A note on Taylor expansion

Shun Tang

Abstract. Let $f(x)$ be a real function which has $(n+1)$-th derivative on an interval $[a, b]$. For any point $x_0 \in (a, b)$ and any integer $0 \leq k \leq n$, denote by $S_{k,x_0}(x)$ the $k$-th truncation of the Taylor expansion of $f(x)$ at $x_0$, i.e.

$$S_{k,x_0}(x) = \sum_{i=0}^{k} \frac{f^{(i)}(x_0)}{i!}(x-x_0)^i.$$

In this note, we consider the $L^2$-approximation of $f(x)$ by polynomials of degree $\leq k$, we show that $S_{k,x_0}(x)$ is the limit of the best approximations of $f(x)$ on $[x_0 - \varepsilon, x_0 + \varepsilon]$ as $\varepsilon \to 0$.

2010 Mathematics Subject Classification: 41A10, 41A52, 41A58

1 Introduction

Given a polynomial $P(x)$ with real coefficients, the $L_2$-distance between $f(x)$ and $P(x)$ in the space of continuous functions on $[a, b]$ is defined as

$$\|f(x) - P(x)\|_2 := \left( \int_a^b |f(x) - P(x)|^2 \, dx \right)^{1/2}.$$ 

Let $\mathbb{R}[x]_{\leq k}$ denote the set of real polynomials of degree less than or equal to $k$. It is a classical problem to find the best approximation of $f(x)$ in $\mathbb{R}[x]_{\leq k}$. To the $L_2$-approximation of a periodic function by trigonometric polynomials, it is well known that the corresponding truncation of its Fourier expansion provides the best approximation of any fixed degree on a period because the set of functions $\{\frac{1}{2}, \cos x, \sin x, \cdots, \cos kx, \sin kx, \cdots \}$ provides an orthogonal basis with respect to the $L_2$ inner product. The Taylor expansion does not have similar property. For instance, consider the Taylor expansion of $f(x) = e^x$ at the point $x_0 = 0$, the constant term $e^0 = 1$ does not provide the best $L_2$-approximation of $e^x$ by degree-0 polynomials on a closed neighborhood of $x_0 = 0$ e.g. on $[-1, 1]$ we have

$$\left( \int_{-1}^{1} (e^x - \frac{e - 1}{2})^2 \, dx \right)^{1/2} < \left( \int_{-1}^{1} (e^x - 1)^2 \, dx \right)^{1/2}.$$ 

However, the Taylor expansion of a differentiable function has deep relation with its best local $L_2$-approximation by polynomials. We shall prove that when the closed neighborhood of $x_0$ gets...
smaller and smaller, on which the best $L_2$-approximation gets closer and closer to the truncation of the Taylor expansion. Our main theorem is the following.

**Theorem 1.1.** Let $f(x)$ be a real function which has $(n+1)$-th derivative on an interval $[a, b]$. Fix a point $x_0 \in (a, b)$ and fix an integer $0 \leq k \leq n$. Write the best $L_2$-approximation of $f(x)$ in $\mathbb{R}[x]_{\leq k}$ on a closed neighborhood $[x_0 - \varepsilon, x_0 + \varepsilon] \subset [a, b]$ with $\varepsilon > 0$ in the form

$$P_{k, \varepsilon}(x) = \sum_{i=0}^{k} a_{i, \varepsilon}(x - x_0)^i,$$

then we have $\lim_{\varepsilon \to 0} a_{i, \varepsilon} = \frac{f^{(i)}(x_0)}{i!}$ for any $0 \leq i \leq k$.

The existence of the best $L_2$-approximation of $f(x)$ in $\mathbb{R}[x]_{\leq k}$ on any subinterval of $[a, b]$ is a well-known fact. The most convenient way to find them out is to transform the linear independent set $\{1, x, x^2, \ldots\}$ to an orthogonal basis via the Gram-Schmidt process with respect to the $L_2$ inner product. The resulting orthogonal polynomials are just the shifted Legendre polynomials up to a constant multiplication factor. In this note, we shall give a direct proof of Theorem 1.1 without using Legendre polynomials.

