):

$$\sqrt{\alpha_J} = O(1), \quad \sqrt{\alpha_J^R} = O(N_p) \tag{84}$$

Further considering the order of $|\gamma_{\min}|$:

$$|\gamma_{\min}| \approx O(\hbar^{-s}) \tag{85}$$

Hence, the estimate in (82) becomes

$$\mathrm{Cond}_{\mathrm{Eff}}(\mathbf{F}) \leqslant C_1 \|f\|_{0,\Omega} + C_2 N_p^{1/s} \|q_2\|_{0,\Gamma_R} \tag{86}$$

This implies that when the domain term is dominating, $\mathrm{Cond}_{\mathrm{Eff}}(\mathbf{F}) \approx O(1)$.

Example 4.1

We re-evaluate the stability of the RKCM linear system associated with the Dirichlet BVP in (41)–(43) using the effective condition number as shown in Table V. It is clear that using the effective

DOI: 10.1002/nme

Table V. Stability of RKCM linear system associated with the BVP in (41)–(43) and the corresponding effective condition number.

	$n=2,\ Np=21$	$n=3,\ Np=21$
$\text{Cond}_{\text{Eff}}(\mathbf{F})$	3.54	3.54
$\|\Delta\mathbf{F}\|$	1×10^{-13}	1×10^{-13}
$\|\mathbf{F}\|$	9453.6	6599.1
$\|\mathbf{P}\Delta\mathbf{r}\|$	0.15×10^{-13}	0.17×10^{-13}
$\|\mathbf{Pr}\|$	62.4	62.4
$\|\Delta\mathbf{y}\|$	8.62×10^{-15}	2.54×10^{-15}
$\|\mathbf{y}\|$	3.16	3.16

Table VI. The condition numbers for RKCM linear system of (41)–(43) with reproducing degree $n=2$.

Np	$h=1/(Np-1)$	$\text{Cond}(\mathbf{F})$	$\text{Cond}_{\text{Eff}}(\mathbf{F})$	$\sigma_{\max}$	$\sigma_{\min}$	$\lvert\gamma_{\min}\rvert$
6	1/5	60.42	7.24	207.3	3.43	4.20
11	1/10	184.69	4.98	848.9	4.59	8.83
21	1/20	549.75	3.54	3420.7	6.22	17.63
41	1/40	1640.32	2.56	13715.2	8.36	34.43
101	1/100	7263.23	1.75	85788.4	11.81	79.56

Table VII. The condition numbers for RKCM linear system of (41)–(43) with reproducing degree $n=3$.

Np	$h=1/(Np-1)$	$\text{Cond}(\mathbf{F})$	$\text{Cond}_{\text{Eff}}(\mathbf{F})$	$\sigma_{\max}$	$\sigma_{\min}$	$\lvert\gamma_{\min}\rvert$
6	1/5	67.99	7.08	233.2	3.43	4.38
11	1/10	191.45	4.97	879.8	4.59	8.88
21	1/20	565.44	3.54	3518.1	6.22	17.64
41	1/40	1683.07	2.56	14072.5	8.36	34.43
101	1/100	7446.50	1.75	87953.1	11.81	79.56

condition in conjunction with the perturbation analysis in (79) yields a much tighter bound for a more precious stability estimation compared with the results in Table IV of Example 3.2 using the traditional condition numbers.

More detailed comparisons of the traditional and effective condition numbers of the matrix $\mathbf{F}$ resulting from RKCM discretization of (41)–(43) using various reproducing order and discretization points are given in Tables VI and VII. We observe from Tables VI and VII that $\text{Cond}(\mathbf{F})$'s are much larger than $\text{Cond}_{\text{Eff}}(\mathbf{F})$'s for larger $|\gamma_{\min}|$. Further, the numerical data confirm the estimated asymptotic behavior of $\text{Cond}(\mathbf{F})\approx O(Np^2)\approx O(h^{-2})$ and $\text{Cond}_{\text{Eff}}(\mathbf{F})\approx O(1)$.

Example 4.2
Consider another BVP:

$$\frac{\mathrm{d}^2u}{\mathrm{d}x^2}=e^x,\quad 0<x<1 \tag{87}$$

$$u(0)=1 \tag{88}$$

$$u(1)=e \tag{89}$$

In comparison with the problem (41)–(43), their linear systems have the same matrices but with different right-hand side vectors.

We list the condition numbers for reproducing degree $n=2$ and $n=3$ in Tables VIII and IX, respectively. The results show that the traditional condition numbers are about the same order as

DOI: 10.1002/nme

Table VIII. The condition numbers for RKCM linear system of (87)–(89) with reproducing degree $n=2$.

Np	$h=1/(Np-1)$	Cond(F)	$\text{Cond}_{\text{Eff}}(F)$	$\sigma_{\max}$	$\sigma_{\min}$	$\lvert\gamma_{\min}\rvert$
6	1/5	60.41	1.29	207.3	3.43	14.67
11	1/10	184.70	1.28	848.9	4.59	26.37
21	1/20	549.76	1.28	3420.7	6.22	49.07
41	1/40	1640.30	1.32	13715.2	8.36	91.65
101	1/100	7263.23	1.46	85788.4	11.81	200.65

Table IX. The condition numbers for RKCM linear system of (87)–(89) with reproducing degree $n=3$.

Np	$h=1/(Np-1)$	Cond($\mathbf{F}$)	$\text{Cond}_{\text{Eff}}(\mathbf{F})$	$\sigma_{\max}$	$\sigma_{\min}$	$\lvert\gamma_{\min}\rvert$
6	1/5	67.99	1.29	233.2	3.43	14.66
11	1/10	191.44	1.28	879.8	4.59	26.37
21	1/20	565.44	1.28	3518.1	6.22	49.07
41	1/40	1683.07	1.32	14072.5	8.36	91.65
101	1/100	7446.51	1.46	87953.1	11.81	200.64

those given in Tables VI and VII, whereas the effective condition numbers are much smaller in the linear system of (87)–(89) compared with those of the linear system of (41)–(43) given in Tables VI and VII. These results are consistent with the definition of traditional condition number and effective condition number, in that the traditional condition considers only the property of the matrix, whereas the effective condition number takes the properties of both the matrix and the right-hand side vector into account. We also note that the results show an asymptotic behavior $\text{Cond}_{\text{Eff}}(\mathbf{F})\approx O(1)$ different from $\text{Cond}(\mathbf{F})\approx O(h^{-2})$ as shown in (68).

