Available online at www.sciencedirect.com

ScienceDirect

Journal of Computational Physics 224 (2007) 477–488

JOURNAL OF COMPUTATIONAL PHYSICS

www.elsevier.com/locate/jcp

Short Note

On the numerical convergence with the inverse polynomial reconstruction method for the resolution of the Gibbs phenomenon

Jae-Hun Jung [a,*], Bernie D. Shizgal [b,1]

[a] Department of Mathematics, University of Massachusetts at Dartmouth, North Dartmouth, MA 02747-2300, United States
[b] Institute of Applied Mathematics, University of British Columbia, 6356 Agricultural Road, Vancouver, BC, Canada V6T 1Z2

Received 18 October 2006; received in revised form 16 January 2007; accepted 19 January 2007
Available online 2 February 2007

Keywords: Inverse polynomial reconstruction method; Matrix ill-posedness; Truncated Gaussian elimination; Gibbs phenomenon

1. Introduction

Spectral methods yield exponential convergence in the approximation of a globally smooth function [3,4,13]. However, if the function has a local discontinuity such that the function is only piecewise smooth, spectral accuracy is no longer manifested as the convergence is at most O(1) in the L_∞ norm. This is known as the Gibbs phenomenon. If the function has a discontinuity or is non-periodic in the given domain, its Fourier representation shows spurious oscillations near the discontinuity or the domain boundaries. These oscillations may deteriorate the stability when applied to the solution of time dependent problems. Several methods have been developed to deal with the Gibbs phenomenon. These methods roughly fall into two different theories; projection theory and direct-inverse theory.

Projection methods include spectral filtering methods [12], adaptive filtering method [7], spectral mollification methods [24], and the Gegenbauer reconstruction method [14,15]. Spectral filtering or spectral mollification methods map the original spectral approximation into the filtered space. The adaptive filtering method and the spectral mollification technique often utilize optimal filtered spaces. The Gegenbauer reconstruction method projects the spectral data with the inherent Gibbs phenomenon onto the polynomial spaces which satisfy the Gibbs complementary condition [14]. Direct-inverse methods include Fourier–Padé method [8], the inverse polynomial reconstruction method (IPRM) [17,18,22] and the statistical filter method [23]. The Fourier–Padé and the IPRM seek the reconstruction of the given spectral data by inverting the corresponding transformation matrix. We provide the essential aspects of the IPRM.

Suppose that $f(x)$ is *analytic* in $x \in [-1, 1]$ but not necessarily periodic. Let $f_N(x)$ be its finite Fourier representation and be given a priori such that,

[*] Corresponding author. Tel.: +1 5089998342; fax: +1 5089106917.
E-mail addresses: jjung@umassd.edu (J.-H. Jung), shizgal@chem.ubc.ca (B.D. Shizgal).
[1] Also with the Department of Chemistry, University of British Columbia, 2036 Main Mall, Vancouver, BC, Canada V6T 1Z1.

doi:10.1016/j.jcp.2007.01.018

$$f_N(x) = \sum_{k=-N}^{N} \hat{f}_k \exp(\mathrm{i}k\pi x), \tag{1}$$

where $\hat{f}_k$ are the Fourier coefficients defined with the Fourier inner product

$$\hat{f}_k := (f(x), \exp(\mathrm{i}k\pi x))_F = \frac{1}{2}\int_{-1}^{1} f(x)\exp(-\mathrm{i}k\pi x)\,\mathrm{d}x. \tag{2}$$

We seek a reconstruction, $\tilde{f}_m(x)$, in a polynomial basis set, $\{\phi_l(x)|l = 0,\ldots,m\}$, that is,

$$\tilde{f}_m(x) = \sum_{l=0}^{m} \tilde{g}_l \phi_l(x), \tag{3}$$

where $\tilde{g}_l$ are the expansion coefficients and $\phi_l(x)$ is a polynomial of degree l. The polynomial set $\{\phi_l(x)|l = 0,\ldots,m\}$ can be either orthogonal or non-orthogonal and are often chosen as the Gegenbauer polynomials, $C_l^{\lambda}(x)$, orthogonal with weight function $w(x,\lambda) = (1-x^2)^{\lambda-1/2}$, that is

$$\int_{-1}^{1} w(x,\lambda) C_l^{\lambda}(x) C_{l'}^{\lambda}\,\mathrm{d}x = \sqrt{\pi} C_l^{\lambda}(1)\frac{\Gamma(\lambda+\frac{1}{2})}{\Gamma(\lambda)(l+\lambda)}\delta_{ll'},$$

where $\Gamma(\lambda)$ is the Gamma function and

$$C_l^{\lambda}(1) = \frac{\Gamma(l+2\lambda)}{l!\Gamma(2\lambda)},$$

[1,16]. These have often been the polynomials of choice for the Gegenbauer reconstruction method [14] and the IPRM [17,22]. For the Gegenbauer reconstruction, they satisfy the important Gibbs complementary condition [14]. As discussed at length in the previous papers [17,22], the IPRM is independent of basis set [22] and one can use any polynomial set such as the Legendre polynomials, $L_l(x) = C_l^{\frac{1}{2}}(x)$, or the Chebyshev polynomials, $T_l(x) = C_l^{1}(x)$.

The expansion coefficients are obtained by making the residue between the projection of $\tilde{f}_m(x)$ in Eq. (3) onto the Fourier space and the Fourier representation, $f_N(x)$, of the original unknown function $f(x)$ orthogonal to the Fourier space. This results in equating the Fourier coefficients of $\tilde{f}_m(x)$ and the given Fourier coefficients $\hat{f}_k$ such that

$$(\tilde{f}_m(x), \exp(\mathrm{i}k\pi x))_F = \hat{f}_k, \quad k = -N,\ldots,N. \tag{4}$$

With the expansion of $\tilde{f}_m(x)$ as in Eq. (3), Eq. (4) yields the linear system

$$\mathbf{W}\cdot\tilde{\mathbf{g}} = \hat{\mathbf{f}}, \tag{5}$$

where $\tilde{\mathbf{g}}$ is the column vector composed of $m+1$ expansion coefficients, i.e. $\tilde{\mathbf{g}} = (\tilde{g}_0,\ldots,\tilde{g}_m)^{\mathrm{T}}$ and $\hat{\mathbf{f}}$ is the column vector composed of $2N+1$ Fourier coefficients, i.e. $\hat{\mathbf{f}} = (\hat{f}_{-N},\ldots,\hat{f}_N)^{\mathrm{T}}$. The matrix $\mathbf{W}$ is defined as the transformation from the polynomial space to the Fourier space, i.e.

$$W_{kl} := (\phi_l(x), \exp(\mathrm{i}k\pi x))_F = \frac{1}{2}\int_{-1}^{1} \phi_l(x)\exp(-\mathrm{i}k\pi x)\,\mathrm{d}x. \tag{6}$$

The Legendre polynomials are predominantly used in this paper as the matrix $\mathbf{W}$ is known exactly for this choice. Jung and Shizgal [17] showed that the square transformation matrix with $m = 2N$ is non-singular. Thus, the expansion coefficient vector $\tilde{\mathbf{g}}$ is given by

$$\tilde{\mathbf{g}} = \mathbf{W}^{-1}\cdot\hat{\mathbf{f}}, \tag{7}$$

and the reconstructed function is calculated from Eq. (3).

