Jin and Luo *Boundary Value Problems* (2019) 2019:63
https://doi.org/10.1186/s13661-019-1176-2

Boundary Value Problems
a SpringerOpen Journal

RESEARCH Open Access

A reduced-order extrapolating collocation spectral method based on POD for the 2D Sobolev equations

Shiju Jin[1] and Zhendong Luo[2*]

[*]Correspondence: zhdluo@163.com
[2]School of Mathematics and Physics, North China Electric Power University, Beijing, China
Full list of author information is available at the end of the article

Abstract

In this paper, we mainly use a proper orthogonal decomposition (POD) to reduce the order of the coefficient vectors of the solutions for the classical collocation spectral (CS) method of two-dimensional (2D) Sobolev equations. We first establish a reduced-order extrapolating collocation spectral (ROECS) method for 2D Sobolev equations so that the ROECS method includes the same basis functions as the classic CS method and the superiority of the classic CS method. Then we use the matrix means to discuss the existence, stability, and convergence of the ROECS solutions so that the procedure of theoretical analysis becomes very concise. Lastly, we present two set of numerical examples to validate the effectiveness of theoretical conclusions and to illuminate that the ROECS method is far superior to the classic CS method, which shows that the ROECS method is quite valid to solve Sobolev equations. Therefore, both theory and method of this paper are completely different from the existing reduced-order methods.

MSC: 65N30; 65N12; 65M15

Keywords: Proper orthogonal decomposition; Classic collocation spectral method; Sobolev equations; Reduced-order extrapolating collocation spectral method; Existence, stability, and convergence

1 Introduction

Let $\Omega \subset \mathbb{R}^2$ be an open bounded domain with boundary $\partial\Omega$. We consider the following two-dimensional (2D) Sobolev equation:

$$\begin{cases} u_t - \varepsilon\Delta u_t - \gamma\Delta u = f(x,y,t), & (x,y,t) \in \Omega \times (0,T), \\ u(x,y,t) = \varphi(x,y,t), & (x,y,t) \in \partial\Omega \times (0,T), \\ u(x,y,0) = u_0(x,y), & (x,y) \in \Omega, \end{cases} \tag{1}$$

where $u_t = \partial u/\partial t$, $\Delta u = \partial^2 u/\partial x^2 + \partial^2 u/\partial y^2$, $\Delta u_t = \partial^2 u_t/\partial x^2 + \partial^2 u_t/\partial y^2$, T represents the final moment, $\varepsilon > 0$ and $\gamma > 0$ are two given constants, $f(x,y,t)$ is the known source item, and $u_0(x,y)$ and $\varphi(x,y,t)$ are the known initial and boundary values, respectively. For simplicity but without loss of generality, we further assume that $\varphi(x,y,t) = 0$.

The system of Sobolev equations (1) is a class of significant partial differential equations (PDEs) with practical physical background, which favorably simulated many engineering problems (see [1, 2]). Particularly, it can be applied to simulate the porous phenomenons (see [3, 4]). Nevertheless, in actual applications, Sobolev equations frequently contain intricate boundary and initial values, complicated source items, or discontinuous constants. As a result, we cannot generally seek analytic solutions, so that we can only rest upon numerical methods.

Currently, finite difference (FD), finite volume element (FVE), finite element (FE), and spectral methods are four famous computational techniques. However, the spectral method gives the highest accuracy among the four computational methods since the unknown functions in the spectral methods are approximated with some sufficiently smooth functions, such as Chebyshev polynomials, trigonometric functions, Legendre polynomials, or Jacobi polynomials, whereas the unknown functions in the FVE and FE methods are commonly approximated with some standard polynomials, and the derivatives in the FD scheme are approximated with some difference quotients. The spectral method is commonly sorted into the collocation spectral (CS) method, Galerkin's spectral method, and the spectral tau method. It has been applied to settle various PDEs including second-order elliptic, parabolic, hyperbolic, telegraph, and hydromechanical equations (see [5–8]).

However, Sobolev equations are mainly solved with the FD scheme, the FE method, and the FVE method (see, e.g., [3, 4, 9–14]), except that one-dimensional Sobolev equations have been settled by the Fourier spectral method [15], and 2D Sobolev equations have recently been settled by the classic CS method [16]. Though the classic CS method (see [16]) for 2D Sobolev equations can attain higher accuracy than the FD scheme, FE method, and FVE method, it also contains a lot of degrees of freedom (unknowns). In this way, because of the round-off error accumulation in numerical calculations, after several computational steps, there generally occurs a floating-point overflow such that we cannot obtain the desired consequences. Hence, to ensure a sufficiently high precision of the classic CS solutions, the crucial question is how to lessen the unknowns (i.e., degrees of freedom) of the CS method to ease the round-off error accumulation in the calculations, which is also central task in this paper.

Many examples have proven that the proper orthogonal decomposition (POD) can significantly reduce the order of numerical methods (see [17–20]). It can vastly decrease the degrees of freedom in the numerical methods. It has been applied in many fields including pattern recognition and signal analysis [21], statistical calculations [22], and computational fluid mechanics [23]. For the past few years, it has successfully been used to the order reduction for the Galerkin methods [24, 25], the FE methods [26, 27], the FD schemes [28–30], the FVE methods [31, 32], and the reduced basis methods for PDEs [33–35]. Nevertheless, the existing POD-based reduced-order methods (see [17–27, 29–35]) are mostly created with the POD bases produced by the classic solutions at all the time nodes on $[0, T]$, before repeatedly finding the order reduction solutions on the same time nodal points. In fact, they belong to some undesirable repeated computations. To get rid of the repeated computations, several reduced-order extrapolated approaches based on POD have been proposed [36–41].

Nevertheless, to our knowledge, there is no reduced-order extrapolating CS (ROECS) method for 2D Sobolev equations created by reducing the order for the coefficient vectors of the CS solutions of the classic CS method via POD. Hence, in this paper, utilizing POD to reduce the order of the coefficient vectors of the CS solutions for the classic CS method, we construct a ROECS method only holding few degrees of freedom. We employ matrix means to discuss the existence, convergence, and stability of the ROECS solutions so that the theoretical means here becomes very concise. In particular, we only employ the classic CS solutions on the first several time nodal points to form the snapshots, and then we use them to produce the POD bases and create the ROECS format so as to obtain the ROECS solutions on all the time nodal points. Thus, we avoid the repeated computations. Moreover, in this paper, we adopt the error estimates to serve as the suggestion of choice of POD bases. The ROECS format contains both advantages that the POD method can reduce the unknowns and the CS method has higher accuracy, so it is an innovation and development of the existing reduced-order methods.

The main merits of the ROECS method hare the following. First, we only reduce the order of the coefficient vectors of the solutions for the classic CS method by POD and have not altered the basis functions for the classic CS method so that the ROECS method holds simultaneously both virtues that POD can reduce the unknowns and the classic CS method has higher accuracy. Second, the classic CS method is totally different from the Galerkin spectral method, and the Sobolev equations not only include a first-order derivative term of time and two 2nd-order derivative terms of spacial variables but also contain two mixed derivative terms of the first order with respect to time and of the second order with respect to spacial variables, that is, the 2D Sobolev equations are more complex than the hyperbolic and parabolic equations in [42, 43]. So the ROECS method is totally different from the methods in [42, 43], but 2D Sobolev equations have some special applications as stated before. Third, we use the matrix means to discuss the existence, convergence, and stability of the ROECS solutions so that theoretical analysis becomes very concise and our theory and methods are totally different from the other existing order reduction methods. Therefore, our method is totally new and superior over the existing order reduction methods.

The rest of this paper is organized the following. In Sect. 2, we first retrospect the classic CS format of the 2D Sobolev equations and gain snapshots from the initial few classic CS solutions. Then, in Sect. 3, we produce a cluster of POD basis from the snapshots, develop the ROECS format, prove the existence, convergence, and stability of the ROECS solutions by the matrix means, and supply the flow-chart for settling the ROECS format. Next, in Sect. 4, we use two sets of numerical examples to illuminate that the ROECS format is distinctly superior to the classic CS model, to validate that the numeric computational conclusions accord with the theoretical ones and that the ROECS format is quite valid to solve Sobolev equations, and to confirm that the ROECS format can greatly lessen the unknowns (i.e., degrees of freedom), the calculation load, the CPU elapsed time, and the required storage volumes in numerical computations. Finally, in Sect. 5, we provide the chief conclusions and discussions.

2 The classic CS method for 2D Sobolev equations

Because any bounded closed domain $\overline{\Omega}$ in $\mathbb{R}^2$ can be approximately covered with several rectangles $[a_i, b_i] \times [c_i, d_i]$ $(i = 1, 2, \ldots, I)$, for simplicity and without loss of generality, let $\overline{\Omega} = [a, b] \times [c, d] \subset \mathbb{R}^2$. Moreover, using the transforms $x' = -1 + 2(x - a)/(b - a)$ and $y' = -1 + 2(y - c)/(d - c)$, we can ensure $[a, b] \leftrightarrow [-1, 1]$ and $[c, d] \leftrightarrow [-1, 1]$, respectively. Thus, for convenience, we can further assume that $a = c = -1$ and $b = d = 1$.