Example 4.3

Here, we validate the stability analysis by considering a two-dimensional problem given below

$$\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}=-2\pi^2\sin(\pi x)\sin(\pi y),\quad 0<x<1,\ \ 0<y<1 \tag{90}$$

$$u=0\quad \text{on}\quad \begin{cases} y=0, & 0<x<1\\ y=1, & 0<x<1\\ x=0, & 0<y<1\\ x=1, & 0<y<1 \end{cases} \tag{91}$$

We first consider pure Dirichlet boundary conditions in (91). This problem is solved by RKCM with reproducing degree $n=2$ and kernel function constructed by the multiplication of one-dimensional quintic B-spline functions. In each direction of the two-dimensional domain, the density of collocation points is twice the density of the nodal points in each direction and thus yields an overdetermined discrete system. The weights for the boundary collocation equations are selected according to Hu *et al.* [26] for balanced domain error and boundary error. The results in Table X again show a much smaller effective condition numbers than the traditional condition numbers. It is also noticed that the effective condition numbers do not increase as the number of nodal points Np increases. The perturbation of RKCM solution in response to the perturbation of the matrix and right-hand side vector, as shown in Tables X and XI, shows the excellent stability of RKCM and the good performance of the effective condition number in estimating the stability of the discrete system. The numerical results demonstrate good agreement with the estimated asymptotic behavior $\text{Cond}_{\text{Eff}}(\mathbf{F})\approx O(1)$ that is different from $\text{Cond}(\mathbf{F})\approx O(h^{-2})$ as shown in (68).

DOI: 10.1002/nme

Table X. The condition numbers for RKCM linear system associated with (90) and (91) with reproducing degree $n=2$.

Np	$\mathrm{Cond}(\mathbf{F})$	$\mathrm{Cond}_{\mathrm{Eff}}(\mathbf{F})$	$\sigma_{\max}$	$\sigma_{\min}$	$\lvert\gamma_{\min}\rvert$
6^2	6.89	1.49	201.9	29.28	65.02
11^2	22.49	1.06	841.5	37.41	186.43
21^2	87.22	1.01	3415.7	39.16	391.56
41^2	347.97	1.001	13722.2	39.43	788.68
81^2	1392.18	1.0001	54953.5	39.47	1578.90

Table XI. Stability of RKCM linear system associated with the BVP in (90) and (91) and the corresponding effective condition number.

	$n=2,\ Np=21^2$
$\mathrm{Cond}_{\mathrm{Eff}}(\mathbf{F})$	1.01
$\lVert\Delta\mathbf{F}\rVert$	1×10^{-13}
$\lVert\mathbf{F}\rVert$	3415.7
$\lVert\mathbf{P}\Delta\mathbf{r}\rVert$	0.9×10^{-13}
$\lVert\mathbf{Pr}\rVert$	394.76
$\lVert\Delta\mathbf{y}\rVert$	1.86×10^{-14}
$\lVert\mathbf{y}\rVert$	9.99

5. CONCLUSION

In this paper, we discuss the stability of RKCM for solving BVPs. Using RKCM for solution of BVPs leads to sparse linear systems as opposed to the popularly used strong form collocation based on RBFs, collectively called the RBCM [5–7, 16, 17]. Although RKCM approach offers algebraic convergence (compared with exponential convergent RBCM), the method is very stable according to our stability analysis and offers a significant stability enhancement over RBCM. More specifically, we proved that the condition numbers of the matrix in RKCM are of the order of:

$$\mathrm{Cond}(\mathbf{F})\approx O(h^{-2}) \tag{92}$$

This stability property is similar to that of the finite element or finite difference method.

From the perturbation analysis, we show how the stability of an overderetmined linear system $\mathbf{Fy}=\mathbf{r}$ is closely related to the condition number of the overdetermined matrix $\mathbf{F}$. We obtained the bound of the perturbation in solution $\mathbf{y}$ due to the perturbation of overdetermined matrix $\mathbf{F}$ and vector $\mathbf{r}$

$$\frac{\lVert\Delta\mathbf{y}\rVert}{\lVert\mathbf{y}\rVert}\leqslant\frac{\left(\dfrac{\lVert\mathbf{r}\rVert}{\lVert\mathbf{Pr}\rVert}+\dfrac{\lVert\Delta\mathbf{r}\rVert}{\lVert\mathbf{Pr}\rVert}\right)\dfrac{\lVert\Delta\mathbf{F}\rVert}{\lVert\mathbf{F}\rVert}+\mathrm{Cond}(\mathbf{F})\dfrac{\lVert\mathbf{P}\Delta\mathbf{r}\rVert}{\lVert\mathbf{Pr}\rVert}}{1-\left(\dfrac{\lVert\Delta\mathbf{F}\rVert}{\lVert\mathbf{F}\rVert}^2+(1+\mathrm{Cond}(\mathbf{F}))\dfrac{\lVert\Delta\mathbf{F}\rVert}{\lVert\mathbf{F}\rVert}\right)} \tag{93}$$

where $\mathbf{P}$ is the projection operator. Further, we study an alternative measure of the stability of a linear system based on the effective condition number $\mathrm{Cond}_{\mathrm{Eff}}(\cdot)$. The effective condition number offers a better measure of conditioning of a linear system than the traditional condition numbers, where both matrix and right-hand side vector are taken into account in the effective condition number, that is

$$\mathrm{Cond}(\cdot)=\mathrm{Cond}(\mathbf{F}) \tag{94}$$

$$\mathrm{Cond}_{\mathrm{Eff}}(\cdot)=\mathrm{Cond}(\mathbf{F},\mathbf{Pr}) \tag{95}$$

