A note on the Sinc approximation
with boundary treatment

Tomoaki Okayama*1

1 Graduate School of Economics, Hitotsubashi University, 2-1, Naka, Kunitachi, Tokyo 186-8601, Japan
E-mail: tokayama@econ.hit-u.ac.jp
Received April 23, 2012, Accepted June 27, 2012

Abstract
The original form of the Sinc approximation is efficient for functions whose boundary values are zero, but not for other functions. The typical way to treat general boundary values is to introduce auxiliary basis functions, and in fact such an approach has been taken commonly in the literature. However, the approximation formula in each research is not exactly the same, and still other formulas can be derived as variants of existing formulas. The purpose of this paper is to sum up those existing formulas and new ones, and to give explicit proofs of those convergence theorems.

Keywords Sinc approximation, Sinc-collocation, boundary treatment

Research Activity Group Scientific Computation and Numerical Analysis

1. Variants of the Sinc approximation
The Sinc approximation on the real axis is expressed as

\[ F(x) \approx \sum_{j=-M}^{N} F(jh)S(j,h)(x), \quad x \in \mathbb{R}, \] (1)

where \( S(j,h)(x) \) is the so-called Sinc function defined by \( S(j,h)(x) = \sin(\pi(x/h - j))/[\pi(x/h - j)] \), and \( h, M, N \) are suitably selected with respect to \( n \). The approximation (1) requires two conditions on \( F \):

(i) \( F \) must be defined on the entire real axis, and
(ii) \( F(x) \) must tend to zero as \( x \to \pm \infty \).

There is a typical remedy for the first condition (i). If we consider the approximation of \( f \) that is defined only on the finite interval \( \Gamma = (a, b) \), choose a proper variable transformation \( \psi \) that maps \( \mathbb{R} \) onto \( \Gamma \), then put \( F(x) = f(\psi(x)) \) and use (1). The standard transformation is

\[ t = \psi(x) = \frac{b - a}{2} \tanh \left( \frac{x}{2} \right) + \frac{b + a}{2}, \]

which is called the Single-Exponential (SE) transformation, and the explicit approximation form is:

\[ f(t) \approx \sum_{j=-M}^{N} f(t_j^{SE})S(j,h)(\phi(t)) \quad t \in \Gamma, \] (2)

where \( t_j^{SE} = \psi_j SE(jh) \) and \( \phi(t) = \{ \psi_j SE \}^{-1}(t) \). The formula (2) is called the SE-Sinc approximation.

The condition (ii) still remains; \( f(a) = f(b) = 0 \) is required in (2). The common remedy is to construct a function with zero boundary values by the operator \( T \):

\[ Tf = f(t) - f(a)W_a(t) - f(b)W_b(t), \]

where \( W_a \) and \( W_b \) are auxiliary basis functions defined by \( W_a(t) = (b - t)/(b - a), \ W_b(t) = (t - a)/(b - a) \). Then \( Tf \) can be approximated by (2). The explicit form is:

\[ f(t) \approx f(a)W_a(t) + f(b)W_b(t) \]

\[ + \sum_{j=-M}^{N} T[f](t_j^{SE})S(j,h)(\phi(t)) \] (3)

This formula has been used by some authors for deriving Sinc-collocation methods for differential/integral equations [1–3]. However, the interpolating points of (3) are not consistent; \( t = a, t_j^{SE}, t_{j-M}^{SE}, \ldots, t_N^{SE} \), which have two exceptions \( t = a \) and \( t = b \). These exceptions make implementation more complicated, especially in the case that the target is a system of many equations.

In order to correct the defect, the following formula:

\[ f(t) \approx f(t_j^{SE})W_a(t) + f(t_N^{SE})W_b(t) \]

\[ + \sum_{j=-M}^{N} T^{SE}[f](t_j^{SE})S(j,h)(\phi(t)) \] (4)

has been used [4, §6–7], where

\[ T^{SE}[f] = f(t) - f(t_j^{SE})W_a(t) - f(t_N^{SE})W_b(t). \]

This formula works fine since \( f(a) \approx f(t_j^{SE}) \) and \( f(b) \approx f(t_N^{SE}) \), and the interpolating points are consistent and simple: \( t = t_j^{SE}, \ldots, t_N^{SE} \).