Although the IPRM is accurate for small N (see Features 1 and 2 discussed later) it was shown that the matrix $\mathbf{W}$ [17,20] is ill-conditioned as the total number of Fourier modes $2N+1$ increases, resulting in the exponential growth of the error once round-off errors begin to be dominant. In fact, the condition number of $\mathbf{W}$ is increasing “exponentially” fast. The numerical examples with the Chebyshev or Legendre basis functions [17,20] showed that the maximum error $L_\infty := \max_{-1\leqslant x\leqslant 1}|f(x) - \tilde{f}_m(x)|$ decays exponentially. However, with increasing N, L_∞ decreases to some minimum value at $N \simeq N_p$ and then increases exponentially for $N \geqslant N_p$. For example, for the function $f(x) = x$, the L_∞ error is approximately machine accuracy, ϵ_M with $m = 1$. As $m(= 2N)$ increases, the error grows until it becomes $\mathrm{O}(1)$; (Fig. 4 of [17]). Thus, although the inverse method is very accurate and exact when $f(x)$ is a polynomial, the method yields only $\mathrm{O}(1)$ for $N \gg N_p$, which is the same order of accuracy as $f_N(x)$. The exponential growth of the L_∞ error for large N was originally attributed to the exponential growth of the condition number $\kappa(\mathbf{W})$ of the transformation matrix $\mathbf{W}$ [17,20]. The main objective of this note is to reconsider the increase of the L_∞ error for $N \geqslant N_p$ and report a modification of the IPRM that inhibits this behavior.

The success of both the projection and direct-inverse methods is that they reconstruct the spectral approximation with high accuracy in sufficiently smooth regions or in the region sufficiently far from the discontinuity. However, the crucial point is the convergence behavior near or up to the discontinuity points. Fourier filtering methods and spectral mollification methods intrinsically yield only $\mathrm{O}(1)$ convergence near the singular points although one can obtain fast convergence in the smooth region from the given Fourier data. This is because these methods still use the Fourier space where the original spectral data is obtained. The Gegenbauer reconstruction method, and IPRM and the Fourier–Padé method utilize polynomials and rational functions respectively, to reconstruct the function represented with the Fourier data. By seeking a reconstruction in a different space, these methods obtain fast convergence even near the discontinuity. Here we refer to *the resolution of the Gibbs phenomenon* to the methods which recover spectral convergence from the Fourier data up to the discontinuity. These methods are different from the filtering or mollification methods as they overcome the $\mathrm{O}(1)$ convergence up to the discontinuity if the exact location of the discontinuity is known. However, the accurate reconstruction with these methods near or at the singularity are very sensitive to the numerical conditioning of the projection or inversion operators. As the number of the given Fourier data $2N + 1$ increases, the accuracy in the interior region away from the singularity is highly enhanced, but the accuracy near the singular points decreases if N is larger than a certain value.

Gelb [9] discussed the round-off errors associated with the Gegenbauer reconstruction method. In [9], it was shown that the main source of the problem comes from the exponentially increasing value of the Gegenbauer polynomial $C_l^\lambda(x)$ at $x = \pm 1$ as λ and l increase while the corresponding expansion coefficients decay also exponentially. Boyd [2] studied more extensively the ill-conditioning of the Gegenbauer reconstruction method in connection with the Runge phenomenon and showed that the original Gegenbauer reconstruction method is not free of round-off errors even though the parameter λ is optimally chosen such that a sufficiently small ratio of order m to N is ensured. Also he pointed out the lack of convergence of the expansion which behaves as a power series for high values of λ resulting in the non-convergence of the Gegenbauer method for high values of m and λ. In fact, for $\lambda = \alpha N$ and $m = \beta N$ with α, $\beta < 1$, it is inevitable for the Gegenbauer method to have poor convergence for arbitrarily large N. Gelb and Tanner [10] used the Freund polynomials with the weight function $w(x) = \exp(-cx^{2n})$ for $n \in \mathbf{Z}^+$ as an alternate Gibbs complementary basis (termed a robust Gibbs complementary basis) to overcome the ill-conditioning of the projection matrix from the Fourier to the polynomial spaces. With this new polynomial basis, the convergence results have been shown to be spectral even for large N as the theory predicts. The key feature of the new robust Gibbs complementary basis is that the weight function becomes wider as N increases while the original Gegenbauer polynomial has a weight function which becomes narrower as N and consequently λ increase. We here note that non-classical polynomials with arbitrary weight functions including Gaussian type weight functions were utilized for the Gegenbauer reconstruction method, and provide better convergence than the Gegenbauer polynomials [22].

Driscoll and Fornberg [8] demonstrated that the inversion of the transformation matrix (denoted by matrix C in [8]) from the Fourier to rational spaces is ill-conditioned and thus limits the L_∞ error to no less than 10^{-8} in double precision (see Fig. 6 in [8]). It appears to be very difficult to obtain the Padé polynomial coefficients to machine accuracy. Min et al. [19], also showed that the Fourier–Padé method provides for automatic detection of the edge locations (poles) but the convergence near the poles is rather slow in practical applications.

In this note, we first use Gaussian elimination to solve the linear system obtained with the IPRM. The expansion coefficients of an analytic function expanded in Legendre polynomials decay exponentially. We solve the linear equations, Eq. (5), with Gaussian elimination and partial pivoting which transforms $\mathbf{W}$ to an upper triangular matrix **U** [6,11]. These same algebraic procedures transform the vector of Fourier coefficients, $\hat{\mathbf{f}}$ to a new vector **h**. We show that the variation of the components of the new vector, h_k, decay spectrally and justifies the truncation of the components just below machine accuracy. This simple procedure prohibits the growth of the L_∞ error. We then show numerically that the L_∞ error with the IPRM decays exponentially with N without any growth.

The structure of this paper is as follows. It was previously suggested [17,20] that the exponential growth of the error arising from round-off errors is due to the exponential growth of the condition number of the matrix **W**. In Section 2, we discuss the main features of the IPRM and the ill-posedness of the transformation matrix **W** and show that the direct inversion yields the growth in L_∞ error. As explained earlier, we use Legendre polynomials as **W** can be written analytically for this case as well as Gegenbauer polynomials. In Section 3, we discuss the truncation procedure of the vector **h** the RHS of the linear equation following Gaussian elimination. A summary is presented in Section 4.