DOI: 10.1002/nme

The numerical results verified that the use of effective condition number gives a tighter bound of perturbation properties in Section 4, and offers a more precious estimate of the stability of linear systems. The traditional conditional number, on the other hand, could lead to a significant overexaggeration of the ill-conditioning of a linear system under certain conditions. In general, we obtain the relationship $\mathrm{Cond}_{\mathrm{Eff}}(\cdot) < \mathrm{Cond}(\cdot)$ and its effectiveness in improving the stability estimation of the linear systems has been analytically proved and numerically validated. Further, the effective condition number shows that the stability of RKCM is not affected by the refinement in the discretization:

$$\mathrm{Cond}_{\mathrm{Eff}}(\mathbf{F}) \approx O(1) \tag{96}$$

This estimation agrees well with the numerical results in the one- and two-dimensional verification problems.

APPENDIX A

The detail perturbation analysis for overdetermined linear systems is derived in this section. The results are in the alternative forms of those given in [21].

Theorem A.1
Let $\mathbf{F}\mathbf{y} = r$, $\mathbf{F} \in \mathbb{R}^{m \times n}$, $m \geqslant n$, be a full rank non-square matrix, $\mathbf{y} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^m$. Consider a perturbed problem $\mathbf{F}(\mathbf{y} + \Delta\mathbf{y}) = (\mathbf{r} + \Delta\mathbf{r})\Delta\mathbf{y} \in \mathbb{R}^n$, $\Delta\mathbf{r} \in \mathbb{R}^m$, then

$$\frac{\|\Delta\mathbf{y}\|}{\|\mathbf{y}\|} \leqslant \mathrm{Cond}(\mathbf{F}) \frac{\|\mathbf{P}\Delta\mathbf{r}\|}{\|\mathbf{P}\mathbf{r}\|} \tag{A1}$$

where

$$\mathrm{Cond}(\mathbf{F}) = \frac{\sigma_{\max}}{\sigma_{\min}} \tag{A2}$$

and $\mathbf{P} = \mathbf{F}\mathbf{F}^{+}$ is the orthogonal projection.

Proof
Suppose $\mathbf{F}$ has an SVD, $\mathbf{F} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathrm{T}}$, where $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_i)$, σ_i are singular values, with $\sigma_{\max} = \sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_n = \sigma_{\min} > 0$. The matrices $\mathbf{U}$, $\mathbf{V}$ are orthonormal, i.e., $\mathbf{U}^{\mathrm{T}}\mathbf{U} = I$, $\mathbf{V}^{\mathrm{T}}\mathbf{V} = I$, and $\mathbf{F}\mathbf{v}_i = \sigma_i \mathbf{u}_i, \forall i$. Let $\mathbf{y}$ and $\hat{\mathbf{y}} = \mathbf{y} + \Delta\mathbf{y}$ are the optimal solution of system $\mathbf{F}\mathbf{y} = \mathbf{r}$ and $\mathbf{F}\hat{\mathbf{y}} = \mathbf{r} + \Delta\mathbf{r}$, respectively. We have

$$\Delta\mathbf{y} = \hat{\mathbf{y}} - \mathbf{y} = \mathbf{F}^{+}(\mathbf{r} + \Delta\mathbf{r}) - \mathbf{F}^{+}\mathbf{r} = \mathbf{F}^{+}\Delta\mathbf{r} = \mathbf{F}^{+}(\mathbf{P}\Delta\mathbf{r} + (\mathbf{I} - \mathbf{P})\Delta\mathbf{r}) \tag{A3}$$

where $\mathbf{F}^{+} = (\mathbf{F}^{\mathrm{T}}\mathbf{F})^{-1}\mathbf{F}^{\mathrm{T}}$ is the pseudoinverse of $\mathbf{F}$.

Using the relation (48) in Section 3.2 yields

$$\Delta\mathbf{y} = \mathbf{F}^{+}\mathbf{P}\Delta\mathbf{r} + \mathbf{F}^{+}(\mathbf{I} - \mathbf{P})\Delta\mathbf{r} = \mathbf{F}^{+}\mathbf{P}\Delta\mathbf{r} \tag{A4}$$

Consequently, we have

$$\|\Delta\mathbf{y}\| \leqslant \|\mathbf{F}^{+}\| \|\mathbf{P}\Delta\mathbf{r}\| \tag{A5}$$

As $\mathbf{y} = \mathbf{F}^{+}\mathbf{P}\mathbf{r}$, thus $\mathbf{F}\mathbf{y} = \mathbf{P}\mathbf{r}$, and therefore

$$\|\mathbf{P}\mathbf{r}\| = \|\mathbf{F}\mathbf{y}\| \leqslant \|\mathbf{F}\| \|\mathbf{y}\| \tag{A6}$$

Rewrite (A6) as

$$\frac{1}{\|\mathbf{y}\|} \leqslant \frac{\|\mathbf{F}\|}{\|\mathbf{P}\mathbf{r}\|} \tag{A7}$$

DOI: 10.1002/nme

Combining (A5) and (A7), we have

$$\frac{\|\Delta\mathbf{y}\|}{\|\mathbf{y}\|}\leqslant\|\mathbf{F}^{+}\|\|\mathbf{F}\|\frac{\|\mathbf{P}\Delta\mathbf{r}\|}{\|\mathbf{P}\mathbf{r}\|}=:\mathrm{Cond}(\mathbf{F})\frac{\|\mathbf{P}\Delta\mathbf{r}\|}{\|\mathbf{P}\mathbf{r}\|} \tag{A8}$$

