Journal of Computational Mathematics, Vol.16, No.1, 1998, 152–164.

THE DEFECT ITERATION OF THE FINITE ELEMENT FOR ELLIPTIC BOUNDARY VALUE PROBLEMS AND PETROV-GALERKIN APPROXIMATION[*1)]

Jun-bin Gao
(*Department of Mathematics, Huazhong University of Science and Technology, Wuhan 430074, China*)

Yi-du Yang
(*Department of Mathematics, Guizhou Normal University, Guiyang 550001, China*)

T.M. Shih
(*Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, China*)

Abstract

In this paper we introduce a Petrov-Galerkin approximation model to the solution of linear and semi-linear elliptic boundary value problems in which piecewise quadratic polynomial space and piecewise linear polynomial space are used as the shape function space and the test function space, respectively. We prove that the approximation order of the standard quadratic finite element can be attained in this Petrov-Galerkin model. Based on the so-called "contractivity" of the interpolation operator, we further prove that the defect iterative sequence of the linear finite element solution converge to the proposed Petrov-Galerkin approximate solution.

Key words: Petrov-Galerkin approximation, defect iteration correction, interpolation operator

1. Introduction

Frank etc. cf. [1] established the iterated defect correction scheme for finite element of elliptic boundary problems. For linear elliptic boundary value problem [2–5] have discussed the efficiency of the scheme by using superconvergence and asymptotic expansion under the conditions that the partition is uniform or strongly regular. It is proven that for the given linear finite element solution as initial approximation the first iterated correction can achieve the approximation order that the standard quadratic finite element solution has. However, for example, when the partition is only piecewise uniform, the approximation order of the quadratic finite element can not be obtained by the first iterated correction under the natural smoothness assumption. Moreover

[*] Received July 19, 1995.

[1)] This author is supported by the Postdoctoral Science Foundation of China, the National Natural Science Foundation of China and the Natural Science Foundation for Youths provided by HUST.

numerical results present that the approximation order is lower around the crossnode of bigger element; cf. [2, 3], although the exact solution is sufficiently smooth. On the other hand, the results in [2, 3] point out that the iterated corrections after many times can make up for a loss of approximation defect. That is, the iterated defect correction scheme is efficient. How can one give a theorectical explanation?

For the linear two-point boundary problem, it has been shown [7] that the iterated defect correction of finite element solution converges to the Petrov-Galerkin approximation solution. Can one further study the convergence of the iterated defect correction scheme for the finite element of the elliptic boundary problem by the aid of ideas given in [7]? To answer the question, we should establish a suitable Petrov-Galerkin scheme for the elliptic boundary problem. Although the theoretical analysis for Petrov-Galerkin approxiamtion of the elliptic boundary problem has been established in [8], to construct a practicable Petrov-Galerkin scheme and prove its convergence and error estimation is important work. This paper will be dedicated to this problem.

The remainder of the paper is organized as follows. We establish the so-called contractivity (cf. Theorem 2.1) of the interpolation operator in the next section. Then, in Section 3 we consider the linear elliptic boundary problem and establish a scheme of Petrov-Galerkin approximation. Furthermore we prove that the iterated defect correction for the linear finite element solution geometrically converges to the solution of the proposed Petrov-Galerkin scheme. Finally, in Section 4 we report the similar results as in Section 3 for the semi-linear elliptic boundary problem.

2. Approximation property of interpolation operator

Given a triangle T with vertices at $P_i = (x_i, y_i)$, $i = 1, 2, 3$, denote by Δ the area of T and set

$$\begin{gathered}
\xi_1 = x_2 - x_3,\ \xi_2 = x_3 - x_1,\ \xi_3 = x_1 - x_2 \\
\eta_1 = y_2 - y_3,\ \eta_2 = y_3 - y_1,\ \eta_3 = y_1 - y_2 \qquad (2.1) \\
r_1(T) = \frac{1}{\Delta}(\xi_2\xi_3 + \eta_2\eta_3),\quad r_2(T) = \frac{1}{\Delta}(\xi_3\xi_1 + \eta_3\eta_1),\ r_3(T) = \frac{1}{\Delta}(\xi_2\xi_3 + \eta_2\eta_3), \\
t_1(T) = \frac{1}{\Delta}(\xi_1^2 + \eta_1^2),\ t_2(T) = \frac{1}{\Delta}(\xi_2^2 + \eta_2^2),\ t_3(T) = \frac{1}{\Delta}(\xi_3^2 + \eta_3^2) \qquad (2.2) \\
l_1(T)^2 = \xi_1^2 + \eta_1^2,\ l_2(T)^2 = \xi_2^2 + \eta_2^2,\ l_3(T)^2 = \xi_3^2 + \eta_3^2, \qquad (2.3)
\end{gathered}$$

where l_i is the length of the edge $P_{i-1}P_{i+1}$ opposite to the vertex P_i (with $i = 1, 2, 3$ and $i-1, i, i+1 \in Z_3$ similarly defined in the following) and it is obvious that $r_i(T) \leq 0$ and

$$t_1(T) = -r_2(T) - r_3(T),\ t_2(T) = -r_3(T) - r_1(T),\ t_3(T) = -r_1(T) - r_2(T) \qquad (2.4)$$

