# A Padé-based algorithm for overcoming the Gibbs phenomenon

Tobin A. Driscoll * Bengt Fornberg **

*Department of Applied Mathematics, University of Colorado, Boulder, CO, 80309*

Truncated Fourier series and trigonometric interpolants converge slowly for functions with jumps in value or derivatives. The standard Fourier–Padé approximation, which is known to improve on the convergence of partial summation in the case of periodic, globally analytic functions, is here extended to functions with jumps. The resulting methods (given either expansion coefficients or function values) exhibit exponential convergence globally for piecewise analytic functions when the jump location(s) are known. Implementation requires just the solution of a linear system, as in standard Padé approximation. The new methods compare favorably in experiments with existing techniques.

**Keywords:** Fourier series, Padé approximation, interpolation, Gibbs phenomenon

**AMS Subject classification:** 65B10,42A20,41A21

## 1. Introduction

Approximation of an analytic, periodic function by Fourier partial sums or trigonometric interpolants is highly effective. Such approximations converge spectrally (that is, exponentially in the amount of data used). On the other hand, the difficulties in using these methods to represent a nonsmooth function are notorious. A discontinuity causes the Gibbs phenomenon, which has two important consequences for the Fourier partial sum of length N:

1. failure to converge at the jump, and
2. pointwise convergence elsewhere at the rate $O(N^{-1})$.

* Corresponding author. Supported by NSF VIGRE grant DMS-9810751.
** Supported in part by NSF DMS-9706919.

More generally, if the function $f$ and its derivatives up to order $p-1$ are continuous, but $f^{(p)}$ is discontinuous (i.e., $f$ has a *jump of order* $p$), the global convergence rate is $O(N^{-p})$. Our goal in this paper is to describe methods that

1. exhibit global spectral convergence for piecewise analytic functions, and
2. improve on straightforward summation even in the periodic, globally analytic case.

Methods that improve convergence for nonsmooth functions require knowledge of the locations of any jumps in value or derivative. This information may be immediately available in some situations, such as the Fourier analysis of smooth but nonperiodic data. Gottlieb and Shu [1] describe one spectral method exploiting this information. They project the Fourier partial sum onto a space spanned by Gegenbauer polynomials. The associated weight functions increasingly emphasize information away from the irregularities as the number of included modes grows. One can also try "subtracting off" the jumps from the Fourier data [2–5]; this additionally requires knowledge of the sizes of the jumps in value and in derivatives. For situations in which no advance knowledge of the singularities is given, both locations and strengths can be estimated by nonlinear optimizations [6,2,5]. An approach based on conjugate series is detailed in [7].

Our methods are based on Padé approximation. The starting point is the *Fourier–Padé* (FP) approximation [8–11] for a truncated Fourier series, which transforms the dependent variable and converts the Fourier coefficients into Padé approximants in the complex plane. By itself, this technique achieves spectral convergence away from jumps and improves the convergence for analytic functions. It does not, however, eliminate the Gibbs overshoot—although it does reduce it to 2.5% [11].

After considering the source of the difficulty created by jumps, we introduce a *singular Fourier–Padé* (SFP) approximation that achieves spectrally accurate approximation throughout the interval. This method augments the Padé process to allow more efficient representation of the generic form of singularities. To obtain global convergence, one need know only the locations of jumps.

A closely related problem occurs when equispaced function values, rather than exact Fourier coefficients, are given. In this case the discrete

Fourier coefficients computed by the FFT are contaminated by aliasing error, which can mask the singularity information that the exact coefficients hold. Instead of working with these coefficients, we adapt the SFP technique to rational interpolation.

All of the methods we describe reduce to linear matrix problems that are easily and quickly solved using standard algorithms. In this work we aim only to describe the computational procedures, implementation, and experimental performance, and do not present any theoretical error analysis.

## 2. Notation and test examples

We are concerned with approximation of a function $f$, defined over $[-\pi,\pi)$, assumed to be piecewise analytic with at most a finite number of jump locations. We do not assume that $f$ is $2\pi$-periodic; nevertheless, we will be making use of information from the Fourier series of $f$ as if it were extended periodically. A lack of periodicity is thus equivalent to discontinuities in $f, f', f'', \ldots$ at $x = \pm\pi$. In sections 3–4 we assume that we are given $2N+1$ exact Fourier coefficients $c_{-N}, \ldots, c_N$, where

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{inx}, \qquad c_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)e^{-inx}\,dx. \tag{2.1}$$

In section 7 we shall instead assume that we know $2N$ discrete values of $f$ at evenly spaced points in $[-\pi,\pi)$. In either case our goal is to accurately reconstruct $f$ throughout the interval.

If $f \in C^{p-1}[-\pi,\pi)$ and $f$ has at least one jump of order $p$, then the Fourier partial sum defined by

$$f_N(x) = \sum_{n=-N}^{N} c_n e^{inx} \tag{2.2}$$

has an error of approximation satisfying

$$\|f - f_N\|_\infty = O(N^{-p}). \tag{2.3}$$

If $f$ is not continuous, the partial sums converge to the average of $f$ from the left and right. We would like to improve this error to be *exponential* in $N$, including the one-sided limits at the jump locations.

We will be using the four test functions listed in Table 2.1 throughout the paper to study a variety of situations. They are pictured in Figure 2.1.

Table 2.1
Test functions for the methods of the paper.

| | |
|---|---|
| $f_a(x) = \exp\big(\sin(3x) + \cos(x)\big)$ | analytic and periodic |
| $f_b(x) = \frac{\lvert x\rvert}{\pi}$ | continuous and periodic, with first-order jumps at $\pm\pi$ and $0$ |
| $f_c(x) = \frac{1}{\pi}\max\{0, x\}$ | first-order jump at $0$ and jumps of order zero and one at $\pm\pi$ |
| $f_d(x) = \exp\big(\sin(2.7x) + \cos(x)\big)$ | analytic but nonperiodic (jumps of all orders at $\pm\pi$) |