2.1 The variational formulation for the 2D Sobolev equations

The Sobolev spaces and norms used in this paper are standard, whose detailed descriptions can be found in [44]. For example, we set $\omega = 1/\sqrt{(1-x^2)(1-y^2)}$, and $L^2_\omega(\Omega)$ denotes the set of all square-integrable functions on Ω equipped with inner product and norm

$$(u,v)_\omega = \int_\Omega uv\omega \,\mathrm{d}x\,\mathrm{d}y, \qquad \|u\|_{0,\omega} = \left(\int_\Omega |u|^2\omega\,\mathrm{d}x\,\mathrm{d}y\right)^{1/2}, \quad \forall u,v \in L^2_\omega(\Omega),$$

whereas $H^m_\omega(\Omega) := \{u \in L^2_\omega(\Omega) : D^\alpha u \in L^2_\omega(\Omega), 0 \le |\alpha| \le m\}$ denotes the weighted Sobolev space on Ω with the CGL quadrature weight function, equipped with the norm

$$\|u\|_{m,\omega} = \left(\sum_{0\le|\alpha|\le m} \left\|D^\alpha u\right\|^2_{0,\omega}\right)^{\frac{1}{2}}.$$

Furthermore, set $H^1_{0,\omega}(\Omega) = \{u \in H^1_\omega(\Omega) : u|_{\partial\Omega} = 0\}$, and let $\|\cdot\|_{H^l(H^m_\omega)}$ be the norm in the space

$$H^l\bigl(0,T;H^m_\omega(\Omega)\bigr) \equiv \left\{v(t) \in H^m_\omega(\Omega) : \|v\|^2_{H^l(H^m_\omega)} \equiv \int_0^T \sum_{i=0}^{l} \left\|\frac{\mathrm{d}^i}{\mathrm{d}t^i}v(t)\right\|^2_{m,\omega} \mathrm{d}t < \infty\right\}.$$

We consider the following variational formulation for 2D Sobolev equations.

Problem 1 For $t \in (0,T)$, find $u \in H^1_{0,\omega}(\Omega)$ such that

$$\begin{cases} (u_t,v)_\omega + \varepsilon(\nabla u_t,\nabla v)_\omega + \gamma(\nabla u,\nabla v)_\omega = (f,v)_\omega, & \forall v \in H^1_{0,\omega}(\Omega), \\ u(x,y,0) = u_0(x,y), \quad (x,y) \in \Omega. \end{cases} \tag{2}$$

The following result on the existence, uniqueness, and stability of the generalized solution for Problem 1 has been provided in [16].

Theorem 2 *If $f \in L^2(0,T;H^{-1}_\omega(\Omega))$ and $u_0 \in H^1_\omega(\Omega)$, then there exists a unique generalized solution for the variational formulation* (2) *satisfying the following stability*:

$$\|u\|_{1,\omega} \le \tilde{c}\bigl(\|u_0\|_{1,\omega} + \|f\|_{L^2(H^{-1}_\omega)}\bigr), \tag{3}$$

where $\tilde{c} = \sqrt{\max\{1,\varepsilon,1/(\gamma c_p^2)\}/\min\{1,\varepsilon\}}$, and c_p is the Poincaré coefficient.

2.2 The classic CS method for the 2D Sobolev equations

Too solve time-dependent PDEs by the CS format, it is necessary to discretize u_t by means of the difference quotient and spatial variables by means of the CS method. The CS method consists in seeking some approximate solutions at time and spatial nodes. In this paper, we take the Chebyshev–Gauss–Lobatto (CGL) type interpolation points (see [8]) as the space nodes, namely, let $\{x_j\}_{j=0}^N$ and $\{y_k\}_{k=0}^N$ be the space nodes in the x and y directions, respectively, with

$$x_j = -\cos\frac{j\pi}{N}, \qquad y_k = -\cos\frac{k\pi}{N},$$

where the positive integer N denotes the number of nodes in a certain direction. For integer $K > 0$, let $\Delta t = T/K$ be the time step, that is, $K\Delta t = T$. We approximate $u(x,y,n\Delta t)$ with u^n, the time derivative u_t of $u(x,y,t)$ at time $t_n = n\Delta t$ with $(u^{n+1} - u^n)/\Delta t$, and $u^n(x,y)$ with $u_N^n(x,y)$, namely,

$$u^n(x,y) \approx u_N^n(x,y) = \sum_{j=0}^{N}\sum_{k=0}^{N} u_N^n(x_j,y_k)h_j(x)h_k(y), \quad 0 \le n \le K,$$

where $\{h_j(x)\}_{j=0}^N$ and $\{h_k(y)\}_{j=0}^N$ are the Lagrange basis polynomials associated with the sets of the CGL points $\{x_j\}_{j=0}^N$ and $\{y_k\}_{k=0}^N$, respectively.

Define the H_ω^1-orthogonal projection $R_N : H_{0,\omega}^1(\Omega) \to P_N$, that is, for any $u \in H_{0,\omega}^1(\Omega)$, it satisfies

$$\big(\nabla(R_N u - u), \nabla v\big)_\omega = 0, \quad \forall v \in P_N.$$

Thus, R_N has the following important property (see [8]).

Theorem 3 *For any $u \in H_\omega^q(\Omega)$ with $q \ge 1$, we have*

$$\|\nabla R_N u\|_{0,\omega} \le \|\nabla u\|_{0,\omega}, \qquad \big\|\partial^k(R_N u - u)\big\|_{0,\omega} \le CN^{k-q}, \quad 0 \le k \le q \le N+1,$$

where C is a general positive constant independent of N and Δt.

Now, we obtain the following CS format for 2D Sobolev equations.

Problem 4 Find $u_N^n \in U_N \equiv H_{0,\omega}^1(\Omega) \cap P_N$ such that

$$\begin{cases} (u_N^{n+1} - u_N^n, v_N)_\omega + \varepsilon(\nabla u_N^{n+1} - \nabla u_N^n, \nabla v_N)_\omega + \gamma\Delta t(\nabla u_N^{n+1}, \nabla v_N)_\omega \\ \quad = \Delta t(f(t_{n+1}), v_N)_\omega, \quad \forall v_N \in U_N, 0 \le n \le K, \\ u_N^0(x,y) = R_N u_0(x,y), \quad (x,y) \in \Omega, \end{cases} \tag{4}$$

where $f(t_n) = f(x,y,t_n)$.

The result on the existence, uniqueness, stability, and convergence about the CS solutions for Problem 4 is given in [16].

Theorem 5 *If $f \in L^2(0,T;L_\omega^2(\Omega))$ and $u_0 \in H_\omega^1(\Omega)$, then there exists a unique series of solutions $u_N^n \in U_N$ $(n = 1,2,\dots,K)$ for the CS format* (4) *satisfying the following stability*:

$$\big\|\nabla u_N^n\big\|_{0,\omega}^2 \le \|\nabla u_0\|_{0,\omega}^2 + \frac{\Delta t}{\gamma}\sum_{j=1}^{n}\big\|f(t_j)\big\|_{0,\omega}^2, \quad n = 1,2,\dots,K. \tag{5}$$

Furthermore, when $\Delta t = O(N^{-1})$ and solutions of Problem 1 $u(t_n) \in H_\omega^q(\Omega)$ $(2 \le q \le N+1)$, the errors between the solution for Problem 1 and the series of solutions of Problem 4 have the following estimates:

$$\big\|\nabla\big(u(t_n) - u_N^n\big)\big\|_{0,\omega} \le C\big(\Delta t + N^{-q+1}\big), \quad n = 1,2,\dots,K, \tag{6}$$

$$\left\| u(t_n) - u_N^n \right\|_{0,\omega} \le C\left(\Delta t^2 + N^{-q}\right), \quad n = 1, 2, \ldots, K. \tag{7}$$

Remark 1 The error estimates in Theorem 5 attain an optimal order. Theorem 5 shows that the classic CS format, that is, Problem 4 for 2D Sobolev equations has a unique series of solutions, which is stable and continuously depends on the initial value and source functions. This theoretically ensures that Problem 4 is effective and reliable for solving 2D Sobolev equations.

2.3 The matrix representation of the classic CS format

To understand more easily the classic CS format for 2D Sobolev equations, we will rewrite the classic CS format (4) in the matrix form so that it can be easily programmed and computed by a computer. Let $u_{N_{j,k}}^n$ $(0 \le j, k \le N)$ denote the spectral approximate values of $u(x_j, y_k, n\Delta t)$, namely

$$u(x, y, n\Delta t) \approx u_N^n = \sum_{j=0}^{N} \sum_{k=0}^{N} u_{N_{j,k}}^n h_j(x) h_k(y). \tag{8}$$

Taking $v_N = h_m(x) h_l(y) \in U_N$ $(0 \le m, l \le N)$ in scheme (4), we come to the conclusion

$$\begin{aligned}
\left(u_N^{n+1}, v_N\right)_\omega &= \sum_{j=0}^{N} \sum_{k=0}^{N} u_{N_{j,k}}^{n+1} \left(h_j(x) h_k(y), h_m(x) h_l(y)\right)_\omega, \\
\left(\nabla u_N^{n+1}, \nabla v_N\right)_\omega &= \sum_{j=0}^{N} \sum_{k=0}^{N} u_{N_{j,k}}^{n+1} \left(h_j'(x) h_k(y), h_m'(x) h_l(y)\right)_\omega \\
&\quad + \sum_{j=0}^{N} \sum_{k=0}^{N} u_{N_{j,k}}^{n+1} \left(h_j(x) h_k'(y), h_m(x) h_l'(y)\right)_\omega.
\end{aligned}$$

From Problem 4 we obtain

$$\begin{aligned}
A_{jm,kl} &= \left(h_j(x) h_k(y), h_m(x) h_l(y)\right)_\omega \\
&= \sum_{p=0}^{N} \sum_{q=0}^{N} h_j(x_p) h_m(x_p) \omega_p h_k(y_q) h_n(y_q) \omega_q,
\end{aligned} \tag{9}$$

$$\begin{aligned}
B_{jm,kl} &= \left(h_j'(x) h_k(y), h_m'(x) h_l(y)\right)_\omega + \left(h_j(x) h_k'(y), h_m(x) h_l'(y)\right)_\omega \\
&= \sum_{p=0}^{N} \sum_{q=0}^{N} h_j'(x_p) h_m'(x_p) \omega_p h_k(y_q) h_l(y_q) \omega_q \\
&\quad + \sum_{p=0}^{N} \sum_{q=0}^{N} h_j(x_p) h_m(x_p) \omega_p h_k'(y_q) h_l'(y_q) \omega_q,
\end{aligned} \tag{10}$$

where $0 \le j, m, k, l \le N$.

