processing and data compression. In this paper we prove new theoretical results and develop numerical algorithms for constructing such approximations. Since our numerical results are far better than our current proofs indicate, we also point out unresolved issues in this emerging theory.

Since our formulation is somewhat unusual, we first provide two examples. Let us consider the identity

$$\frac{1}{x} = \int_0^\infty e^{-xt}\, dt \tag{1}$$

for $x > 0$. This integral representation readily leads to an approximation of the function $\frac{1}{x}$ as a sum of exponentials. In fact, for any fixed $\varepsilon > 0$ there exist positive weights and nodes (exponents) of the generalized Gaussian quadrature such that

$$\left| \frac{1}{x} - \sum_{m=1}^{M} w_m e^{-t_m x} \right| \leqslant \frac{\varepsilon}{x} \tag{2}$$

for all x in a finite interval, $0 < \delta \leqslant x \leqslant 1$, and where the number of terms is $M = \mathcal{O}(\log \delta^{-1})$. Theoretically the existence of such approximations follows from [19–22]. This particular example has been examined in [27] with the goal of using (2) for constructing fast algorithms. Specific exponents and weights are provided there for several intervals and values of ε, so that (2) can be verified explicitly. The approximation (2) has important applications to fast algorithms that we will consider below.

The second example is the Bessel function $J_0(bx)$, where $b > 0$ is a parameter and $x \in [0, 1]$. Using the approach developed in this paper, we obtain for all x on $[0, 1]$,

$$\left| J_0(bx) - \sum_{m=1}^{M} w_m e^{\tau_m x} \right| \leqslant \varepsilon \tag{3}$$

where w_m and τ_m are now complex numbers and the number of terms, M, is remarkably small and increases with b and ε as $M = \mathcal{O}(\log b) + \mathcal{O}(\log \varepsilon^{-1})$. In the sum (3) we will refer to the coefficients w_m as weights and to the values e^{τ_m} as nodes; such terminology is natural since, as it turns out, e^{τ_m} are zeros of a certain polynomial as is usually the case for quadratures. We illustrate (3) in Figs. 1 and 2 by showing the error of the approximation and the location of the weights w_m and (normalized) nodes $e^{\tau_m/b}$ corresponding to $b = 100\pi$ and $\varepsilon \simeq 10^{-11}$. The number of nodes is $M = 28$ and they accumulate at e^{i} and e^{-i} as expected from the form of the approximation in (3) and the asymptotics of J_0 for large argument,

$$J_0(bx) \sim \frac{(1-i)e^{ibx} + (1+i)e^{-ibx}}{2\sqrt{\pi b x}}.$$

Also, since the real part of the exponents is always negative, $\mathrm{Re}(\tau_m) < 0$, all nodes belong to the unit disk. The approximation (3) with these 28 terms is remarkable in that there is no obvious integral, as in (1), to represent the function and, thus, by some quadrature, obtain so few terms for a given accuracy and parameter b. Clearly, there are many possible integrals in the complex plane to represent the Bessel function but, unfortunately, there is no obvious criteria to choose a particular integral or contour. Finding such a contour may be attempted via the steepest descent method, in this case starting from, e.g.,

$$J_0(bx) = \frac{1}{\pi} \int_{-1}^{1} \frac{e^{ibxt}}{\sqrt{1-t^2}}\, dt \tag{4}$$