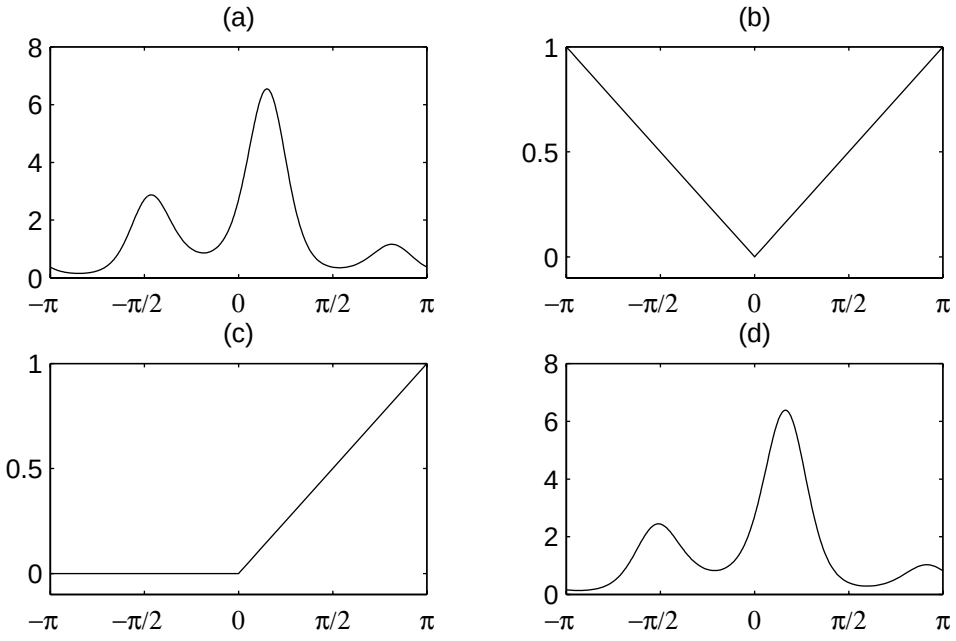


Figure 2.1. Test functions listed in Table 2.1.

The errors of partial sums and other approximations tend to cross zero frequently in the interval $[-\pi, \pi)$, and they vary over many orders of magnitude as N changes. To display such errors over the interval, we use a

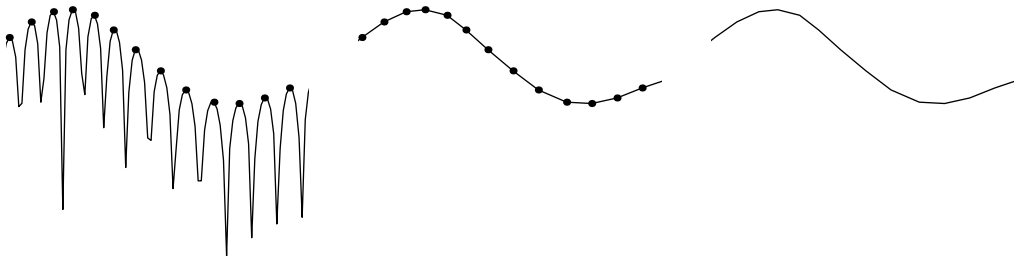

Figure 2.2. Smoothing of error curves. The true error curve on a log scale (left) has many cusps that theoretically extend to $-\infty$. We extract the approximate locations of maxima (dots, left and center) and connect them to form an envelope (center and right). This envelope is what appears in all error graphs.

smoothing technique illustrated in Figure 2.2. The curves are sampled at 257 points, and discrete local maxima are picked out. These maxima are connected to form an envelope which represents the true size of the local error. For all figures which display such error curves, we are really displaying these envelopes.

Figure 2.3 shows the errors of their Fourier partial sums for varying N. To make these and similar pictures, we compute the actual error, which is highly oscillatory, at many points throughout the interval. We then extract the local maxima of the error and interpolate these (in log scale) by a cubic spline. For the analytic, periodic function $f_a$, the convergence is spectral. In the other cases, the convergence is globally very slow and not occurring at all at the jumps.

## 3. Fourier–Padé

We define the new variable $z = e^{ix}$, transforming the interval $[-\pi, \pi)$ to the unit circle. The Fourier expansion (2.1) becomes a Laurent expansion in $z$ which can be split into two parts:

$$f(z) = \sum_{n=-\infty}^{\infty} c_n z^n = \sum_{n=0}^{\infty}{}' c_n z^n + \sum_{n=0}^{\infty}{}' c_{-n} z^{-n} = f^+(z) + f^-(z^{-1}), \qquad (3.1)$$

where the primed sums indicate that the zeroth term should be halved. If $f$ is real, then $c_{-n} = \bar{c}_n$, and one need only work with $f^+$ in practice.

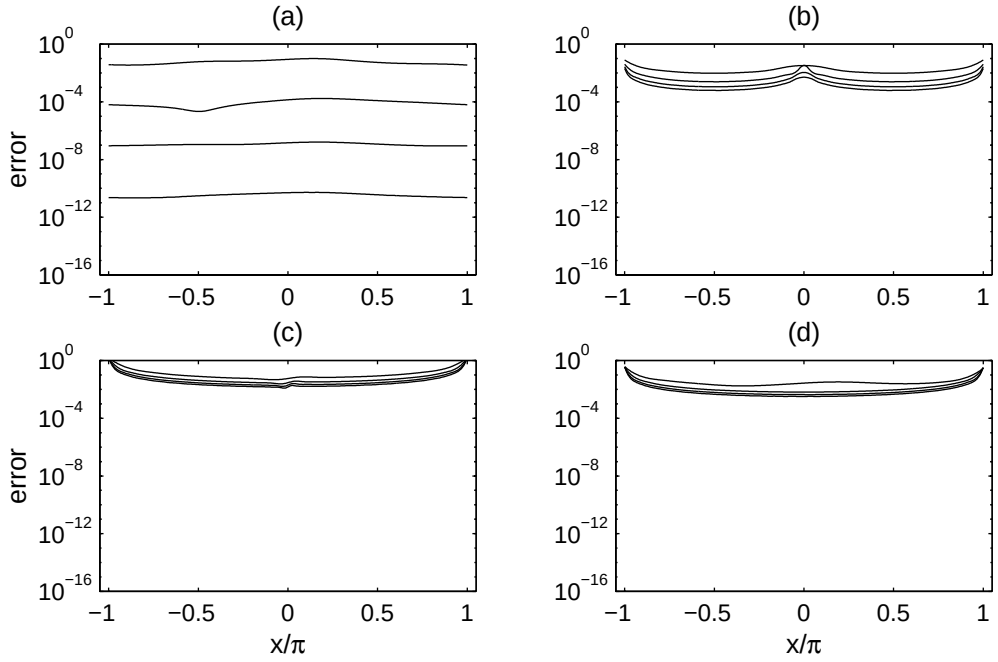


Figure 2.3. Error envelopes of Fourier partial sums for $N = 8, 16, 24, 32$. (The points shown represent local maxima of the errors, and a spline is used to interpolate these values.) Whenever a jump is present, the convergence is globally slow.