Now let λ_i be the area coordinates related to the vertices P_i, i.e.,

$$\begin{cases} x = x_1\lambda_1 + x_2\lambda_2 + x_3\lambda_3 \\ y = y_1\lambda_1 + y_2\lambda_2 + y_3\lambda_3 \\ 1 = \lambda_1 + \lambda_2 + \lambda_3 \end{cases} \tag{2.5}$$

such that the triangle T is transformed into the standard simplex $\widehat{T} = \{(\lambda_1, \lambda_2, \lambda_3) \mid \lambda_1 + \lambda_2 + \lambda_3 = 1,\ \lambda_i \geq 0\}$, where $(\lambda_1, \lambda_2, \lambda_3)$ are called the barycentric coordinates of (x, y) with respect to the triangle T. By the transformation (2.5) any function $u(x, y)$ defined on T can be associated with a function $\widehat{u}(\lambda_1, \lambda_2, \lambda_3)$ defined on $\widehat{T}$ such that

$$u(x, y) \equiv \widehat{u}(\lambda_1, \lambda_2, \lambda_3) \tag{2.6}$$

On the other hand it is not difficult to prove that, with (2.5),

$$\begin{aligned} &\frac{\partial\lambda_1}{\partial x} = \frac{\eta_1}{2\Delta},\ \frac{\partial\lambda_2}{\partial x} = \frac{\eta_2}{2\Delta},\ \frac{\partial\lambda_3}{\partial x} = \frac{\eta_3}{2\Delta}, \\ &\frac{\partial\lambda_1}{\partial y} = -\frac{\xi_1}{2\Delta},\ \frac{\partial\lambda_2}{\partial y} = -\frac{\xi_2}{2\Delta},\ \frac{\partial\lambda_3}{\partial y} = -\frac{\xi_3}{2\Delta}, \end{aligned} \tag{2.7}$$

Hence we can conclude from (2.6) and (2.7) that

$$\begin{aligned} \frac{\partial u}{\partial x} &= \frac{1}{2\Delta}\Big(\eta_1\frac{\partial\widehat{u}}{\partial\lambda_1} + \eta_2\frac{\partial\widehat{u}}{\partial\lambda_2} + \eta_3\frac{\partial\widehat{u}}{\partial\lambda_3}\Big) \\ \frac{\partial u}{\partial y} &= -\frac{1}{2\Delta}\Big(\xi_1\frac{\partial\widehat{u}}{\partial\lambda_1} + \xi_2\frac{\partial\widehat{u}}{\partial\lambda_2} + \xi_3\frac{\partial\widehat{u}}{\partial\lambda_3}\Big) \end{aligned} \tag{2.8}$$

For any triangle T with vertices at $P_i = (x_i, y_i)$ $(i = 1, 2, 3)$, four sub-triangles T_0, T_1, T_2 and T_3 can be obtained by connecting the midpoint $\overline{P_i}$ of each edge $P_{i-1}P_{i+1}$ opposite to the vertex P_i, where $T_0 = \overline{P_1}\overline{P_2}\overline{P_3}$, $T_1 = P_1\overline{P_3}\overline{P_2}$, $T_2 = \overline{P_3}P_2\overline{P_1}$ and $T_3 = \overline{P_2}\overline{P_1}P_3$. We can define similarly $r_i(T_j)$ to (2.2) and it can be proved that $r_i(T_j) = r_i(T)$ and $t_i(T_j) = t_i(T)$ for $i = 1, 2, 3$ and $j = 0, 1, 2, 3$. Then we also denote by $r_i \equiv r_i(T_j) = r_i(T)$ and $t_i \equiv t_i(T_j) = t_i(T)$ without any confusion.

Let I_2u be the quadratic Lagrange interpolation polynomial defined on the nodes $\{P_i, \overline{P_i}, i = 1, 2, 3\}$ and I_1u be the piecewise linear interpolation polynomial with respect to $\{P_i, \overline{P_i}, i = 1, 2, 3\}$ on the four sub-triangles T_0, T_1, T_2 and T_3.