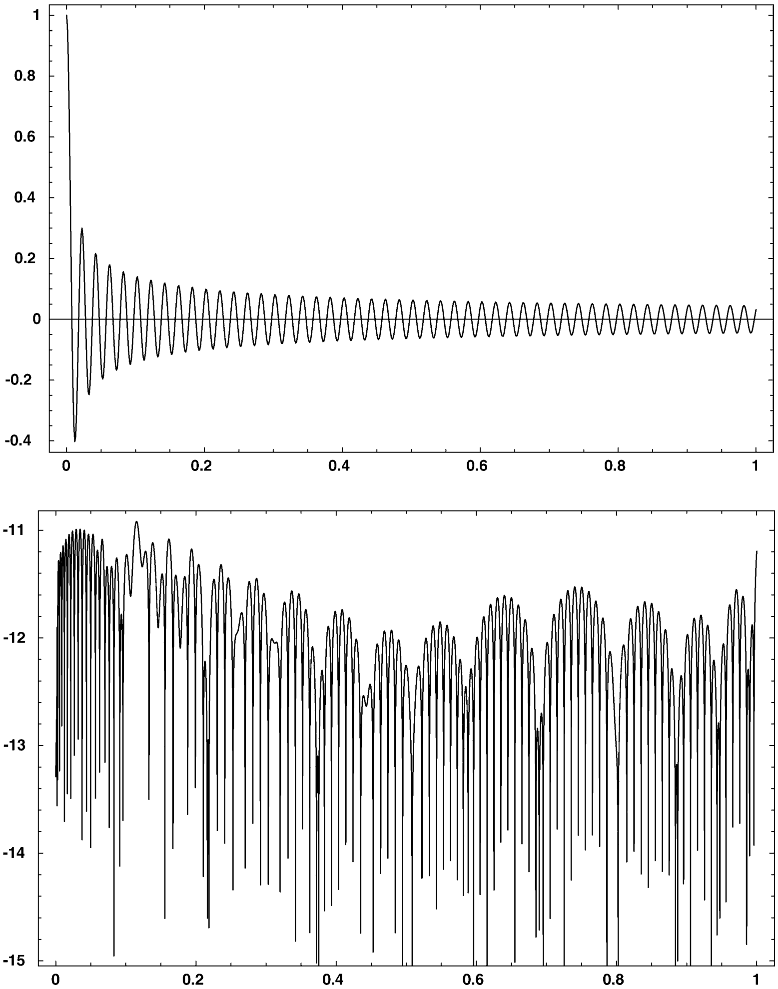


Fig. 1. The function $_0(100$) and the error (in logarithmic scale) of its 28-term approximation via (3).

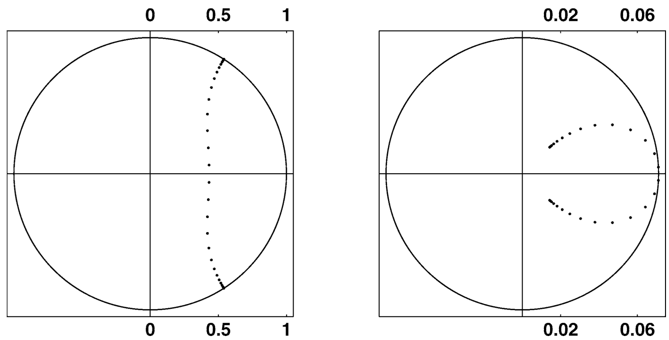


Fig. 2. The complex nodes (left) and weights (right) for the approximation of $_0$ in the interval [0 100].

However, different changes of variables in (4) will result in different contours with no a priori guidance for the choice. Using, for example, $\tau = \sin(t)$, we have

$$J_0(cx) = \frac{1}{\pi}\int_{-\pi/2}^{\pi/2} e^{icx\sin(t)}\,dt \tag{5}$$

and, with $t = x + iy$, we obtain the steepest descent path as the solution to $\sin x\cosh(y) = \pm 1$, where $|x| \leq \pi/2$ and $y \geq 0$. The discretization of the integral along this path yields (3) but with more terms than via our method. On the other hand, upon examination of the weights and nodes in Fig. 2, it is clear that their location is not accidental. It appears as if our algorithm selects a contour on which a possible integrand is least oscillatory, since that would reduce the number of necessary nodes.

We note that by optimizing the location of the nodes, we reduce their number to keep it well below the number of terms needed in Fourier expansions or in more general approximations like those discussed in [11]. We do not have a precise estimate for the optimal number of terms but we have observed that it only depends logarithmically on the parameter c and on the accuracy.

We have obtained similar results for a great variety of functions. The functions may be oscillatory, periodic, nonperiodic, or singular. For a given accuracy, we have developed algorithms to obtain the approximation with optimal or nearly optimal number of nodes and weights.

These examples motivate us to formulate the following approximation problem. Given the accuracy $\varepsilon > 0$, for a smooth function $f(x)$ find the minimal (or nearly minimal) number of complex weights w_m and complex nodes e^{t_m} such that

$$\left| f(x) - \sum_{m=1}^{M} w_m e^{t_m x} \right| \leq \varepsilon \quad \forall x \in [0, 1]. \tag{6}$$

For functions singular at $x = 0$, we formulate (6) on the interval $[\delta, 1]$, where $\delta > 0$ is a small parameter. Depending on the function and/or problem under consideration, we may measure the approximation error in (6) in a different way, e.g., we may use relative error.