Then we can rewrite the classic CS format (4) for the 2D equations as the following matrix form with $(N+1)^2$ equations for $\{u_{N_z}^n\}_{n=0}^K$:

$$\begin{cases}
(\boldsymbol{A} + \varepsilon \boldsymbol{B} + \gamma \Delta t \boldsymbol{B}) \boldsymbol{U}_N^{n+1} = \Delta t \boldsymbol{F}^{n+1} + (\boldsymbol{A} + \varepsilon \boldsymbol{B}) \boldsymbol{U}_N^n, \quad 0 \le n \le K-1, \\
\boldsymbol{U}_N^0 = \boldsymbol{U}_0,
\end{cases} \tag{11}$$

where

$$\begin{aligned}
&\boldsymbol{A} = [A_{jm,kl}]_{(N+1)^2\times(N+1)^2}, \qquad \boldsymbol{B} = [B_{jm,kl}]_{(N+1)^2\times(N+1)^2},\\
&\boldsymbol{U}_N^{n+1} = \left[u_{N0,0}^{n+1}, u_{N1,0}^{n+1}, \ldots, u_{NN,0}^{n+1}, u_{N0,1}^{n+1}, u_{N1,1}^{n+1}, \ldots, u_{NN,1}^{n+1}, \ldots, u_{N0,N}^{n+1}, \ldots, u_{NN,N}^{n+1}\right]^T,\\
&\boldsymbol{F}^{n+1} = \left[F_{0,0}^{n+1}, F_{1,0}^{n+1}, \ldots, F_{N,0}^{n+1}, F_{0,1}^{n+1}, \ldots, F_{N,1}^{n+1}, \ldots, F_{0,N}^{n+1}, \ldots, F_{N,N}^{n+1}\right]^T,\\
&F_{m,l}^{n+1} = f\big(x_m, y_l, (n+1)\Delta t\big), \quad 0 \le n \le N-1,\\
&\boldsymbol{U}_0 = \big[u_0(x_0,y_0), u_0(x_1,y_0), \ldots, u_0(x_N,y_0), u_0(x_0,y_1), \ldots, u_0(x_N,y_1), \ldots,\\
&\qquad u_0(x_0,y_N), \ldots, u_0(x_N,y_N)\big]^T.
\end{aligned}$$

Remark 2 Because the classic CS format adopts the Chebyshev polynomials as basic functions, it has a higher accuracy than general numerical methods, such as the FE method, FD scheme, and FVE method, but it also contains as many unknowns as the general numerical methods, so that it has to bear a lot of computing load. Thus, reducing the order for the classic CS format is more significative than for other numerical methods. For this purpose, we extract the initial L coefficient vectors $\boldsymbol{U}_N^1, \boldsymbol{U}_N^2, \ldots, \boldsymbol{U}_N^L$ ($L \ll K$) in the series of coefficient vectors $\{\boldsymbol{U}_N^n\}_{n=1}^K$ for the classic CNCS matrix format (11) to form a set of snapshots.

3 The ROECS method based on POD for 2D Sobolev equations

3.1 Formulation of POD basis

We use the set of snapshots obtained by Sect. 2.3 to form a snapshot matrix $\boldsymbol{P} = (\boldsymbol{U}_N^1, \boldsymbol{U}_N^2, \ldots, \boldsymbol{U}_N^L)$ of volume $(2N+1)^2 \times L$. Let $\lambda_j > 0$ ($j = 1, 2, \ldots, r =: \operatorname{rank}(\boldsymbol{P})$) be the positive eigenvalues of $\boldsymbol{P}\boldsymbol{P}^T$ arranged nonincreasingly, and let $\boldsymbol{U} = (\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_r) \in \mathbb{R}^{(2N+1)^2\times r}$ be the associated orthonormal eigenvectors of $\boldsymbol{P}\boldsymbol{P}^T$. Thus, the POD basis $\boldsymbol{\Phi} = (\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_d)$ ($d \le r$) is formed by the first d vectors in $\boldsymbol{U}$ and satisfies (see [36])

$$\left\|\boldsymbol{P} - \boldsymbol{\Phi}\boldsymbol{\Phi}^T\boldsymbol{P}\right\|_{2,2} = \sqrt{\lambda_{d+1}}, \tag{12}$$

where $\|\boldsymbol{P}\|_{2,2} = \sup_{\boldsymbol{\chi}\neq\boldsymbol{0}} \|\boldsymbol{P}\boldsymbol{\chi}\|_2/\|\boldsymbol{\chi}\|_2$, and $\|\boldsymbol{\chi}\|_2$ is the norm of a vector $\boldsymbol{\chi}$. Further, we obtain

$$\begin{aligned}
\left\|\boldsymbol{U}_N^n - \boldsymbol{\Phi}\boldsymbol{\Phi}^T\boldsymbol{U}_N^n\right\|_2 &= \left\|\big(\boldsymbol{P} - \boldsymbol{\Phi}\boldsymbol{\Phi}^T\boldsymbol{P}\big)\boldsymbol{e}_n\right\|_2\\
&\le \left\|\boldsymbol{P} - \boldsymbol{\Phi}\boldsymbol{\Phi}^T\boldsymbol{P}\right\|_{2,2}\left\|\boldsymbol{e}_n\right\|_2 \le \sqrt{\lambda_{d+1}}, \quad n = 1, 2, \ldots, L,
\end{aligned} \tag{13}$$

where $\boldsymbol{e}_n = (0, \ldots, 0, 1, 0, \ldots, 0)^T$ ($n = 1, 2, \ldots, L$) with the nth component equal to 1. Hence, $\boldsymbol{\Phi} = (\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_d)$ is an optimal POD basis.

Remark 3 Since the order $(2N+1)^2$ of the matrix $\boldsymbol{P}\boldsymbol{P}^T$ is far larger than the order L of the matrix $\boldsymbol{P}^T\boldsymbol{P}$, the number of the nodes of spatial meshes $(2N+1)^2$ is far larger than that of extracted snapshots L. Nevertheless, both positive eigenvalues λ_i ($i = 1, 2, \ldots, r$) are the same, and thus we may first search out the eigenvalues λ_i ($i = 1, 2, \ldots, r$) of $\boldsymbol{P}^T\boldsymbol{P}$ and the associated eigenvectors $\boldsymbol{\varphi}_i$ ($i = 1, 2, \ldots, r$), and then by the formula $\boldsymbol{\phi}_i = \boldsymbol{P}\boldsymbol{\varphi}_i/\sqrt{\lambda_i}$ ($i = 1, 2, \ldots, r$) we can gain the eigenvectors $\boldsymbol{\phi}_i$ ($i = 1, 2, \ldots, r$) associated with the positive eigenvalues λ_i ($i = 1, 2, \ldots, r$) of $\boldsymbol{P}\boldsymbol{P}^T$ and such that we can expediently obtain the POD basis.

3.2 Establishment of the ROECS model

By (13) in Sect. 3.1, we can obtain the first L ($L \leq K$) coefficient vectors of ROESE solutions: $\boldsymbol{U}_d^n = \boldsymbol{\Phi}\boldsymbol{\Phi}^T\boldsymbol{U}_N^n =: \boldsymbol{\Phi}\boldsymbol{\beta}_d^n$ ($n = 1, 2, \ldots, L$), where $\boldsymbol{U}_d^n = (u_{d,0,0}^n, u_{d,1,0}^n, \ldots, u_{d,N,0}^n, u_{d,0,1}^n, u_{d,1,1}^n, \ldots, u_{d,N,1}^n, \ldots, u_{d,0,N}^n, u_{d,1,N}^n, \ldots, u_{d,N,N}^{n+1})^T$ and $\boldsymbol{\beta}_d^n = (\beta_1^n, \beta_2^n, \ldots, \beta_d^n)^T$. When the coefficient vectors $\boldsymbol{U}_N^n$ in (11) are replaced with $\boldsymbol{U}_d^n = \boldsymbol{\Phi}\boldsymbol{\beta}_d^n$ ($n = L+1, L+2, \ldots, K$), we can obtain the following ROECS format:

