Spectrally Accurate Solution of Non-Periodic Boundary-Value Problems by Gegenbauer Expansions

A. Weill [*] M. Israeli [†] L. Vozovoi [‡]

Abstract

In this paper we apply the Fourier-Gegenbauer (FG) method, introduced in [4], to evaluate spatial derivatives of discontinuous but piecewise analytic functions. The basic conception of this method consists of the reexpansion of the partial sum of Fourier series of a function, which does not converge in the maximum norm (Gibbs phenomenon), into a rapidly convergent Gegenbauer series. This technique is extended in order to construct the Gegenbauer series for the derivatives. Although the derivatives of discontinuous functions are not in L_2, the exponential convergence of truncated Gegenbauer series can be proved, and the rate of convergence can be estimated. parameters

When the FG method is applied to the solution of a boundary-value problem with a modified Helmholtz operator, an intermediate solution may have steep profiles near the boundaries. These steep regions introduce a large error into the final solution, wich has (presumably) a smooth profile. A method which compensates for this loss of accuracy by using appropriately constructed boundary Green's functions. is proposed.

Key words: Gibbs phenomenon, Gegenbauer polynomials.

AMS subject classifications: 65N35, 65B99.

1 Introduction

Fourier spectral methods, dealing with the approximation of functions by trigonometric series, are highly efficient for the solution of differential equations, for the following reasons: first, the differential operators are represented in the transform space by diagonal matrices, therefore decoupling harmonics with different wave numbers. Then, the pseudo-spectral Fourier method is compatible with a fast transform (FFT) on an regularly spaced grid. Finally, for time-dependent problems the use of a uniform spatial grid permits larger stability bounds on the time step than in polynomial methods([5]).

It is known, however, that trigonometric series converge exponentially fast only for analytic and periodic functions. For non-periodic functions, having a discontinuous periodic extension, Fourier series do not converge uniformly in the interval. Away from the boundaries the rate of convergence is $O(1/N)$, while near the boundaries oscillations of order $O(1)$, which do not decrease with N, appear (the Gibbs phenomenon) .

The Fourier method can be successfully used, however, for the solution of non-periodic problems if the functions are preliminary smoothed. In [2] the trigonometric basis was employed along with a smoothing procedure, using an appropriately constructed bell function. However, such a smoothing procedure requires the knowledge of the function on an extended domain, which is not possible in case of non-periodicity.

In [6, 4] it was shown that the first Fourier coefficients $\hat{f}(k)$, $|k| \leq N$ of an analytic but not periodic function $f(x), x \in [-1,1]$ contain enough information to construct a spectrally accurate approximation to this function by a Gegenbauer expansion. This expansion is spectrally accurate on the whole interval, including the point of discontinuity itself ($x = \pm 1$). It was proven that if the number of terms and the parameter λ of the Gegenbauer polynomials $C_l^{\lambda}(x)$ are proportional to the number of Fourier modes, then this series converges exponentially with N.

[*] Geophysical Fluid Dynamics Lab., Forrestal Campus, Princeton University, Princeton, NJ 08542

[†] Computer Science Department, Technion Israel Institute of Technology, Haifa 32000, Israel

[‡] Computer Science Department, Technion Israel Institute of Technology, Haifa 32000, Israel

ICOSAHOM'95: Proceedings of the Third International Conference on Spectral and High Order Methods. ©1996 Houston Journal of Mathematics, University of Houston.

In the present paper we extend the Fourier-Gegenbauer (FG) method of [6, 4] to evaluate, within spectral accuracy, the derivatives of an analytic but non-periodic function. The convergence of truncated Gegenbauer series with N is not ensured automatically for the derivatives, since the Fourier series for the derivatives are not necessarily bounded. which is We will demonstrate in this paper that there exists a parametric region where Gegenbauer series for the derivatives converge exponentially.

The application of the FG method to the solution of differential equations, in particular, to the modified Helmholtz equation

$$u'' - \mu^2 u = -\mu^2 f(x), \qquad x \in [a,b] \tag{1}$$

which is frequently used in CFD applications, faces additional difficulties. For $\mu \gg 1$, a particular solution, which is obtained in an intermediate step of the numerical method, has a large gradient near the boundaries. This gradient cannot be resolved accurately by the present method. Thus, a large error is introduced into the final solution, even if it is smooth and does not contain boundary layers. We propose a correction procedure, using appropriately constructed homogeneous solutions, in order to recover the spectral accuracy.

The outline of this paper is as follows: in section 2 we set up the stage for the rest of the paper with a brief description and notations for the Fourier-Gegenbauer method. In section 3, estimates for the accuracy of Gegenbauer integration and differentiation are given. In the next section, a method for improving the convergence of the FG method, based on successive smoothings of the original function, is described. Finally, in section 5 we apply the FG method to the solution of non-periodic boundary-value problems while preserving the spectral accuracy.

2 The Fourier-Gegenbauer method

In this section we briefly describe the Fourier-Gegenbauer method of [6, 4]. Consider an analytic but not periodic function $f(x)$ defined in $[-1,1]$. Such a function has discontinuities at the boundaries $x = \pm 1$ if it is extended periodically with period 2. The Fourier coefficients of $f(x)$ are defined by

$$\hat{f}(k) = \frac{1}{2}\int_{-1}^{1} f(x)e^{-ik\pi x}dx \tag{2}$$

Assume that the first $2N+1$ Fourier coefficients $\hat{f}(k)$ are given. Our objective is to recover the function $f(x)$ on $x \in [-1,1]$ with exponential accuracy in the maximum norm.

The truncated Fourier series for a discontinuous function $f(x)$

$$f_N(x) = \sum_{k=-N}^{N} \hat{f}(k)e^{ik\pi x} \tag{3}$$

converges slowly, like $O(\frac{1}{N})$, inside the interval and exhibits $O(1)$ spurious oscillations near the boundaries $x = \pm 1$ known as Gibbs phenomenon. Thus there is no convergence in the maximum norm.

The basic approach of [4] consists of reexpanding Eq. (3) into rapidly convergent Gegenbauer series

$$f(x) = \sum_{l=0}^{\infty} \hat{f}^{\lambda}(l)C_l^{\lambda}(x) \tag{4}$$

where $C_l^{\lambda}(x)$ is the two-parametric family of the Gegenbauer polynomials (l is the order of the polynomial, λ is a parameter. The formula for computation of the polynomials $C_l^{\lambda}(x)$ can be found in [1], page 782).