2. Main features of the IPRM

As discussed in the previous papers [17,22], the IPRM has the following important properties (1) it is exact if the given function $f(x)$ is a polynomial, that is, if $f(x)=\sum_{l=0}^{m'} g_l\phi_l(x)$, the expansion coefficients $\tilde{g}_l$ of the reconstructed function $\tilde{f}_m(x)$ are determined exactly such that,

$$\tilde{g}_l=\begin{cases} g_l & \text{for } l=0,\dots,m', \\ 0 & \text{for } l>m'. \end{cases} \tag{8}$$

(2) The method is independent of the basis set, that is, the final inverse reconstruction is uniquely determined regardless of the polynomial basis used. If $\tilde{f}^1_m(x)$ and $\tilde{f}^2_m(x)$ are the reconstructions with two different basis functions $\phi_l(x)$ and $\psi_l(x)$, respectively, then

$$\tilde{f}^1_m(x)=\tilde{f}^2_m(x), \tag{9}$$

for $m=2N$ and $\tilde{f}^1_m(x)$ and $\tilde{f}^2_m(x)$ have a non-vanishing maximum polynomial order m. For this reason, we can choose any polynomial set for the inverse reconstruction; (3) The method yields spectral convergence for a single domain, that is,

$$L_\infty := \max_{-1\leqslant x\leqslant 1}|f(x)-\tilde{f}_m(x)| \leqslant Cq^{\alpha N}, \tag{10}$$

where C is a constant independent of N and $0<q<1$ and $\alpha>0$. Here α is function dependent. Here we note that spectral convergence of the IPRM, Eq. (10), has been demonstrated numerically and the proof was given for the finite Fourier space [17]. The full proof of spectral convergence is under consideration and the current note focuses only on the numerical issues of convergence.

The numerical results with the Gegenbauer polynomial basis functions show that the L_∞ error decays with N up to some $N\sim N_p$ and then grows exponentially with N due to round-off errors as discussed in the Introduction. The behavior was demonstrated for different $f(x)$ (see for example Fig. 4 of [17]), and also shown later in this paper (see Fig. 2). From these results we thus note two major characteristics of the variation of the L_∞ error with N; Feature 1, and Feature 2. These features are due to the round-off errors and this note aims to resolve these features in the inverse reconstruction, especially Feature 2.

- Feature 1: For a given computing precision, ϵ_M, define $L_\infty^{\min}$ as the minimum error for $\forall N$;

  $$L_\infty^{\min}=\min_{N=0,1,\dots} L_\infty(N).$$

  Then there exists a minimum $L_\infty^{\min}\geqslant\epsilon_M$ for each $f(x)$ and which occurs for $N\sim N_p$.
- Feature 2: Once the L_∞ error reaches $L_\infty^{\min}$ at $N\sim N_p$, the L_∞ error starts to grow exponentially for $N>N_p$. Here note that the different behavior in the L_∞ versus N curves for $N>N_p$ depends on the nature of the round-off errors that arise in the inversion of **W** for different basis sets.

In order to look at the ill-posedness features of the IPRM, we consider in some detail the spectral convergence of the IPRM as given by Eq. (10).

Assume that $f(x)$ is analytic in $x \in [-1,1]$ such that it can be expanded with the infinite Legendre series, i.e.

$$f(x) = \sum_{l=0}^{\infty} g_l L_l(x), \tag{11}$$

where $L_l(x)$ is the Legendre polynomial of degree l and $g_l(x)$ is the corresponding expansion coefficient. Here note that $g_l \neq \tilde{g}_l$ in general. Then from Eqs. (10) and (11),

$$L_\infty = \max_{-1\leqslant x\leqslant 1} \left|f(x) - \tilde{f}_m(x)\right| = \max_{-1\leqslant x\leqslant 1} \left|\sum_{l=0}^{m}(g_l - \tilde{g}_l)L_l(x) + \sum_{l=m+1}^{\infty} g_l L_l(x)\right|. \tag{12}$$

We define the vector $\mathbf{g} = (g_0,\ldots,g_m)^{\mathrm{T}}$ and the matrix $\mathbf{W}^c$ and the vector $\mathbf{g}^c$,

$$W^c_{lk} := W_{lk}, \quad -N \leqslant k \leqslant N, \quad l = m+1,\ldots,$$
$$g^c_l := g_l, \quad l = m+1,\ldots$$

Then using the definitions of $\hat{\mathbf{f}}$, Eq. (2), and $\mathbf{g}$ we can write Eq. (5) in the form

$$\hat{\mathbf{f}} = \mathbf{W}\cdot\mathbf{g} + \mathbf{W}^c\cdot\mathbf{g}^c. \tag{13}$$

Thus from Eq. (7) we have

$$\tilde{\mathbf{g}} = \mathbf{V}\cdot\mathbf{W}\cdot\mathbf{g} + \mathbf{V}\cdot\mathbf{W}^c\cdot\mathbf{g}^c, \tag{14}$$

where $\mathbf{V}$ is the inverse matrix of $\mathbf{W}$ numerically evaluated with some known inversion algorithms, that is,

$$\mathbf{V} \neq \mathbf{W}^{-1}.$$

Then from Eq. (12), and using $|L_l(x)| \leqslant 1, \forall l$, we have

$$\begin{aligned} L_\infty &\leqslant \|\mathbf{g}-\tilde{\mathbf{g}}\|_1 + \|\mathbf{g}^c\|_1 = \|\mathbf{g}-\mathbf{V}\cdot\mathbf{W}\cdot\mathbf{g}-\mathbf{V}\cdot\mathbf{W}^c\cdot\mathbf{g}^c\|_1 + \|\mathbf{g}^c\|_1 \\ &\leqslant \|\mathbf{g}-\mathbf{V}\cdot\mathbf{W}\cdot\mathbf{g}\|_1 + \|\mathbf{V}\cdot\mathbf{W}^c\cdot\mathbf{g}^c\|_1 + \|\mathbf{g}^c\|_1 \leqslant \|\mathbf{g}-\mathbf{V}\cdot\mathbf{W}\cdot\mathbf{g}\|_1 + (\|\mathbf{V}\cdot\mathbf{W}^c\|_1+1)\|\mathbf{g}^c\|_1. \end{aligned} \tag{15}$$

Here $\|\cdot\|_1$ denotes the vector or matrix 1-norm. If $\mathbf{V} = \mathbf{W}^{-1}$ exactly, the first term on the RHS of (15) vanishes and we find that the convergence of the IPRM is solely determined by $\|\mathbf{V}\cdot\mathbf{W}^c\|_1$ and $\|\mathbf{g}^c\|_1$. It is well known that $\|\mathbf{g}^c\|_1$ decays exponentially if $f(x)$ is analytic in $x \in [-1,1]$. Our numerical results in this and previous papers with the IPRM suggest that the errors decay exponentially with N as given by Eq. (10). Thus $\|\mathbf{V}\cdot\mathbf{W}^c\|_1$ decays or at least grows slower than $\|\mathbf{g}^c\|_1$ decays. Since the second term decays fast, the first term on the RHS, $\|\mathbf{g}-\mathbf{V}\cdot\mathbf{W}\cdot\mathbf{g}\|_1$, is the main source of the ill-posedness of the IPRM.