$$\begin{cases} \boldsymbol{\Phi}\boldsymbol{\beta}_d^n = \boldsymbol{\Phi}\boldsymbol{\Phi}^T\boldsymbol{U}_N^n, \quad 1 \leq n \leq L, \\ (\boldsymbol{A} + \varepsilon\boldsymbol{B} + \gamma\Delta t\boldsymbol{B})\boldsymbol{\Phi}\boldsymbol{\beta}_d^{n+1} = (\boldsymbol{A} + \varepsilon\boldsymbol{B})\boldsymbol{\Phi}\boldsymbol{\beta}_d^n + \Delta t\boldsymbol{F}^{n+1}, \quad L \leq n \leq K-1, \\ \boldsymbol{U}_d^n = \boldsymbol{\Phi}\boldsymbol{\beta}_d^n, \quad n = 1, 2, \ldots, K, \end{cases} \tag{14}$$

where $\boldsymbol{U}_N^n$ ($n = 1, 2, \ldots, L$) are the initial L coefficient vectors in (11), and the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are provided in (11). Further, due to the reversibility of the matrix ($\boldsymbol{A} + \varepsilon\boldsymbol{B} + \gamma\Delta t\boldsymbol{B}$), the format (14) is abbreviated as follows:

$$\begin{cases} \boldsymbol{\beta}_d^n = \boldsymbol{\Phi}^T\boldsymbol{U}_N^n, \quad 1 \leq n \leq L, \\ \boldsymbol{\beta}_d^{n+1} = \boldsymbol{\beta}_d^n - \gamma\Delta t\boldsymbol{\Phi}^T(\boldsymbol{A} + \varepsilon\boldsymbol{B} + \gamma\Delta t\boldsymbol{B})^{-1}\boldsymbol{B}\boldsymbol{\Phi}\boldsymbol{\beta}_d^n \\ \qquad + \Delta t\boldsymbol{\Phi}^T(\boldsymbol{A} + \varepsilon\boldsymbol{B} + \gamma\Delta t\boldsymbol{B})^{-1}\boldsymbol{F}^{n+1}, \quad L \leq n \leq K-1, \\ \boldsymbol{U}_d^n = \boldsymbol{\Phi}\boldsymbol{\beta}_d^n, \quad n = 1, 2, \ldots, K. \end{cases} \tag{15}$$

Remark 4 As equation (11) contains $(N+1)^2$ unknowns at each time node, but the ROECS model, that is, the format (15) at the same time node only involves d unknowns ($d \leq L \ll (N+1)^2$, for example, $d = 6$, but $(N+1)^2 = 10{,}201$ in Sect. 4), the format (15) is obviously superior to equation (11). After we have gained $\boldsymbol{U}_d^n = (u_{d,0,0}^n, u_{d,1,0}^n, \ldots, u_{d,N,0}^n, u_{d,0,1}^n, u_{d,1,1}^n, \ldots, u_{d,N,1}^n, \ldots, u_{d,0,N}^n, u_{d,1,N}^n, \ldots, u_{d,N,N}^{n+1})^T$ ($1 \leq n \leq K$) via (15), we can obtain the ROECS solutions by the formula $u_d^n(x,y) = \sum_{j=0}^N \sum_{k=0}^N u_{d,j,k}^n h_j(x) \times h_k(y)$ ($n = 1, 2, \ldots, K$).

3.3 The existence, stability, and convergence for the ROECS solutions

To discuss the existence, stability, and convergence of the ROECS solutions, we think about the max-norms of a matrix and vector (for more detail, see [45]), which are, respectively, defined dy

$$\|\boldsymbol{D}\|_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^{l} |d_{ij}|, \quad \forall \boldsymbol{D} = (d_{ij})_{m \times l} \in \mathbb{R}^m \times \mathbb{R}^l,$$

$$\|\boldsymbol{\chi}\|_\infty = \max_{1 \leq j \leq m} |\chi_j|, \quad \forall \boldsymbol{\chi} = (\chi_1, \chi_2, \ldots, \chi_m)^T \in \mathbb{R}^m.$$

We also employ the following discrete Gronwall inequality (see [46, Lemma 3.4] or [36, Lemma 1.4.1]).

Lemma 6 *If* $\{a_n\}$ *and* $\{b_n\}$ *are two nonnegative sequences and* $\{c_n\}$ *is a positive monotone sequence satisfying*

$$a_n + b_n \leq c_n + \bar{\lambda} \sum_{i=0}^{n-1} a_i \quad (\bar{\lambda} > 0), \qquad a_0 + b_0 \leq c_0,$$

then

$$a_n + b_n \le c_n \exp(n\bar{\lambda}), \quad n = 0, 1, 2, \ldots.$$

We have the following main result of the existence, stability, and convergence of the ROECS solutions for the format (15).

Theorem 7 *Under the conditions of Theorem 5, there exists a unique series of ROECS solutions u_d^n $(n = 1, 2, \ldots, K)$ satisfying the following stability*:

$$\|\nabla u_d^n\|_{0,\omega} \le C(u_0, f), \quad n = 1, 2, \ldots, L, L+1, L+2, \ldots, K, \tag{16}$$

where $C(u_0, f)$ are positive constants dependent on u_0 and f but independent of N and Δt. Moreover, when $u(t_n) \in H_\omega^q(\Omega)$ $(2 \le q \le N+1)$, the errors between the solutions for Problem 1 and the ROECS solutions u_d^n $(n = 1, 2, \ldots, K)$ have the following estimates:

$$\|u(t_n) - u_d^n\|_{0,\omega} \le C\bigl(\Delta t^2 + N^{-q} + \sqrt{\lambda_{d+1}}\bigr), \quad 1 \le n \le K. \tag{17}$$

Proof (1) *The existence and stability for the ROECS solutions.*

Due to the reversibility of the matrix $(\boldsymbol{A} + \varepsilon \boldsymbol{B} + \gamma \Delta t \boldsymbol{B})$, from the format (15) and Remark 4 we can conclude that the format (15) has a unique series of the ROECS solutions.

From (14) we can recover the following format:

$$\boldsymbol{U}_d^n = \boldsymbol{\Phi}\boldsymbol{\Phi}^T \boldsymbol{U}_N^n, \quad 1 \le n \le L; \tag{18}$$

$$\begin{aligned} \boldsymbol{U}_d^{n+1} &= \boldsymbol{U}_d^n - \gamma \Delta t (\boldsymbol{A} + \varepsilon \boldsymbol{B} + \gamma \Delta t \boldsymbol{B})^{-1} \boldsymbol{B} \boldsymbol{U}_d^n \\ &\quad + \Delta t (\boldsymbol{A} + \varepsilon \boldsymbol{B} + \gamma \Delta t \boldsymbol{B})^{-1} \boldsymbol{F}^n, \quad L \le n \le K-1. \end{aligned} \tag{19}$$

Write $\boldsymbol{H}(x,y) = (h_0(x)h_0(y), h_1(x)h_0(y), \ldots, h_N(x)h_0(y), h_0(x)h_1(y), h_1(x)h_1(y), \ldots, h_N(x) \times h_1(y), \ldots, h_0(x)h_N(y), h_1(x)h_N(y), \ldots, h_N(x)h_N(y))^T$. Then we denote the solutions for Problem 4 by $u_N^n = (\boldsymbol{U}_N^n)^T \boldsymbol{H}(x,y) = \boldsymbol{U}_N^n \cdot \boldsymbol{H}(x,y)$. Similarly, $u_d^n = (\boldsymbol{U}_d^n)^T \boldsymbol{H}(x,y) = \boldsymbol{U}_d^n \cdot \boldsymbol{H}(x,y)$.

When $n = 1, 2, \ldots, L$, we have

$$\begin{aligned} \|u_d^n\|_{0,\omega} &= \|\boldsymbol{\Phi}\boldsymbol{\Phi}^T \boldsymbol{U}_N^n \cdot \boldsymbol{H}(x,y)\|_{0,\omega} \le \|\boldsymbol{\Phi}\boldsymbol{\Phi}^T\|_\infty \|\boldsymbol{U}_N^n \cdot \boldsymbol{H}(x,y)\|_{0,\omega} \\ &\le \|u_N^n\|_{0,\omega}, \quad n = 1, 2, \ldots, L. \end{aligned} \tag{20}$$

Furthermore, by Theorem 5 we conclude that (16) is correct for $n = 1, 2, \ldots, L$.