Consider the matrix 2-norm and use the unitary invariant properties

$$\|\mathbf{F}\|=\|\mathbf{F}\|_2=\|\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathrm{T}}\|_2=\|\boldsymbol{\Sigma}\mathbf{V}^{\mathrm{T}}\|_2=\|\boldsymbol{\Sigma}\|_2=\sigma_{\max} \tag{A9}$$

and

$$\|\mathbf{F}^{+}\|=\|\mathbf{F}^{+}\|_2=\|\mathbf{U}\boldsymbol{\Sigma}^{+}\mathbf{V}^{\mathrm{T}}\|_2=\|\boldsymbol{\Sigma}^{+}\mathbf{V}^{\mathrm{T}}\|_2=\|\boldsymbol{\Sigma}^{+}\|_2=\frac{1}{\sigma_{\min}} \tag{A10}$$

Thus, the condition number is

$$\mathrm{Cond}(\mathbf{F})=\frac{\sigma_1}{\sigma_n}=\frac{\sigma_{\max}}{\sigma_{\min}} \tag{A11}$$

Theorem A.2
Let $\mathbf{F}\mathbf{y}=\mathbf{r}$, $\mathbf{F}\in\mathbb{R}^{m\times n}$, $m\geqslant n$, be full rank, $\mathbf{y}\in\mathbb{R}^n$, $\mathbf{b}\in\mathbb{R}^m$, and consider the perturbed problem $(\mathbf{F}+\Delta\mathbf{F})(\mathbf{y}+\Delta\mathbf{y})=(\mathbf{r}+\Delta\mathbf{r})\Delta\mathbf{F}\in\mathbb{R}^{m\times n}$, $\Delta\mathbf{y}\in\mathbb{R}^n$, $\Delta\mathbf{r}\in\mathbb{R}^m$, then

$$\frac{\|\Delta\mathbf{y}\|}{\|\mathbf{y}\|}\leqslant\frac{\left(\dfrac{\|\mathbf{r}\|}{\|\mathbf{P}\mathbf{r}\|}+\dfrac{\|\Delta\mathbf{r}\|}{\|\mathbf{P}\mathbf{r}\|}\right)\dfrac{\|\Delta\mathbf{F}\|}{\|\mathbf{F}\|}+\mathrm{Cond}(\mathbf{F})\dfrac{\|\mathbf{P}\Delta\mathbf{r}\|}{\|\mathbf{P}\mathbf{r}\|}}{1-\left(\dfrac{\|\Delta\mathbf{F}\|^2}{\|\mathbf{F}\|^2}+(1+\mathrm{Cond}(\mathbf{F}))\dfrac{\|\Delta\mathbf{F}\|}{\|\mathbf{F}\|}\right)} \tag{A12}$$

The condition number is defined in Theorem B.1.

Proof
Let $\mathbf{y}$ and $\tilde{\mathbf{y}}=\mathbf{y}+\Delta\mathbf{y}$ be the optimal least-squares solution to $\mathbf{F}\mathbf{y}=\mathbf{r}$ and $(\mathbf{F}+\Delta\mathbf{F})\tilde{\mathbf{y}}=\mathbf{r}+\Delta\mathbf{r}$, respectively. We have

$$\begin{aligned}\Delta\mathbf{y}&=\tilde{\mathbf{y}}-\mathbf{y}=(\mathbf{F}+\Delta\mathbf{F})^{+}(\mathbf{r}+\Delta\mathbf{r})-\mathbf{F}^{+}\mathbf{r}\\&=((\mathbf{F}+\Delta\mathbf{F})^{\mathrm{T}}(\mathbf{F}+\Delta\mathbf{F}))^{-1}(\mathbf{F}+\Delta\mathbf{F})^{\mathrm{T}}(\mathbf{r}+\Delta\mathbf{r})-\mathbf{F}^{+}\mathbf{r}\end{aligned} \tag{A13}$$

where $\mathbf{F}^{+}=(\mathbf{F}^{\mathrm{T}}\mathbf{F})^{-1}\mathbf{F}^{\mathrm{T}}$ is the pseudoinverse of $\mathbf{F}$. Furthermore, we have

$$\begin{aligned}\Delta\mathbf{y}&=(\mathbf{F}^{\mathrm{T}}\mathbf{F}+\Delta\mathbf{F}^{\mathrm{T}}\mathbf{F}+\mathbf{F}^{\mathrm{T}}\Delta\mathbf{F}+\Delta\mathbf{F}^{\mathrm{T}}\Delta\mathbf{F})^{-1}(\mathbf{F}+\Delta\mathbf{F})^{\mathrm{T}}(\mathbf{r}+\Delta\mathbf{r})-\mathbf{F}^{+}\mathbf{r}\\&=[\mathbf{I}+(\mathbf{F}^{\mathrm{T}}\mathbf{F})^{-1}\Delta\mathbf{F}^{\mathrm{T}}\mathbf{F}+(\mathbf{F}^{\mathrm{T}}\mathbf{F})^{-1}\mathbf{F}^{\mathrm{T}}\Delta\mathbf{F}+(\mathbf{F}^{\mathrm{T}}\mathbf{F})^{-1}\Delta\mathbf{F}^{\mathrm{T}}\Delta\mathbf{F}]^{-1}\\&\quad\times(\mathbf{F}^{\mathrm{T}}\mathbf{F})^{-1}(\mathbf{F}+\Delta\mathbf{F})^{\mathrm{T}}(\mathbf{r}+\Delta\mathbf{r})-\mathbf{F}^{+}\mathbf{r}\end{aligned} \tag{A14}$$