Now consider the inverse reconstruction of a simple test function $f(x) = x$ with Legendre polynomials to clearly see the ill-posedness of the IPRM. By definition, $\mathbf{g}^c = \mathbf{0}$ and $\mathbf{g} = (0,1,0,\ldots,0)^{\mathrm{T}}$, and from Eq. (15), we have

$$L_\infty \leqslant \|\mathbf{g}-\mathbf{V}\cdot\mathbf{W}\cdot\mathbf{g}\|_1 + (\|\mathbf{V}\cdot\mathbf{W}^c\|_1+1)\|\mathbf{g}^c\|_1 \tag{16}$$
$$= \|\mathbf{g}-\mathbf{V}\cdot\mathbf{W}\cdot\mathbf{g}\|_1$$
$$\leqslant \|\mathbf{I}-\mathbf{V}\cdot\mathbf{W}\|_1\|\mathbf{g}\|_1 = \|\mathbf{I}-\mathbf{V}\cdot\mathbf{W}\|_1. \tag{17}$$

It is interesting that the decay or growth rate of the L_∞ error for $f(x) = x$ is determined only by $\mathbf{W}$ and its inverse. The RHS of Eq. (17) measures the error of the inverse matrix $\mathbf{V}$. Since the error of the inverse matrix is roughly given by the product of the condition number κ of $\mathbf{W}$ and machine accuracy ϵ_M, we have

$$L_\infty \leqslant \|\mathbf{I}-\mathbf{V}\cdot\mathbf{W}\|_1 \leqslant A\epsilon_M\kappa(\mathbf{W}), \tag{18}$$

where A is a constant independent of N. From Eq. (18), we know that L_∞ error is bounded by the condition number of $\mathbf{W}$.

Fig. 1 shows the ill-posedness features of the IPRM. Fig. 1(A) shows the singular values of $\mathbf{W}$ for $N = 64$. As shown in the figure, the smallest singular value of $\mathbf{W}$ is already as small as machine accuracy ϵ_M which

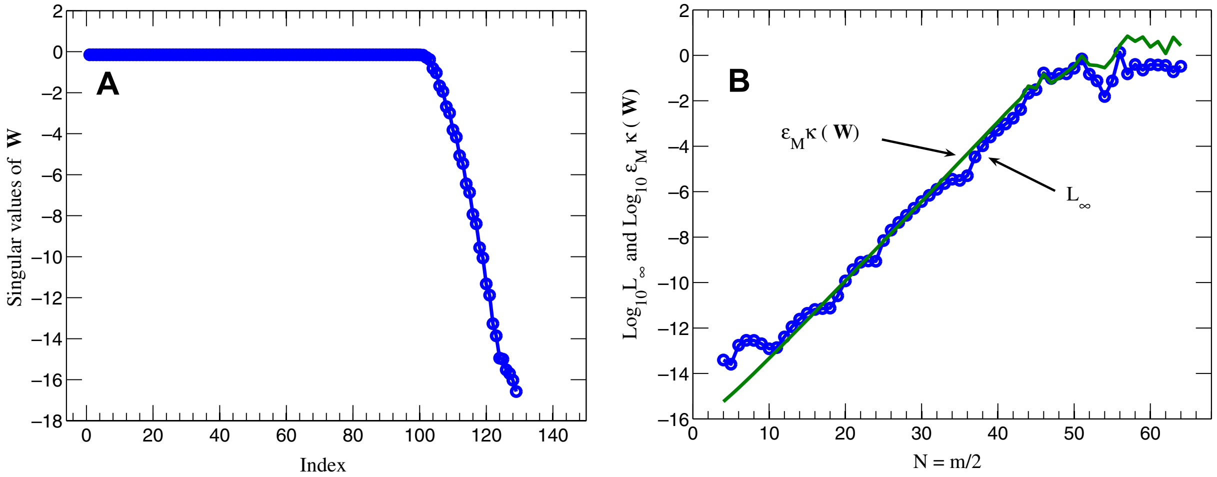

Fig. 1. (A) The singular values of **W** for $N = 64$. (B) The L_∞ errors and $\epsilon_M \kappa(\mathbf{W})$ versus N. For figure (B) the test function $f(x) = x$ is used.

yields a large condition number and the linear system Eq. (7) is ill-posed. For the inversion of **W**, MATLAB® backslash operation ('\') is used. Here we note that the MATLAB® backslash operation uses Gaussian elimination. Fig. 1(B) shows the variation of L_∞ and the value of $A\epsilon_M \kappa(\mathbf{W})$ with N on a logarithmic scale for the inverse reconstruction of $f(x) = x$. For the reconstruction, the Legendre polynomials and the exact $\hat{\mathbf{f}}$ and **W** are used. For Eq. (18), we choose $A = 1$ and $\epsilon_t = 10^{-16}$. As shown in the figure, the L_∞ error and the condition number grow exponentially and both have similar growth rates as Eq. (18) indicates. Let

$$L_\infty = C_1 p^N,$$

and

$$\kappa(\mathbf{W}) = C_2 q^N,$$

where $p, q > 1$ and C_1 and C_2 are constants. From a linear fit to the numerical results of Fig. 1(B), we find that

$$p \sim 1.8, \quad q \sim 2.0,$$

which shows that the estimate Eq. (18) agrees with the numerical results for $f(x) = x$. Eq. (18) and Fig. 1(B) shows that the condition number $\kappa(\mathbf{W})$ accounts for the exponential growth of the L_∞ error.

3. Gaussian elimination with truncation

For the numerical experiments, we use MATLAB® 'inv' command for the direct inversion and backslash command '\' which uses Gaussian elimination with partial pivoting to solve Eq. (5). Both inversion algorithms show the ill-posedness features, Features 1 and 2.

Fig. 2(A) shows the exponential decay ($N \leqslant 12$) followed by the exponential growth of L_∞ versus N for the inverse reconstruction of $f(x) = \cos[1.4\pi(x + 1)]$ with Legendre polynomials. The error for large N remains O(1). Fig. 2(B) shows the pointwise errors ($N = 2, 4, 8, 12$, and 32) of the final reconstruction with the MATLAB® backslash operation. The figure shows that the exponential increase of the maximum error L_∞ arises primarily from the boundaries (see the pointwise errors with $N = 32$). This also implies that if there exists a jump discontinuity inside the domain, the error also grows until its magnitude reaches O(1) in the neighborhood of the discontinuity. It is useful to keep in mind that for the IPRM, the size of the transformation matrix **W** is determined by the maximum polynomial order m of the final reconstruction $\tilde{f}_m(x)$. Let N' be the largest Fourier mode used for the reconstruction out of the first $2N + 1$. Then $m + 1 = 2N' + 1$, i.e., the total number of the Fourier modes used for the reconstruction is a function of the polynomial order m. Thus the polynomial order m is the crucial factor that determines the extent of the round-off errors of the transformation matrix **W**.