As in our paper [11], we reformulate the continuous problem (6) as a discrete problem. Namely, given $2N + 1$ values of $f(x)$ on a uniform grid in $[0, 1]$ and a target accuracy $\varepsilon > 0$, we find the minimal number M of complex weights w_m and complex nodes γ_m such that

$$\left| f\left(\frac{k}{2N}\right) - \sum_{m=1}^{M} w_m \gamma_m^k \right| \leq \varepsilon \quad \forall k,\ 0 \leq k \leq 2N. \tag{7}$$

The sampling rate $2N$ has to be chosen as to oversample $f(x)$ and guarantee that the function can be accurately reconstructed from its samples. The nodes and weights in (7) depend on ε and N. Once they are obtained, the continuous approximation (6) is defined using the same weights while the exponents are set as

$$t_m = 2N \log \gamma_m$$

to match the form in (6). The nonlinear problem of finding the nodes and weights in (7) is split into two problems: to obtain the nodes, we solve a singular value problem and find M roots of a polynomial; to obtain the weights, we use the nodes to solve a well-conditioned linear Vandermonde system.

2. Preliminary considerations: properties of Hankel matrices

2.2. Fast application of Hankel matrices

For any vector $\mathbf{x} = (_0 \quad)$ denote by $\mathbf{P}_\mathbf{x}$ the polynomial $\mathbf{P}_\mathbf{x}( ) = \sum_{\geqslant 0}$ of degree at most We want to compute the vector $\mathbf{Hx}$, where $\mathbf{H}$ is the Hankel matrix defined by the vector $\mathbf{h}$ in $\mathbb{C}^{2 +1}$. Let be an integer, $2 + 1$ and $=$ a root of unity. We write

$$= \frac{1}{\ }$$

Using (18) we expand the left-hand side of (26)

$$\sum_{l=0}^{L} d_l^{(m)} u_{k+l} = \frac{1}{N}\sum_{n=0}^{N-1} \tilde{\sigma}_n \gamma_n^{k} \sum_{l=0}^{L} d_l \gamma_n^{l} = \frac{1}{N}\sum_{n=0}^{N-1} \tilde{\sigma}_n \mathbf{P_u}(\gamma_n)\gamma_n^{k}$$

and, due to (17), the last term equals

$$\frac{1}{N}\sum_{n=0}^{N-1} \mathbf{P_{\bar{u}}}(\bar{\gamma}_n)\gamma_n^{k} = \bar{u}_k.$$

Finally, since $|\tilde{\sigma}_n| = 1$ for all n, the l^2-norm of $\mathbf{d}^{(m)}$ equals 1. □

Next, we prove Theorem 3.

Proof. Part (1) is a direct consequence of (19), while part (3) follows from the first two. For part (2), (26) implies

$$\mathbf{H_d u} = \bar{\mathbf{u}}$$

and with the notation $\|\cdot\|$ for both the matrix 2-norm and the vector l^2-norm, we derive $\|\mathbf{H_d}\| \geq \frac{\|\mathbf{H_d u}\|}{\|\mathbf{u}\|} = 1$; thus, the norm is at least one. To see that it is at most one, let $\mathbf{v} \in \mathbb{C}^{N+1}$ and use (13) and (18) to write for $0 \leq k \leq L$

$$(\mathbf{H_d v})_k = \frac{1}{\sqrt{N}}\sum_{n=0}^{N-1}\left(\frac{\tilde{\sigma}_n \mathbf{P_v}(\gamma_n)}{\sqrt{N}}\right)\gamma_n^{k}$$

The right-hand side of the last equation is well defined for $0 \leq k \leq N-1$ and corresponds to the DFT of the vector $\frac{\tilde{\sigma}_n \mathbf{P_v}(\gamma_n)}{\sqrt{N}}$. Since the DFT is unitary and $|\tilde{\sigma}_n| = 1$, we obtain

$$\|\mathbf{H_d v}\|^2 \leq \left\|\frac{\tilde{\sigma}_n \mathbf{P_v}(\gamma_n)}{\sqrt{N}}\right\|^2 = \|\mathbf{v}\|^2$$

The last inequality holds for any vector $\mathbf{v}$

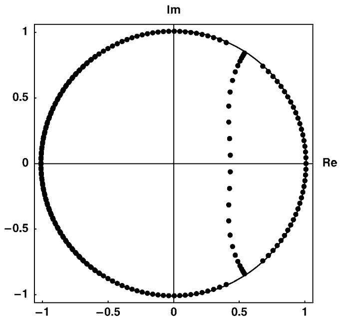


Fig. 3. Locations of all roots of the c-eigenpolynomial corresponding to the singular value $_{28}$ in the approximation of $_0$ in [0 100]. In practice, we only use the 28 roots inside the unit disk.