Note that $f^+$ and $f^-$ are expressed as Taylor series at zero and infinity, respectively. This naturally suggests replacing truncated Taylor polynomials with Padé approximants, which converge much more rapidly in general. Specifically, we seek four polynomials $p^+(z)$, $q^+(z)$, $p^-(z)$, $q^-(z)$, each of degree $N/2$, such that

$$\begin{aligned} p^+(z) - q^+(z)f^+(z) &= O(z^{N+1}), \qquad z \to 0 \\ p^-(z) - q^-(z)f^-(z) &= O(z^{N+1}), \qquad z \to 0. \end{aligned} \tag{3.2}$$

(Technically, the polynomials may not have degree $N/2$, but we will not be concerned with such Padé degeneracies.) If such polynomials can be found, the *Fourier–Padé* (FP) approximant to $f$ is

$$f(x) \approx \frac{p^+(z)}{q^+(z)} + \frac{p^-(z^{-1})}{q^-(z^{-1})} = \frac{p^+(e^{ix})}{q^+(e^{ix})} + \frac{p^-(e^{-ix})}{q^-(e^{-ix})}. \tag{3.3}$$

Figure 3.1 displays the errors of Fourier–Padé approximants for the test functions of Figure 2.1. We observe dramatic improvement over the Fourier

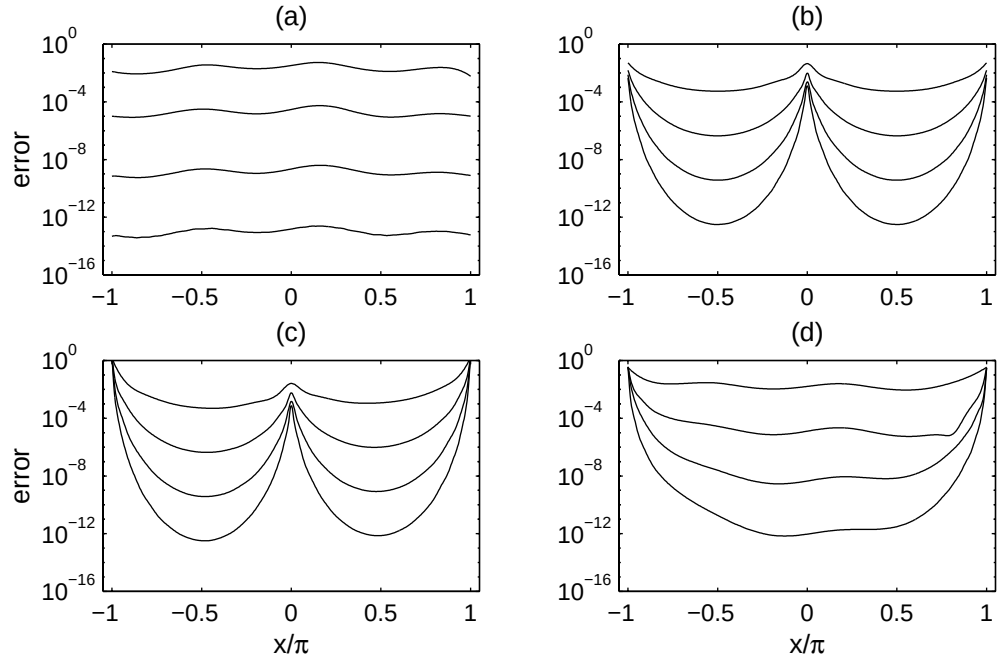


Figure 3.1. Error envelopes of Fourier–Padé approximants for $N = 8, 16, 24, 32$. At jumps, convergence is not occurring. Elsewhere, however, spectral convergence is observed. Compare to Figure 2.3.

partial sums of Figure 2.3. Even for the analytic function $f_a$, the errors have been significantly reduced. In the other cases, spectral convergence has been achieved, except at jumps. A Gibbs phenomenon still occurs at a jump, although the magnitude of the overshoot is about 2.5% instead of the usual 9% [11]. However, unlike the case with Fourier partial sums, the convergence is not degraded globally.

The fundamental limitation of the Fourier–Padé method is that irregularities in $f$ introduce features in $f^+$ and $f^-$ that Padé approximations do not handle well. Specifically, the use of poles to approximate branch cuts is inefficient. One general technique for better approximating branch cuts is the *quadratic Hermite–Padé* method [12,13], in which square-root singularities complement the usual Padé poles. However, as we shall soon see, the branch cuts in $f^+$ and $f^-$ are logarithmic, not square-root, in nature. Our experiments indicate that a Fourier–Hermite–Padé method improves only slightly on standard FP.

## 4. Singular Fourier–Padé

To analyze what is holding back the Fourier–Padé method, we consider the simplest example of a jump, the nonperiodic function $f(x) = x$. (Jumps at arbitrary locations can be analyzed similarly.) In the $z$ variable, the Laurent splitting defined in (3.1) becomes

$$\begin{aligned} f^+(z) &= -i\left(z - \frac{z^2}{2} + \frac{z^3}{3} - \cdots\right) = -i\log(z+1) \\ f^-(z) &= i\left(z - \frac{z^2}{2} + \frac{z^3}{3} - \cdots\right) = i\log(z+1). \end{aligned} \tag{4.1}$$

Note that $f^+(z) + f^-(z^{-1}) = -i\log(z) = x$, as required. By the linearity of the Fourier transform, every zeroth-order jump in a generic $f$ is represented by a logarithmic singularity in $f^+$ and $f^-$ at the conjugate points on the unit circle corresponding to the location of the jump. A multiplicative constant adjusts the size of the jump. This logarithmic singularity on the unit circle is very difficult for the Padé approximants to simulate. More generally, a function with jumps in derivatives of all orders at $x = \pm\pi$ can be seen to have

$$f^\pm(z) = R_0^\pm \log(z+1) + R_1^\pm (z+1)\log(z+1) + R_2^\pm (z+1)^2 \log(z+1) + \cdots + g^\pm(z), \tag{4.2}$$

where $g^\pm(z)$ is analytic near $z = -1$. The coefficients $R_0^\pm, R_1^\pm, R_2^\pm, \ldots$ are respectively associated with jumps in $f, f', f'', \ldots$.