Let $\mathbf{F}^{*}:=(\mathbf{F}^{\mathrm{T}}\mathbf{F})^{-1}\Delta\mathbf{F}^{\mathrm{T}}$, and $\mathbf{I}<(\mathbf{I}+\mathbf{F}^{*}\mathbf{F}+\mathbf{F}^{+}\Delta\mathbf{F}+\mathbf{F}^{*}\Delta\mathbf{F})^{-1}$, we have

$$\begin{aligned}\Delta\mathbf{y}&=(\mathbf{I}+\mathbf{F}^{*}\mathbf{F}+\mathbf{F}^{+}\Delta\mathbf{F}+\mathbf{F}^{*}\Delta\mathbf{F})^{-1}(\mathbf{F}^{+}+\mathbf{F}^{*})(\mathbf{r}+\Delta\mathbf{r})-\mathbf{F}^{+}\mathbf{r}\\&\leqslant(\mathbf{I}+\mathbf{F}^{*}\mathbf{F}+\mathbf{F}^{+}\Delta\mathbf{F}+\mathbf{F}^{*}\Delta\mathbf{F})^{-1}(\mathbf{F}^{+}\Delta\mathbf{r}+\mathbf{F}^{*}\mathbf{r}+\mathbf{F}^{*}\Delta\mathbf{r})\\&=(\mathbf{I}+\mathbf{F}^{*}\mathbf{F}+\mathbf{F}^{+}\Delta\mathbf{F}+\mathbf{F}^{*}\Delta\mathbf{F})^{-1}(\mathbf{F}^{+}\mathbf{P}\Delta\mathbf{r}+\mathbf{F}^{*}\mathbf{r}+\mathbf{F}^{*}\Delta\mathbf{r})\end{aligned} \tag{A15}$$

Consider the following conditions:

$$\|\mathbf{F}^{*}\|\leqslant\frac{\|\Delta\mathbf{F}\|}{\|\mathbf{F}\|^2}\quad\text{and}\quad\|\mathbf{F}^{*}\|\|\mathbf{F}\|\leqslant\frac{\|\Delta\mathbf{F}\|}{\|\mathbf{F}\|} \tag{A16}$$

DOI: 10.1002/nme

and

$$\|(\mathbf{I}+\mathbf{F}^*\mathbf{F}+\mathbf{F}^+\Delta\mathbf{F}+\mathbf{F}^*\Delta\mathbf{F})^{-1}\| \leqslant \frac{1}{1-\|\mathbf{F}^*\mathbf{F}+\mathbf{F}^+\Delta\mathbf{F}+\mathbf{F}^*\Delta\mathbf{F}\|} \leqslant \frac{1}{1-(\|\mathbf{F}^*\|\|\mathbf{F}\|+\|\mathbf{F}^+\|\|\Delta\mathbf{F}\|+\|\mathbf{F}^*\|\|\Delta\mathbf{F}\|)}. \tag{A17}$$

From (A16) and (A17), we obtain

$$\|\Delta\mathbf{y}\| \leqslant \frac{\|\mathbf{F}^+\|\|\mathbf{P}\Delta\mathbf{r}\|+\|\mathbf{F}^*\|\|\mathbf{r}\|+\|\mathbf{F}^*\|\|\Delta\mathbf{r}\|}{1-(\|\mathbf{F}^*\|\|\mathbf{F}\|+\|\mathbf{F}^+\|\|\Delta\mathbf{F}\|+\|\mathbf{F}^*\|\|\Delta\mathbf{F}\|)} \tag{A18}$$

Further, we have

$$\frac{1}{\|\mathbf{y}\|} \leqslant \frac{\|\mathbf{F}\|}{\|\mathbf{P}\mathbf{r}\|} \tag{A19}$$

Combining (A18) and (A19) leads to

$$\frac{\|\Delta\mathbf{y}\|}{\|\mathbf{y}\|} \leqslant \frac{\|\mathbf{F}^+\|\|\mathbf{P}\delta\mathbf{r}\|+\|\mathbf{F}^*\|\|\mathbf{r}\|+\|\mathbf{F}^*\|\|\delta\mathbf{r}\|}{1-(\|\mathbf{F}^*\|\|\mathbf{F}\|+\|\mathbf{F}^+\|\|\Delta\mathbf{F}\|+\|\mathbf{F}^*\|\|\Delta\mathbf{F}\|)} \cdot \frac{\|\mathbf{F}\|}{\|\mathbf{P}\mathbf{r}\|} \tag{A20}$$

By using (A16), (A20) becomes

$$\frac{\|\Delta\mathbf{y}\|}{\|\mathbf{y}\|} \leqslant \frac{\left(\frac{\|\mathbf{r}\|}{\|\mathbf{P}\mathbf{r}\|}+\frac{\|\Delta\mathbf{r}\|}{\|\mathbf{P}\mathbf{r}\|}\right)\frac{\|\Delta\mathbf{F}\|}{\|\mathbf{F}\|}+\mathrm{Cond}(\mathbf{F})\frac{\|\mathbf{P}\Delta\mathbf{r}\|}{\|\mathbf{P}\mathbf{r}\|}}{1-\left(\frac{\|\Delta\mathbf{F}\|^2}{\|\mathbf{F}\|^2}+(1+\mathrm{Cond}(\mathbf{F}))\frac{\|\Delta\mathbf{F}\|}{\|\mathbf{F}\|}\right)} \tag{A21}$$

The condition number is defined in (A11).

□

APPENDIX B

Here, we provide the detailed discussions about the effective condition numbers for both square and overdetermined linear systems.

Theorem B.1
Consider $A\mathbf{x}=\mathbf{b}$, $A\in\mathbb{R}^{n\times n}$, $\mathbf{x}\in\mathbb{R}^n$, $\mathbf{b}\in\mathbb{R}^n$. Let $\hat{\mathbf{x}}=\mathbf{x}+\Delta\mathbf{x}$ be the solution of the perturbed problem: $A(\mathbf{x}+\Delta\mathbf{x})=\mathbf{b}+\Delta\mathbf{b}$, $\Delta\mathbf{b}\in\mathbb{R}^n$. The perturbation of solution $\mathbf{x}$ due to the perturbation of $\mathbf{b}$ is given by

$$\frac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|} \leqslant \mathrm{Cond}_{\mathrm{Eff}}(\mathbf{A}) \times \frac{\|\Delta\mathbf{b}\|}{\|\mathbf{b}\|} \tag{B1}$$

where

$$\mathrm{Cond}_{\mathrm{Eff}}(\mathbf{A}) = \frac{\|\mathbf{b}\|}{\sqrt{\frac{\|\mathbf{b}\|^2-\beta_{\min}^2}{\mathrm{Cond}(\mathbf{A})^2}+\beta_{\min}^2}} \tag{B2}$$

and $\mathbf{u}_n$ is the nth (last) eigenvector of matrix $\mathbf{A}$.