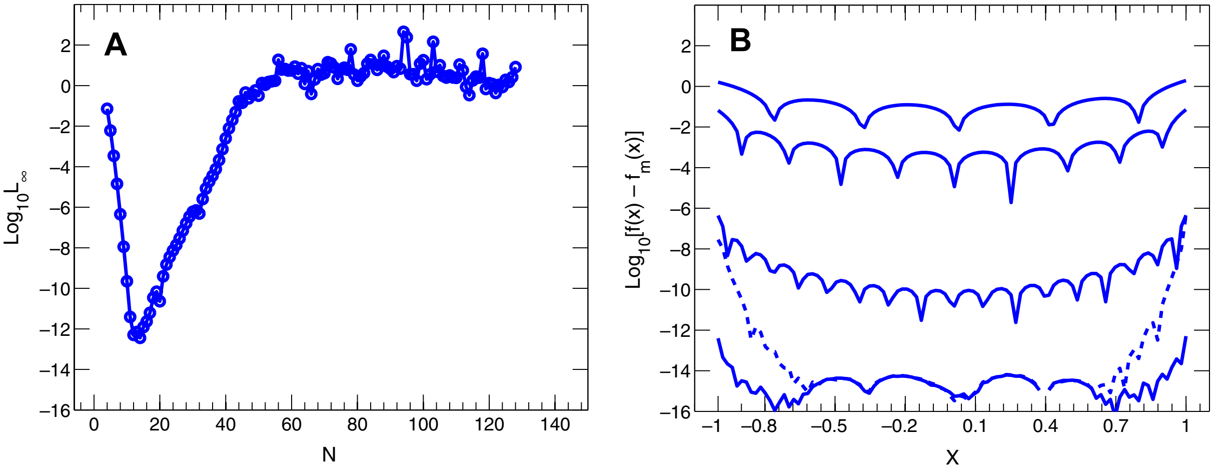

Fig. 2. The inverse reconstruction with the Legendre polynomials for $f(x) = \cos[1.4\pi(x+1)]$. (A) L_∞ error versus N. (B) Pointwise errors with $N = 2, 4, 8, 12, 32$, from top to bottom. The dotted line denotes the pointwise errors for $N = 32$ and the other solid lines for $N = 2, 4, 8, 12$ from top to bottom. Note that we clearly see the exponential convergence even at the boundaries for $N = 2, 4, 8, 12$. However, after $N \sim 12$, the error with the Legendre polynomials grows from the boundaries while the errors in the region away the boundaries are still small (see figure B with $N = 32$).

We note in presenting these results that complement similar results in the previous papers that the IPRM gives a very good reconstruction for non-polynomial functions.

We propose a truncation method with Gaussian elimination in order to prevent the exponential growth of L_∞. To explain the truncation method, consider a function $f(x)$ which is analytic but not periodic in $x \in [-1, 1]$. Then its Fourier coefficients $\hat{\mathbf{f}}$ decay algebraically slowly. However, the elements of the expansion coefficient vector $\mathbf{g}$ when $f(x)$ is expanded in some basis set decay exponentially fast. In particular, if $f(x)$ is a polynomial of order p, we expect that the first p components of $\tilde{\mathbf{g}}$ are non-zero and the remainder are identically zero if the inversion, Eq. (7), could be performed exactly. Since Gaussian elimination transforms $\mathbf{W}$ to an upper triangular matrix $\mathbf{U}$ we first define the Gaussian elimination operation $\mathbf{P}$ [21] such that

$$\mathbf{U} := \mathbf{P} \cdot \mathbf{W} \tag{19}$$

and the equation to be inverted by backward substitution is

$$\mathbf{U} \cdot \tilde{\mathbf{g}} = \mathbf{h}. \tag{20}$$

After Gaussian elimination, the Fourier coefficient vector $\hat{\mathbf{f}}$ is mapped into $\mathbf{h}$

$$\mathbf{h} := \mathbf{P} \cdot \hat{\mathbf{f}}, \tag{21}$$

which decays exponentially as shown in Fig. 3 for $f(x) = \cos[1.4\pi(x+1)]$. To explain why $\mathbf{h}$ decays also exponentially as does $\mathbf{g}$, consider a test function which is a polynomial of degree m. Then we know that $g_l = 0$ for $\forall l > m$. Since $\mathbf{U}$ is an upper triangular matrix, from Eq. (20) $h_l = 0$ identically for $\forall l > m$.

Fig. 3 shows the variation of the components of the vectors of $\mathbf{g}$, $\tilde{\mathbf{g}}$, $\mathbf{h}$ and $\hat{\mathbf{f}}$ decay. We use the same test function $f(x) = \cos[1.4\pi(x+1)]$ and Legendre polynomials for the inverse reconstruction. The exact Legendre expansion coefficients are given by

$$g_l = \frac{2l+1}{2} \int_{-1}^{1} f(x) L_l(x)\, \mathrm{d}x.$$

Fig. 3 shows that the expansion coefficients g_l and the elements of the RHS vector $\mathbf{h}$ decay exponentially and they have almost the same decay rate. In the figure, the exact Legendre expansion coefficients are denoted by g_l which are evaluated using quadrature rules [5]. Note that both $\mathbf{g}$ and $\mathbf{h}$ do not change but remain about 10^{-16} once the minimum value reaches ϵ_M. Components $\hat{f}_k$ of the Fourier coefficient vector $\hat{\mathbf{f}}$, the RHS vector in the linear system with the IPRM, Eq. (5) are also shown in Fig. 3 to decay algebraically slowly. After Gaussian

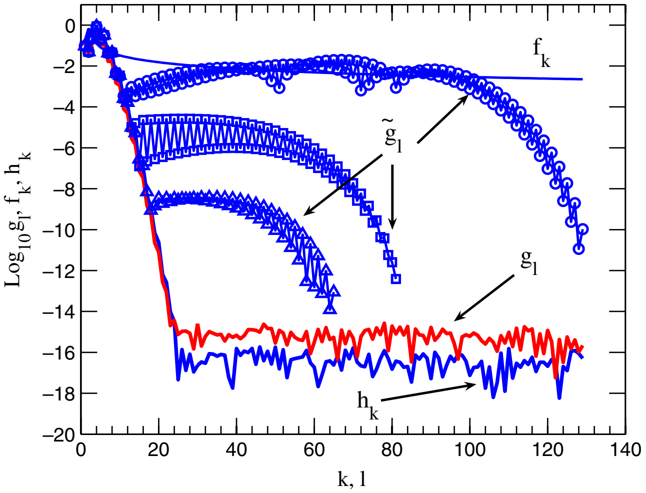

Fig. 3. The variation of f_k, g_l, h_k and $\tilde{g}_l$ versus k or l for $f(x) = \cos[1.4\pi(x+1)]$. $\hat{f}_k$ denotes the Fourier coefficients, g_l the exact Legendre expansion coefficients, $\tilde{g}_l$ the approximated expansion coefficients with the IPRM and h_k the mapped Fourier coefficients. The symbols △, □ and ○ denote $\tilde{g}_l$ for $N = 32, 40$, and 64, respectively.

elimination, the Fourier coefficient vector $\hat{\mathbf{f}}$ is mapped into $\mathbf{h}$ and the decay rate is dramatically changed such that h_k decays exponentially. The figure also shows the reconstructed expansion coefficient vector $\tilde{\mathbf{g}}$ with the IPRM denoted by $\tilde{g}_l$ for different $N = 32$ (△), 40 (□), and 64 (○). As the figure shows, the expansion coefficients with the IPRM also decay exponentially with the same rate as the exact expansion coefficients up to a certain polynomial degree m_p, with $m_p = 17, 15, 11$ for $N = 32, 40, 64$, respectively. For $m > m_p$, the decay rates of the reconstructed expansion coefficients $\tilde{g}_l$ for each N depart from the exponential rates.