The Gegenbauer coefficients are defined by

$$\hat{f}^{\lambda}(l) = \frac{1}{h_l^{\lambda}}\int_{-1}^{1}(1-x^2)^{\lambda-\frac{1}{2}}f(x)C_l^{\lambda}(x)dx \tag{5}$$

where

$$h_l^{\lambda} = \pi^{\frac{1}{2}}C_l^{\lambda}(1)\frac{\Gamma(\lambda+1/2)}{\Gamma(\lambda)(l+\lambda)} \tag{6}$$

As we do not know the function $f(x)$, but rather its truncated Fourier series Eq. (3), we have only an approximation to $\hat{f}^{\lambda}(l)$ which we denote by $\hat{g}_N^{\lambda}$:

$$\hat{g}_N^{\lambda}(l) = \frac{1}{h_l^{\lambda}}\int_{-1}^{1}(1-x^2)^{\lambda-\frac{1}{2}}f_N(x)C_l^{\lambda}(x)dx. \tag{7}$$

It is a remarkable fact that the approximate Gegenbauer coefficients $\hat{g}_N^{\lambda}(l)$ can be explicitly expressed in terms of the Fourier coefficients $\hat{f}(k)$ as follows:

$$\hat{g}_N^{\lambda}(l) = \delta_{0l}\hat{f}(0) + \Gamma(\lambda)i^l(l+\lambda)\sum_{0<|k|\le N} J_{l+\lambda}(\pi k)\left(\frac{2}{\pi k}\right)^{\lambda}\hat{f}(k) \tag{8}$$

where $\Gamma(\lambda)$ and $J_n(x)$ are the Gamma and the Bessel functions. The corresponding Gegenbauer expansion, based on the approximate coefficients $g_N^{\lambda}(l)$ will be then:

$$f_{M,N}^{\lambda}(x) = \sum_{l=0}^{M}\hat{g}_N^{\lambda}(l)C_l^{\lambda}(x) \tag{9}$$

We shall refer to Eqs. (9, 9) as the *Fourier-Gegenbauer* (FG) approximation of $f(x)$. The transformation from $f(x)$ to $\hat{g}_N^\lambda(l)$ will be denoted $\mathcal{G}$.

The difference between the Gegenbauer partial sum with M terms of the function $f(x)$

$$f_M^\lambda(x) = \sum_{l=0}^{M} \hat{f}^\lambda(l) C_l^\lambda(x) \tag{10}$$

and that of the truncated Fourier series $f_N(x)$ is called the *truncation error*:

$$\begin{aligned} TE(x,f,\lambda,M,N) &= \left| f_M^\lambda(x) - f_{M,N}^\lambda(x) \right| \\ &= \left| \sum_{l=0}^{M} (\hat{f}^\lambda(l) - \hat{g}_N^\lambda(l)) C_l^\lambda(x) \right| \end{aligned} \tag{11}$$

It measures the error in the finite Gegenbauer expansion due to the truncation of the Fourier series. This error decays exponentially with N provided that both λ and M are proportional to (but less than) N. For example, the relations $M = \lambda = N/4$ can guarantee such a decay.
The total error of the FG approximation

$$E(x,f,\lambda,M,N) = \mid f(x) - f_{M,N}^\lambda(x) \mid$$

can be split into two components as follows:

$$\begin{aligned} E(x,f,\lambda,M,N) &= \mid f(x) - f_M^\lambda(x) + f_M^\lambda(x) - f_{M,N}^\lambda(x) \mid \\ &\le \mid f(x) - f_M^\lambda(x) \mid + \\ &\quad \mid f_M^\lambda(x) - f_{M,N}^\lambda(x) \mid . \end{aligned} \tag{12}$$

The second component is the truncation error (11). The first component

$$RE(x,f,\lambda,M,N) = \left| \sum_{l=0}^{\infty} \hat{f}^\lambda(l) C_l^\lambda(x) - \sum_{l=0}^{M} \hat{f}^\lambda(l) C_l^\lambda(x) \right|$$

arises due to truncation of the Gegenbauer series. It is called the *regularization error.*

3 Convergence of the Fourier-Gegenbauer series for derivatives and integrals

Our purpose is to construct a spectrally accurate approximation to the derivatives (integrals) of an analytic and not periodic function $f(x)$. As in the case of interpolation, we are given only the first $2N+1$ Fourier coefficients $\hat{f}(k)$ defined in Eq. (2). Knowing $\hat{f}(k)$, we can represent the derivatives (integrals) of the function $f(x)$ in the spectral space. For the r-th derivative $f^{(r)}(x)$, $r = 1, 2, ...,$ we have:

$$\hat{f}_r(k) \equiv (i\pi k)^r \hat{f}(k), \qquad |k| \le N \tag{13}$$

and, similarly, for the integral $I(x) = \int_{-1}^{x} f(t)dt$:

$$\hat{I}(k) \equiv \frac{\hat{f}(k)}{i\pi k}, \qquad |k| \le N. \tag{14}$$

A "natural" way to construct an approximation to the derivatives or to the integral is to implement the FG algorithm, using the coefficients (13) or (14) instead of $\hat{f}(k)$ in Eq. (9).

We consider first the case of the derivatives. The Fourier partial sum for the rth derivative of a function $f(x)$ is defined by:

$$f_N^{(r)}(x) = \sum_{k=-N}^{N} \hat{f}_r(k) e^{ik\pi x} \tag{15}$$

The Gegenbauer coefficients for $f^{(r)}$ and $f_N^{(r)}$ are given respectively by:

$$\hat{f}_r^\lambda(l) = \frac{1}{h_l^\lambda} \int_{-1}^{1} (1-x^2)^{\lambda-\frac{1}{2}} f^{(r)}(x) C_l^\lambda(x) dx \tag{16}$$

$$\hat{g}_{N,r}^\lambda(l) = \frac{1}{h_l^\lambda} \int_{-1}^{1} (1-x^2)^{\lambda-\frac{1}{2}} f_N^{(r)}(x) C_l^\lambda(x) dx \tag{17}$$

where h_l^λ is defined in (6). Then the FG approximation to the rth derivative of $f(x)$ will be:

$$f_{M,N}^{(r)}(x) = \sum_{l=0}^{M} \hat{g}_{N,r}^\lambda(l) C_l^\lambda(x) \tag{18}$$

(note that $f_{M,N}^{(r)}(x)$ depends also on λ).

An estimate of the truncation error of this approximation is given by the following lemma:

Lemma 3.1 *Given a function $f(x)$ in $L^2(-1,1)$, there exists a constant $\bar{A}$ independent of λ, M, N such that the truncation error in the FG expansion of the r-th derivative of $f(x)$ satisfies the following estimate:*

$$TE(x, f^{(r)}, \lambda, M, N) \le \bar{A} \Phi_r(M,\lambda) \left(\frac{2}{\pi N}\right)^{\lambda-r-1} \tag{19}$$

We start with the case $r = 1$. We shall prove Lemma 3.1 for this case and show the generalization to $r > 1$.

Proof Using the definitions (16), (17) of the Gegenbauer coefficients for $f'(x)$ and $f'_N(x)$ ($r = 1$), the definition (11) of the truncation error, and the equality $\max_{-1\leq x\leq 1} | C_l^\lambda(x) | = C_l^\lambda(1)$ (see [3], page 206) we have:

$$TE(x, f', \lambda, M, N) \leq$$

$$M \max_{0\leq l\leq M} \max_{-1<x<1} \left|(\hat{f}_1^\lambda(l) - \hat{g}_{N,1}^\lambda(l))\right| \left|C_l^\lambda(x)\right| \tag{20}$$

$$\leq M \max_{0\leq l\leq M} \frac{C_l^\lambda(1)}{h_l^\lambda}$$

$$\left|\int_{-1}^1 (f'(x) - f'_N(x))\, C_l^\lambda(x)(1-x^2)^{\lambda-1/2}dx\right|$$

We have to find now a bound for the integral :

$$\mathcal{I}_1 = \int_{-1}^1 (f'(x) - f'_N(x))\, C_l^\lambda(x)(1-x^2)^{\lambda-1/2}dx \tag{21}$$

It is convenient to introduce the following notations :

$$\begin{aligned} T_N(x) &= f(x) - f_N(x) \\ T'_N(x) &= f'(x) - f'_N(x) \\ W_l(x) &= C_l^\lambda(x)(1-x^2)^{\lambda-\frac{1}{2}} \\ W'_l(x) &= (1-x^2)^{\lambda-\frac{3}{2}}(l+2\lambda-1)\left(-xC_l^\lambda(x) + C_{l-1}^\lambda(x)\right) \end{aligned}$$

(in the last expression we used the differential relation for $C_l^\lambda(x)$ - see [1], page 783).