When $n = L+1, L+2, \ldots, K$, we rewrite (19) as follows:

$$\begin{aligned} \|\boldsymbol{U}_d^{n+1}\|_\infty &\le \|\boldsymbol{U}_d^n\|_2 + \gamma \Delta t \|(\boldsymbol{A} + \varepsilon \boldsymbol{B} + \gamma \Delta t \boldsymbol{B})^{-1} \boldsymbol{B}\|_\infty \|\boldsymbol{U}_d^n\|_\infty \\ &\quad + \Delta t \|(\boldsymbol{A} + \varepsilon \boldsymbol{B} + \gamma \Delta t \boldsymbol{B})^{-1}\|_\infty \|\boldsymbol{F}^n\|_\infty, \quad L \le n \le K-1. \end{aligned} \tag{21}$$

Moreover, from the FE method (see, e.g., [46, Lemmas 1.18 and 1.22]), the CS method (see, e.g., [6, Chapters II and III]), and the properties of matrix norms we can attain the inequalities

$$\|\boldsymbol{A}\|_\infty \le C; \qquad \|\boldsymbol{A}^{-1}\|_\infty \le C; \qquad \|\boldsymbol{B}\|_\infty \le CN; \qquad \|\boldsymbol{B}^{-1}\|_\infty \le CN^{-1}. \tag{22}$$

Furthermore, by the properties of matrixes (see [45, Lemma 1.4.1]) and (22) we obtain

$$\begin{aligned}\big\|(\boldsymbol{A}+\varepsilon\boldsymbol{B}+\gamma\Delta t\boldsymbol{B})^{-1}\big\|_\infty &= \frac{1}{\varepsilon+\gamma\Delta t}\big\|\big(\gamma\Delta t\boldsymbol{A}\boldsymbol{B}^{-1}+\boldsymbol{I}\big)^{-1}\boldsymbol{B}^{-1}\big\|_\infty\\ &\le \frac{1}{\varepsilon+\gamma\Delta t}\big\|\boldsymbol{B}^{-1}\big\|_\infty \le CN^{-1},\end{aligned} \tag{23}$$

$$\big\|(\boldsymbol{A}+\varepsilon\boldsymbol{B}+\gamma\Delta t\boldsymbol{B})^{-1}\boldsymbol{B}\big\|_\infty = \frac{1}{\varepsilon+\gamma\Delta t}\big\|\big(\gamma\Delta t\boldsymbol{A}\boldsymbol{B}^{-1}+\boldsymbol{I}\big)^{-1}\big\|_\infty \le \frac{1}{\varepsilon+\gamma}. \tag{24}$$

Thus, from (21), (23), and (24) we get

$$\big\|\boldsymbol{U}_d^{n+1}\big\|_\infty \le \big\|\boldsymbol{U}_d^n\big\|_2 + C\Delta t\big\|\boldsymbol{U}_d^n\big\|_\infty + C\Delta tN^{-1}\big\|\boldsymbol{F}^n\big\|_\infty, \quad L\le n\le K-1. \tag{25}$$

Summing (25) from L to n, we obtain

$$\big\|\boldsymbol{U}_d^{n+1}\big\|_\infty \le \big\|\boldsymbol{U}_d^L\big\|_2 + C\Delta t\sum_{i=L}^{n}\big\|\boldsymbol{U}_d^i\big\|_\infty + C\Delta tN^{-1}\sum_{i=L}^{n}\big\|\boldsymbol{F}^i\big\|_\infty, \quad L\le n\le K-1. \tag{26}$$

Applying the discrete Gronwall lemma (Lemma 6) to (26), we get

$$\big\|\boldsymbol{U}_d^{n+1}\big\|_\infty \le \left(\big\|\boldsymbol{U}_d^L\big\|_2 + C\Delta tN^{-1}\sum_{i=L}^{n}\big\|\boldsymbol{F}^i\big\|_\infty\right)\exp\big[C(n-L)\Delta t\big], \tag{27}$$

where $L\le n\le K-1$. Thus, we obtain

$$\begin{aligned}\big\|u_d^n\big\|_{0,\omega} &= \big\|\boldsymbol{\Phi}\boldsymbol{\Phi}^T\boldsymbol{U}_N^n\cdot\boldsymbol{H}(x,y)\big\|_{0,\omega}\\ &\le \big\|\boldsymbol{\Phi}\boldsymbol{\Phi}^T\big\|_\infty\big\|\boldsymbol{U}_N^n\big\|_\infty\big\|\boldsymbol{H}(x,y)\big\|_{0,\omega} \le C(u_0,f), \quad L+1\le n\le K,\end{aligned} \tag{28}$$

which shows that (16) is correct for $n=L+1,L+2,\dots,K$.

(2) *Error estimates* (17).

Set $\boldsymbol{e}^n=\boldsymbol{U}_N^n-\boldsymbol{U}_d^n$. For $n=1,2,\dots,L$, from (13) we get

$$\begin{aligned}\big\|\boldsymbol{e}^n\big\|_\infty &\le \big\|\boldsymbol{e}^n\big\|_2 = \big\|\boldsymbol{U}_N^n-\boldsymbol{U}_d^n\big\|_2\\ &= \big\|\boldsymbol{U}_N^n-\boldsymbol{\Phi}\boldsymbol{\Phi}^T\boldsymbol{U}_N^n\big\|_2 \le \sqrt{\lambda_{d+1}}, \quad n=1,2,\dots,L.\end{aligned} \tag{29}$$

For $n=L+1,L+2,\dots,K$, from (11) and (19) we obtain

$$\boldsymbol{e}^{n+1}=\boldsymbol{e}^n-\gamma\Delta t(\boldsymbol{A}+\varepsilon\boldsymbol{B}+\gamma\Delta t\boldsymbol{B})^{-1}\boldsymbol{B}\boldsymbol{e}^n, \quad L\le n\le K-1. \tag{30}$$

Further, from (24) we get

$$\begin{aligned}\big\|\boldsymbol{e}^{n+1}\big\|_\infty &\le \big\|\boldsymbol{e}^n\big\|_\infty + \gamma\Delta t\big\|(\boldsymbol{A}+\varepsilon\boldsymbol{B}+\gamma\Delta t\boldsymbol{B})^{-1}\boldsymbol{B}\big\|_\infty\big\|\boldsymbol{e}^n\big\|_\infty\\ &\le \big\|\boldsymbol{e}^n\big\|_\infty + C\Delta t\big\|\boldsymbol{e}^n\big\|_\infty, \quad L\le n\le K-1.\end{aligned} \tag{31}$$

Summing (31) from L to n, we obtain

$$\big\|\boldsymbol{e}^{n+1}\big\|_\infty \le \big\|\boldsymbol{e}^L\big\|_\infty + C\Delta t\sum_{i=L}^{n}\big\|\boldsymbol{e}^i\big\|_\infty, \quad L\le n\le K-1. \tag{32}$$

Applying the discrete Gronwall lemma to (32), from (29) we get

$$\left\|\boldsymbol{e}^n\right\|_\infty \le \left\|\boldsymbol{e}^L\right\|_\infty \exp\bigl[C\Delta t(n-L)\bigr] \le C\sqrt{\lambda_{d+1}}, \quad L+1 \le n \le K. \tag{33}$$

Thus, by $u_N^n = \boldsymbol{U}_N^n \cdot \boldsymbol{H}$, $u_d^n = \boldsymbol{U}_d^n \cdot \boldsymbol{H}$, and $\|\boldsymbol{H}\|_{0,\omega} \le 1$, with the orthogonality of elements in $\boldsymbol{H}(x,y)$ and the inverse estimate theorem, we attain

$$\begin{aligned}\left\|u_N^n - u_d^n\right\|_{0,\omega} &= \left\|\boldsymbol{e}^n \cdot \boldsymbol{H}(x,y)\right\|_{0,\omega} \le \left\|\boldsymbol{e}^n\right\|_\infty \left\|\boldsymbol{H}(x,y)\right\|_{0,\omega} \\ &\le C\sqrt{\lambda_{d+1}}, \quad 1 \le n \le K.\end{aligned} \tag{34}$$

Combining Theorem 5 and (34), we gain (17). This finishes the argument of Theorem 7. □

Remark 5 We have two annotations for the ROECS format.

(1) The factor $\sqrt{\lambda_{d+1}}$ in Theorem 7 is caused by the order reduction for the CS format and can serve as the criterion of choice of the POD basis, that is, it is necessary to choose the number of the POD basis d and L satisfying $\sqrt{\lambda_{d+1}} \le \max\{\Delta t^2, N^{-2}\}$.

(2) We clearly can get that the matrix representation of the classic CS format (11) contains $(N+1)^2$ unknowns at each time node; nevertheless, the ROECS format (15) has only d unknowns ($d \le L \ll (N+1)^2$, for instance, $d = 6$, but $(N+1)^2 = 10{,}201$ in Sect. 4) at the same time node. Therefore, in comparison with the classic CS format, the ROECS format can greatly lessen unknowns, so that it can alleviate the calculation load and save the CPU consuming time and the storage requirements in the computational process for solving 2D Sobolev equations.

3.4 The flowchart for solving the ROECS format

We further provide the flowchart of finding the ROECS solutions for 2D Sobolev equations, which consists of the following five steps.

Step 1. For given parameters ε and γ, the source term $f(x,y,t)$ and the initial function $u_0(x,y)$, the number of nodes N in the direction of x or y, the nodes $\{x_m\}_{m=0}^N = -\cos(m\pi/N)$ and $\{y_l\}_{l=0}^N = -\cos(l\pi/N)$, the time increment Δt. Solving the classic CS format (11) on the first L steps obtains the numerical solutions $\boldsymbol{U}_N^n$ $(1 \le n \le L)$.

Step 2. Put $\boldsymbol{P} = (\boldsymbol{U}_N^1, \boldsymbol{U}_N^2, \ldots, \boldsymbol{U}_N^L)_{(N+1)^2 \times L}$ and seek the positive eigenvalues $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_\kappa > 0$ ($r = \dim\{u_N^n : 1 \le n \le L\}$) and the associated eigenvectors $\boldsymbol{\varphi}_i$ $(i = 1,2,\ldots,r)$ of $\boldsymbol{P}^T\boldsymbol{P}$.

Step 3. Determine the number d of POD basis by means of the inequality $\lambda_{d+1} \le \max\{\Delta t^4, N^{-2q}\}$ and produce the POD basis $\boldsymbol{\Phi} = (\phi_1, \phi_2, \ldots, \phi_d)$ by the formula $\boldsymbol{\phi}_i = \boldsymbol{P}\boldsymbol{\varphi}_i/\sqrt{\lambda_i}$ $(1 \le i \le d)$.