Here we note that the inversion of $\mathbf{U}$ is not involved for the evaluation of $\mathbf{h}$. The major observation from the numerical experiments is that $\mathbf{h}$ has an exponential decay rate but this deteriorates after the inversion of $\mathbf{U}$ which is a highly ill-conditioned matrix analogous to $\mathbf{W}$. That is, the mapping $\mathbf{P}$ transforms $\hat{\mathbf{f}}$ into $\mathbf{h}$ that decays exponentially but the following mapping $\mathbf{U}^{-1}$ fails to make $\mathbf{h}$ exponential because $\mathbf{U}$ is ill-conditioned:

$$\mathbf{P} : \hat{\mathbf{f}} \mapsto \mathbf{h},$$
$$\mathbf{U}^{-1} : \mathbf{h} \mapsto \tilde{\mathbf{g}}.$$

Note that $\kappa(\mathbf{U}) = \kappa(\mathbf{W})$.

Based on this observation, a simple truncation method is proposed with which the round-off errors due to the ill-conditioned matrix $\mathbf{U}$ is reduced by exploiting the exponentially decaying $\mathbf{h}$. We truncate $\mathbf{h}$ with a certain tolerance level ϵ_t which is taken to be slightly larger than ϵ_M such that

$$h_k = 0 \quad \text{if } |h_k| \leqslant \epsilon_t.$$

The tolerance level ϵ_t should be taken so as to eliminate the round-off errors due to $\mathbf{U}^{-1}$ when the absolute value of the elements of $\mathbf{h}$ reaches ϵ_M. If $\epsilon_t < \epsilon_M$, the inaccurate h_k are used for the computation of $\tilde{g}_l$ and consequently there exists the growth of L_∞ error. If $\epsilon_t \gg \epsilon_M$, the errors due to the inaccurate h_k are reduced, but the minimum L_∞ error is only bounded by ϵ_t although there is no growth in L_∞ error,

$$\min_N L_\infty \sim O(\epsilon_t).$$

We suggest that ϵ_t be taken close to but larger than ϵ_M, for example, $\epsilon_t \approx 10\epsilon_M$.

Fig. 4(A) and (B) shows the L_∞ errors versus N for $f(x) = \cos[1.4\pi(x+1)]$ and $f(x) = \exp[\sin(2.7x) + \cos(x)]$, respectively, both of which show the results of Gaussian elimination *with* and *without* truncation of $\mathbf{h}$. These functions were previously used in [22]. The maximum value of N is 64 with which we obtain the expansion coefficient up to $\tilde{g}_{128}$. For the truncation of $\mathbf{h}$ we use a tolerance level, $\epsilon_t = 10^{-15}$. The line marked with circle is the plot of the L_∞ errors with N *without* the truncation of $\mathbf{h}$ and the line marked with □ is the plot

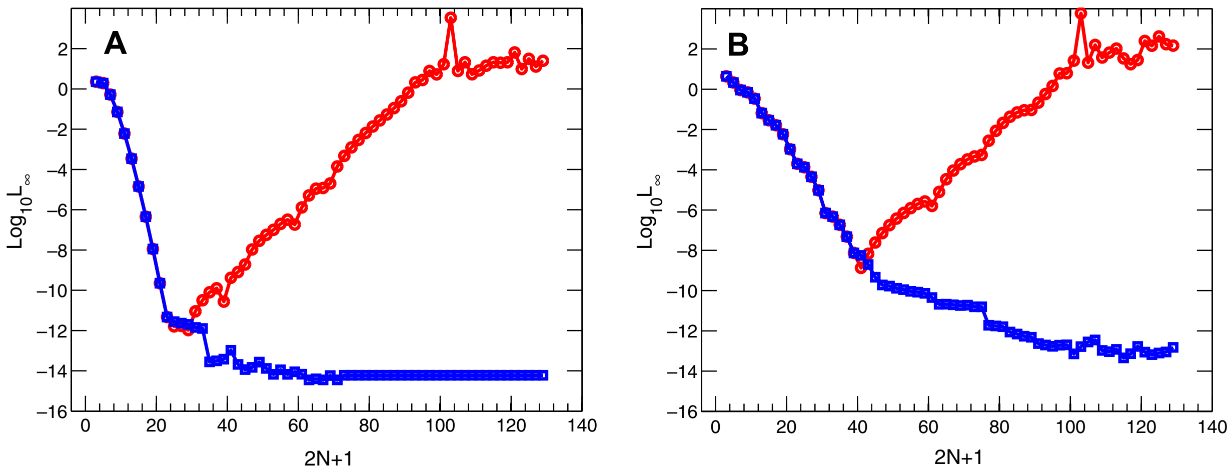

Fig. 4. The variation of L_∞ versus N arising from Gaussian elimination with partial pivoting *with* and *without* truncating **h**. The filled circle symbol denotes the results *without* the truncation and the square symbol *with* the truncation. (A) $f(x)=\cos[1.4\pi(x+1)]$. (B) $f(x)=\exp[\sin(2.7x)+\cos(x)]$.

with the truncation of **h**. As shown in the figures, the results are remarkable. The figure shows that the L_∞ errors decay *with* N without any growth if **h** is truncated. It is also shown that the predicted spectral convergence with large N is obtained with machine accuracy even for $f(x)=\exp[\sin(2.7x)+\cos(x)]$ which yields a slow convergence with the Legendre polynomial expansion. Compared to the numerical results with this function in [8,22], the truncation method proposed in this paper yields a remarkable performance without any growth in L_∞ error.