Suppose we know in advance the location of an irregularity at $z = -1$ (i.e., $x = \pm\pi$), although not the sizes or orders of any jumps there. Rewriting (4.2) as

$$f^\pm(z) = \log(z+1) g_1^\pm(z) + g_2^\pm(z) \tag{4.3}$$

suggests that we replace (3.2) by

$$\frac{p^\pm(z)}{q^\pm(z)} + \frac{r^\pm(z)}{q^\pm(z)}\log(z+1) = f^\pm(z) + O(z^{N+1}) \tag{4.4}$$

for polynomials $p^\pm$, $q^\pm$, and $r^\pm$ (assuming $q^\pm(0) \neq 0$). We call the resulting approximation the *singular Fourier–Padé* (SFP) approximant. The choice of a *common* denominator for the analytic and logarithmic parts is algorithmically significant; this issue and the assignment of the polynomials' degrees is discussed in section 5.

In general, if singularities are located at $\zeta_1, \ldots, \zeta_s = e^{ix_s}$ on the unit circle, the SFP method uses the approximations

$$p^{\pm}(z) + r_1^{\pm}(z)\log(1 - z/\zeta_1^{\pm 1}) + \cdots + r_s^{\pm}(z)\log(1 - z/\zeta_s^{\pm 1}) = q^{\pm}(z)f^{\pm}(z) + O(z^{N+1}). \tag{4.5}$$

The logarithms in this formula are understood to use the principal branch; their arguments have been manipulated to ensure that the branch cuts point away from the origin. If $s = 0$, the SFP method is just the standard Fourier–Padé method. Dividing (4.5) through by $q^{\pm}(z)$ produces the needed approximation to $f^{\pm}$.

We defer application of the SFP method to our test functions until section 8, in which SFP is compared to other methods for Fourier series of functions with jumps.

## 5. Algorithmic description

We take the most straightforward approach to computing the polynomial coefficients needed in the SFP method. Rewriting (4.4) (with some superscripts dropped for clarity), we have

$$p(z) + r(z)\log(z + 1) = q(z)f^{\pm}(z) + O(z^{N+1}). \tag{5.1}$$

Both $\log(z + 1)$ and $f^{\pm}(z)$ have Taylor expansions known to order $N$. The resulting equation turns out to be a linear system for the polynomial coefficients. The choice of a common denominator in (4.4) was crucial to the linearity.

Rescaling all the polynomials $p(z)$, $q(z)$, and $r(z)$ by a constant leaves (5.1) unchanged. Designating the degrees of the polynomials by $n_p$, $n_q$, and $n_r$, respectively, we therefore must have $n_p + n_q + n_r = N - 1$ in order to make the coefficients well-determined.

Note that $q(z)$ and $r(z)$ are determined by the terms of order greater than $n_p$ alone. Thus we seek a solution to

$$\Big[C - L\Big]\begin{bmatrix}\mathbf{q}\\ \mathbf{r}\end{bmatrix} = \mathbf{0}. \tag{5.2}$$

Here C is the $(n_q + n_r + 1) \times (n_q + 1)$ Toeplitz matrix

$$\begin{bmatrix} c_{N/2+1} & c_{N/2} & \cdots & c_1 \\ c_{N/2+2} & c_{N/2+1} & \cdots & c_2 \\ \vdots & \vdots & & \vdots \\ c_N & c_{N-1} & \cdots & c_{N/2} \end{bmatrix}, \tag{5.3}$$

and L has size $(n_q+n_r+1)\times(n_r+1)$ and is defined similarly using the Taylor coefficients of $\log(z+1)$. The vectors $\mathbf{q}$ and $\mathbf{r}$ hold the unknown polynomial coefficients in order of increasing degree. Because the matrix in (5.2) has column dimension one greater than its row dimension, at least one nonzero solution exists. Usually this can be made into a square system by choosing, say, $q(0) = 1$, but if one does not want to assume that any particular coefficient is nonzero, one can solve (5.2) by a singular value decomposition. Finally, the coefficients of $p(z)$ are found by multiplication, via

$$\mathbf{p} = \begin{bmatrix} \frac{1}{2}c_0 & 0 & \cdots & 0 \\ c_1 & \frac{1}{2}c_0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ c_{N/2} & c_{N/2-1} & \cdots & \frac{1}{2}c_0 \end{bmatrix} \mathbf{q} - \begin{bmatrix} \ell_0 & 0 & \cdots & 0 \\ \ell_1 & \ell_0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ \ell_{N/2} & \ell_{N/2-1} & \cdots & \ell_0 \end{bmatrix} \mathbf{r}. \tag{5.4}$$

If the original function is real, only $f^+$ needs to be considered in (5.1). If there is more than one jump location in the interval and (4.5) is to be used, the equation (5.2) is modified to have an L matrix and a vector of coefficients for each location, and (5.4) changes similarly.

We have no rigorous formula for choosing the degrees $n_p$, $n_q$, and $n_r^{(1)}, \ldots, n_r^{(s)}$. Because the denominator polynomial $q(z)$ is shared, we allow $n_q$ to be the largest, with the others equal so far as possible. For the case of just one jump location, $n_q$ is roughly 40% of the total available degrees of freedom. Experiments suggest that these choices can affect the observed accuracy—occasionally by as much as an order of magnitude—but on average there is little variation within a broad range of choices.

It is well known that the matrix C in (5.3) becomes severely ill-conditioned as $N \to \infty$ [12], and a similar trend is observed for the matrix of (5.2). This suggests that there may be a great deal of error in the coefficients determined for the polynomials in the methods. We observe that the values of the resulting approximations are still useful—indeed, the "value

problem" is commonly believed to be much better conditioned than the "coefficient problem" in Padé approximation [12].