## 2 Proof of Theorem 1.1

The extreme points of $\|f(x) - P(x)\|_2$ are the same as that of $\|f(x) - P(x)\|_2^2$. We write

$$J(a_0, \varepsilon, a_1, \varepsilon, \ldots, a_k, \varepsilon) = \int_{x_0 - \varepsilon}^{x_0 + \varepsilon} [f(x) - \sum_{i=0}^{k} a_{i, \varepsilon}(x - x_0)^i]^2 dx,$$

this is a differentiable function of $k + 1$ variables. To find the potential extreme points of $J(a_0, \varepsilon, a_1, \varepsilon, \ldots, a_k, \varepsilon)$ is to solve the following system of equations

$$\begin{cases} 
\frac{\partial J}{\partial a_0, \varepsilon} = 0 \\
\frac{\partial J}{\partial a_1, \varepsilon} = 0 \\
\vdots \\
\frac{\partial J}{\partial a_k, \varepsilon} = 0.
\end{cases}$$

We compute

$$\frac{\partial J}{\partial a_{i, \varepsilon}} = -2 \int_{x_0 - \varepsilon}^{x_0 + \varepsilon} [f(x) - \sum_{j=0}^{k} a_{j, \varepsilon}(x - x_0)^j](x - x_0)^i dx$$

$$= -2 \int_{x_0 - \varepsilon}^{x_0 + \varepsilon} f(x)(x - x_0)^i dx + 2 \int_{x_0 - \varepsilon}^{x_0 + \varepsilon} \sum_{j=0}^{k} a_{j, \varepsilon}(x - x_0)^{i+j} dx$$
\[
= 2 \sum_{j=0}^{k} \frac{a_{j,\varepsilon}}{t + j + 1} [e^{t+j+1} - (-\varepsilon)^{t+j+1}] - 2 \int_{x_0 - \varepsilon}^{x_0 + \varepsilon} f(x)(x - x_0)^t dx.
\]

So Equation (1) is a system of linear equations of \( k + 1 \) variables. Write it as a matrix equation
\[
A_{k+1} \cdot X = W_{k+1},
\]
then \( A_{k+1} \) is a \((k+1) \times (k+1)\)-matrix in which \( a_{rs} = \frac{1}{\varepsilon^{r+s-1}} [e^{r+s-1} - (-\varepsilon)^{r+s-1}] \) and \( W_{k+1} \) is a \((k + 1)\)-vector in which \( w_r = \int_{x_0 - \varepsilon}^{x_0 + \varepsilon} f(x)(x - x_0)^{r-1} dx \). For example, for \( k = 4 \),
\[
A_5 = \begin{pmatrix}
2\varepsilon & 0 & \frac{2\varepsilon^3}{3} & 0 & \frac{2\varepsilon^5}{5} \\
0 & \frac{2\varepsilon^3}{3} & 0 & \frac{2\varepsilon^5}{5} & 0 \\
0 & \frac{2\varepsilon^3}{3} & 0 & \frac{2\varepsilon^5}{5} & 0 \\
0 & \frac{2\varepsilon^3}{3} & 0 & \frac{2\varepsilon^5}{5} & 0 \\
\frac{2\varepsilon^5}{5} & 0 & \frac{2\varepsilon^5}{5} & 0 & \frac{2\varepsilon^5}{9}
\end{pmatrix}.
\]

Note that \( A_{k+1} \) is also the Hessian matrix of \( J(a_0,\varepsilon, a_1,\varepsilon, \ldots, a_k,\varepsilon) \). The positive definiteness of \( A_{k+1} \) follows from its inner-product expression \( a_{rs} = \langle (x-x_0)^{r-1}, (x-x_0)^{s-1} \rangle_{L_2} \) and the fact that \{1, \( (x-x_0), (x-x_0)^2, \ldots, (x-x_0)^k \} \) are linear independent. For the needs of mathematics teaching, we include a direct proof (which is perhaps an interesting one). We shall directly prove that \( \det(A_{k+1}) > 0 \) for all \( k \geq 0 \).