Various Gegenbauer polynomials $C_l^\lambda(x)$ with different values of λ are also considered with the truncation method. We use $\lambda = 0.5, 20, 60$, and 120 with $N = 64$. Fig. 5(A) and (B) shows the L_∞ errors for $f(x)=\cos[1.4\pi(x+1)]$ and $f(x)=\exp[\sin(2.7x)+\cos(x)]$ respectively. In the figures, the symbols ○, □, ×, and ◇ denote the results with the Gegenbauer polynomials with $\lambda = 0.5, 20, 60$, and 120, respectively. The tolerance levels for Fig. 5(A) and (B) were chosen to be $\epsilon_t = 10^{-14}$ and $\epsilon_t = 5\times10^{-13}$, respectively. It is well known that with finite precision the round-off errors are severe in the computation of the Gegenbauer polynomials with large λ [9] for which the tolerance level is slightly increased from $\epsilon_t = 10^{-15}$. Both figures show that the L_∞

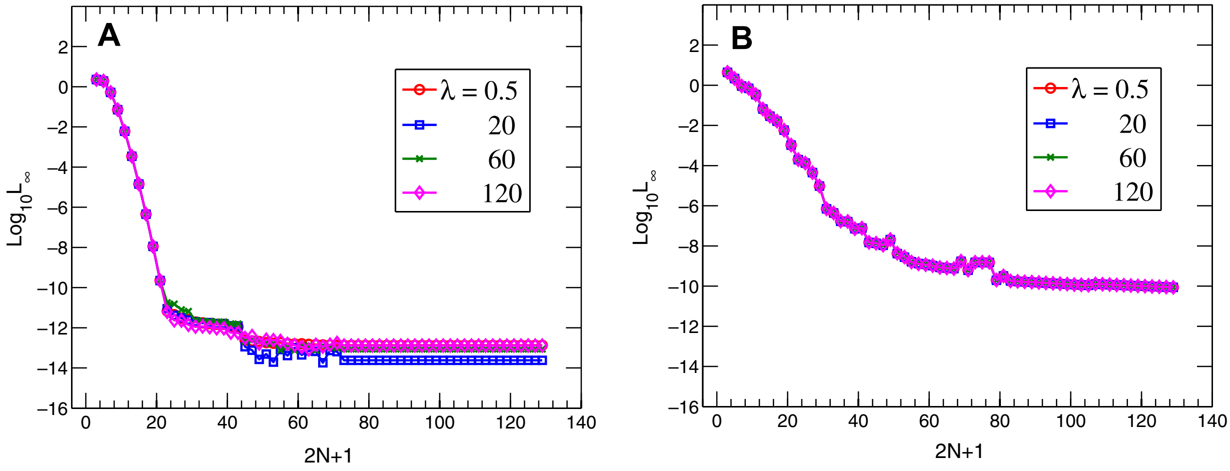

Fig. 5. The L_∞ errors with different Gegenbauer polynomials of $\lambda = 0.5$ (○), 20 (□), 60 (×), and 120 (◇). (A) $f(x)=\cos[1.4\pi(x+1)]$. (B) $f(x)=\exp[\sin(2.7x)+\cos(x)]$.

errors still decay fast without any growth even with a value of λ as large as $\lambda = 120$. Fig. 5(A) shows that the inverse reconstruction with $\lambda = 20$ yields the best results for large N. It also shows that the inverse reconstruction with $\lambda = 60$ provides better results than the case with the Legendre polynomials, i.e. $\lambda = 0.5$ for large N. Fig. 5(B) shows that the results are almost the same for all λ. These numerical results show that the proposed truncation method yields the results of non-growing L_∞ errors with various Gegenbauer polynomials as well.

We also compare the truncated IPRM with a filtering method which involves applying a filter function, $\sigma(k)$, to the RHS $\mathbf{h}$ in Eq. (20) such that

$$\mathbf{U} \cdot \tilde{\mathbf{g}} = \mathbf{\Sigma} \cdot \mathbf{h},$$

where $\mathbf{\Sigma} = \mathrm{diag}(\sigma(0), \sigma(1), \ldots, \sigma(2N))$. We use a simple exponential filter function $\sigma(k) = \exp(\alpha(k/2N)^p)$ with $\alpha = \log \epsilon_M$ and $k = 0, 1, \ldots, 2N$. Fig. 6 shows the L_∞ errors with N for $f(x) = e^{x^2+x^8}$ with the IPRM, truncated IPRM and filtered IPRM. Here note that the x-axis is N as the even test function chosen is represented by the Fourier cosine series and can be expanded in polynomials of only even order. The symbols × and ○ denote the L_∞ errors with the IPRM and the truncated IPRM, respectively. The truncation level ϵ_t for the truncated IPRM is chosen as $\epsilon_t = 10^{-15}$. The filtering orders p used are, $p = 32$ (+), 8 (□), and 4 (△) for the filtered IPRM. Although the filtered IPRM gives better results with increasing p, the L_∞ error grows with N for all filtering orders. The IPRM with $p \to \infty$ is equivalent to the IPRM without the filter. The figure shows that the truncated IPRM provides the best result.

Finally, we consider the Runge function

$$f(x) = \frac{1}{1 + 25x^2}, \quad x \in [-1, 1].$$

It is well known that this function yields the so-called Runge phenomenon in the polynomial interpolation with equidistance collocation points. Here note that the Runge function is continuous and periodic in $x \in [-1, 1]$. Thus there is no Gibbs phenomenon for this function. It is interesting, however, to apply the IPRM with the polynomial basis to the Fourier approximation of the Runge function. In [17], it was shown that if the function is analytic, the IPRM reconstruction of the Fourier approximation yields exponential convergence in the Fourier space for large N. Fig. 7 shows the IPRM reconstruction of the Runge function. With the same reasoning as with the test function used for the filtered IPRM, only the cosine series is used and the reconstruction is only with the polynomials of even order. The symbols × and ○ denote the L_∞ errors with the IPRM and the truncated IPRM. For the truncated IPRM, the truncation level $\epsilon_t = 10^{-15}$ is used. The figure

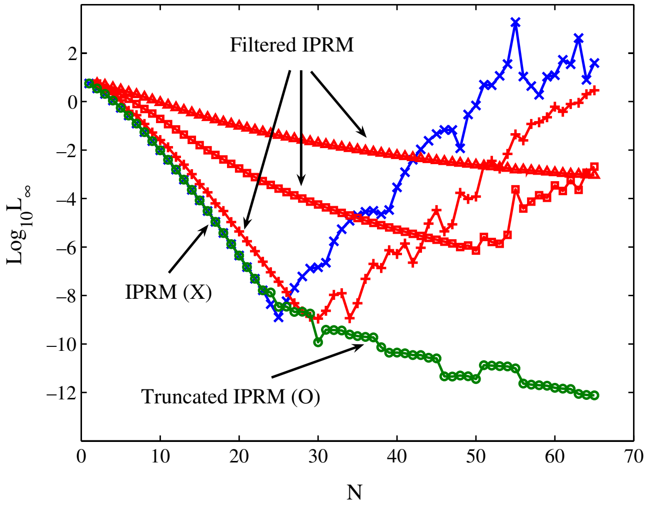

Fig. 6. The L_∞ errors versus N for $f(x) = e^{x^2+x^8}$ with the IPRM, truncated IPRM and filtered IPRM. The symbols × and ○ denote the L_∞ errors with the IPRM and the truncated IPRM with $\epsilon_t = 10^{-15}$, respectively. For the filtered IPRM, the filtering orders p are $p = 32$ (+), 8 (□), and 4 (◇). The truncated IPRM provides the best result.