Replacing the relevant terms in Eq. (21) and performing integration by parts, we obtain:

$$\begin{aligned} \mathcal{I}_1 &= \int_{-1}^1 T'_N(x)W_l(x)dx \\ &= [T_N(x)W_l(x)]_{-1}^1 - \int_{-1}^1 T_N(x)W'_l(x)dx \end{aligned} \tag{22}$$

The first component vanishes because $W_l(x)$ is zero at the end points ±1. Substituting the expression from Eq. (22) and using the following relation for the Gegenbauer polynomials

$$xC_l^\lambda(x) - C_{l-1}^\lambda(x) = \frac{l+1}{2(\lambda-1)}C_{l+1}^{\lambda-1}(x) \tag{23}$$

(it can be derived after some manipulations with the recursion formulas in [1], page 782) we have:

$$\begin{aligned} \mathcal{I}_1 &= \int_{-1}^1 T'_N(x)W(x)dx \\ &= \Psi_1(l,\lambda)\int_{-1}^1 T_N(x)(1-x^2)^{\lambda-\frac{3}{2}}C_{l+1}^{\lambda-1}(x)dx \end{aligned}$$

where

$$\Psi_1(l,\lambda) = \frac{(l+2\lambda-1)(l+1)}{2(\lambda-1)} \tag{24}$$

Combining (21), (21) and (24), we obtain:

$$TE(x, f', \lambda, M, N) \leq$$

$$M \max_{0\leq l\leq M} \Psi_1(l,\lambda)\frac{C_l^\lambda(1)}{h_l^\lambda}$$

$$\left|\int_{-1}^1 T_N(x)C_{l+1}^{\lambda-1}(x)(1-x^2)^{\lambda-\frac{3}{2}}dx\right|$$

At this point we substitute the expression

$$T_N(x) = f(x) - f_N(x) = \sum_{|k|>N} \hat{f}(k)e^{ik\pi x} \tag{25}$$

into (25), and using the relation

$$\frac{1}{h_l^\lambda}\int_{-1}^1 e^{in\pi x}C_l^\lambda(x)(1-x^2)^{\lambda-\frac{1}{2}}dx =$$

$$\Gamma(\lambda)i^l(l+\lambda)J_{l+\lambda}(\pi n)(\frac{2}{\pi n})^\lambda \tag{26}$$

(see [3], page 178) together with the boundedness of the Fourier coefficients of $f(x)$ ($f(x) \in L^2[-1,1]$).

$$| \hat{f}(k) | \quad \leq \quad A \tag{27}$$

we are left with the following estimate:

$$TE(x, f', \lambda, M, N) \leq$$

$$MA \max_{0\leq l\leq M} \Psi_1(l,\lambda)\,\frac{h_{l+1}^{\lambda-1}}{h_l^\lambda}\,\Gamma(\lambda-1)(l+\lambda)$$

$$C_l^\lambda(1)\sum_{|k|>N}\left(\frac{2}{\pi k}\right)^{\lambda-1} |J_{l+\lambda}(\pi k)|$$

Since $| J_\nu(x) |\leq 1$ for all x and $\nu \geq 0$, we obtain, after some algebra:

$$TE(x, f', \lambda, M, N) \leq \tag{28}$$

$$\bar{A}\max_{0\leq l\leq M} \Phi_1(l,\lambda)\sum_{|k|>N}\left(\frac{2}{\pi|k|}\right)^{\lambda-1}.$$

where

$$\Phi_1(l,\lambda) = 2\,\frac{\Gamma(\lambda)}{\Gamma(2\lambda)}\,\frac{(l+\lambda)\,\Gamma(l+2\lambda)}{l!}. \tag{29}$$

Here we also used the notation $\bar{A} = MA$, Eq. (6) for h_l^λ and the relation

$$C_l^\lambda(1) \quad = \quad \frac{\Gamma(l+2\lambda)}{l!\Gamma(2\lambda)} \tag{30}$$

(see [3], page 206).

It is easily seen that $\Phi_1(l,\lambda)$ is an increasing function of l. Thus, we have:

$$TE(x,f',\lambda,M,N) \leq \bar{A}\Phi_1(M,\lambda)\sum_{|k|>N}\left(\frac{2}{\pi|k|}\right)^{\lambda-1}$$

$$\leq \bar{A}\Phi_1(M,\lambda)\left(\frac{2}{\pi N}\right)^{\lambda-2} \tag{31}$$

The truncation error for this case is $O(\frac{1}{N^{\lambda-2}})$ □

We have found that the estimate (31) is valid for the FG expansion of the first derivative of $f(x)$.

The truncation error for the approximation of $f^{(r)}(x)$ is given by :

$$\begin{aligned}\mathcal{I}_r &\underset{\cdot}{=} \int_{-1}^{1} T_N^{(r)}(x)W_l(x)dx \\ &= \left[T_N^{(r-1)}(x)W_l(x)\right]_{-1}^{1} - \int_{-1}^{1} T_N^{(r-1)}(x)W_l'(x)dx\end{aligned}$$

where $T_N^{(r-1)}(x)$ is the truncation error for the approximation of the $(r-1)$th derivative. The integration by parts can be repeated $r-1$ times,
yielding:

$$\mathcal{I}_r = \int_{-1}^{1} T_N^{(r)}(x)W_l(x)dx$$

$$= (-1)^r\int_{-1}^{1} T_N(x)W_l^{(r)}(x)dx \tag{32}$$

since the expressions in brackets vanish at the end points if $\lambda > r$.

It can be shown that:

$$W_l^{(s)}(x) = (1-x^2)^{\lambda-\frac{1}{2}-s}C_{l+s}^{\lambda-s}(x)\Psi_s(l,\lambda) \tag{33}$$

$$\Psi_s(l,\lambda) = \frac{\prod_{p=1}^{s}\left[(l+p)(l+2\lambda-p)\right]}{2^s\prod_{p=1}^{s}(\lambda-p)} \tag{34}$$

(for $s=1$ it coincides with Ψ_1 of Eq. (24)).