Firstly, set
\[
\tilde{A}_{k+1} := \begin{pmatrix}
1 & 0 & \frac{1}{3} & \cdots & \frac{1-(-1)^{k+1}}{2(k+1)^2} \\
0 & \frac{1}{3} & 0 & \cdots & \frac{1-(-1)^{k+2}}{2(k+2)^2} \\
\frac{1}{3} & 0 & \frac{1}{5} & \cdots & \frac{1-(-1)^{k+3}}{2(k+3)^2} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\frac{1-(-1)^{k+1}}{2(k+1)} & \frac{1-(-1)^{k+2}}{2(k+2)} & \frac{1-(-1)^{k+3}}{2(k+3)} & \cdots & \frac{1-(-1)^{2k+1}}{2(2k+1)^2}
\end{pmatrix},
\]
one computes
\[
\det(A_{k+1}) = \sum_{\sigma \in S_{k+1}} (-1)^{\text{sgn}(\sigma)} a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{(k+1)\sigma(k+1)}
= 2^{k+1}\varepsilon^{\frac{k+1}{2}} \left( \sum_{i=1}^{k+1} (i+\sigma(i)-1) \right) \det(A_{k+1})
= 2^{k+1}\varepsilon^{\frac{k+1}{2}} \left( 2^{\sum_{i=1}^{k+1} i} - (k+1) \right) \det(A_{k+1})
= 2^{k+1}\varepsilon^{\frac{k+1}{2}} \det(A_{k+1})
\]
where \( S_{k+1} \) denotes the \((k + 1)\)-th symmetric group. Since \( \varepsilon > 0 \), we only need to prove that \( \det(\tilde{A}_{k+1}) > 0 \) for all \( k \geq 0 \).

Set
\[
u = \begin{cases}
\frac{k}{2}, & \text{if } k \text{ is even} \\
\frac{k+1}{2}, & \text{if } k \text{ is odd}
\end{cases}
\]
\[
u = \begin{cases}
\frac{k+1}{2}, & \text{if } k \text{ is even} \\
\frac{k+1}{2}, & \text{if } k \text{ is odd}
\end{cases}
\]
and choose a permutation $\sigma \in S_{k+1}$ which sends $(1, 2, \cdots, k+1)$ to $(2, 4, \cdots, 2u, 1, 3, \cdots, 2v-1)$.

Apply the elementary row operations corresponding to $\sigma$ to the matrix $A_{k+1}$ and apply the elementary column operations corresponding to $\sigma$ to the resulting matrix, $A_{k+1}$ is then transformed to a partitioned matrix

$$
\begin{pmatrix}
B_u & 0 \\
0 & C_v
\end{pmatrix}
$$

whose determinant is equal to $\det(A_{k+1})$. The block $B_u$ is a $u \times u$-matrix, $C_v$ is a $v \times v$-matrix. They are of the following forms

$$
B_u = \begin{pmatrix}
\frac{1}{1} & \frac{1}{2} & \cdots & \frac{1}{2u+1} \\
\frac{1}{2} & \frac{1}{3} & \cdots & \frac{1}{2u+3} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{1}{2u+1} & \frac{1}{2u+3} & \cdots & \frac{1}{2u+2v-1}
\end{pmatrix},
C_v = \begin{pmatrix}
\frac{1}{1} & \frac{1}{3} & \cdots & \frac{1}{2v-1} \\
\frac{1}{3} & \frac{1}{5} & \cdots & \frac{1}{2v+1} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{1}{2v-1} & \frac{1}{2v+1} & \cdots & \frac{1}{2v+2v-1}
\end{pmatrix}.
$$

If we can prove $\det(B_u) > 0$ and $\det(C_v) > 0$ for all $u$ and all $v$, then we get $\det(A_{k+1}) > 0$ for all $k \geq 0$. This fact follows from the following lemma.