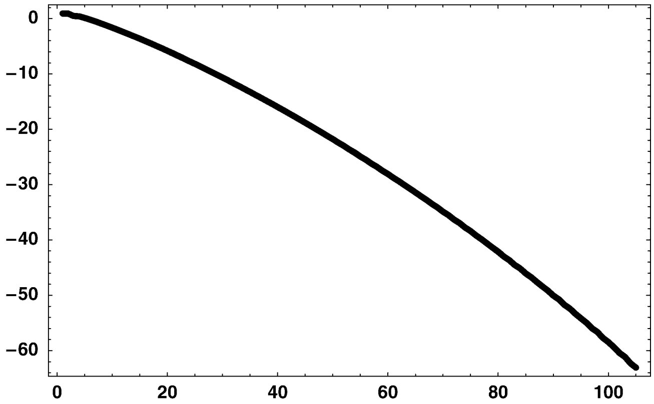
$_0$( ) in the interval [0 100]. The 28 significant weights (see Fig. 2) are associated with the nodes inside the unit disk. We note that the nodes corresponding to the discarded terms are located outside but very close to the unit circle. The error of the 28-terms approximation is displayed in Fig. 1.

By keeping only the terms with significant weights, the singular value index provides a -term approximation of the sequence with error of the order of . This behavior matches that of indices of the singular values in AAK theory, where the th singular value of the Hankel operator equals the distance from that operator to the set of Hankel operators of rank at most .

Currently we do not have a characterization of the conditions under which finite Hankel matrices may satisfy the results of the infinite theory. We only note that assuming fast decay of the singular values and that $-$ terms have small weights in (19), the approximation $= \sum_{=1}$ has the optimal number of terms. Indeed, let $\mathbf{H_b}$ be the corresponding Hankel matrix for $\mathbf{b}$. Since $\mathbf{H_b}$ has rank , we have

$$\|\mathbf{H} - \mathbf{H_b}\| + \qquad (27)$$

for some 0. Under the assumptions of $-$ small weights and of fast decay of the singular values, it is reasonable to expect small enough so that $+$ b


0
−10
−20
−30
−40
−50
−60
0
20
40
60
80
100

$$h_k = f\left(\frac{k}{2N}\right), \quad 0 \le k \le 2N \tag{28}$$

of the function to be approximated in the interval [0, 1], our goal is to find an optimal (minimal) number of nodes γ_m and weights w_m such that

$$\left| h_k - \sum_{m=1}^{M} w_m \gamma_m^k \right| \le \varepsilon \quad \forall k, \ 0 \le k \le 2N. \tag{29}$$

If the function $f(x)$ is properly oversampled, we also obtain the continuous approximation (7) of $f(x)$ over the interval [0, 1].

Let us describe the steps of the algorithm to obtain an approximation of the function f with accuracy ε.

(1) Sample the function f as in (28) by choosing appropriate N to achieve the necessary oversampling. Using those samples define the corresponding $(N + 1) \times (N + 1)$ Hankel matrix $\mathbf{H}_{kk'} = h_{k+k'}$.

(2) Find a c-eigenpair $\{\sigma, \mathbf{u}\}$, $\mathbf{H}\mathbf{u} = \sigma \bar{\mathbf{u}}$, with the c-eigenvalue σ close to the target accuracy ε. We use

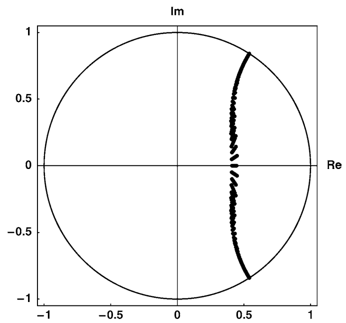


Fig. 5. Nodes corresponding to singular values in the range $[4\ 7 \times 10^{-15}\ \ 6\ 7 \times 10^{-9}]$ for the approximation of $\ _0(100\ \ )$.

at and $^{-}$. It is instructive to observe how roots in the first region stay some distance away from these accumulation points. As we have mentioned in the Introduction, this accumulation can be expected from the asymptotics of the Bessel function. More important from a computational perspective is that the nodes slowly change their locations as we modify either the approximation interval (parametrized by the constant) or the accuracy ε (parametrized by the singular values). In this way, computation of roots can be performed efficiently by, if necessary, obtaining first the nodes for a small and using them as starting points in Newton's method. To illustrate this property, in Fig. 5 we display the nodes for a range of singular values varying from $6\ 7 \times 10^{-9}$ to $4\ 7 \times 10^{-15}$.