Substituting Eq. (34) into Eq. (32), using again the bound Eq. (27), and combining Eqs. (26), (6) and (30), we have finally:

$$TE(x,f^{(r)},\lambda,M,N) \leq$$

$$\bar{A}\max_{0\leq l\leq M}\Psi_r(l,\lambda)\frac{C_l^\lambda(1)}{h_l^\lambda}$$

$$\left|\int_{-1}^{1}\left(\sum_{|k|>N}e^{ik\pi x}\right)C_{l+r}^{\lambda-r}(x)(1-x^2)^{\lambda-\frac{1}{2}-r}dx\right|$$

$$\leq \bar{A}\Phi_r(M,\lambda)\sum_{|k|>N}\left(\frac{2}{\pi|k|}\right)^{\lambda-1}$$

$$\leq \bar{A}\Phi_r(M,\lambda)\left(\frac{2}{\pi N}\right)^{\lambda-2}$$

where

$$\Phi_r(l,\lambda) = 2^r\,\frac{\Gamma(\lambda)}{\Gamma(2\lambda)}\,\frac{(l+\lambda)\,\Gamma(l+2\lambda)}{l!} \tag{35}$$

(compare with Eq. (29) for $r=1$). Therefore for the rth derivative the truncation error will be of the order $O(\frac{1}{N^{\lambda-r-1}})$. □

A similar proof applies for the case of the integration, (which is a simpler one, since $I(x)$ is in $L^2[-1,1]$). For the integral of $f(x)$, the truncation error $TE(x,I,\lambda,M,N)$ is of order $O(\frac{1}{N^\lambda})$. Likewise, successive integrations can be performed on the Fourier coefficients, gaining a power of $1/N$ at each integration in the bound for the truncation error.

Following the demonstration of [4] it can be shown that the truncation error in the approximation of the derivatives and integrals of $f(x)$ becomes exponentially small when there is a linear relation between M,λ and N.

The results are shown in Figs. 1 and 2 for the function $f(x)=x^3$.

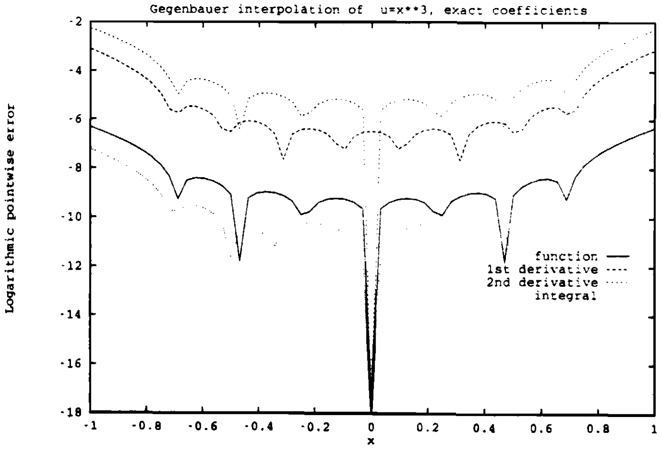

Figure 1: Effect of differentiation and integration on the pointwise error for Gegenbauer interpolation

The accuracy of the FG approximation increases with N. For fixed N, the error increases with the number of derivations: we obtain a larger error for the second derivative than for the first derivative, which itself is less accurate than the interpolation . Integration is more accurate than interpolation. This is in agreement with the theoretical

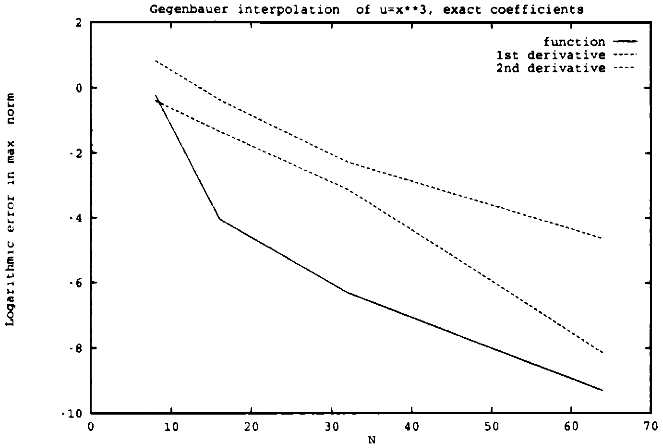

Figure 2: Effect of differentiation and integration on the maximum error for Gegenbauer interpolation

expectations, since each coefficient $\hat{f}(k)$ is multiplied by a factor of k for each differentiation, and divided by the same factor for integration.

It should be noted that practically the same results are obtained when the Fourier coefficients are computed exactly (spectral methods) and in the pseudo-spectral implementation. when the Fourier coefficients are computed using a FFT routine.

4 Improving the convergence

In this section we discuss a way of accelerating the convergence of the FG algorithm. As it is shown in [6], a fairly large number of modes per wave is needed to achieve spectral accuracy in the Gegenbauer expansion. Therefore, the FG approximation of highly oscillating components ($k \gg 1$) of the function requires the computation of a large number of terms in the expansion (9). As the implementation of the FG method for large m, λ is subject to roundoff errors, the FG approximation looses its high accuracy.

The convergence of the FG method can be substantially improved by a preliminary smoothing of the original function, in order to reduce the relative amplitude of the high harmonics in the Fourier spectrum and thus their impact on the accuracy. A simple way to do this is to subtract a linear function from the original function as follows :

$$\begin{aligned} \tilde{f}(x) &= f(x) - \sigma_1(x), \qquad x \in (-l, l) \\ \tilde{f}(\pm l) &= 0, \\ \sigma_1(x) &= \frac{f_l + f_{-l}}{2} + \frac{f_l - f_{-l}}{2}\frac{x}{l} \end{aligned} \tag{36}$$

in order to remove the jump on the boundaries (the function $\tilde{f}$ obtained is a C^0- continuous function). The relations between the smoothed derivatives $\tilde{f}', \tilde{f}''$ and the original ones are easily found . The results are shown in Table 1 for $N = 32, m = 8$ (second column). The approximation obtained from the subtraction procedure is 1 to 2 orders of magnitude more accurate , both for the function and for the first and second derivatives.

Now we can use the computed values of the first derivative on the boundaries $\tilde{f}'(\pm l)$, to construct a third order subtraction polynomial

$$\tilde{\tilde{f}}(x) = \tilde{f}(x) - \sigma_3(x)$$

$$\tilde{\tilde{f}}(\pm l) = \tilde{\tilde{f}}'(\pm l) = 0$$

$$\sigma_3(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3$$

$$a_0 = \frac{l}{4}\left(\tilde{f}'_{-l} - \tilde{f}'_l\right); a_1 = -\frac{1}{4}\left(\tilde{f}'_{-l} + \tilde{f}'_l\right)$$

$$a_2 = -a_0/l^2; a_3 = -a_1/l^2$$

For the doubly smoothed (C^1- continuous) function $\tilde{\tilde{f}}$ the error is several orders of magnitude less than for the $\tilde{f}$ (third column in Table 1). In the rightmost column results after three smoothing steps are shown. The 5-th order subtraction polynomial is constructed using the second derivative $\tilde{\tilde{f}}''(\pm l)$ on the boundaries.

error	without subtrac.	subtract. linear	subtract. cubic	subtract. quintic
$\varepsilon(f)$	4.5 (-6)	3.4 (-7)	1.6 (-9)	7.0 (-12)
$\varepsilon(f')$	4.1 (-5)	3.5 (-6)	1.8 (-7)	3.2 (-10)
$\varepsilon(f'')$	2.5 (-3)	1.8 (-4)	1.4 (-6)	4.4 (-9)

Table 1: The effect of preliminary smoothing of the original function $f(x) = x^7 + x^4$ on the convergence of the FG method.