**Lemma 2.1.** For any integer $i \geq 1$ and any integer $t \geq 0$, the determinant of the following matrix

$$
D(i, t) = \begin{pmatrix}
\frac{1}{1} & \frac{1}{i+2} & \frac{1}{i+4} & \cdots & \frac{1}{i+2t} \\
\frac{1}{i+2} & \frac{1}{i+4} & \frac{1}{i+6} & \cdots & \frac{1}{i+2(t+1)} \\
\frac{1}{i+4} & \frac{1}{i+6} & \frac{1}{i+8} & \cdots & \frac{1}{i+2(t+2)} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\frac{1}{i+2t} & \frac{1}{i+2(t+1)} & \frac{1}{i+2(t+2)} & \cdots & \frac{1}{i+4t}
\end{pmatrix}
$$

is always positive.

**Proof.** For any integer $i \geq 1$, $\det(D(i, 0)) = \frac{1}{i} > 0$ and $\det(D(i, 1)) = \frac{1}{i(i+4)} - \frac{1}{(i+2)^2} > 0$. For general $t \geq 2$, we firstly multiply the $j$-th row of $D(i, t)$ by a constant $i + 2(j - 1) > 0$ from $j = 1$ to $j = t + 1$, then we get a matrix

$$
\begin{pmatrix}
1 & \frac{1}{i+2} & \frac{1}{i+4} & \cdots & \frac{1}{i+2(t+2)} \\
1 & \frac{1}{i+4} & \frac{1}{i+6} & \cdots & \frac{1}{i+2(t+1)} \\
1 & \frac{1}{i+6} & \frac{1}{i+8} & \cdots & \frac{1}{i+2(t+2)} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & \frac{1}{i+2(t+2)} & \frac{1}{i+2(t+1)} & \cdots & \frac{1}{i+4t}
\end{pmatrix}
$$

whose determinant has the same sign as $\det(D(i, t))$. Next, we multiply the $j$-th column of this matrix by a constant $i + 2(j - 1)$ from $j = 1$ to $j = t$, then we get a matrix

$$
\begin{pmatrix}
1 & \frac{1}{i+2} & \frac{1}{i+4} & \cdots & \frac{1}{i+2(t+2)} \\
1 & \frac{1}{i+4} & \frac{1}{i+6} & \cdots & \frac{1}{i+2(t+1)} \\
1 & \frac{1}{i+6} & \frac{1}{i+8} & \cdots & \frac{1}{i+2(t+2)} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & \frac{1}{i+2(t+2)} & \frac{1}{i+2(t+1)} & \cdots & \frac{1}{i+4t}
\end{pmatrix}
$$

whose determinant has the same sign as $\det(D(i, t))$. Therefore, $\det(D(i, t)) > 0$ for all $t \geq 0$. This completes the proof of Lemma 2.1.
matrix by a constant \( i + 2(j - 1) > 0 \) from \( j = 2 \) to \( j = t + 1 \), then we get the following matrix

\[
\begin{pmatrix}
1 & i & i & \cdots & i \\
1 & \frac{(i+2)^2}{i+4} & \frac{(i+2)(i+4)}{i+6} & \cdots & \frac{(i+2)(i+2t)}{i+2(t+1)} \\
1 & \frac{(i+2)(i+4)}{i+6} & \frac{(i+4)^2}{i+8} & \cdots & \frac{(i+4)(i+2t)}{i+2(t+2)} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & \frac{(i+2)(i+2t)}{i+2(t+1)} & \frac{(i+4)(i+2t)}{i+2(t+2)} & \cdots & \frac{(i+2t)^2}{i+4t}
\end{pmatrix}
\]

whose determinant also has the same sign as \( \det(D(i, t)) \).