Similar, but different, another formula with the simple interpolating points has been proposed [5]:

\[ f(t) \approx f(t_j^{SE})W_a(t) + f(b)W_b(t) \]

\[ + \sum_{j=-M}^{N} \tilde{T}^{SE}[f](t_j^{SE})S(j,h)(\phi(t)), \] (5)

where \( f_a(t) = f(t)/W_a(t), f_b(t) = f(t)/W_b(t) \), and

\[ \tilde{T}^{SE}[f] = f(t) - f_a(t_j^{SE})W_a(t) - f_b(t_N^{SE})W_b(t). \]
Stenger [5] has introduced the following notations:
\[
\omega_{-M}^S(x) = (1 + \rho^S(-Mh)) \left( \frac{1}{1 + \rho^S(x)} \right),
\]
\[
- \sum_{k=-M+1}^{N} \frac{1}{1 + \rho^S(kh)} S(k, h)(x),
\]
\[
\omega_j^S(x) = S(j, h)(x), \quad -M < j < N,
\]
\[
\omega_N^S(x) = 1 + \rho^S(Nh) \left( \frac{\rho^S(x)}{1 + \rho^S(x)} \right),
\]
\[
- \sum_{k=-M}^{N-1} \frac{\rho^S(kh)}{1 + \rho^S(kh)} S(k, h)(x),
\]
where \( \rho^S(x) = e^x \). Then (5) can be rewritten as
\[
f(t) \approx \sum_{j=-M}^{N} f(t^S_j) \omega_j^S(\phi^S(t)).
\]
In this form we easily see the interpolating points since \( \omega_j^S(x) = \delta_{ij} \). The formulas through (3)–(5) should be called generalized SE-Sinc approximations in the sense that they can handle general boundary values. The first purpose of this paper is to sum up convergence theorems of the formulas with proofs (not explicitly given so far).

If we return our attention to the condition (i), there is another famous variable transformation:
\[
t = \psi^D(x) = \frac{b - a}{2} \tanh \left( \frac{\pi}{2} \sinh x \right) + \frac{b + a}{2},
\]
which is called the Double-Exponential (DE) transformation. This transformation also maps \( \mathbb{R} \) onto \( \Gamma \), and we can consider the DE-Sinc approximation by replacing ‘SE’ with ‘DE’ in (2). Accordingly the formula (3) can be modified as
\[
f(t) \approx f(a) W_a(t) + f(b) W_b(t)
\]
\[
+ \sum_{j=-M}^{N} \mathcal{T}[f](t^D_j) S(j, h)(\phi^D(t)),
\]
and in fact this formula has also been used [2, 3] as a generalized SE-Sinc approximation. In addition, we can derive two new generalized DE-Sinc approximations:
\[
f(t) \approx f(t^D_M) W_a(t) + f(t^D_N) W_b(t)
\]
\[
+ \sum_{j=-M}^{N} \mathcal{T}^D[f](t^D_j) S(j, h)(\phi^D(t)),
\]
(7)
\[
f(t) \approx f_a(t^D_M) W_a(t) + f_b(t^D_N) W_b(t)
\]
\[
+ \sum_{j=-M}^{N} \mathcal{T}^D[f](t^D_j) S(j, h)(\phi^D(t)),
\]
(8)
by replacing ‘SE’ with ‘DE’ in (4) and (5), respectively. If we define \( \rho^D(x) = e^{x \sinh z} \), and replace ‘SE’ with ‘DE’ in Stenger’s notations, the latter formula (8) can be rewritten as
\[
f(t) \approx \sum_{j=-M}^{N} f(t^D_j) \omega_j^D(\phi^D(t)).
\]
In addition to deriving (7) and (8) for simple interpolating points, this paper gives explicit proofs of the convergence theorems for (6)–(8), which is the second purpose.

This paper is organized as follows. The convergence theorems for (3)–(8) are stated in Section 2, and it turns out the convergence rate of the formulas (3)–(5) is the same, \( O(\sqrt{n}e^{-c\sqrt{n}}) \). The convergence rate of the formulas (6)–(8) is also the same: \( O(e^{-c'n/\log n}) \), but much higher than SE’s rate. The result is confirmed numerically in Section 3. Section 4 is devoted to proofs.

2. Convergence theorems

The following function space is crucial in this section.