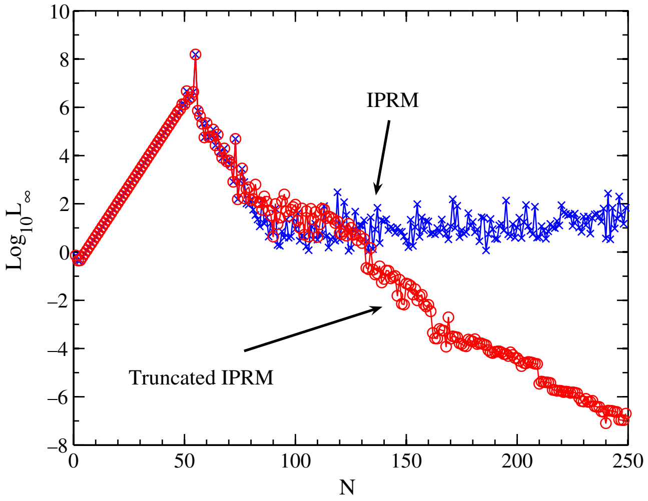

Fig. 7. The IRPM reconstruction of the Runge function $f(x) = \frac{1}{1+25x^2}$. (×): without truncation. (○): with truncation with $\epsilon_t = 10^{-15}$.

shows that the L_∞ error actually grows with N, but it decays exponentially beyond a certain N. This implies that the IPRM indeed yields exponential convergence as $N \to \infty$. The truncated IPRM clearly shows that the IPRM reconstruction decays without any L_∞ growth.

4. Summary

In this note, we use the Gaussian elimination for the inversion of the transformation matrix $\mathbf{W}$ between Fourier and Legendre basis functions, Eq. (6). We show that the exponential growth of the L_∞ errors when the round-off errors become dominant with the inverse polynomial reconstruction method is due to the ill-posedness of the transformation matrix $\mathbf{W}$ and consequently due to the numerical calculation of $\mathbf{W}^{-1}$. We show that the algebraically decaying Fourier coefficients $\hat{f}_k$ are mapped to h_k by the Gaussian upper triangularization procedure which exhibit an exponential decay rate. The h_k are mapped into $\tilde{g}_l$ by the inversion of the upper triangular matrix $\mathbf{U}$ and do not maintain the desired exponential decay rate due to the ill-conditioned matrix $\mathbf{U}$. Based on this observation, a simple truncation method has been proposed with which we use the truncation of $\mathbf{h}$ with a certain tolerance level ϵ_t and show that the L_∞ errors decay exponentially with N without any growth. It is also shown that the proposed truncation method yields the same performance with various Gegenbauer polynomials of different λ and that the inverse reconstruction provides accurate results even with large λ. Additional numerical examples are provided for the filtered IPRM and the IPRM reconstruction of the Runge function. From the numerical results, it is confirmed that the truncated IPRM yields the best results for both examples. Numerical examples used in our previous works are used to confirm the resolution of the L_∞ error growth due to the ill-posedness of the IPRM with the proposed truncation method of $\mathbf{h}$.

Acknowledgements

JHJ gratefully acknowledges the support from NSF under Grant No. DMS-0608844. The research of BDS is supported in part by a grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada. We are grateful to Chen Greif for many useful discussions. A portion of this work was carried out during a visit by BDS to the Mathematics Department of the University of Nice.

References

[1] H. Bateman, Higher Transcendental Functions, vol. 2, McGraw-Hill, New York, 1953.

[2] J.P. Boyd, Trouble with Gegenbauer reconstruction for defeating Gibbs phenomenon: Runge phenomenon in the diagonal limit of Gegenbauer polynomial approximations, J. Comput. Phys. 204 (2005) 253–264.

[3] J.P. Boyd, Chebyshev and Fourier Spectral Methods, Dover, New York, 2000.

[4] C. Canuto, M.Y. Hussaini, A. Quarteroni, T.A. Zang, Spectral Methods in Fluid Dynamics, Springer Series in Computational Physics, Springer, New York, 1988.

[5] P.J. Davis, P. Rabinowitz, Methods of Numerical Integration, Academic Press, New York, 1989.

[6] J. Demmel, Applied Numerical Linear Algebra, SIAM, Philadelphia, 1997.

[7] W.-S. Don, D. Gottlieb, J.-H. Jung, A multidomain spectral method for supersonic reactive flows, J. Comput. Phys. 192 (2003) 325–354.

[8] T.A. Driscoll, B. Fornberg, A Padé-based algorithm for overcoming the Gibbs phenomenon, Numer. Algorithms 26 (2001) 77–92.

[9] A. Gelb, Parameter optimization and reduction of round off error for the Gegenbauer reconstruction method, J. Sci. Comput. 20 (2004) 433–459.

[10] A. Gelb, J. Tanner, Robust reprojection methods for the resolution of Gibbs phenomenon, Appl. Comput. Harmon. Anal. 20 (2006) 3–25.

[11] G.H. Golub, C.F. Van Loan, Matrix Computations, third ed., Johns Hopkins University Press, Baltimore, 1996.

[12] D. Gottlieb, J.S. Hesthaven, Spectral methods for hyperbolic problems, J. Comput. Appl. Math. 128 (2001) 83–131.

[13] D. Gottlieb, S. Orszag, Numerical Analysis of Spectral Methods: Theory and Applications, SIAM, Philadelphia, 1977.

[14] D. Gottlieb, C.-W. Shu, On the Gibbs phenomenon and its resolution, SIAM Rev. 39 (1997) 644–668.

[15] D. Gottlieb, C.-W. Shu, A. Solomonoff, H. Vandeven, On the Gibbs phenomenon I: Recovering exponential accuracy from the Fourier partial sum of a nonperiodic analytic function, J. Comput. Appl. Math. 43 (1992) 81–92.

[16] I.S. Gradshteyn, I.M. Ryzhik, Table of Integrals, Series, and Products, sixth ed., Academic Press, San Diego, 2000.

[17] J.-H. Jung, B.D. Shizgal, Generalization of the inverse polynomial reconstruction method in the resolution of the Gibbs phenomena, J. Comput. Appl. Math. 172 (2004) 131–151.

[18] J.-H. Jung, B.D. Shizgal, Inverse polynomial reconstruction of Two Dimensional Fourier images, J. Sci. Compt. 25 (2005) 367–399.

[19] M.S. Min, S.M. Kaber, W.-S. Don, Fourier–Padé approximations and filtering for the spectral simulations of incompressible Boussinesq convection problem, Math. Comput. in press.

[20] R. Pasquetti, On the inverse methods for the resolution of the Gibbs phenomenon, J. Comput. Appl. Math. 170 (2004) 303–315.

[21] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, Numerical Recipes in Fortran 77: The Art of Scientific Computing, Cambridge University Press, Cambridge, 1997.

[22] B.D. Shizgal, J. -H Jung, Towards the resolution of the Gibbs phenomena, J. Comput. Appl. Math. 161 (2003) 41–65.

[23] A. Solomonoff, Reconstruction of a discontinuous function from a few fourier coefficients using Bayesian estimation, J. Sci. Comput. 10 (1995) 29–80.

[24] E. Tadmor, J. Tanner, Adaptive mollifiers for high resolution recovery of piecewise smooth data from its spectral information, Found. Comput. Math. 2 (2002) 155–189.