Proof
Let $\mathbf{A}$ be a symmetric matrix with full rank and has a diagonal canonical form $\mathbf{U}^{\mathrm{T}}\mathbf{A}\mathbf{U}=\mathbf{D}=\mathrm{diag}(\lambda_1,\lambda_2,\ldots,\lambda_n)$, where λ_i are the eigenvalues of $\mathbf{A}$ with the order $\lambda_{\max}=\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_n=$

DOI: 10.1002/nme

$\lambda_{\min} > 0$. The column vectors of $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n]$ are the corresponding orthonormalized eigenvectors, i.e., $\mathbf{u}_i^{\mathrm{T}} \mathbf{u}_j = \delta_{ij}$. By definition, we have:

$$\mathbf{A}\mathbf{u}_i = \lambda_i \mathbf{u}_i, \quad \mathbf{A}^{-1}\mathbf{u}_i = \frac{1}{\lambda_i}\mathbf{u}_i, \quad i = 1, 2, \ldots, n \tag{B3}$$

Consider eigenvector expansions of $\mathbf{b}$ and $\Delta\mathbf{b}$

$$\mathbf{b} = \sum_{i=1}^{n} \beta_i \mathbf{u}_i, \quad \Delta\mathbf{b} = \sum_{i=1}^{n} \alpha_i \mathbf{u}_i \tag{B4}$$

then $\beta_i = \mathbf{u}_i^{\mathrm{T}}\mathbf{b}$, $\alpha_i = \mathbf{u}_i^{\mathrm{T}}\Delta\mathbf{b}$ and

$$\|\mathbf{b}\| = \sqrt{\sum_{i=1}^{n} \beta_i^2}, \quad \|\Delta\mathbf{b}\| = \sqrt{\sum_{i=1}^{n} \alpha_i^2} \tag{B5}$$

where $\|\|$ is a 2-norm.

By means of (B3), we have

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b} = \mathbf{A}^{-1} \sum_{i=1}^{n} \beta_i \mathbf{u}_i = \sum_{i=1}^{n} \beta_i \mathbf{A}^{-1}\mathbf{u}_i = \sum_{i=1}^{n} \left(\frac{\beta_i}{\lambda_i}\right) \mathbf{u}_i \tag{B6}$$

It follows that

$$\|\mathbf{x}\| = \sqrt{\sum_{i=1}^{n} \left(\frac{\beta_i}{\lambda_i}\right)^2} \tag{B7}$$

Moreover, $\Delta\mathbf{x} = \mathbf{A}^{-1}\Delta\mathbf{b}$, then

$$\|\Delta\mathbf{x}\| = \sqrt{\sum_{i=1}^{n} \left(\frac{\alpha_i}{\lambda_i}\right)^2} \tag{B8}$$

Combining (B8) with (B5), we have

$$\|\Delta\mathbf{x}\|^2 = \sum_{i=1}^{n} \frac{\alpha_i^2}{\lambda_i^2} \leqslant \frac{\sum_{i=1}^{n} \alpha_i^2}{\lambda_n^2} = \frac{\|\Delta\mathbf{b}\|^2}{\lambda_n^2} \tag{B9}$$

and

$$\|\Delta\mathbf{x}\| \leqslant \frac{\|\Delta\mathbf{b}\|}{\lambda_n} \tag{B10}$$

Similarly, Equation (B7) yields

$$\|\mathbf{x}\|^2 = \sum_{i=1}^{n-1} \frac{\beta_i^2}{\lambda_i^2} + \frac{\beta_n^2}{\lambda_n^2} \geqslant \frac{\sum_{i=1}^{n-1} \beta_i^2}{\lambda_1^2} + \frac{\beta_n^{y} 2}{\lambda_n^2} = \frac{1}{\lambda_n^2}\left\{\frac{\|\mathbf{b}\|^2 - \beta_n^2}{(\lambda_1/\lambda_n)^2} + \beta_n^2\right\}, \tag{B11}$$

and consequently

$$\frac{1}{\|\mathbf{x}\|} \leqslant \frac{\lambda_n}{\sqrt{\dfrac{\|\mathbf{b}\|^2 - \beta_n^2}{\mathrm{Cond}(A)^2} + \beta_n^2}} \tag{B12}$$

where $\mathrm{Cond}(\mathbf{A}) = \lambda_n/\lambda_1$. Combining (B10) and(B12), we have

$$\frac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|} \leqslant \frac{\|\mathbf{b}\|}{\sqrt{\dfrac{\|\mathbf{b}\|^2 - \beta_n^2}{\mathrm{Cond}(\mathbf{A})^2} + \beta_n^2}} \cdot \frac{\|\Delta\mathbf{b}\|}{\|\mathbf{b}\|} =: \mathrm{Cond}_{\mathrm{Eff}} \frac{\|\Delta\mathbf{b}\|}{\|\mathbf{b}\|} \tag{B13}$$