Concerns over numerical ill-conditioning are also muted by the ill-posedness of the original problem. Any method which takes N Fourier coefficients and attempts to recover "the" function which produced them must necessarily make assumptions about the remaining coefficients. (In forming partial sums, we assume that they are zero, which is in fact a stable regularization of the problem.) In the Fourier–Padé method (and by extension SFP), we extend the coefficients by assuming a rational form in the $z$-plane for $f^{\pm}$. But, to take an extreme example, the Padé approximant $(1+bz)^{-1}$ is determined by its first two coefficients at $z=0$. If $|b| \approx 1$, then a small perturbation in the second coefficient can determine whether the extrapolated Taylor coefficients grow or decay, causing a huge (potentially infinite) change in the result on the unit circle. In other words, we could never hope to get the Padé polynomial coefficients accurate enough anyway. This does not invalidate the practical utility of the method, but one must use accurate data, as in any extrapolation method.

## 6. Classical examples

We have assumed that the locations of singularities in the function being reconstructed are known in advance. In light of the formula (4.2) that describes the form of the half-functions $f^{\pm}$ for a singularity at $z=-1$, one might try differentiating their series to convert logarithmic singularities into poles. (See also [5] on differentiation of the original series.) These poles would then be located by standard Padé approximation. However, if $f'$ is also discontinuous, logarithmic singularities will persist and interfere with attempts to locate the poles accurately. In experiments we have been unable to get more than four or five accurate digits in general for the singularity location by this method. (The resulting approximations are quite accurate anyway, except very near the jumps.)

It is interesting, though, to apply this idea to classical examples of the Gibbs phenomenon. For the function $f(x)=x$, which has a jump in value only at $\pm\pi$, we noted in (4.1) that $f^{+}(z) = -f^{-}(z) = -i\log(z+1)$. The Padé approximant for the derivative in each case is the exact derivative, $\pm i/(z+1)$, provided $N \geq 2$. Hence these approximants locate singularity exactly. Given

this location, from (4.4) we see that consequently the singular Fourier–Padé approximant is also exact provided $N \geq 2$. A similar analysis holds for the function $f(x) = \mathrm{sgn}(x)$, because in this case

$$f^+(z) = -f^-(z) = -\frac{i}{\pi} \log\left(\frac{1+z}{1-z}\right).$$

We conclude that for these classical examples of the Gibbs phenomenon, using no information besides the first $N$ Fourier series coefficients, the SFP method is exact (in exact arithmetic) for $N \geq 4$.

## 7. The interpolation problem

Often one may not have access to the exact Fourier coefficients of a function but rather to the values of the function at equally spaced points. In this section we use the points

$$x_n = -\pi + \frac{2n+1}{2N}\pi, \qquad n = 0, 1, \ldots, 2N-1.$$

(Note that $-\pi$, $0$ and $\pi$ are excluded. We choose this convention because the singularities in our test functions occur exactly at these points, and coincidence with a grid point would lead to ambiguity in the value of the function to assign there.) Corresponding to the grid points are the values $f_n = f(x_n)$.

One approach to this situation is to apply the FFT to $f_0, \ldots, f_{2N-1}$ to obtain approximate values for the Fourier coefficients:

$$\tilde{c}_n = \frac{1}{2N} \sum_{m=0}^{2N-1} f_m e^{-inx_m}, \qquad n = -N, \ldots, N. \tag{7.1}$$

By definition, $\tilde{c}_{-N} = -\bar{\tilde{c}}_N$. The Fourier partial sum (2.2) in this context becomes the trigonometric interpolant $\tilde{f}_N$,

$$\tilde{f}_N(x) = \sum_{n=-N}^{N}{}'' \, \tilde{c}_n e^{inx}, \tag{7.2}$$

where the doubly primed sum indicates that the first and last terms are to be halved. (The contribution from the sawtooth mode is divided between equivalent positive and negative modes to preserve symmetry.)

In the presence of a jump singularity, the convergence properties of the trigonometric interpolant are identical to those for the exact truncated Fourier partial sums. The problem is made more difficult by the fact that $\tilde{c}_n$

is an aliased approximation to $c_n$. For modest values of N, aliasing error can contaminate the precise singularity information that is encoded in the exact coefficients. For this reason, application of SFP to the discrete coefficients is much less successful than in the exact Fourier case.

Therefore, we avoid trying to extract Fourier data and instead work with the function values directly. The relationship between Fourier partial sums and trigonometric interpolants is analogous to the relationship between Padé approximants and rational interpolants (multipoint Padé approximants). Thus the interpolation form of the Fourier–Padé technique, which we call *Fourier-rational interpolation* (FRI), is the approximation $p(z)/q(z)$, where polynomials $p$ and $q$ have degrees $N-1$ and $N$ and satisfy

$$p(z_n) - f_n q(z_n) = 0, \qquad n = 0, \dots, 2N-1. \tag{7.3}$$

Here $z_n = e^{ix_n}$. To generalize the SFP method, we first note that there is no longer a natural splitting of the function into two complementary parts. Accordingly, to get a jump we combine the $\log(1-z/\zeta_1)$ and $\log(1-\zeta_1/z)$ terms to get $-i\log(z/\zeta_1)$. Similarly, a jump in the $m$th derivative can be created by $z^m \log(z/\zeta_1)$. The natural adaptation of (4.5), then, is

$$p(z_n) + r_1(z_n)\log(\zeta_1^{-1}z_n) + \cdots + r_s(z_n)\log(\zeta_s^{-1}z_n) - f_n q(z_n) = 0, \tag{7.4}$$

$$n = 0, \dots, 2N-1. \tag{7.5}$$

Again, one must allocate degrees of freedom amongst the polynomials so that they sum to 2N. Then the singular Fourier-rational interpolant (SFRI) for $f$ is

$$f(x) \approx \frac{p(z) + r_1(z)\log(\zeta_1^{-1}z) + \cdots + r_s(z)\log(\zeta_s^{-1}z)}{q(z)}, \qquad z = e^{ix}.$$

Both (7.3) and (7.4) reduce to linear problems. Let $V_M$ be a Vandermonde matrix of degree $M$; that is,

$$V_{nm} = (z_n)^m, \qquad n = 0, \dots, 2N-1, \quad m = 0, \dots, M.$$

The polynomial coefficients satisfy an equation of the form

$$\left[V_{n_p} - \mathrm{diag}(f_0, \ldots, f_{2N-1}) V_{n_q} L_1 V_{n_1} \cdots L_s V_{n_s}\right] \begin{bmatrix} \mathbf{p} \\ \mathbf{q} \\ \mathbf{r}_1 \\ \vdots \\ \mathbf{r}_s \end{bmatrix} = \mathbf{0}.$$

The matrix has column dimension one larger than row dimension, so a nonzero solution exists.