As we noted for Fig. 2, the locations of nodes and weights suggest the existence of some integral representation of $\ _0$ on a contour in the complex plane where the integrand is least oscillatory; integration over such contour yields an efficient discretization that would correspond to the output of our algorithm.

The final approximation (6) exhibits an interesting property that we also have observed for other oscillatory functions. Suppose that we would like to obtain a decreasing function (an envelope) that touches each of the local maxima of the Bessel function and, similarly, a increasing function going through each of the local minima. The approximation (6) provides such functions in a natural way. Estimating the absolute value of an exponential sum, we define its positive envelope env( ) as

$$\left|\sum_{=1}^{} \right| \leqslant \sum_{=1}^{} | \ | \ |^{\mathrm{Re}(\ )} = \mathrm{env}(\ )$$

and its *negative envelope* as $-\mathrm{env}(\ )$. In Fig. 6 we display the Bessel function $\ _0(100\ \ )$ together with its envelopes. We note that we are not aware of any other simple method to obtain such envelopes.

5.1. The Dirichlet kernel

Another representative example is the periodic Dirichlet kernel,

$$(\ ) = \frac{1}{}\sum_{=-}^{} {}^{2\ i} = \frac{\sin}{\ \sin} \tag{30}$$

where $= 2\ + 1$. We would like to construct an approximation (6) of on the interval $[0\ \ 1]$. Since is an even function about $1\ 2$ and it approaches 1 near $= 1$ (see Fig. 8), decaying exponentials

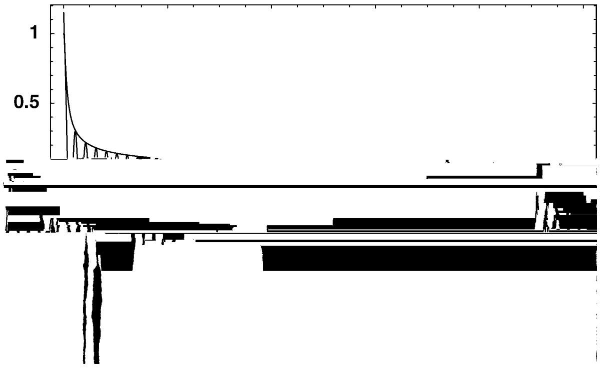


Fig. 6. The Bessel function $_0(100$) together with its envelope functions.

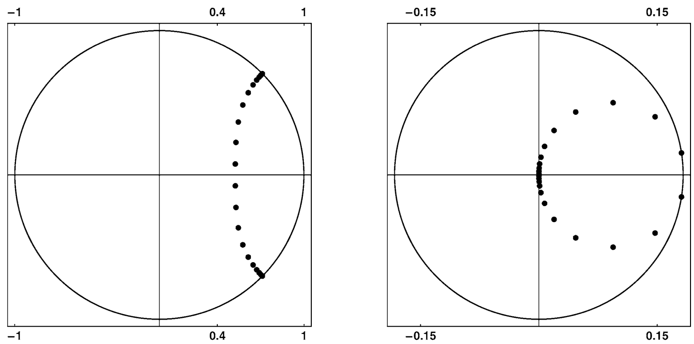


Fig. 7. The 22 nodes (left) and weights (right) for the approximation of the auxiliary function $_{50}$ in [0 1].

We note that $| \ ^{-} \ |$ 1 and, thus, the final approximation of has nodes both inside and outside of the unit disk. In Fig. 8 we display the Dirichlet kernel $_{50}$ and the error of the approximation with 44 terms given by this construction. For $_{200}$ we need 50 terms.

5.2. *The kernels* $\log\sin^2(\pi\ )$ *and* $\cot(\pi\ )$

Let us consider two examples of important kernels in harmonic analysis. The function $\log\sin^2(\pi\ )$ is the kernel of the Neumann to Dirichlet map on the unit circle for functions harmonic outside the unit disk whereas $\cot(\pi\ )$ is the Hilbert kernel for functions on the unit circle. We note that the Hilbert kernel represents a singular operator.