Define

$$\|u\|_T = \sqrt{\int_T \Big(\Big(\frac{\partial u}{\partial x}\Big)^2 + \Big(\frac{\partial u}{\partial y}\Big)^2\Big)dxdy} \tag{2.9}$$

Then with (2.8) we have

$$\begin{aligned} \|u\|_T &= \int_T \Big(\Big(\frac{\partial u}{\partial x}\Big)^2 + \Big(\frac{\partial u}{\partial y}\Big)^2\Big)dxdy \\ &= \int_{\widehat{T}} \frac{1}{4\Delta}\Big(\Big(\eta_1\frac{\partial\widehat{u}}{\partial\lambda_1} + \eta_2\frac{\partial\widehat{u}}{\partial\lambda_2} + \eta_3\frac{\partial\widehat{u}}{\partial\lambda_3}\Big)^2 + \Big(\xi_1\frac{\partial\widehat{u}}{\partial\lambda_1} + \xi_2\frac{\partial\widehat{u}}{\partial\lambda_2} + \xi_3\frac{\partial\widehat{u}}{\partial\lambda_3}\Big)^2\Big)d\Delta \end{aligned}$$

$$= \frac{1}{4}\int_T \Big(\frac{\partial \widehat{u}}{\partial \lambda_1}, \frac{\partial \widehat{u}}{\partial \lambda_2}, \frac{\partial \widehat{u}}{\partial \lambda_3}\Big) \begin{pmatrix} -r_2 - r_3 & r_3 & r_2 \\ r_3 & -r_3 - r_1 & r_1 \\ r_2 & r_1 & -r_1 - r_2 \end{pmatrix} \begin{pmatrix} \dfrac{\partial \widehat{u}}{\partial \lambda_1} \\ \dfrac{\partial \widehat{u}}{\partial \lambda_2} \\ \dfrac{\partial \widehat{u}}{\partial \lambda_3} \end{pmatrix} d\Delta \tag{2.10}$$

Now we are in a position to state a main theorem about the contractivity of the interpolation operator.

Theorem 2.1.

$$\|I_2u - I_1u\|_T \le \sqrt{\frac{2}{3}}\|I_1u\|_T \tag{2.11}$$

$$\|I_2u - I_1u\|_T \le \sqrt{\frac{3}{4}}\|I_2u\|_T \tag{2.12}$$

Proof. It is not difficult to show that $\widehat{I_2u}$ can be represented as follows by the barycentric coordinates with respect to the triangle T,

$$\widehat{I_2u} = \sum_{i=1}^{3} u(P_i)\lambda_i^2 + \sum_{i=1}^{3}(4u(\overline{P_i}) - u(P_{i-1}) - u(P_{i+1}))\lambda_{i-1}\lambda_{i+1} \tag{2.13}$$

From (2.10) we can obtain the following, after some tedious computation,

$$\begin{aligned}\|I_2u\|_T^2 =& \frac{1}{4}\int_{\widehat{T}} \Big(\frac{\partial \widehat{I_2u}}{\partial \lambda_1}, \frac{\partial \widehat{I_2u}}{\partial \lambda_2}, \frac{\partial \widehat{I_2u}}{\partial \lambda_3}\Big) \begin{pmatrix} -r_2 - r_3 & r_3 & r_2 \\ r_3 & -r_3 - r_1 & r_1 \\ r_2 & r_1 & -r_1 - r_2 \end{pmatrix} \begin{pmatrix} \dfrac{\partial \widehat{I_2u}}{\partial \lambda_1} \\ \dfrac{\partial \widehat{I_2u}}{\partial \lambda_2} \\ \dfrac{\partial \widehat{I_2u}}{\partial \lambda_3} \end{pmatrix} d\Delta \\ =& -\frac{1}{24}\sum_{i=1}^{3} r_i[3(u(P_{i-1}) - u(\widehat{P_i}))^2 + 3(u(P_{i+1}) - u(\widehat{P_i}))^2 \\ &+ 2(u(P_{i-1}) - u(\widehat{P_i}))(u(P_{i+1}) - u(\widehat{P_i})) + 8(u(\widehat{P_{i-1}}) - u(\widehat{P_{i+1}}))^2]\end{aligned}$$