## 8. Comparisons

In this section we compare the SFP and SFRI methods to other available techniques. The Gegenbauer projection technique of [1] consists of projecting the partial Fourier sum onto a space spanned by Gegenbauer polynomials $C_m^\lambda$. The weight parameter $\lambda$ and the number of included Gegenbauer polynomials each grow linearly with N, the number of included Fourier modes. We have examined two numerical implementations: Direct calculation by quadrature of the inner products of the partial Fourier sum with Gegenbauer polynomials; and a cruder, weighted (with the appropriate Gegenbauer weight) least-squares projection of the partial sum evaluated at many points using a basis of Chebyshev polynomials. Both methods produced essentially identical results in our tests.

We remind the reader that we are actually plotting smoothed envelopes of the error, as described in Figure 2.2.

Eckhoff [2] described a method of singularity removal for functions with jump discontinuities. The idea is to subtract off the singular part of the function, in the form of a combination with unknown coefficients of prototype functions with known jumps in value and derivatives. The coefficients are determined by solving an overdetermined linear least-squares problem derived from the ansatz that the singular part of the function matches the exact Fourier data asymptotically.

Both the Gegenbauer and singularity removal methods have discrete variants. In the Gegenbauer case, equispaced function values are converted to the trigonometric interpolant, which is then used in the Gegenbauer projection. For singularity removal, Eckhoff advocates matching discrete coef-

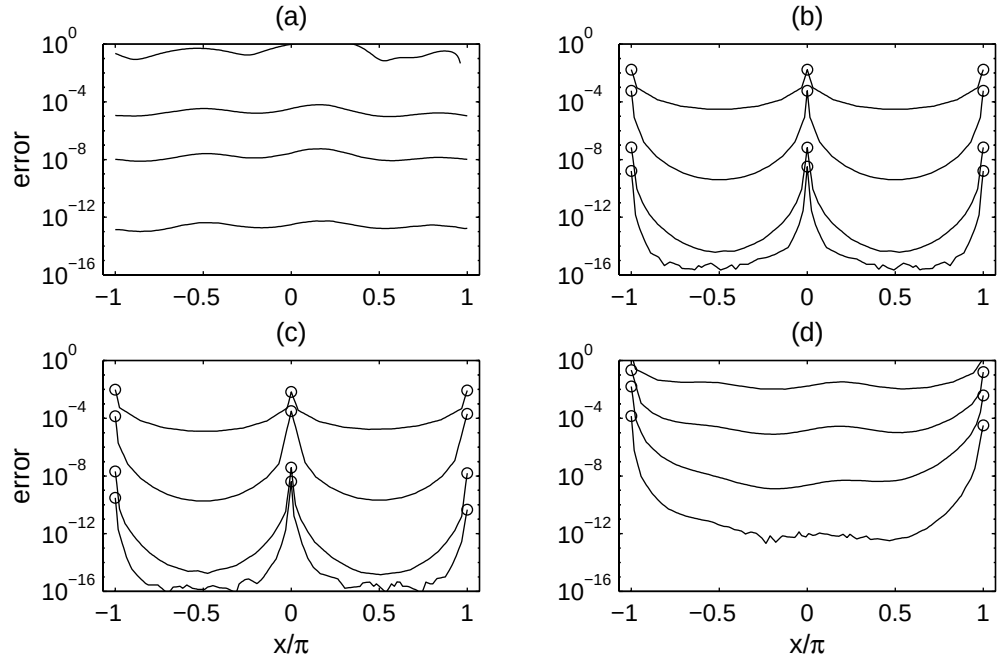


Figure 8.1. Error envelopes for the singular Fourier–Padé method with $N = 8, 16, 24, 32$. Convergence has been restored throughout the interval in each case, including at the jumps.

ficients of the function with discrete coefficients of the singular prototype functions. As both methods essentially involve some least-squares projection, neither results in an interpolant of the given data.

In Figure 8.1 we show the results of the SFP method for our four test functions. Convergence now occurs at the jumps, although accuracy is lessened near them. We comment that the MATLAB code implementing the general SFP method is only about 100 lines, and that all the data for Figure 8.1 was generated in seconds on a PC. To better see the convergence at a jump location, in Figure 8.2 we isolate the error at $x = \pi$ for the smooth nonperiodic function $f_d$. In arbitrarily high precision, the convergence is clearly spectral. However, in double precision a common Padé ill-conditioning problem (see section 5) has a definite effect on the accuracy of the solution, ultimately causing a plateau at around $10^{-8}$.

To compare methods, we return to the analytic, nonperiodic function $f_d(x)$ (see Figure 2.1 and Table 2.1). We let $N = 16, 24, 32$ be the number of modes in the Fourier partial sum, and we choose parameters to approxi-

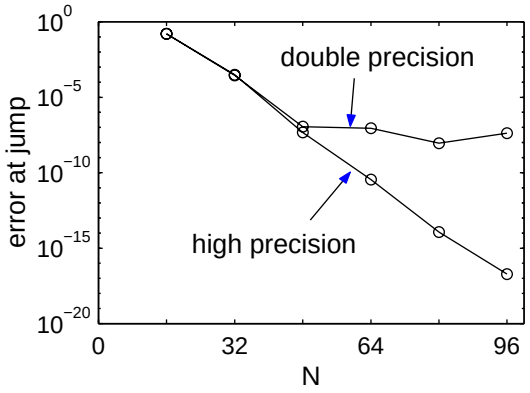


Figure 8.2. Error at the jump singularity using SFP for the function $f_d$. Using arbitrarily high precision (as performed by Mathematica), the convergence is spectral. For a double-precision implementation, the effects of Padé ill-conditioning ultimately limit the accuracy obtained.