We first find identities similar to (31). Using the reflection formula for the gamma function,

$$(\ )\ (1-\ ) = \frac{\pi}{\sin(\pi\ )} \tag{34}$$

we obtain

$$\log\ (\ ) + \log\ (1-\ ) = \log\pi\ - \frac{1}{}$$

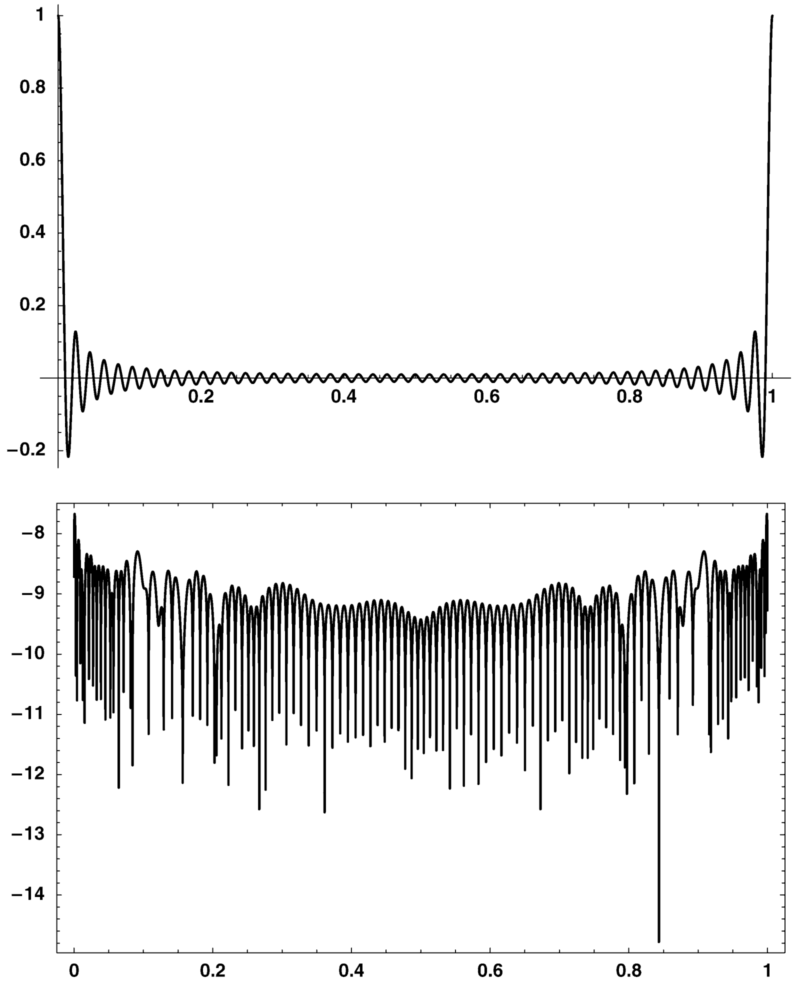


Fig. 8. Dirichlet kernel $_{50}$ (top) and the error (in logarithmic scale) of its 44-term approximation via (33).

and

$$\left| \cot - \sum_{=1} 1 \; {}^{-1} + \sum_{=1} 1 \; {}^{-1(1-)} \right| \le 2\varepsilon \qquad (38)$$

5.3. Fast evaluation of one-dimensional kernels

Let us consider computing

$$(\;) = \int_0^1 (\; - \;)(\;)\, \mathrm{d} \qquad (39)$$

at points $\{\ \}_{=1}$, $\in [0\ 1]$. In practice, we need to compute the sum

$$(\ ) = \sum_{=1} (\ - \ )\ (\ ) \tag{40}$$

where we assume that the discretization of the integral (39) has already been performed by some appropriate quadrature and we include the quadrature weights in $(\ )$.

The direct computation of (40) requires $\cdot$ operations. If we first obtain an -term exponential approximation of the kernel, an elegant algorithm [28] computes the sum with accuracy ε in