**Definition 1** Let \( \mathcal{D} \) be a bounded and simply-connected domain (or Riemann surface) that contains the interval \( \Gamma \). Let \( \alpha \) and \( \beta \) be positive constants with \( \alpha \leq 1 \) and \( \beta \leq 1 \). Then \( M_{\alpha, \beta}(\mathcal{D}) \) denotes the family of all functions \( f \) that are analytic and bounded on \( \mathcal{D} \), and satisfy the following inequalities with a constant \( C \):
\[
|f(z) - f(a)| \leq C|z - a|^\alpha,
\]
\[
|f(b) - f(z)| \leq C|b - z|^\beta,
\]
for all \( z \in \mathcal{D} \).

In the subsequent theorems, \( \mathcal{D} \) is either \( \psi^S(\mathcal{D}_d) \) or \( \psi^D(\mathcal{D}_d) \), where \( \mathcal{D}_d = \{ z \in \mathcal{C} : \Im \zeta < d \} \) for \( d > 0 \). Let us define \( e^S_n \) and \( e^D_n \) as \( e^S_n = \sqrt{n}e^{-c'\sqrt{n}} \) and \( e^D_n = e^{-c'dn/\log(2dn)} \) for short.

First three theorems are for the formula (3)–(5).

**Theorem 2** (Well-known, cf. Stenger [4, §4])

Let \( f \in M_{\alpha, \beta}(\psi^S(\mathcal{D}_d)) \) for \( d \in (0, \pi) \). Let \( \mu = \min\{\alpha, \beta\} \), \( n \) be a positive integer, and \( h \) be selected by the formula
\[
h = \sqrt{\frac{\pi d}{\mu n}}. \tag{9}
\]
Moreover, let \( M \) and \( N \) be positive integers defined by
\[
\begin{align*}
M &= n, \quad N = \lceil\alpha n/\beta\rceil & (\text{if } \mu = \alpha) \\
N &= n, \quad M = \lceil\beta n/\alpha\rceil & (\text{if } \mu = \beta)
\end{align*}
\]
(10)
respectively. Then there exists a constant \( C \) independent of \( n \) such that
\[
\sup_{t \in \Gamma} \left| \mathcal{T}[f](t) - \sum_{j=-M}^{N} \mathcal{T}[f](t^S_j) S(j, h)(\phi^S(t)) \right| \leq C e^S_n.
\]
**Theorem 3** Let the assumptions in Theorem 2 be fulfilled. Then there exists a constant \( C \) independent of \( n \) such that
\[
\sup_{t \in \Gamma} \left| \mathcal{T}^S[f](t) - \sum_{j=-M}^{N} \mathcal{T}^S[f](t^D_j) S(j, h)(\phi^D(t)) \right| \leq C e^D_n.
\]
**Theorem 4** (Stenger [5, Theorem 4.2]) Let the assumptions in Theorem 2 be fulfilled. Then there exists a constant \( C \) independent of \( n \) such that
\[
\sup_{t \in \Gamma} \left| f(t) - \sum_{j=-M}^{N} f(t^S_j) \omega_j^S(\phi^S(t)) \right| \leq C e^S_n.
\]
The following three theorems are for (6)–(8).

**Theorem 5** Let \( f \in \mathcal{M}_{\delta}^{\alpha,\beta}(\mathcal{D}_d) \) for \( d \in (0, \pi/2) \). Let \( \mu = \min\{\alpha, \beta\} \), \( n \) be a positive integer, and \( h \) be selected by the formula
\[
h = \frac{\log(2dn/\mu)}{n}. \tag{11}
\]
Moreover, let \( M \) and \( N \) be positive integers defined by
\[
\begin{align*}
M &= n, \quad N = n - \lfloor \log(\beta/\alpha)/h \rfloor \quad \text{(if } \mu = \alpha) \\
N &= n, \quad M = n - \lfloor \log(\alpha/\beta)/h \rfloor \quad \text{(if } \mu = \beta)
\end{align*} \tag{12}
\]
respectively. Then there exists a constant \( C \) independent of \( n \) such that
\[
\sup_{t \in \Gamma} \left| f(t) - \sum_{j=-M}^{N} T[j;h;\phi^{\mathcal{S}}(t)](\phi^{\mathcal{D}}(t)) \right| \leq C \epsilon_n^{\mathcal{S}}.
\]