This proves (B1). □

DOI: 10.1002/nme

Theorem B.2
Consider $\mathbf{F}\mathbf{y}=\mathbf{r}$, $\mathbf{F}\in\mathbb{R}^{m\times n}$, $\mathbf{y}\in\mathbb{R}^n$, $\mathbf{r}\in\mathbb{R}^m$. Let $\hat{\mathbf{y}}=\mathbf{y}+\Delta\mathbf{y}$ be the solution of the perturbed problem: $\mathbf{F}(\mathbf{y}+\Delta\mathbf{y})=\mathbf{r}+\Delta\mathbf{r}$, $\Delta\mathbf{r}\in\mathbb{R}^m$. The perturbation of solution $\mathbf{y}$ due to the perturbation of $\mathbf{r}$ is given by

$$\frac{\|\Delta\mathbf{y}\|}{\|\mathbf{y}\|}\leqslant \mathrm{Cond}_{\mathrm{Eff}}(\mathbf{F})\times\frac{\|\mathbf{P}\Delta\mathbf{r}\|}{\|\mathbf{P}\mathbf{r}\|} \tag{B14}$$

where

$$\mathrm{Cond}_{\mathrm{Eff}}(\mathbf{F})=\frac{\|\mathbf{P}\mathbf{r}\|}{\sqrt{\dfrac{\|\mathbf{P}\mathbf{r}\|^2-\gamma_n^2}{\mathrm{Cond}(\mathbf{F})^2}+\gamma_n^2}},\quad \gamma_n=\mathbf{u}_n^{\mathrm{T}}\mathbf{P}\mathbf{r} \tag{B15}$$

and $\mathbf{u}_n$ is the nth (last) left eigenvector of matrix F.

Proof
Let matrix $\mathbf{F}$ be full ranked and has an SVD: $\mathbf{U}^{\mathrm{T}}\mathbf{F}\mathbf{V}=\sum=\mathrm{diag}(\sigma_1,\sigma_2,\ldots,\sigma_n)$, where σ_i are the singular values of $\mathbf{F}$, $\sigma_{\max}=\sigma_1\geqslant\sigma_2\cdots\geqslant\sigma_n=\sigma_{\min}>0$, and the matrix $\sum$ is of dimension $m\times n$. The columns $\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_m$ of matrix $\mathbf{U}$ are the left singular vectors, and columns $\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n$ of matrix $\mathbf{V}$ are the right singular vectors. Further, we have $\mathbf{F}\mathbf{V}=\mathbf{U}\sum$, i.e.

$$\mathbf{F}\mathbf{v}_i=\sigma_i\mathbf{u}_i,\quad \mathbf{F}^{+}\mathbf{u}_i=\frac{1}{\sigma_i}\mathbf{v}_i,\quad i=1,2,\ldots,n \tag{B16}$$

Consider the following eigenvector expansion:

$$\mathbf{P}\mathbf{r}=\sum_{i=1}^{n}\gamma_i\mathbf{u}_i,\quad \mathbf{P}\Delta\mathbf{r}=\sum_{i=1}^{n}\tau_i\mathbf{u}_i \tag{B17}$$

It follows that $\gamma_i=\mathbf{u}_i^{\mathrm{T}}\mathbf{P}\mathbf{r}$, $\tau_i=\mathbf{u}_i^{\mathrm{T}}\mathbf{P}\Delta\mathbf{r}$, and

$$\|\mathbf{P}\mathbf{r}\|=\sqrt{\sum_{i=1}^{n}\gamma_i^2},\quad \|\mathbf{P}\Delta\mathbf{r}\|=\sqrt{\sum_{i=1}^{n}\tau_i^2} \tag{B18}$$

By means of (B16), we have

$$\mathbf{y}=\mathbf{F}^{+}\mathbf{r}=\mathbf{F}^{+}\mathbf{P}\mathbf{r}=\mathbf{F}^{+}\sum_{i=1}^{n}\gamma_i\mathbf{u}_i=\sum_{i=1}^{n}\gamma_i\mathbf{F}^{+}\mathbf{u}_i=\sum_{i=1}^{n}\gamma_i\frac{1}{\sigma_i}\mathbf{v}_i=\sum_{i=1}^{n}\frac{\gamma_i}{\sigma_i}\mathbf{v}_i \tag{B19}$$

Consequently, we have the following 2-norms:

$$\|\mathbf{y}\|=\sqrt{\sum_{i=1}^{n}\left(\frac{\gamma_i}{\sigma_i}\right)^2} \tag{B20}$$

$$\|\Delta\mathbf{y}\|=\sqrt{\sum_{i=1}^{n}\left(\frac{\tau_i}{\sigma_i}\right)^2} \tag{B21}$$

Combining (B18) and (B21), we obtain

$$\|\Delta\mathbf{y}\|^2=\sum_{i=1}^{n}\frac{\tau_i^2}{\sigma_i^2}\leqslant\frac{\sum_{i=1}^{n}\tau_i^2}{\sigma_n^y 2}=\frac{\|\mathbf{P}\Delta\mathbf{r}\|^2}{\sigma_n^2} \tag{B22}$$

and thus

$$\|\Delta\mathbf{y}\|\leqslant\frac{\|\mathbf{P}\Delta\mathbf{r}\|}{\sigma_n} \tag{B23}$$

DOI: 10.1002/nme

From (B20), we have

$$\|\mathbf{y}\|^2=\sum_{i=1}^{n-1}\frac{\gamma_i^2}{\sigma_i^2}+\frac{\gamma_n^2}{\sigma_n^2}\geqslant\frac{\sum_{i=1}^{n-1}\gamma_i^2}{\sigma_1^2}+\frac{\gamma_n^2}{\sigma_n^2}=\frac{1}{\sigma_n^2}\left\{\frac{\|Pr\|^2-\gamma_n^2}{(\sigma_1/\sigma_n)^2}+\gamma_n^y 2\right\} \tag{B24}$$

Consequently

$$\frac{1}{\|\mathbf{y}\|}\leqslant\frac{\sigma_n}{\sqrt{\dfrac{\|\mathbf{Pr}\|^2-\gamma_n^2}{\mathrm{Cond}(\mathbf{F})^2}+\gamma_n^2}} \tag{B25}$$

where $\mathrm{Cond}(\mathbf{F})=\sigma_n/\sigma_1$. Combining (B23) and (B25), we have

$$\frac{\|\Delta\mathbf{y}\|}{\|\mathbf{y}\|}\leqslant\frac{\|\mathbf{Pr}\|}{\sqrt{\dfrac{\|\mathbf{Pr}\|^2-\gamma_n^2}{\mathrm{Cond}(\mathbf{F})^2}+\gamma_n^2}}\cdot\frac{\|\mathbf{P}\Delta\mathbf{r}\|}{\|\mathbf{Pr}\|}=:\mathrm{Cond}_{\mathrm{Eff}}(\mathbf{F})\frac{\|\mathbf{P}\Delta\mathbf{r}\|}{\|\mathbf{Pr}\|} \tag{B26}$$

This proves (B14). □

ACKNOWLEDGEMENTS

The support of this work by National Science Council of Taiwan, R. O. C., under project number NSC 98-2115-M-029-001-MY2 to the first author, and the support by US Army ERDC under contract W912HZ-07-C-0019 to the second and third authors are greatly acknowledged.