Step 4. First, obtain the ROECS solutions $\boldsymbol{U}_d^n = (u_{d,0,0}^n, u_{d,1,0}^n, \ldots, u_{d,N,0}^n, u_{d,0,1}^n, u_{d,1,1}^n, \ldots, u_{d,N,1}^n, \ldots, u_{d,0,N}^n, u_{d,1,N}^n, \ldots, u_{d,N,N}^{n+1})^T$ $(1 \le n \le K)$ by solving the ROECS format, that is, the format (15), and then we can obtain the ROECS solutions by the formula $u_d^n(x,y) = \sum_{j=0}^N \sum_{k=0}^N u_{d,j,k}^n h_j(x) h_k(y)$ $(n = 1,2,\ldots,K)$.

Step 5. If $\|u_d^n - u_d^{n+1}\|_{0,\omega} \le \max\{\Delta t^2, N^{-q}\}$ $(n = L, L+1, \ldots, K-1)$, then end. Else, let $\boldsymbol{U}_N^i = \boldsymbol{U}_d^{n-L-i}$ $(i = 1,2,\ldots,L)$ and return to *Step* 2.

4 Some numerical examples

In this section, we present several sets of comparative numerical examples to show the advantage of the ROECS method for the 2D Sobolev equation.

Example 1 In the Sobolev equations, we take $f(x,y,t) = 2(\cos 2\pi x \cos 2\pi y - 1)\exp(-2t)$, $u_0(x,y) = 1 - \cos 2\pi x \cos 2\pi y$ (depicted in Fig. 1), $\varphi(x,y,t) = u_0(x,y)\exp(-2t)$, $\varepsilon = 1/\pi^2$,

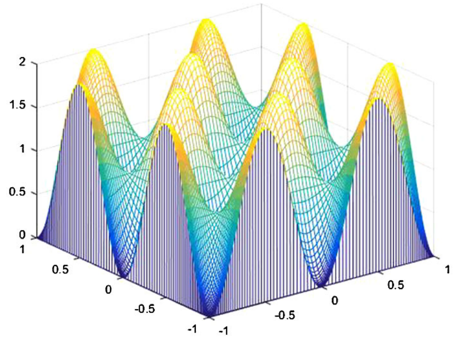


Figure 1 Initial value function when $t = 0$

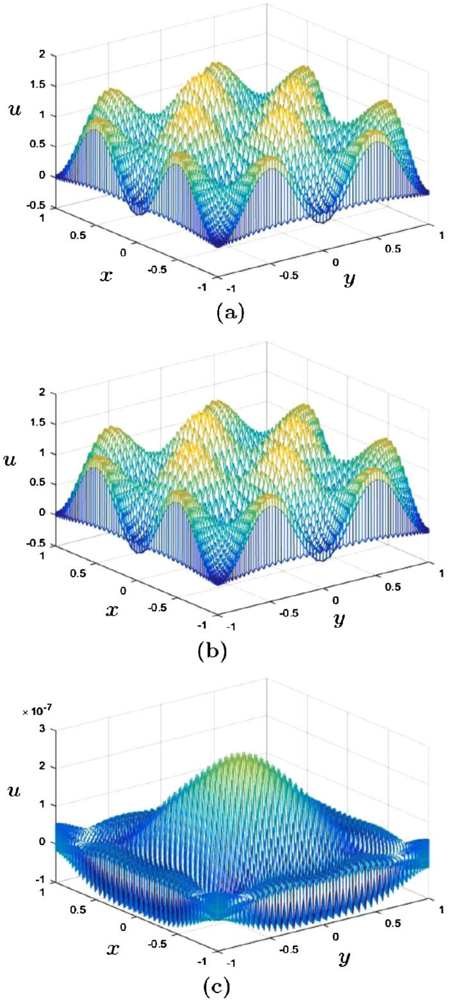
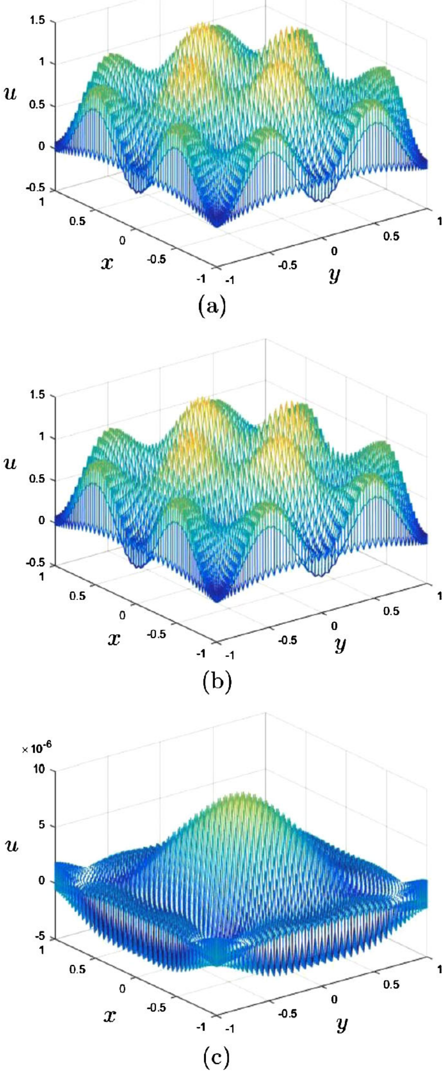


Figure 2 (**a**) The classic CS solution; (**b**) the ROECS solution; (**c**) the error between between the ROECS solution and the CS solution when $t = 0.3$

Figure 3 (**a**) The classic CS solution; (**b**) the ROECS solution; (**c**) the error between the ROECS solution and the CS solution when $t = 0.6$



$\gamma = 2/\pi^2$, the time step $\Delta t = 0.01$, and the number of nodes in two directions $N = 100$, $q = 2$.

We first compute the initial $L = 20$ coefficient vectors $\boldsymbol{U}_N^n$ of the CS solutions of (11) at time nodes t_n ($n = 1, 2, \dots, 20$) to form the snapshot matrix $\boldsymbol{P} = (\boldsymbol{U}_N^1, \boldsymbol{U}_N^2, \dots, \boldsymbol{U}_N^{20})$. Then we find the eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{20} \geq 0$ and the associated eigenvectors $\boldsymbol{\varphi}_i$ ($i = 1, 2, \dots, r$) of $\boldsymbol{P}^T\boldsymbol{P}$ according to Step 2 in Sect. 3.4. By computing we obtain that the error factor $\sqrt{\lambda_7} \leq 4 \times 10^{-4}$. Thus, we only need to produce the POD basis $\boldsymbol{\Phi} = (\phi_1, \phi_2, \dots, \phi_d)$ by the formula $\boldsymbol{\phi}_i = \boldsymbol{P}\boldsymbol{\varphi}_i/\sqrt{\lambda_i}$ ($1 \leq i \leq d$). Then, by the ROECS format, we find the ROECS solutions at $T = 0.3, 0.6, 0.9$, respectively, depicted in (b) of Figs. 2 to 4, respectively.

To make a reasonable comparison, we also compute out the classic CS solutions at $T = 0.3, 0.6, 0.9$ by the CS format (4), respectively, depicted in (a) in Figs. 2 to 4, respectively.

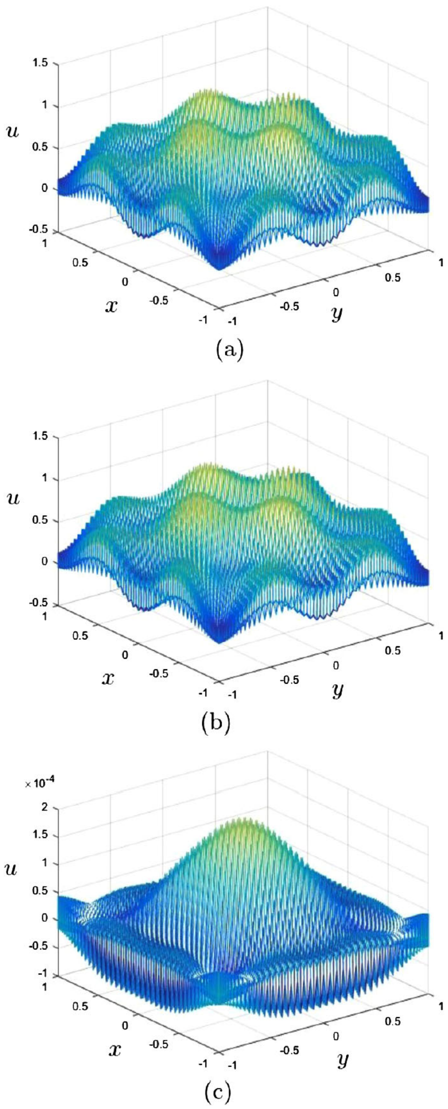


Figure 4 (**a**) The classic CS solution; (**b**) the ROECS solution; (**c**) the error between the ROECS solution and the CS solution when $t = 0.9$

The comparison of (a) and (b) in Figs. 2 to 4 exhibits a quasi-identical similarity. In the computational process, the ROECS format at each time level only contains six unknowns, whereas the classic CS format has 10,201 unknowns. Therefore, the ROECS format can not only alleviate the calculation load and reduce the accumulation of round-off errors, but can also save CPU elapsed time and resources in the computational process. Photos (c) in Figs. 2 to 4 are, respectively, the errors between the ROECS solutions and the CS solutions when $t = 0.3, 0.6, 0.9$, which accord with the theoretical errors, because both errors are $O(10^{-4})$.