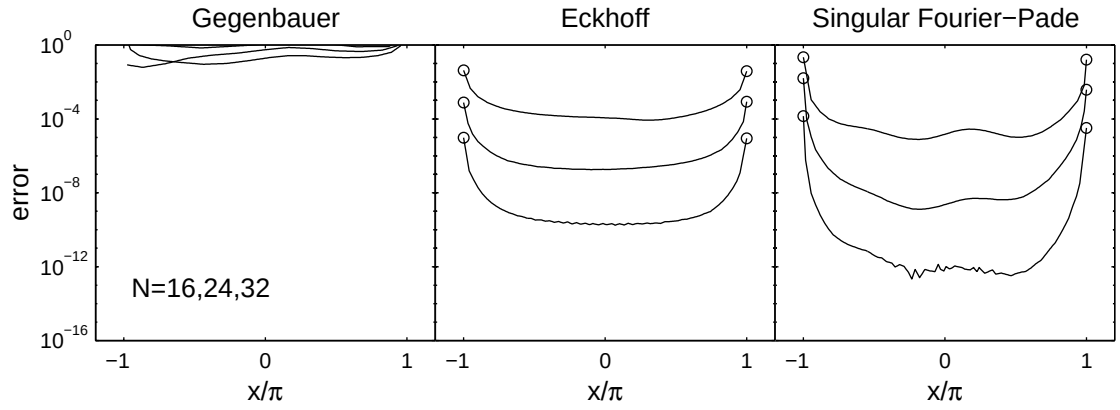


Figure 8.3. Comparison of Gegenbauer projection, Eckhoff's singularity removal, and the SFP methods for an analytic, nonperiodic test function ($f_d$ in Table 2.1).

mately optimize the Gegenbauer and Eckhoff method. The Gegenbauer projection method performs poorly on this example. In fact, the reason has nothing to do with Fourier series; the results are not changed if we project the exact function onto Gegenbauer polynomials of these orders (up to degree 16 with weight parameter 2 in the last case). Ultimately, convergence ought to be spectral, but for these values of N little happens. Eckhoff's singularity removal does much better, and the values at the jump are even more

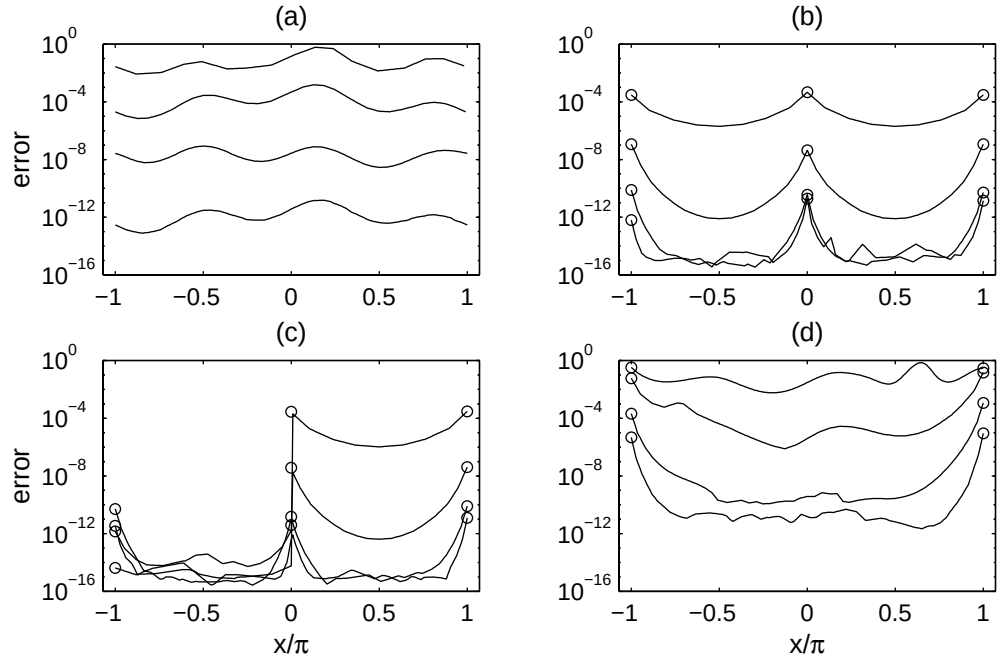


Figure 8.4. Error envelopes in collocated singular Fourier–Padé approximations for $N = 8, 16, 24, 32$. Global spectral convergence has been restored.

accurate than those obtained by SFP. Overall, though, SFP is better by several orders of magnitude. It is possible that a least-squares approach in SFP might overcome ill-conditioning, as it does for Eckhoff's method.

For the interpolation problem, Eckhoff [2] suggests using the discrete coefficients of both the given data and the singular prototype functions in the least-squares problem. We compare methods for the function $f(x) = (1 + 25x^2)^{-1}$ over $[-1, 1]$ using 32 equispaced function values. This function is analytic on the real line but has poles near enough to the interval that equispaced polynomial interpolation is unstable due to the Runge phenomenon [14].

Figure 8.4 displays the results of the SFRI algorithm on the four test functions. It is apparent that global spectral convergence is restored in all cases, including at jump locations. The convergence of SFRI at the jump for $f_d$ is shown in Figure 8.5. As was the case for singular Fourier-Padé, numerical ill-conditioning can cause the convergence to plateau well above machine precision. This can also be seen in the interior of the approximation

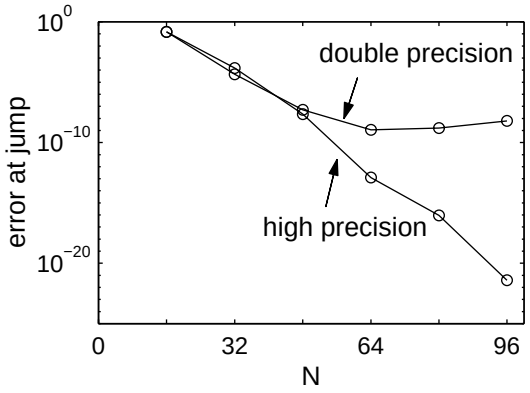


Figure 8.5. Error at the jump singularity for the collocation problem for $f_d$. The effects of Padé ill-conditioning are again clear in double precision.

to $f_d$ with $N = 32$ in Figure 8.4.