REFERENCES

1. Liu WK, Belytschko T, Oden JT. Special issue on meshless methods. *Computer Methods in Applied Mechanics and Engineering* 1996; **139**.
2. Chen JS, Liu WK. Special issue on meshfree methods: recent advances and new applications. *Computer Methods in Applied Mechanics and Engineering* 2004; **193**:933–1321.
3. Dolbow J, Belytschko T. Numerical integration of Galerkin weak form in meshfree methods. *Computational Mechanics* 1999; **23**:219–230.
4. Chen JS, Wu CT, Yoon S, You Y. A stabilized conforming nodal integration for Galerkin meshfree method. *International Journal for Numerical Methods in Engineering* 2001; **50**:435–466.
5. Kansa EJ. Multiqudrics—a scattered data approximation scheme with applications to computational fluid dynamics I. Surface approximations and partial derivatives. *Computers and Mathematics with Applications* 1990; **19**:127–161.
6. Kansa EJ. Multiquadrics—a scattered data approximation scheme with applications to computational fluid dynamics I. Solution to parabolic, hyperbolic and elliptic partial differential equations. *Computers and Mathematics with Applications* 1990; **19**:127–161.
7. Schaback R. Error estimates and condition numbers for radial basis function interpolation. *Advances in Computational Mathematics* 1995; **3**:251–264.
8. Aluru NR. A point collocation method based on reproducing kernel approximation. *International Journal for Numerical Methods in Engineering* 2000; **47**:1083–1121.
9. Kim DW, Kim Y. Point collocation method using the fast moving least-square reproducing kernel approximation. *International Journal for Numerical Methods in Engineering* 2003; **56**:1445–1464.
10. Hu HY, Chen JS, Hu W. Error analysis of collocation method based on reproducing kernel approximation. *Numerical Methods for Partial Differential Equations* 2010; DOI: 10.1002/num20539.
11. Lancuster P, Salkauskas K. Surface generated by moving least squares methods. *Mathematics of Computation* 1981; **37**:141–158.
12. Belytschko T, Lu YY, Gu L. Element-free Galerkin methods. *International Journal for Numerical Methods in Engineering* 1994; **37**:229–256.
13. Liu WK, Jun S, Zhang YF. Reprodicing kernel particle methods. *International Journal for Numerical Methods in Fluids* 1995; **20**:1081–1106.
14. Chen JS, Pan C, Wu CT, Liu WK. Reproducing kernel particle methods for large deformation analysis of nonlinear structures. *Computer Methods in Applied Mechanics and Engineering* 1996; **139**:195–227.
15. Chen JS, Han W, You Y, Meng X. A reproducing kernel method with nodal interpolation property. *International Journal for Numerical Methods in Engineering* 2003; **56**:935–960.

16. Frank C, Schaback R. Solving partial differential equations by collocation using radial basis function. *Applied Mathematics and Computation* 1998; **93**:73–82.
17. Li ZC, Lu TT, Hu HY, Cheng AH-D. *Trafftz and Collocation Methods*. WIT Press: Southampton, U.K., 2008.
18. Hu HY, Lai CK, Chen JS. A study on convergence and complexity of reproducing kernel collocation method. *Interaction and Multiscale Mechanics* 2009; **2**(3):295–319.
19. Atkinson KE. *An Introduction to Numerical Analysis*. Wiley: New York, 1988.
20. Horn RA, Johnson CR. *Matrix Analysis*. Cambridge University Press: Cambridge, 1990.
21. Demmel JW. *Applied Numerical Linear Algebra*. SIAM: Philadelphia, PA, 1997.
22. Chan FC, Foulser DE. Effectively well-conditioned linear systems. *SIAM Journal on Scientific and Statistical Computing* 1988; **9**:963–969.
23. Christiansen S, Hansen PC. The effective condition number applied to error analysis of certain boundary collocation methods. *Journal of Computational and Applied Mathematics* 1994; **54**:15–36.
24. Li ZC, Chien CS, Huang HT. Effective Condition number for finite difference method. *Journal of Computational and Applied Mathematics* 2007; **198**:208–235.
25. Huang HT, Li ZC. Effective condition number and superconvergence of the Trefftz method coupled with high order FEM for singularity problems. *Engineering Analysis with Boundary Elements* 2006; **30**:270–283.
26. Hu HY, Chen JS, Hu W. Weighted radial basis collocation method for boundary value problem. *International Journal for Numerical Methods in Engineering* 2007; **69**:2736–2757.
27. Lu TT, Hu HY, Li ZC. Highly accurate solutions of Motz's problem and the cracked beam problem. *Engineering Analysis with Boundary Elements* 2004; **28**(11):1387–1403.
28. Li ZC, Lu TT, Hu HY. Collocation Trefftz methods for biharmonic equations with crack singularities. *Engineering Analysis with Boundary Elements* 2004; **28**(1):79–96.
29. Hu HY, Li ZC. Collocation methods for Poisson's equation. *Computer Methods in Applied Mechanics and Engineering* 2006; **195**:4139–4160.

DOI: 10.1002/nme