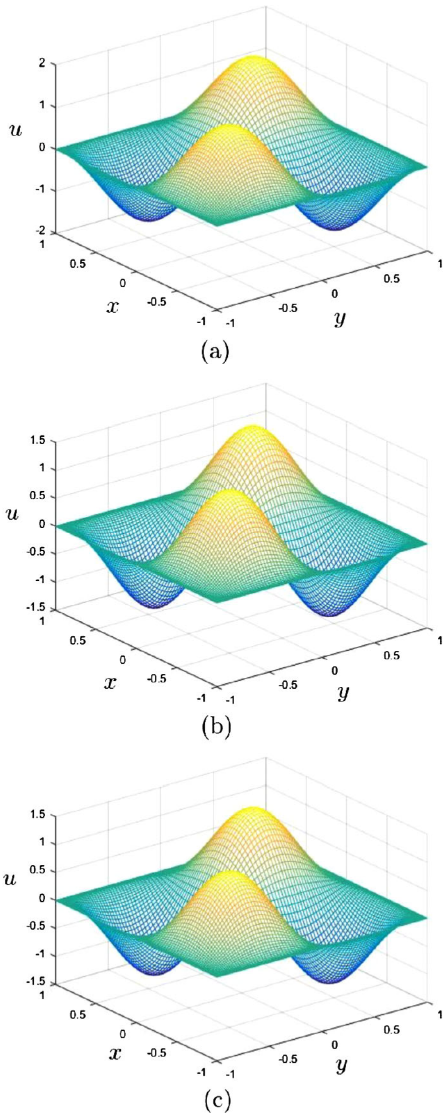
By operating records from solving the classic CS format and the ROECS format in the same Laptop (Microsoft Surface Book: Int Core i7 Processor, 16 GB RAM), we find that the CPU elapsed time for solving the classic CS format on $0 \leq t \leq 0.9$ is about 6059 seconds, but the CPU elapsed time for solving the ROECS format is about 48 seconds, that is, the CPU consuming time for solving the classic CS format is 125 times greater than for solving the ROECS format. This shows that the ROECS format is far superior to the classic CS format.

Example 2 To compare the CS and ROECS methods, we give another example, which also has an analytical solution for a 2D Sobolev equation. In the 2D Sobolev equation (1), we take $\varepsilon = 1$, $\gamma = 10$, $f(x,y,t) = (\frac{1}{2} + \varepsilon\pi^2 + 2\gamma\pi^2)\sin\pi x\sin\pi y\exp(t/2)$, $u_0(x,y) = \sin\pi x\sin\pi y$, $\varphi(x,y,t) = 0$. Then this Sobolev equation has an exact solution $u(x,y,t) = \sin\pi x\sin\pi y\exp(t/2)$.

First, we depict the exact solution $u(x,y,t) = e^{t/2}\sin\pi x\sin\pi y$ at $T = 0.9$ in (a) of Fig. 5.

Next, we take time step $\Delta t = 0.01$ and the number $N = 100$ of nodes in the x and y directions. By the classic CS format (4) we compute out the classic CS solution at $T = 0.9$, depicted in (b) of Fig. 5.

Finally, to make a reasonable comparison, in solving the ROECS format, we adopt the same time step and number of nodes as in the classic CS format. We also compute the initial $L = 20$ coefficient vectors $\boldsymbol{U}_N^n$ of the CS solutions with the classic CS format (11) at time nodes t_n $(n = 1, 2, \ldots, 20)$ to form the snapshot matrix $\boldsymbol{P} = (\boldsymbol{U}_N^1, \boldsymbol{U}_N^2, \ldots, \boldsymbol{U}_N^{20})$. Then we find the eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{20} \geq 0$ and the associated eigenvectors $\boldsymbol{\varphi}_i$ $(i = 1, 2, \ldots, r)$



(a)

(b)

(c)

Figure 5 (**a**) The analytical solution when $t = 0.9$; (**b**) the CS solution when $t = 0.9$, (**c**) The ROECS solution when $t = 0.9$

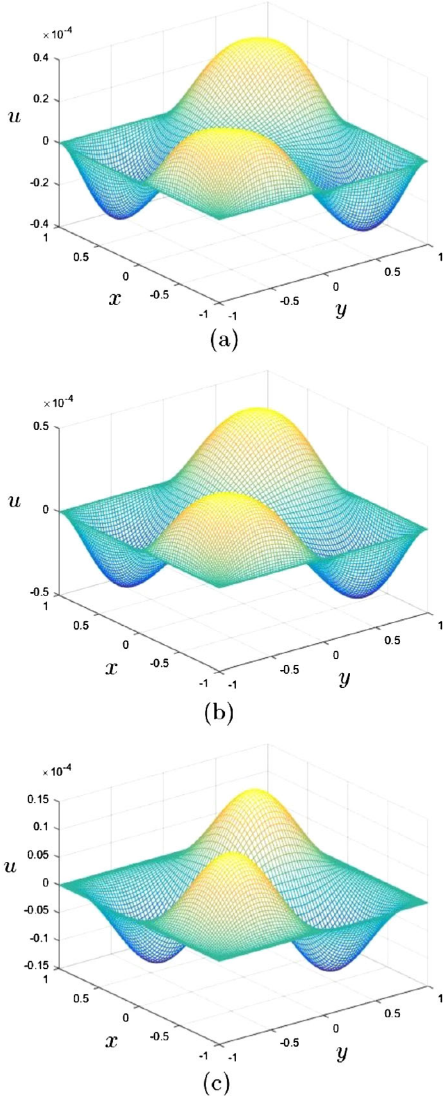


Figure 6 (**a**) The error between the CS solution and the analytical solution solution when $t = 0.9$; (**b**) the error between the ROECS solution and the analytical solution solution when $t = 0.9$; (**c**) the error between the ROECS solution and the CS solution when $t = 0.9$

of $\boldsymbol{P}^T\boldsymbol{P}$ according to Step 2 in Sect. 3.4. By computing we obtain that the error factor $\sqrt{\lambda_7} \leq 2 \times 10^{-4}$, so that it is only necessary to adopt the most main six POD bases $\boldsymbol{\Phi} = (\phi_1, \phi_2, \ldots, \phi_d)$ by the formula $\boldsymbol{\phi}_i = \boldsymbol{P}\boldsymbol{\varphi}_i/\sqrt{\lambda_i}$ $(1 \leq i \leq d)$ when solving the ROECS format. The obtaining ROECS solution at $T = 0.9$ is depicted in (c) of Fig. 5.

Moreover, we depict the errors between the analytical solution and CS solution, the analytical solution and ROECS solution, and the CS solution and ROECS solution at $T = 0.9$ in photos (a), (b), and (c) of Fig. 6, respectively.

By observing three photos of Fig. 5, we clearly find that the photos of the analytical solution, the CS numerical solution, and the ROECS numerical solution are basically identical and that the errors between the analytical solution and CS solution, the analytical solution and ROECS solution, and the CS solution and ROECS solution are less than 2×10^{-4}, which verify the correctness of the theory for error analysis because the theoretical error is $O(10^{-4})$ according to Theorem 7. Especially, the unknowns (only six) in the ROECS format are far fewer than those in the classic CS format (i.e., $(N + 1)^2 = 10{,}201$), so that the CPU consuming time of the ROECS format is far less than that of the classic CS format;

for instance, it takes only about 260 s when computing the ROECS solution at $T = 0.9$, but takes 7033 s when computing the classic CS solution at the same time in the same Laptop (Microsoft Surface Book: Int Core i7 Processor, 16 GB RAM).

5 Conclusions and discussions

In this study, we have studied the reduced-order of the coefficient vectors of the solutions for the classic CS method of 2D Sobolev equations. We have established the ROECS format in matrix form for 2D Sobolev equations via the POD technique, proven the existence, uniqueness, stability, and convergence of the ROECS solutions by the matrix means, and also given the flowchart for solving the ROECS format of 2D Sobolev equations. Moreover, we have supplied two numerical examples to verify the correctness of the theoretical analysis to explain that the ROECS format is far superior to the classic CS format because the unknowns of the ROECS format are far fewer than those of the classic CS format, so that, compared to the classic CS format, the ROECS format can greatly lessen the computational load, retard the round-off error accumulation, and save the CPU consuming time in the operational process.

Especially, the ROECS format for the 2D Sobolev equations is first presented in this paper and is a development and improvement over the existing reduced-order methods because the ROECS format has higher accuracy than other reduced-order methods, such as the reduced-order FE method, FVE method, and FD scheme. Both theory and method of this paper are new and completely different from the existing reduced-order methods.

Although we restrict our ROECS method to Sobolev equations on rectangular domain $\overline{\Omega} = [a, b] \times [c, d]$, our technique can be extended to more general domains and applied in more complex engineering problems. Therefore, our technique has important applied prospect.

Acknowledgements
The authors are thankful to the honorable reviewers and editors for their valuable suggestions and comments, which improved the paper.

Funding
This research was supported by the National Science Foundation of China grant 11671106 and Qian Science Cooperation Platform Talent grant [2017]5726-41.