The procedure described above allows us to improve the convergence of the FG approximation by successive smoothing of the original function. To implement this smoothing procedure we employ derivatives on the boundaries, computed by the same method on a previous iterative step (self-acceleration). For polynomial functions of the form $f(x) = \sum_{k=1}^{m} a_m x^m$, $m \leq 14$. the error remains practically the same as in Table 1. A substantial reduction of the numerical error (by 2-3 orders of magnitude) due to preliminary smoothing is observed for the function $f = \sinh\, \alpha x / \sinh\, \alpha$ at $\alpha \leq 5$ which exhibits a steep profile near the boundaries.

5 Helmholtz equation

We now consider the solution of differential equations by the FG method. We illustrate our approach on a second-order equation:

$$\begin{aligned} u_{xx} - \mu^2 u &= -\mu^2 f(x) \qquad -1 \le x \le 1 \\ u(-1) &= B_1 \\ u(1) &= B_2 \end{aligned} \tag{37}$$

where $f(x)$ is a continuous non-periodic function in the domain $[-1,1]$. We assume that the solution is not dependent on μ. This assumption is accurate for the equations arising from the implicit time discretization of a time-dependent CFD problem (in this case the parameter μ is related to the time step τ as $\mu \propto 1/\sqrt{\tau}$).

The numerical solution process consists of two steps. In the first step we apply the Fourier transform to the Eq.(37) and integrate in the Fourier space, to obtain the coefficients $\hat{u}(k)$:

$$\hat{u}(k) = \frac{\mu^2 \hat{f}(k)}{(\pi k)^2 + \mu^2} \qquad k = -N, \ldots, N \tag{38}$$

Replacing the Fourier coefficients $\hat{f}(k)$ in the FG algorithm (7)-(9) by the coefficients (38) we obtain a particular solution $u_p(x)$ in the physical space .

$$u^{\lambda}_{M.N}(x) = \sum_{l=0}^{M} \hat{v}^{\lambda}_{N}(l) C^{\lambda}_{l}(x) \tag{39}$$

where the coefficients $\hat{v}^{\lambda}_{N}(l)$ are the FG coefficients for $u_p(x)$. We define as well $\hat{u}^{\lambda}(l)$, the Gegenbauer coefficients for $u_p(x)$.

It can be easily shown that the truncation error $u_p(x)$ satisfies the following estimate:

$$TE(u_p, \lambda, m, N) \le \tilde{A}\Phi(m,\lambda)(\frac{2}{\pi N})^{\lambda+1} \tag{40}$$

where $\Phi(M,\lambda) = \frac{(M+\lambda)\Gamma(M+2\lambda)\Gamma(\lambda)}{(M-1)!\Gamma(2\lambda)}$ and $\tilde{A}$ is a constant. It can be made exponentially small for large N and by choosing the parameters M, λ accordingly. Details are given in Appendix A.

The particular solution thus constructed tends to 0 near the boundaries $x = \pm 1$ in accordance with an asymptotic behavior of the Fourier coefficients $\hat{u}(k) \sim \hat{f}(k)/k^2 \sim 1/k^3$ at $k \gg 1$, which is typical of C^1- continuous functions (Gottlieb and Orszag, [7]). Therefore it does not necessarily satisfy the boundary conditions (37). For example, the profile $u_p(x)$ is shown in Fig.1 in the case $f(x) = x$, $\lambda = 5$ (solid line); the dashed line corresponds to the exact solution $u_s(x) = x$.

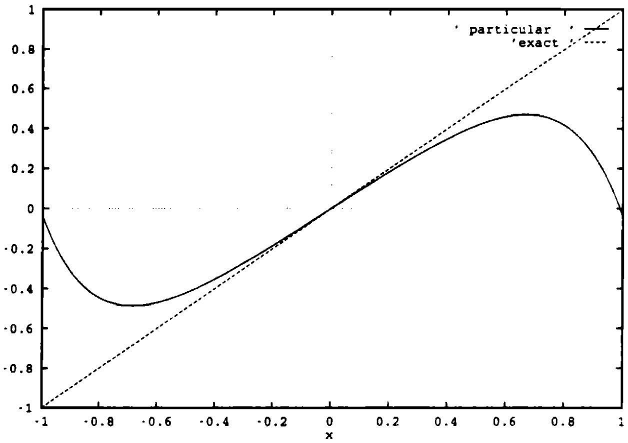

Figure 3: The particular solution $u_p(x)$ at $\lambda = 5$ (solid line) and the exact solution $u_s(x)$ (dashed line)

The purpose of the second step is to correct the particular solution obtained so that it satisfies the given boundary conditions. This can be done by adding two linearly independent homogeneous solutions as follows :

$$u(x) = u_p(x) + D_1 e^{-\mu x} + D_2 e^{\mu x} \tag{41}$$

D_1 and D_2 being uniquely determined by the boundary conditions $B1$ and $B2$.

Equation (37) was solved for the case $u(x) = x^3$. with $N = 64, m = \lambda = 16$. Two cases were implemented : the spectral case (Fourier coefficients are computed exactly) and the pseudospectral case, where they are computed by a FFT procedure. In both cases a linear combination of the exact homogeneous functions $e^{\pm \mu x}$ was added to the particular solution, in order to enforce boundary conditions. The logarithm of the error for $\mu = 1$, is shown in Figs. 4 and 5, for the spectral case. The FG series converges pointwise with exponential accuracy. Results for several values of μ are summarized in Table 2. The logarithm of the maximum error norm is shown for the spectral and pseudospectral Gegenbauer procedure , and compared to the spectral and pseudospectral Fourier expansion.

We can see that for small μ's the Gegenbauer expansion recovers the accuracy lost in the Fourier expansion, both in the spectral and pseudospectral case. However, for $20 \le \mu \le 60$ the spectral accuracy deteriorates, to the extent that in this parameter interval the Fourier expansion gives better results than the FG expansion . The reason for this

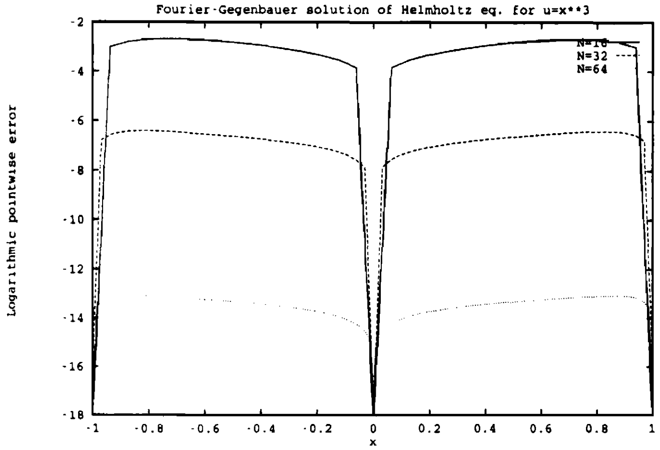

Figure 4: Pointwise error in F-G solution of Helmholtz equation, spectral case

μ	Spectral		Pseudospectral	
	Gegenbauer	Fourier	Gegenbauer	Fourier
1	-13.066	-4.840	-12.925	-3.969
5	-5.930	-4.260	-5.930	-3.413
10	-2.889	-3.567	-2.888	-2.750
20	-1.386	-2.946	-1.383	-2.191
40	-1.874	-2.347	-1.861	-1.708
60	-3.057	-2.006	-3.026	-1.478
80	-4.228	-1.773	-4.176	-1.350

Table 2: *Log* $\| u - u_{ex} \|_\infty$ for $u'' - \mu^2 u = f, u = x^3$

behavior is the presence of exponential components $e^{-\mu x}$ and $e^{\mu x}$ in the particular solution $u_p(x)$ obtained from the procedure . For large μ's the profile $u_p(x)$ coincides with the line $u(x) = x$ inside the interval, except for two thin regions near the boundary, where it abruptly decays to zero.