In Figure 8.6 we compare the performances of polynomial interpolation, cubic spline interpolation, Eckhoff's discrete method, and SFRI. (While $f$ would be trivially representable by Padé approximation in the original variable $x$, this is no longer the case in the transformed variable $z = e^{i\pi x}$.) SFRI is 6–12 orders of magnitude better than the comparable cubic spline and Eckhoff methods.

We are not certain how to evaluate the practical significance of the (discontinuous) interpolation problem, despite its popularity in the literature. If the jump locations are known or can be found, one supposes that piecewise interpolation by standard methods (e.g., rational interpolation) should perform well.

## 9. Conclusions

The standard Fourier–Padé technique can be used to accelerate significantly the convergence of Fourier series of analytic functions. We modify this method to account for the general form of singularity introduced by a discontinuity. The resulting approximations are spectrally accurate in a global sense. When the jump location(s) are known in advance, the method works equally well for truncated Fourier or equispaced value data, and it compares favorably to existing methods.

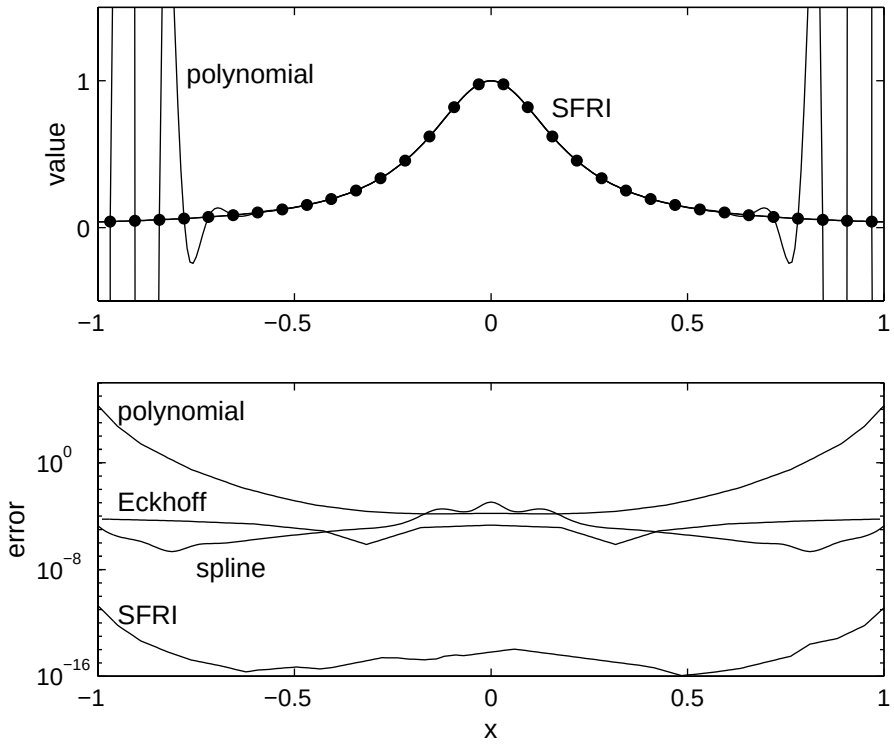


Figure 8.6. Comparison of interpolation methods for $f(x) = (1 + 25x^2)^{-1}$ and 32 points. Equispaced polynomial interpolation is unstable. A cubic spline and Eckhoff's discrete method achieve comparable accuracy. SFRI is many orders of magnitude better throughout, apparently encountering double-precision roundoff.

## References

[1] D. Gottlieb and C.-W. Shu. On the Gibbs phenomenon and its resolution. *SIAM Review*, 39:644–668, 1997.

[2] Knut S. Eckhoff. On a high order numerical method for functions with singularities. *Math. Comp.*, 67:1063–1087, 1998.

[3] James Geer and Nana Saheb Banerjee. Exponentially accurate approximations to piecewise smooth periodic functions. *J. Sci. Computing*, 12:253–287, 1997.

[4] George Kvernadze. *Locating discontinuities of a bounded function by spectral methods*. PhD thesis, University of New Mexico, August 1998.

[5] G. Kvernadze, T. Hagstrom, and H. Shapiro. Locating discontinuities of a bounded function by the partial sums of its Fourier series. *submitted*, 1998.

[6] Nana S. Banerjee and James F. Geer. Exponential approximations using Fourier series partial sums. Technical Report 97-56, ICASE, 1997.

[7] A. Gelb and E. Tadmor. Detection of edges in spectral data. *Appl. Comput. Harmon. Anal.*, 7:101–135, 1999.

[8] J. S. R. Chisholm and A. K. Common. Generalisations of Padé approximation for Chebyshev and Fourier series. In P. L. Butzer and F. Fehér, editors, *E. B. Christoffel: The Influence of His Work on Mathematics and the Physical Sciences*, pages 212–231. Birkhäuser, 1981.

[9] James F. Geer. Rational trigonometric approximations using Fourier series partial sums. *J. Sci. Computing*, 10:325–356, 1995.

[10] W. B. Gragg. Laurent, Fourier, and Chebyshev–Padé tables. In E. B. Saff and R. S. Varga, editors, *Padé and Rational Approximation, Theory and Applications*. Academic Press, 1977.

[11] R. D. Small and R. J. Charron. Continuous and discrete nonlinear approximation based on Fourier series. *IMA J. Num. Analy.*, 8:281–293, 1988.

[12] George A. Baker and Peter Graves-Morris. *Padé Approximants*, volume 59 of *Encyclopedia of Mathematics and Its Applications*. Cambridge University Press, 2nd edition, 1996.

[13] R. G. Brookes. The local behavior of the quadratic Hermite–Padé approximation. In W. L. Hogarth and B. J. Noye, editors, *Computational Techniques and Applications*, pages 569–575, New York, 1990. Hemisphere Publ. Corp.

[14] B. Fornberg. *A Practical Guide to Pseudospectral Methods*. Cambridge Universtiy Press, Cambridge, 1996.