6. Reduction of number of terms

The algorithm in Section 4 allow us to find approximations for a large variety of functions but it is not well suited to deal with the extremely large ranges needed in some applications. Also, we would like to have a mechanism to approximate functions that can be expressed in terms of other functions for which we already have exponential sum approximations. Clearly, the nodes and weights for the sum or product of two known approximations are readily available, but their number is suboptimal. Similarly, an accurate but suboptimal expansion may be available, for example as the result of using some quadrature rule or simply applying the discrete Fourier transform of the data to be approximated. We now show how to take advantage of accurate but suboptimal approximations using a general approach on how to reduce (optimize) the number of terms of a given exponential sum. It consists of applying the algorithm of Section 4 to a function which is already a linear combination of exponentials on the interval $[0, 1]$ and taking advantage of some simplifications which hold for this particular class of functions. We obtain a fast algorithm for the following problem. Given

$$f_0(x) = \sum_{j=1}^{M_0} c_j e^{-\eta_j x} \tag{43}$$

and $\varepsilon > 0$, let us find a function (of the same form),

$$f(x) = \sum_{m=1}^{M} w_m e^{-t_m x} \tag{44}$$

with $M < M_0$ and such that

$$\left|f_0(x) - f(x)\right| \le \varepsilon \quad \text{for } x \in [0, 1]. \tag{45}$$

Without loss of generality, we assume distinct η_j and nonzero c_j in (43). Following the algorithm in Section 4, for some appropriate $N \gg M_0$, we construct the Hankel matrix $\mathbf{H} = \{h_{k+k'}\}_{k,k'=0}^{N}$, where

$$h_k = f_0\left(\frac{k}{2N}\right) = \sum_{j=1}^{M_0} c_j e^{-\eta_j \frac{k}{2N}}. \tag{46}$$

Denoting $\gamma_j = e^{-\frac{\eta_j}{2N}}$, $j = 1, \ldots, M_0$, we have

$$h_k = \sum_{j=1}^{M_0} c_j \gamma_j^k \tag{47}$$

and, therefore, a factorization of the Hankel matrix

$$\mathbf{H} = \mathbf{V}\mathbf{B}\mathbf{V}^{\mathbf{t}} \tag{48}$$

where $\mathbf{V}$ is the $(N + 1) \times M_0$ Vandermonde matrix

$$\mathbf{V}_{kj} = \gamma_j^k \tag{49}$$

and $\mathbf{B}$ is the diagonal matrix with entries $(c_1, \ldots, c_{M_0})$. We note that the matrix $\mathbf{H}$ has a large nullspace of dimension $N + 1 - M_0$. In fact, the nullspace consists of vectors with coordinates given by the coefficients of the polynomials $\prod_{j=1}^{M_0}(z - \gamma_j)\, P(z)$ where $P(z)$ is any polynomial of degree at most $N - M_0$.

For the second part we mimic the steps used to obtain (52) and we also use (60). The last part follows from (56) with $=$,

$$\overline{\mathbf{P_u}(\ )} = \frac{(\mathbf{Av})}{\ } = \ \ \square$$

7. Approximation of power functions and separated representations

Let us discuss how to approximate the power functions $^{-}$, $\ 0$, with a linear combination of Gaussians,

$$\Big| \ ^{-} - \sum_{=1} \ ^{-} \ ^{2[}$$

satisfy

$$(,) = (- {}^{2} {}^{2}), (,)$$

where $( )$ are polynomials of degree .

Before ending this section, we would like to remark on another application of the reduction algorithm to the summation of slowly convergent series. These results will appear separately and here we only note that our approach yields an excellent rational approximation of functions like $^{-}$, 0, providing a numerical tool to obtain best order rational approximations as indicated by Newman [24] (see also [18, p. 169]).

8. Conclusions

We have introduced a new approach, and associated algorithms, for the approximation of functions and sequences by linear combination of exponentials with complex-valued exponents. Such approximations obtained for a finite but arbitrary accuracy may be viewed as representations of functions which are more efficient (significantly fewer terms) than the standard Fourier representations. These representations can be used for a variety of purposes. For example, if used to represent kernels of operators, these approximations yield fast algorithms for applying these operators to functions. For multi-dimensional operators, we have shown how the approximation of $^{-}$, 0 leads to separated representations of Green's functions (e.g., the Poisson kernel).

We note that we just began developing the theory of such approximations and there are still many questions to be answered. We have indicated some of these questions but, in this paper, instead of concentrating on the theoretical aspects we have chosen to emphasize examples and applications of these remarkable approximations.

Acknowledgments

We thank Dr. Brad Alpert (NIST) for pointing out connections with the steepest descent method and Dr. Martin Mohlenkamp (Ohio University) for many useful suggestions.