**Theorem 6** Let the assumptions in Theorem 5 be fulfilled. Then there exists a constant \( C \) independent of \( n \) such that
\[
\sup_{t \in \Gamma} \left| f(t) - \sum_{j=-M}^{N} T[j;h;\phi^{\mathcal{D}}(t)] \right| \leq C \epsilon_n^{\mathcal{D}}.
\]

**Theorem 7** Let the assumptions in Theorem 5 be fulfilled. Then there exists a constant \( C \) independent of \( n \) such that
\[
\sup_{t \in \Gamma} \left| f(t) - \sum_{j=-M}^{N} f(t_j^{\mathcal{D}}) \omega_j^{\mathcal{D}}(\phi^{\mathcal{D}}(t)) \right| \leq C \epsilon_n^{\mathcal{D}}.
\]

3. Numerical results

To confirm the theorems in Section 2 numerically, the generalized SE/DE approximations (3)–(8) were applied to the function \( f_1(t) = \sqrt{1 + t^2} \). The interval is set as \( (a, b) = (-1, 1) \). The assumptions of Theorems 2–7 are satisfied since \( f_1 \in \mathcal{M}_{1.1}^{1.1}(\mathcal{D}_{\pi/2}) \) and \( f_1 \in \mathcal{M}_{1.1}^{1.1}(\mathcal{D}_{\pi/2}) \). The computational programs were written in C with quad-precision. The errors were checked on 1999 equally-spaced points on \((-1, 1)\), i.e., \( t = -0.999, -0.998, \ldots, 0.999 \), and the maximum error among them is plotted in Fig. 1. We can observe the rate of SE’s formulas (3)–(5) is the same: \( O(\epsilon_n^{\mathcal{S}}) \). The rate of DE’s formulas (6)–(8) is also the same: \( O(\epsilon_n^{\mathcal{D}}) \), but it is much higher than SE’s rate.

4. Proofs

Let us introduce the following function space.

**Definition 8** Let \( \mathcal{D} \) be a bounded and simply-connected domain (or Riemann surface) that contains the interval \( \Gamma \), and let \( \alpha \) and \( \beta \) be positive constants. Then \( \mathcal{L}^{\alpha,\beta}(\mathcal{D}) \) denotes the family of all functions \( f \) that are analytic on \( \mathcal{D} \), and satisfy the following inequality with a constant \( C \):
\[
|f(z)| \leq C|z-a|^{\alpha}|b-z|^{\beta},
\]
for all \( z \in \mathcal{D} \).

This function space describes the assumptions for the SE-Sinc approximation (2) and the DE-Sinc approximation, as stated below.

![Fig. 1. Maximum error of the approximations (3)–(8) for \( f_1(t) = \sqrt{1 + t^2} \) on \((-1, 1)\).](image-url)
Lemma 12 (Stenger [4, p. 142]) Let \( h > 0 \). Then

\[
\sup_{x \in \mathbb{R}} \sum_{j=-n}^{n} |S(j, h)(x)| \leq \frac{2}{\pi} (3 + \log n).
\]

Proof of Theorem 3 From \( f \in M_{a, \beta}(\psi^{SE}(\mathcal{D}_d)) \), we have the following bound:

\[
|f(t_{-M}) - f(a)| \leq C(t_{-M} - a)\alpha = \frac{C(b - a)^\alpha}{(1 + e^{\alpha M})^\alpha},
\]

\[
|f(b) - f(t_{+N})| \leq C(b - t_{+N})\beta = \frac{C(b - a)^\beta}{(1 + e^{\alpha N})^\beta}.
\]

Moreover, using \( |W_a(t)| \leq 1 \) and \( |W_b(t)| \leq 1 \), and substituting (9)–(10), we have the following bound:

\[
|f(a) - f(t_{+N})|W_a(t) + (f(b) - f(t_{+N}))W_b(t)|
\leq C_1 e^{-\sqrt{\pi}dn},
\]

for some constant \( C_1 \). From this we also have

\[
|\mathcal{T}[f](t_{+N}) - \mathcal{T}^{SE}[f](t_{+N})| \leq C_1 e^{-\sqrt{\pi}dn}.
\]

Finally using Lemma 12 we obtain (13). (QED)

Similarly, Theorem 6 can be shown as follows.