The operator $\mathcal{G}$ interpolates well the smooth part of the particular solution, but cannot obtain a high accuracy in the interpolation of the steep exponential functions . This is shown in Fig. 6 .

The previous observation gives us the means to *compensate exactly* for the numerical error which arises due to the boundary layers . Instead of using the exact functions $e^{\pm\mu x}$ in the second step of the algorithm, we shall define as new homogeneous solutions the FG expansion of these functions, as follows :

$$u_{h1} = \mathcal{G}^{-1}(e^{\mu x}) \qquad (42)$$

$$u_{h2} = \mathcal{G}^{-1}(e^{-\mu x}) \qquad (43)$$

and thus cancel the approximation error in the intermediate step of the computation of $u_p(x)$.

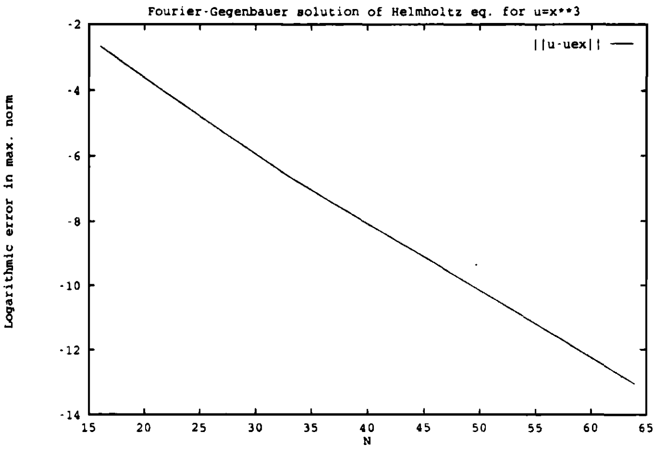

Figure 5: Maximum error in F-G solution of Helmholtz equation, spectral case

As shown in Table 3, we obtain then a good accuracy for every μ in the interval [1, 80] .

μ	Spectral	Pseudospectral
	Gegenbauer	Gegenbauer
1	-9.421	10.078
5	-9.608	-10.212
10	-9.854	-10.480
20	-10.342	-10.970
40	-10.913	-11.498
60	-11.184	-11.835
80	-11.362	-11.932

Table 3: *Log* $\| u - u_{ex} \|_\infty$ for $u'' - \mu^2 u = f, u = x^3$, for approximated homogeneous solutions

The approximation error for the Fourier-Gegenbauer solution of the Helmholtz equation stems from the large regularization error $RE(x, u_p, M, \lambda, N)$ in the homogeneous components of the solution for large μ's. This error is shown in Fig. 7 as a function of $2N$, the number of terms in the Fourier partial sum, and for several values of μ. We have chosen the same values for the parameters as in [4], namely, $M = \lambda = \frac{N}{2}$. When the ratio $r = \frac{M}{\mu}$ reaches some minimum value (meaning that the minimum number of terms per wave is satisfied), we obtain an exponential convergence of the solution, as expected.

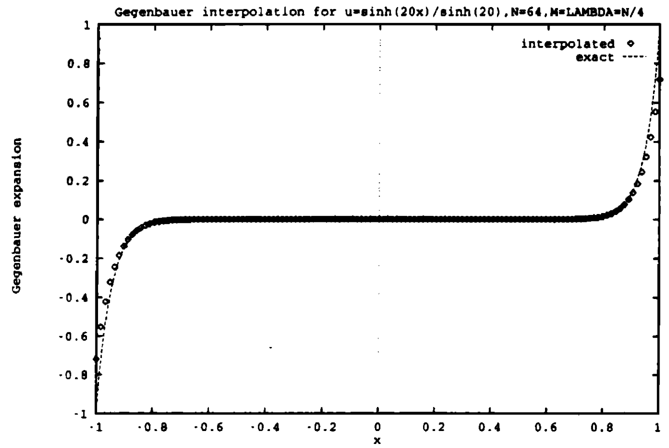

Figure 6: Gegenbauer interpolation of $\frac{\sinh(\mu x)}{\sinh(\mu)}$ for $\mu = 20$

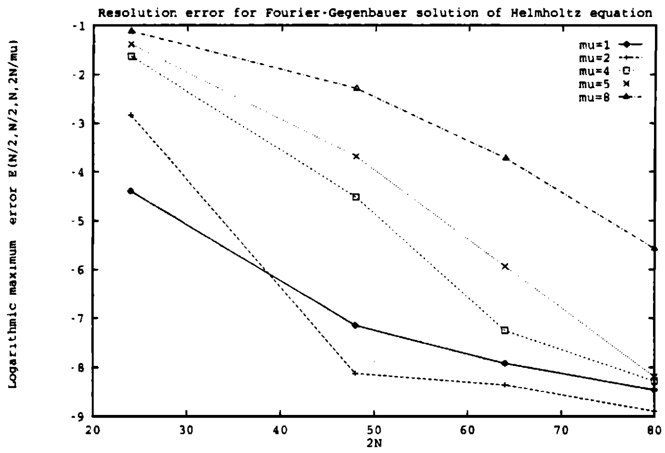

Figure 7: Resolution error $E(\frac{N}{2}, \frac{N}{2}, N, \frac{2N}{\mu})$ in maximum norm for different values of μ.

5.1 Recovering the accuracy

In order to overcome the inaccuracy of the solution of Eq. (37) for large μ, for a L_2 non-periodic function $f(x)$, it is useful to note that the partial sum $u_N(x) = \sum_{k=-N}^{N} \hat{u}_k e^{ik\pi x}$ converges to a periodic function that we shall designate by $u_g(x)$. We propose here a method to recover the accuracy for large μ's, first for an antisymmetric function and then for a symmetric function. As every function $f(x)$ can be written as the sum of a symmetric function and of an antisymmetric function , the following procedure is suitable for every $f(x) \in L_2$ and non-periodic.

Antisymmetric case

We consider the antisymmetric case where :

$$\begin{aligned} u_{xx} - \mu^2 u &= f(x) \quad -1 \le x \le 1 \\ u(-1) &= -u(1) \end{aligned} \tag{44}$$

where $f(x)$ is a continuous antisymmetric function in the domain $(-1, 1)$. As explained before the particular solution obtained by the spectral procedure would have (for large μ) a smooth component and a non-smooth exponential component. If $u_g(x)$ designates such a particular solution, it must be of the form:

$$u_g(x) = u_s(x) - u_s(1)\frac{\sinh(\mu x)}{\sinh(\mu)} \tag{45}$$

since $u_g(1) = 0$. It can be assumed that the homogeneous solution will be in this case a combination of antisymmetric functions, hence the $\sinh(\mu x)$ instead of $e^{\pm\mu x}$. However, the solution obtained from the Gegenbauer procedure is $\mathcal{G}^{-1}(u_g(1))$ and not $(u_g(1))$. As $\mathcal{G}^{-1}(u_g(1)) \neq 0$ we do not have an equality but:

$$\mathcal{G}^{-1}(u_g(1)) \approx \mathcal{G}^{-1}(u_s(1))[1 - u_{h1}(1)] \tag{46}$$

where $u_{h1}(x) = \mathcal{G}^{-1}(\frac{\sinh(\mu x)}{\sinh(\mu)})$.