Adding a \(-1\) multiple of the first row to the \( j\)-th row from \( j = 2 \) to \( j = t + 1 \), we get

\[
\begin{pmatrix}
1 & i & i & \cdots & i \\
0 & \frac{4}{i+4} & \frac{8}{i+6} & \cdots & \frac{4t}{i+2(t+1)} \\
0 & \frac{4}{i+6} & \frac{8}{i+8} & \cdots & \frac{4t}{i+2(t+2)} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & \frac{4t}{i+2(t+1)} & \frac{8t}{i+2(t+2)} & \cdots & \frac{4t^2}{i+4t}
\end{pmatrix}
\]

so that the sign of \( \det(D(i, t)) \) is the same as the sign of the determinant of

\[
\begin{pmatrix}
\frac{4}{i+4} & \frac{8}{i+6} & \cdots & \frac{4t}{i+2(t+1)} \\
\frac{8}{i+6} & \frac{16}{i+8} & \cdots & \frac{8t}{i+2(t+2)} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{4t}{i+2(t+1)} & \frac{8t}{i+2(t+2)} & \cdots & \frac{4t^2}{i+4t}
\end{pmatrix},
\]

and hence is the same as the sign of the determinant of

\[
\begin{pmatrix}
\frac{1}{i+4} & \frac{2}{i+6} & \cdots & \frac{4}{i+2(t+1)} \\
\frac{2}{i+6} & \frac{4}{i+8} & \cdots & \frac{4t}{i+2(t+2)} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{4}{i+2(t+1)} & \frac{4t}{i+2(t+2)} & \cdots & \frac{4t^2}{i+4t}
\end{pmatrix}.
\]

Multiplying the \( j\)-th column of the above matrix by \( \frac{1}{j} \) from \( j = 2 \) to \( j = t \), we may conclude that the sign of \( \det(D(i, t)) \) is the same as the sign of \( \det(D(i+4, t-1)) \). Repeating these manipulations, we know that \( \det(D(i, t)) \) has the same sign as \( \det(D(i+4(t-1), 1)) \) which is positive. So we are done.

To compute the limit of the solutions of \( A_{k+1} \cdot X = W_{k+1} \) as \( \varepsilon \to 0 \), we describe the elements in \( A_{k+1}^{-1} \).

**Lemma 2.2.** For any \( 1 \leq r, s \leq k + 1 \), the element in \( A_{k+1}^{-1} \) at the place \((r, s)\) is of the form \( \alpha_{rs} \cdot \left(\frac{1}{r}\right)^{r+s-1} \) where \( \alpha_{rs} \) is a real number depending only on \( k \). If \( r + s \) is an odd number, then \( \alpha_{rs} = 0 \).
Proof. Denote by $A^{-1}_{k+1}$ the adjugate matrix of $A_{k+1}$. The first statement is easily seen from $A^{-1}_{k+1} = \frac{1}{\det(A_{k+1})} A^T_{k+1}$, $\det(A_{k+1}) = 2^{k+1} \epsilon^{(k+1)^2} \det(\mathbf{A}_{k+1})$ and the cofactor of $a_{rs}$ equals
\[
\sum_{\sigma \in S_{k+1} \atop \sigma(s)=r} (-1)^{\text{sgn}(\sigma)} a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{(s-1)\sigma(s-1)} a_{(s+1)\sigma(s+1)} \cdots a_{(k+1)\sigma(k+1)}
\]
and hence equals $\epsilon^{(k+1)^2-(s+r-1)}$ multiplied by a real number. Furthermore, if $r + s$ is an odd number, then each term in the above summation must contain some $a_{s' r'}$ as a factor such that $s' + r'$ is also odd. This implies that $a_{rs} = 0$. \qed