Availability of data and materials
The authors declare that all data and material in the paper are available and veritable.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
Both authors contributed equally and significantly in writing this article. Both authors wrote, read, and approved the final manuscript.

Author details
[1] School of Control and Computer Engineering, North China Electric Power University, Beijing, China. [2] School of Mathematics and Physics, North China Electric Power University, Beijing, China.

Publisher's Note

Received: 11 October 2018 Accepted: 21 March 2019 Published online: 28 March 2019

References

1. Ting, T.W.: A cooling process according to two-temperature theory of heat conduction. J. Math. Anal. Appl. **45**(1), 23–31 (1974)

2. Shi, D.M.: On the initial boundary value problem of nonlinear equation of the migration of the moisture in soil. Acta Math. Appl. Sin. **13**(1), 31–38 (1990)
3. Liu, Y., Li, H., He, S., Gao, W., Mu, S.: A new mixed scheme based on variation of constants for Sobolev equation with nonlinear convection term. Appl. Math. J. Chin. Univ. **28**(2), 158–172 (2013)
4. Shi, D.Y., Wang, H.H.: Nonconforming H1-Galerkin mixed FEM for Sobolev equations on anisotropic meshes. Acta Math. Appl. Sin. **25**(2), 335–344 (2009)
5. Guo, B.Y.: Some progress in spectral methods. Sci. China Math. **56**(12), 2411–2438 (2013)
6. Guo, B.Y.: Spectral Methods and Their Applications. World Scientific, Singapore (1998)
7. Zhou, Y.J., Luo, Z.D.: A Crank–Nicolson collocation spectral method for the two-dimensional telegraph equations. J. Inequal. Appl. **2018**, 137 (2018)
8. Shen, J., Tang, T.: Spectral and High-Order Methods with Applications. Science Press, Beijing (2006)
9. Luo, Z.D., Teng, F.: A reduced-order extrapolated finite difference iterative scheme based on POD method for 2D Sobolev equation. Appl. Math. Comput. **329**, 374–383 (2018)
10. Gao, F.Z., Qiu, J.X., Zhang, Q.: Local discontinuous Galerkin finite element method and error estimates for one class of Sobolev equation. J. Sci. Comput. **41**, 436–460 (2009)
11. Jiang, Z.W., Chen, H.Z.: Error estimates for mixed finite element methods for Sobolev equation. Northeast. Math. J. **17**(3), 301–314 (2001)
12. Li, H., Luo, Z.D., An, J.: A fully discrete finite volume element formulation for Sobolev equation and numerical simulations. Math. Numer. Sin. **34**(2), 163–172 (2010)
13. Shi, D.Y., Wang, H.H., Guo, C.: Anisotropic rectangular nonconforming finite element analysis for Sobolev equations. Appl. Math. Mech. **29**(9), 1203–1214 (2008)
14. Luo, Z.D., Teng, F., Chen, J.: A POD-based reduced-order Crank–Nicolson finite volume element extrapolating algorithm for 2D Sobolev equations. Math. Comput. Simul. **146**, 118–133 (2018)
15. Lu, W.J., Zhang, F.Y.: Long-time behavior of completely discrete Fourier spectral method of solutions to Sobolev equations. J. Nat. Sci. Heilongjiang Univ. **18**(2), 5–8 (2001)
16. Jin, S.J., Luo, Z.D.: A collocation spectral method for the two-dimensional Sobolev equations. Bound. Value Probl. **2018**, 53 (2018)
17. Cazemier, W., Verstappen, R.W.C.P., Veldman, A.E.P.: Proper orthogonal decomposition and low-dimensional models for driven cavity flows. Phys. Fluids **10**(7), 1685–1699 (1998)
18. Holmes, P., Lumley, J.L., Berkooz, G.: Turbulence, Coherent Structures, Dynamical Systems and Symmetry. Cambridge University Press, Cambridge (1996)
19. Ly, H.V., Tran, H.T.: Proper orthogonal decomposition for flow calculations and optimal control in a horizontal CVD reactor. Q. Appl. Math. **60**(4), 631–656 (1989)
20. Sirovich, L.: Turbulence and the dynamics of coherent structures: parts I–III. Q. Appl. Math. **45**(3), 561–590 (1987)
21. Fukunaga, K.: Introduction to Statistical Recognition. Academic Press, New York (1990)
22. Jolliffe, I.T.: Principal Component Analysis. Springer, Berlin (2002)
23. Selten, F.M.: Baroclinic empirical orthogonal functions as basis functions in an atmospheric model. J. Atmos. Sci. **54**(16), 2099–2114 (1997)
24. Kunisch, K., Volkwein, S.: Galerkin proper orthogonal decomposition methods for parabolic problems. Numer. Math. **90**(1), 117–148 (2001)
25. Kunisch, K., Volkwein, S.: Galerkin proper orthogonal decomposition methods for a general equation in fluid dynamischs. SIAM J. Numer. Anal. **40**(2), 492–515 (2002)
26. Luo, Z.D., Chen, J., Navon, I.M., Yang, X.Z.: Mixed finite element formulation and error estimates based on proper orthogonal decomposition for the non-stationary Navier–Stokes equations. SIAM J. Numer. Anal. **47**(1), 1–19 (2008)
27. Luo, Z.D., Li, H., Zhou, Y.J., Xie, Z.H.: A reduced finite element formulation based on POD method for two-dimensional solute transport problems. J. Math. Anal. Appl. **385**(1), 371–383 (2012)
28. Cao, Y.H., Luo, Z.D.: A reduced-order extrapolating Crank–Nicolson finite difference scheme for the Riesz space fractional order equations with a nonlinear source function and delay. J. Nonlinear Sci. Appl. **11**, 672–682 (2018)
29. Luo, Z.D., Yang, X.Z., Zhou, Y.J.: A reduced finite difference scheme based on singular value decomposition and proper orthogonal decomposition for Burgers equation. J. Comput. Appl. Math. **229**(1), 97–107 (2009)
30. Sun, P., Luo, Z.D., Zhou, Y.J.: Some reduced finite difference schemes based on a proper orthogonal decomposition technique for parabolic equations. Appl. Numer. Math. **60**(1–2), 154–164 (2010)
31. Luo, Z.D., Li, H., Zhou, Y.J., Huang, X.M.: A reduced FVE formulation based on POD method and error analysis for two-dimensional viscoelastic problem. J. Math. Anal. Appl. **385**(1), 310–321 (2012)
32. Luo, Z.D., Xie, Z.H., Shang, Y.Q., Chen, J.: A reduced finite volume element formulation and numerical simulations based on POD for parabolic problems. J. Comput. Appl. Math. **235**(8), 2098–2111 (2011)
33. Benner, P., Cohen, A., Ohlberger, M., Willcox, A.K.: Model Reduction and Approximation: Theory and Algorithm. Computational Science and Engineering. SIAM, Philadelphia (2017)
34. Hesthaven, J.S., Rozza, G., Stamm, B.: Certified Reduced Basis Methods for Parametrized Partial Differential Equations. Springer, Berlin (2016)
35. Quarteroni, A., Manzoni, A., Negri, F.: Reduced Basis Methods for Partial Differential Equations. Springer, Berlin (2016)
36. Luo, Z.D., Chen, G.: Proper Orthogonal Decomposition Methods for Partial Differential Equations. Mathematics in Science and Engineering. Elsevier, Amsterdam (2018). https://www.elsevier.com/books/proper-orthogonal-decomposition-methods-for-partial-differential-equations/luo/978-0-12-816798-4
37. Luo, Z.D., Gao, J.Q.: A POD-based reduced-order finite difference time-domain extrapolating scheme for the 2D Maxwell equations in a lossy medium. J. Math. Anal. Appl. **444**, 433–451 (2016)
38. Luo, Z.D., Li, H.: A POD reduced-order SPDMFE extrapolating algorithm for hyperbolic equations. Acta Math. Sci. Ser. B Engl. Ed. **34**(3), 872–890 (2014)
39. Luo, Z.D., Li, H., Sun, P., Gao, J.Q.: A reduced-order finite difference extrapolation algorithm based on POD technique for the non-stationary Navier–Stokes equations. Appl. Math. Model. **37**(7), 5464–5473 (2013)
40. Xia, H., Luo, Z.D.: An optimized finite difference iterative scheme based on POD technique for the 2D viscoelastic wave equation. Appl. Math. Mech. **38**(12), 1721–1732 (2017)

41. Luo, Z.D., Teng, F.: Reduced-order proper orthogonal decomposition extrapolating finite volume element format for two-dimensional hyperbolic equations. Appl. Math. Mech. **38**(2), 289–310 (2017)
42. An, J., Luo, Z.D., Li, H., Sun, P.: Reduced-order extrapolation spectral-finite difference scheme based on POD method and error estimation for three-dimensional parabolic equation. Front. Math. China **10**(5), 1025–1040 (2015)
43. Luo, Z.D., Jin, S.J.: A reduced-order extrapolation spectral-finite difference scheme based on the POD method for 2D second-order hyperbolic equations. Math. Model. Anal. **22**(5), 569–586 (2017)
44. Adams, R.A.: Sobolev Spaces. Academic Press, New York (1975)
45. Zhang, W.S.: Finite Difference Methods for Partial Differential Equations in Science Computation. Higher Education Press, Beijing (2006) (in Chinese)
46. Luo, Z.D.: Mixed Finite Element Methods and Applications. Science Press, Beijing (2006) (in Chinese)