Therefore we obtain the following approximation for $u_s(1)$:

$$u_s(1) \approx \frac{\mathcal{G}^{-1}(u_g(1))}{[1 - u_{h1}(1)]} \tag{47}$$

so that we can compute an approximation to $u_s(x)$ by Eq. (45).

We expect this approximation to be accurate for large μ as it cancels the inaccuracy present in the Gegenbauer representation of $\frac{\sinh(\mu x)}{\sinh \mu}$. For small μ's ($\mu < 10$) this procedure becomes inaccurate.

Symmetric case

The symmetric case is very similar ; $f(x)$ is a continuous, symmetric function in the domain $(-1, 1)$, and we expect that :

$$u_g(x) = u_s(x) - u_s(1)\frac{\mu\cosh(\mu x)}{\sinh(\mu)} \tag{48}$$

Designating by $v(x)$ the first derivative of $u(x)$, which is antisymmetric, we obtain :

$$v_g(x) = v_s(x) - v_s(1)\frac{\sinh(\mu x)}{\sinh(\mu)} \tag{49}$$

and an estimate to $v_s(1)$ can be found as follows:

$$v_s(1) \approx \frac{\mathcal{G}^{-1}(v_g(1))}{[1 - \mu u_{h2}(1)]} \tag{50}$$

where $u_{h2}(x) = \mathcal{G}^{-1}(\frac{\cosh(\mu x)}{\mu \sinh(\mu)})$. $u_s(x)$ is obtained by integration of $v_s(x)$.

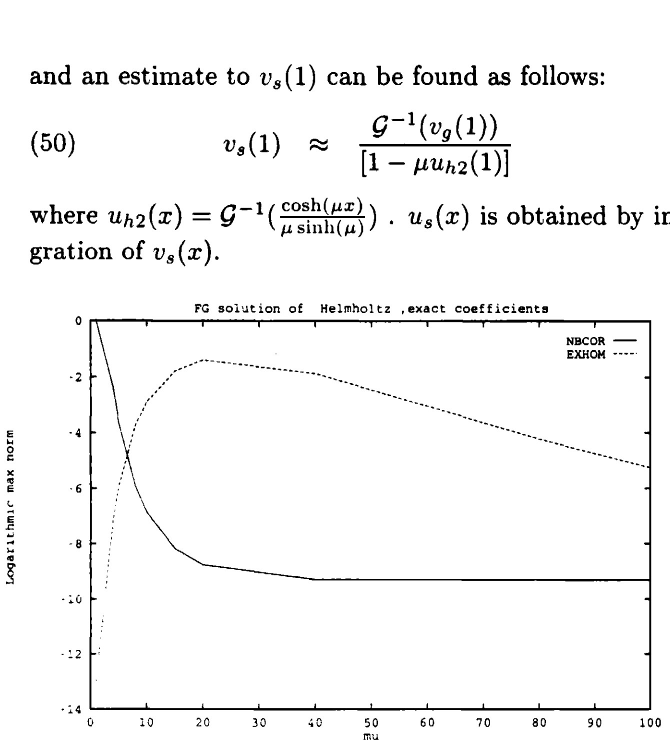

Figure 8: Comparison of the accuracy of two different solutions. NBCOR : correction procedure. EXHOM : exact homogeneous functions and boundary conditions

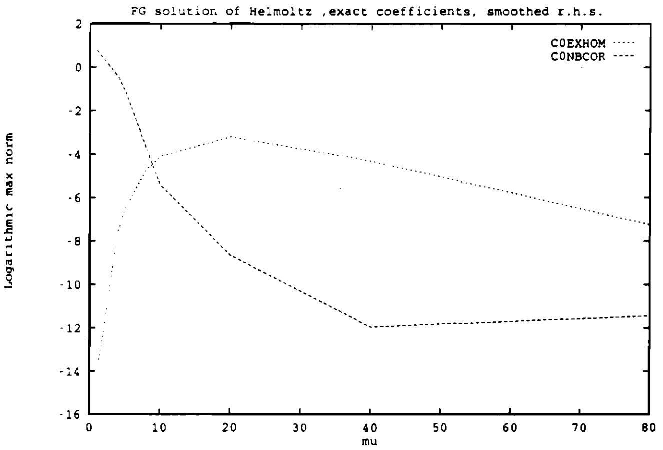

Figure 9: Comparison of the accuracy of two different solutions . C0NBCOR : subtraction +correction procedure. C0EXHOM : subtraction + exact homogeneous functions and boundary conditions

Results

The correction procedure described above was applied to the same test problem as in section 4: $u(x) = x^3, N = 64, M = \lambda = 16$. The comparison between the two methods (the solution procedure with exact boundary conditions and exact homogeneous solution on one hand, and the correction procedure on the other hand) is shown in Fig. 8 . For the latter procedure, the accuracy for small μ's is poor, but improves quickly and an accuracy of 10^{-10} is achieved for $\mu \geq 20$. For the former procedure the situation reverses itself : accurate results are obtained for small μ's, and accuracy degradates when μ increases.

The same comparison was repeated for the same problem with smoothed right hand side (C^0-continuity ensured by subtracting a first-order polynomial from the right hand side). As before the first solution strategy yields spectral accuracy for $\mu \leq 5$, then the accuracy deteriorates while the accuracy of the correction procedure improves. Due to the higher smoothness of the periodic extension of the $f(x)$, the correction procedure achieves an accuracy of 10^{-12} in the maximum norm. However, qualitatively the results of the two tests are similar (Fig. 9).

It is therefore possible to combine the two methods, choosing the correction procedure or the FG method with prescribed boundary conditions, according to the value of μ. Although the accuracy at the intersection point, in the examples shown here, is only of 10^{-5}, a better accuracy can be obtained by successive subtractions of polynomials. For example, If one has to work in the region $5 \leq \mu \leq 10$, it is advisable to work with the cubic subtraction procedure, in order to ensure a good accuracy.

Results for the solution of the symmetric case $u(x) = x^4$ are comparable to the C^0-continuity case. When solving a problem with an arbitrary right hand side ($f(x) = x^3 + x^4$), we obtain the worst-case accuracy, e.g. the same accuracy

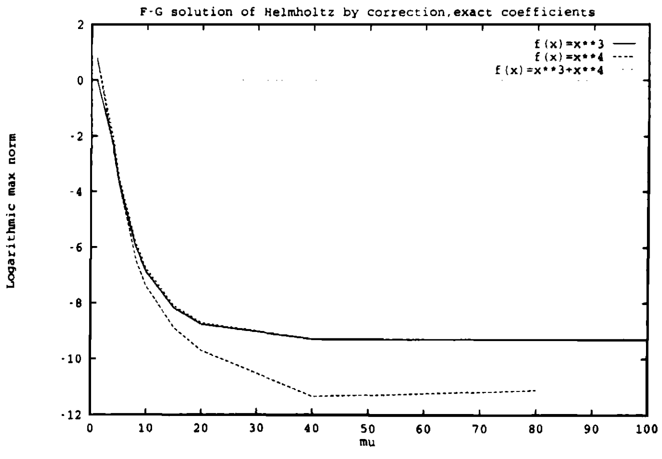

Figure 10: Accuracy of the correction method for antisymmetric, symmetric and arbitrary r.h.s

as for $u(x) = x^3$ (Fig. 10).