Since $f(x)$ has $(n+1)$-th derivative on $[a, b]$ and $k \leq n$, we may write
\[
f(x) = f(x_0) + f'(x_0)(x - x_0) + \cdots + \frac{1}{(k+1)!} f^{(k+1)}(x_0)(x - x_0)^{k+1} + o((x - x_0)^{k+1}).
\]
Set
\[
R_k(x) := f(x) - [f(x_0) + f'(x_0)(x - x_0) + \cdots + \frac{1}{k!} f^{(k)}(x_0)(x - x_0)^k],
\]
then
\[
\lim_{x \to x_0} \frac{R_k(x)}{(x - x_0)^{k+1}} = \frac{1}{(k+1)!} f^{(k+1)}(x_0).
\]
We define a new function
\[
G_k(x) = \begin{cases} 
\frac{R_k(x)}{(x - x_0)^{k+1}}, & \text{if } x \neq x_0 \\
\frac{1}{(k+1)!} f^{(k+1)}(x_0), & \text{if } x = x_0.
\end{cases}
\]
It is a continuous function and it satisfies $G_k(x)(x - x_0)^{k+1} = R_k(x)$. For $1 \leq s \leq k + 1$,
\[
w_s = \int_{x_0 - \epsilon}^{x_0 + \epsilon} f(x) - \frac{x_0 + \epsilon}{x_0 - \epsilon} dx
\]
\[
= \sum_{r=1}^{k+1} \int_{x_0 - \epsilon}^{x_0 + \epsilon} \frac{1}{(r-1)!} f^{(r-1)}(x_0)(x - x_0)^{r+s-2} dx + \int_{x_0 - \epsilon}^{x_0 + \epsilon} G_k(x)(x - x_0)^{s+k} dx
\]
\[
= \sum_{r=1}^{k+1} \left\{ \frac{1}{(r-1)!} f^{(r-1)}(x_0) \frac{1}{r + s - 1} \left[ (r+s-1 - \epsilon)^{r+s-1} - (r+s-1)^{r+s-1} \right] \right\} + \int_{x_0 - \epsilon}^{x_0 + \epsilon} G_k(x)(x - x_0)^{s+k} dx
\]
\[
= \sum_{r=1}^{k+1} \frac{1}{(r-1)!} f^{(r-1)}(x_0) a_{rs} + \int_{x_0 - \epsilon}^{x_0 + \epsilon} G_k(x)(x - x_0)^{s+k} dx.
\]
Denote the unique solution of $A_{k+1} \cdot X = W_{k+1}$ by $(a_{0, \epsilon}, a_{1, \epsilon}, \cdots, a_{k, \epsilon})^T$, then for any $0 \leq i \leq k$,
\[
a_{i, \epsilon} = -\frac{1}{i!} f^{(i)}(x_0) = \sum_{s=1}^{k+1} \left\{ a_{(i+1)s} \frac{1}{\epsilon^{s+i}} \int_{x_0 - \epsilon}^{x_0 + \epsilon} G_k(x)(x - x_0)^{s+k} dx \right\}.
\]
We are left to estimate the absolute value of $a_{i,\varepsilon} - \frac{1}{i!}f^{(i)}(x_0)$. Since $G_k(x)$ is continuous and hence is bounded over $[a, b]$, we may suppose that $|G_k(x)| \leq M$ over $[a, b]$ for some positive number $M$. Then

$$
|a_{i,\varepsilon} - \frac{1}{i!}f^{(i)}(x_0)| \leq \sum_{s=1}^{k+1} |\alpha_{(i+1)s}| \cdot M \cdot \left(\frac{1}{\varepsilon}\right)^{s+i} \int_{x_0-\varepsilon}^{x_0+\varepsilon} |(x-x_0)^{s+k}| \, dx
$$

$$
= \sum_{s=1}^{k+1} |\alpha_{(i+1)s}| \cdot 2M \cdot \left(\frac{1}{\varepsilon}\right)^{s+i} \int_{x_0-\varepsilon}^{x_0+\varepsilon} (x-x_0)^{s+k} \, dx
$$

$$
= \sum_{s=1}^{k+1} |\alpha_{(i+1)s}| \cdot \frac{2M}{s+k+1} \cdot |x_0|^k \cdot \varepsilon^{k+1-i}
$$

which tends to 0 as $\varepsilon \to 0$. This completes the proof of Theorem 1.1.