Proof of Theorem 6 Since \( f \in M_{a, \beta}(\psi^{SE}(\mathcal{D}_d)) \), we have

\[
|f(t_{+M}) - f(a)| \leq C(t_{+M} - a)\alpha = \frac{C(b - a)^\alpha}{(1 + e^{\alpha M})^\alpha},
\]

\[
|f(b) - f(t_{+N})| \leq C(b - t_{+N})\beta = \frac{C(b - a)^\beta}{(1 + e^{\alpha N})^\beta}.
\]

Moreover, using \( |W_a(t)| \leq 1 \) and \( |W_b(t)| \leq 1 \), and substituting (11)–(12), we have the following bound:

\[
|\mathcal{T}[f](t_{+N}) - \mathcal{T}^{DE}[f](t_{+N})| \leq C_2 e^{-\pi dn},
\]

for some constant \( C_2 \). From this we also have

\[
|\mathcal{T}[f](t_{+N}) - \mathcal{T}^{DE}[f](t_{+N})| \leq C_2 e^{-\pi dn}.
\]

Finally using Lemma 12 we obtain (14). (QED)

Let us now proceed to Theorems 4 and 7. We will estimate the difference between (4) and (5), and between (7) and (8), respectively. That is, we show

\[
|f(t_{-M}) - f(a)| - f(t_{-M})|W_a(t) + (f(b) - f(t_{+N}))W_b(t) + \sum_{j=-M}^{N} (\mathcal{T}^{SE}[f](t_{+N}) - \mathcal{T}^{DE}[f](t_{+N}))SE(j, h)(\phi^{SE}(t_{+N}))| \leq C_3 e^{-\pi dn},
\]

where \( \nu = \max\{\alpha, \beta\} \). From these estimates Theorems 4 and 7 are proved, since the convergence rate of (15) is higher than \( \epsilon_n^{SE} \), and the rate of (16) is also higher than \( \epsilon_n^{DE} \) (notice \( -\pi dn/\nu \leq -\pi dn \) since \( \nu \in (0, 1) \)).

Proof of Theorem 4 Firstly we easily have

\[
|f_a(t_{+M}) - f(t_{+N})| = |f(t_{+N}) - e^{-Mh},
\]

\[
|f_b(t_{+N}) - f(t_{+N})| = |f(t_{+N}) - e^{-Nh}.
\]

Moreover, using \( |W_a(t)| \leq 1 \), \( |W_b(t)| \leq 1 \), and substituting (9)–(10), we have:

\[
|f(t_{+M}) - f_a(t_{+M})|W_a(t) + (f_b(t_{+N}) - f_b(t_{+N}))W_b(t)|
\leq C_1 e^{-\sqrt{\pi}dn/\nu},
\]

for some constant \( C_1 \). From this we also have

\[
|\mathcal{T}^{SE}[f](t_{+N}) - \mathcal{T}^{SE}[f](t_{+N})| \leq C_1 e^{-\sqrt{\pi}dn/\nu}.
\]

Finally using Lemma 12 we obtain (15). (QED)

Proof of Theorem 7 Firstly we easily have

\[
|f_a(t_{+M}) - f(t_{+N})| = |f(t_{+N}) - e^{-\pi \sinh(Mh)},
\]

\[
|f_b(t_{+N}) - f(t_{+N})| = |f(t_{+N}) - e^{-\pi \sinh(Nh)}.
\]

Moreover, using \( |W_a(t)| \leq 1 \), \( |W_b(t)| \leq 1 \), and \( \alpha, \beta \in (0, 1) \), and substituting (11)–(12), we have:

\[
|f(t_{+M}) - f_a(t_{+M})|W_a(t) + (f_b(t_{+N}) - f_b(t_{+N}))W_b(t)|
\leq C_1 e^{-\sqrt{\pi}dn/\nu},
\]

for some constant \( C_1 \). From this we also have

\[
|\mathcal{T}^{SE}[f](t_{+N}) - \mathcal{T}^{DE}[f](t_{+N})| \leq C_1 e^{-\sqrt{\pi}dn/\nu}.
\]

Finally using Lemma 12 we obtain (16). (QED)

Acknowledgments

This work was supported by JSPS Grants-in-Aid for Scientific Research.

References