Appendix A

We show how to choose the parameters involved in the approximation of $^{-}$ 0 by linear combination of exponentials as well as estimate the number of terms. Theorem 9 follows by substituting $\mapsto \frac{}{2}$, $\mapsto {}^{2}$, $\mapsto {}^{2}$ and choosing $= \mathcal{O}(\log \varepsilon^{-1})$ in the next

Theorem A.1. *For any* 0, $0 < 1$, *and* $0 < \varepsilon \leq \min\{\frac{1}{2}, \frac{4}{}\}$, *there exist positive numbers and such that*

$$\left| {}^{-} - \sum_{=1} {}^{-} \right| \leq {}^{-}\varepsilon \quad \text{for all} \leq \leq 1 \tag{A.1}$$

with

$$ / \quad \square \quad \frac{,\,(2/\quad + 1)}{,}\Big[,\ ^{-1}\log 4(,\ \varepsilon)^{-1} + \log 2 $$

where $h = \frac{b-a}{N}$ is the step size, $[x]$ is the integer part of the real number x, B_{2k} are the Bernoulli numbers, and $B_k(x)$ the Bernoulli polynomials. For all $x \in [0, 1]$ and $k \geq 1$ we have the inequalities (see, e.g., [14, p. 474]),

$$\frac{|B_{2k}(x)|}{2k!} \leq \frac{|B_{2k}|}{2k!} = \frac{2}{(2\pi)^{2k}} \sum_{n \geqslant 1} n^{-2k} \leq 4(2\pi)^{-2k}.$$

We then estimate the error in (A.6) as

$$\left| \int f(x)\,dx - \cdots \right| \leq 4\left(\frac{h}{2\pi}\right)^{2m} \int \left| f^{(2m)}(x) \right| dx + 4\sum_{k=1}^{m} \left(\frac{h}{2\pi}\right)^{2k} \left(\left| \cdots \right. \right.$$

a condition that follows from Lemma A.3 below. Since $\varepsilon \,\square\, \frac{1}{2}$, assumption (A.4) implies that $\square$ 2 and, therefore, $\left(\frac{\varepsilon}{4}\right)^{-\frac{1}{}} \,\square\, \varepsilon^{-1}$. Therefore, we set the following condition for the right end of the interval of integration,

$$\ln\left(2 \quad ^{-1} \ln\left(\ (\ \varepsilon)^{-1}\right)\right) \quad , \tag{A.21}$$

which also implies (A.16).

Lemma A.3. *Let , and* ε *be positive numbers such that* $^{-1}\varepsilon^{-\frac{1}{}} \,\square\, \frac{1}{2}$ *and define* $_0 = \ln 2 \quad ^{-1} \times \ln(\quad ^{-1}\varepsilon^{-\frac{1}{}})$. *Then the inequality*

$$\varepsilon \tag{A.22}$$

holds for all $\square \ _0$.

Taking the logarithm in both sides of (A.22) we get $- \quad \frac{\ln\varepsilon}{}$, and introducing the new variable $= \quad \square\ 1$, we obtain

$$\ln\left(\quad ^{-1} \quad \right) - \quad \frac{\ln\varepsilon}{} \tag{A.23}$$

or

$$= \ln \quad ^{-1}\varepsilon^{-\frac{1}{}} \qquad - \ln$$

Since $1 - \quad \square \ -\ln$ for positive , we have

$$2 \quad - \ln 2 + (1 - \quad) \,\square\, 2 \quad - \ln(2 \quad)$$

and, thus, (A.23) holds for $\square$ 2 since $- \ln$ is increasing for $\square$ 1.

A.4. Condition for and selection of the step size

Let us show by induction on $\square$ 0, that for all 0

$$\int_{-\infty}^{\infty}$$

[17] R.A. Horn, C.R. Johnson, Matrix Analysis, Cambridge Univ. Press, Cambridge, 1990.
[18] S. Jaffard, Y. Meyer, R.D. Ryan, Wavelets: Tools for Science & Technology, revised ed., SIAM, Philadelphia, PA, 2001.
[19] S. Karlin, W.J. Studden, Tchebycheff Systems: With Applications in Analysis and Statistics, Pure and Applied Mathematics, vol. XV, Interscience/Wiley, New York/London/Sydney, 1966.
[20] M.G. Kreĭn, A.A. Nudel'man, The Markov Moment Problem and Extremal Problems, Ideas and Problems of P.L. Čebyšev and A.A. Markov and Their Further Development, in: Translations of Mathematical Monographs, vol. 50, American Mathematical Society, Providence, RI, 1977.
[21] A.A. Markov, On the limiting values of integrals in connection with interpolation, Zap. Imp. Akad. Nauk Fiz.-Mat. Otd. 8