**Remark 2.3.** (a) The relationship between Taylor expansion and the best $L_2$-approximation by polynomials is much clearer for analytic complex functions. For example, assume that $f$ is a complex function which is analytic in the open unit disk $D: |z| < 1$ and continuous on $\overline{D}: |z| \leq 1$, then for any $k \geq 0$, the $k$-th truncation of the Taylor expansion of $f$ at $z_0 = 0$ is the best $L_2$-approximation of $f$ on the unit circle $C: |z| = 1$. That is

$$
\int_C |f(z) - S_k(z)|^2 \, dz \leq \int_C |f(z) - P(z)|^2 \, dz, \quad \forall P(z) \in \mathbb{C}[z]_{\leq k},
$$

where $S_k(z) = \sum_{i=0}^{k} \frac{f^{(i)}(0)}{i!} z^i$ and $\mathbb{C}[z]_{\leq k}$ denotes the set of all complex polynomials of degree at most $k$. This is simply because $\{1, z, z^2, \ldots, z^k\}$ provides an orthogonal basis with respect the $L_2$ hermitian inner product on $C$, i.e.

$$
\langle z^p, z^q \rangle = \int_C z^p \overline{z^q} \, |dz| = 0, \quad \text{if } p \neq q.
$$

This statement is also true if the line integral (integral on $C$) is replaced by a surface integral (integral on $\overline{D}$), see [Wal] Section 6.1, 6.5.

(b) Replacing the $L_2$-norm by the $L_{\infty}$-norm

$$
\|f(x) - P(x)\|_{\infty} := \max_{a \leq x \leq b} |f(x) - P(x)|,
$$

one can ask whether the local Chebyshev best consistent approximations of $f(x)$ in $\mathbb{R}[x]_{\leq k}$ around $x_0 \in (a, b)$ also converge to the $k$-th truncation of the Taylor expansion of $f(x)$ at $x_0$. The answer is YES, because the Chebyshev best consistent approximating polynomial in $\mathbb{R}[x]_{\leq k}$ is a Lagrange interpolating (and also a Newton interpolating) polynomial of degree $k$. As the interpolating interval gets smaller and smaller, the difference quotients appearing in the interpolating polynomials converge to the Taylor coefficients.

(c) The main theorem in this paper has no contribution to numerical approximation applied in practice, it originates in a simple question from the author’s teaching, whether the Taylor
coefficients can be explained by the idea of least square approximation. Different from the least square approximation on the complex plane, up to now, the author has not found any relevant conclusion or expression about real functions in previous literature. So the author wrote down this note and shared his findings, hoping that the teacher who is teaching related contents of Taylor expansion, could help students understand the significance of Taylor coefficients from the perspective of least square approximation.

(d). Although the author knows that the Taylor polynomial is the limit of the local best $L_2$-approximations (resp. $L_\infty$-approximations), he can not answer the following question: for any $P(x) \in \mathbb{R}[x]_{\leq k}$ which is not equal to the $k$-th Taylor polynomial $S_{k,x_0}(x)$, if there exists a positive number $\delta$ such that for all $0 < \varepsilon < \delta$,

$$
\left( \int_{x_0-\varepsilon}^{x_0+\varepsilon} [f(x) - S_{k,x_0}(x)]^2 \, dx \right)^{1/2} < \left( \int_{x_0-\varepsilon}^{x_0+\varepsilon} [f(x) - P(x)]^2 \, dx \right)^{1/2}
$$

(resp. $\max_{x_0-\varepsilon \leq x \leq x_0+\varepsilon} |f(x) - S_{k,x_0}(x)| < \max_{x_0-\varepsilon \leq x \leq x_0+\varepsilon} |f(x) - P(x)|$).

References

[Wal] J. L. Walsh, *Interpolation and approximation by rational functions in the complex domain*, 5th ed., Amer. Math. Soc. Coll. Publ., vol. 20, Amer. Math. Soc., Providence, RI, 1969.