Finally, the same tests were performed for the pseudospectral method. The exact Fourier coefficients of the Galerkin formulation were replaced by Fourier coefficients obtained from a standard FFT procedure . The procedure was found to be very sensitive to the accuracy to which these coefficients are computed. Very poor results were obtained for the correction method, and the computation of the coefficients by a high-order Romberg procedure was needed in order to achieve a high accuracy. However, the Romberg procedure adds few calculations to the process, and the FFT method plus the extra computation for the Romberg method still are efficient enough for our purpose.

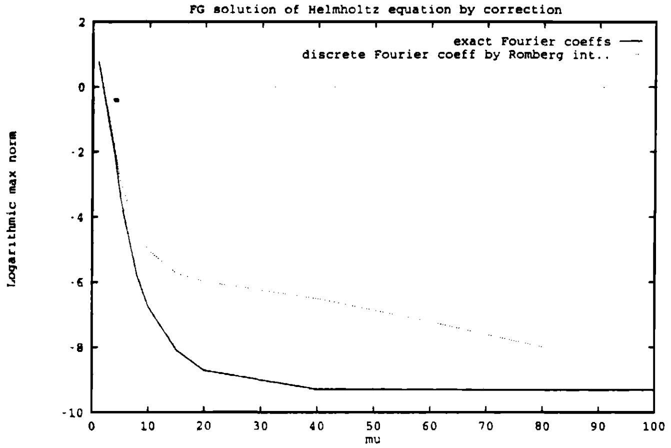

Figure 11: Comparison of the accuracy of the correction method, for exact and discrete Fourier coefficients

Results are shown in Fig. 11.

6 Conclusion

The Fourier-Gegenbauer method was adapted to the solution of Helmholtz like equations in non-periodic domains. This method can be helpful in the multidomain solution of CFD problems, since a good accuracy can be recovered (after a suitable adaptation to improve the approximation of the homogeneous components). No overlapping is needed between the subdomains, thus saving computation time and storage space. However, for oscillatory functions, the resolution requirements of the Gegenbauer expansion are more stringent than for Chebyshev or Fourier expansions, and therefore more collocation points per wave are needed when resolving steep gradients. The method becomes then less efficient. Future directions of research in this topic should be based on the combination of Fourier-Gegenbauer method with other spectral methods, as Chebyshev or Fourier methods.

Convergence estimates for the solution of Helmholtz equation, derivatives and integrals were investigated for the first time. The numerical results seem to match the expectations. It becomes therefore possible to use the Fourier coefficients of non-periodic functions to reconstruct its derivatives with a spectral accuracy. This was not possible in the classical Fourier spectral methods.

A Truncation error for the particular solution

We are interested in evaluating the truncation error in the Gegenbauer expansion for a particular solution of Eq. (37).

Lemma A.1 *Given the equation $u'' - \mu^2 u = f(x)$, if $f(x)$ is a L^2 function on $[-1,1]$, and $u_p(x)$ is a solution of the equation such that it has a continuous periodic extension, there exists a constant $\tilde{A}$ independent of λ, M, N such that the truncation error for $u_p(x)$ satisfies the following estimate:*

$$(51) \quad TE(x, u_p, \lambda, M, N) \;\; \leq \;\; \tilde{A}\Phi(M,\lambda)(\frac{2}{\pi N})^{\lambda+1}$$

where $\Phi(M,\lambda) = \frac{(M+\lambda)\Gamma(M+2\lambda)\Gamma(\lambda)}{(M-1)!\Gamma(2\lambda)}$.

Proof In the definition of the truncation error, we replace f by u to obtain :

$$\begin{aligned} & TE(x, u_p, \lambda, M, N) \leq \\ & \quad M \max_{0\leq l\leq M} \max_{-1\leq x\leq 1} \\ (52) \quad & \quad | (\hat{u}^\lambda(l) - \hat{u}_N^\lambda(l)) C_l^\lambda(x) | \\ & \leq M \max_{0\leq l\leq M} \frac{C_l^\lambda(1)}{h_l^\lambda} \\ & \quad \left| \int_{-1}^{1} (u_p(x) - u_N(x))\, C_l^\lambda(x)(1-x^2)^{\lambda-1/2} dx \right| \end{aligned}$$

$f(x)$ is a L^2 function, which gives :

$$(53) \quad | \hat{f}(k) | \;\; \leq \;\; A \qquad k = N, N+1, N+2 \ldots$$

It follows that $u_p(x) \in L^2[-1,1]$, since it was obtained by successive integrations of $f(x)$. Replacing $\hat{f}(k)$ by $\hat{u}(k)$ we obtain:

$$(54) \qquad | \hat{u}(k) | \;\; \leq \;\; \frac{\mu^2 A}{k^2 + \mu^2}$$

$$\leq \frac{\mu^2 A}{k^2} \tag{55}$$

$$\leq \frac{\mu^2 A}{N^2} \tag{56}$$

Note that the last bound is valid only for $\mu < N$.

As in Paragraph 2, we can combine the equations (25) (26) and (30) and obtain the following bound for the truncation error of u_p:

$$\begin{aligned} TE(x, u_p, \lambda, M, N) &\leq \frac{\mu^2 A}{N^2}\Phi(M,\lambda) \sum_{|k|>N} (\frac{2}{\pi \mid k \mid})^{\lambda} \\ &\leq \tilde{A}\Phi(M,\lambda)(\frac{2}{\pi N})^{\lambda+1} \qquad (57) \end{aligned}$$

□

where $\Phi(M,\lambda) = \frac{(M+\lambda)\Gamma(M+2\lambda)\Gamma(\lambda)}{(M-1)!\Gamma(2\lambda)}$ which can be made exponentially small for large N and by choosing the parameters M, λ accordingly.

References

[1] M. Abramowitz and I.A. Stegun. *Handbook of Mathematical Functions.* Dover, New-York, 1972.

[2] A. Averbuch, M.Israeli, and L.Vozovoi. *A Spectral Multi-Domain Technique with Local Fourier Basis.* J. of Sci. Comput., Vol. 12, No.1-3,pp. 193-212, 1993.

[3] H. Bateman. *Higher Transcendental Functions,* volume 2. McGraw-Hill, 1953.

[4] D.Gottlieb, C.W.Shu, A.Solomonoff, and H.Vandevon. *Recovering Exponential Accuracy in Maximum Norm from the Fourier Partial Sum of a Non-Periodic Analytic Function Using Gegenbauer Polynomials.* J. Comput. Applied Math., **43**, pp. 81-88, 1992.

[5] D.Gottlieb and E.Tadmor. *The CFL Conditions for Spectral Approximations to Hyperbolic Initial-Boundary Value Problems.* ICASE Report 90-42, 1990.

[6] D. Gottlieb and C.W. Shu. *Resolution Properties of the Fourier Method for Discontinuous Waves.* ICASE Report 92-27, 1992.

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