Since $\|u\|_T^2 = \|u\|_{T_0}^2 + \|u\|_{T_1}^2 + \|u\|_{T_2}^2 + \|u\|_{T_3}^2$, we just need to compute $\|u\|_{T_j}$ ($j = 0, 1, 2, 3$), respectively, where u is piecewise defined on T. We first take computing $\|u\|_1$ as an example. On T_1, I_1u can be represented as follows in the barycentric coordinate (μ_1, μ_2, μ_3) with respect to T_1,

$$I_1u|_{T_1} = u(P_1)\mu_1 + u(\widehat{P_3})\mu_2 + u(\widehat{P_2})\mu_3$$

And also

$$\begin{aligned}I_2u =& u(P_1)\mu_1^2 + u(\widehat{P_3})\mu_2^2 + u(\widehat{P_2})\mu_3^2 + \Big(\frac{1}{2}u(P_1) - \frac{1}{2}u(P_2) + 2u(\widehat{P_3})\Big)\mu_1\mu_2 \\ &+ \Big(\frac{1}{2}u(P_1) - \frac{1}{2}u(P_3) + 2u(\widehat{P_2})\Big)\mu_1\mu_3 + \Big(u(\widehat{P_1}) + u(\widehat{P_2}) \\ &+ u(\widehat{P_3}) - \frac{1}{2}u(P_2) - \frac{1}{2}u(P_3)\Big)\mu_2\mu_3\end{aligned}$$

Then we have that

$$\|I_1u\|_{T_1}^2 = -\frac{1}{8}r_1(u(\widehat{P_2}) - u(\widehat{P_3}))^2 - \frac{1}{8}r_2(u(\widehat{P_2}) - u(P_1))^2 - \frac{1}{8}r_3(u(\widehat{P_3}) - u(P_1))^2 \tag{2.14}$$

and

$$\begin{aligned}\|I_2u - I_1u\|_{T_1}^2 =& -\frac{1}{96}\sum_{i=1}^{3} r_i[(u(P_{i-1}) - u(\widehat{P_i}) + u(\widehat{P_{i-1}}) - u(\widehat{P_{i+1}}))^2 \\ &+ (u(P_{i+1}) - u(\widehat{P_i}) + u(\widehat{P_{i+1}}) - u(\widehat{P_{i-1}}))^2] \triangleq A\end{aligned}$$

Similarly, we can also obtain the following results

$$\|I_1u\|_{T_2}^2 = -\frac{1}{8}r_2(u(\widehat{P_3}) - u(\widehat{P_1}))^2 - \frac{1}{8}r_1(u(\widehat{P_1}) - u(P_2))^2 - \frac{1}{8}r_3(u(\widehat{P_3}) - u(P_2))^2 \tag{2.15}$$

$$\|I_1u\|_{T_3}^2 = -\frac{1}{8}r_3(u(\widehat{P_1}) - u(\widehat{P_2}))^2 - \frac{1}{8}r_2(u(\widehat{P_2}) - u(P_3))^2 - \frac{1}{8}r_1(u(\widehat{P_1}) - u(P_3))^2 \tag{2.16}$$

$$\|I_1u\|_{T_0}^2 = -\frac{1}{8}r_1(u(\widehat{P_2}) - u(\widehat{P_3}))^2 - \frac{1}{8}r_2(u(\widehat{P_3}) - u(\widehat{P_1}))^2 - \frac{1}{8}r_3(u(\widehat{P_1}) - u(\widehat{P_2}))^2 \tag{2.17}$$

and

$$\|I_2u - I_1u\|_{T_j}^2 = A \quad (j = 0, 1, 2, 3) \tag{2.18}$$

Thus with (2.14), (2.15)–(2.18) we can write

$$\|I_1u\|_T^2 = -\sum_{i=1}^{3}\frac{1}{8}r_i[2(u(\widehat{P_{i-1}}) - u(\widehat{P_{i+1}}))^2 + (u(\widehat{P_i}) - u(P_{i-1}))^2 + (u(\widehat{P_i}) - u(P_{i+1}))^2]$$

and

$$\begin{aligned}\|I_2u - I_1u\|_T^2 =& 4A = -\frac{1}{24}\sum_{i=1}^{3} r_i[(u(P_{i-1}) - u(\widehat{P_i}) + u(\widehat{P_{i-1}}) - u(\widehat{P_{i+1}}))^2 \\ &+ (u(P_{i+1}) - u(\widehat{P_i}) + u(\widehat{P_{i+1}}) - u(\widehat{P_{i-1}}))^2]\end{aligned}$$

Let

$$\begin{aligned}A_i =& \frac{1}{24}[(u(P_{i-1}) - u(\widehat{P_i}) + u(\widehat{P_{i-1}}) - u(\widehat{P_{i+1}}))^2 \\ &+ (u(P_{i+1}) - u(\widehat{P_i}) + u(\widehat{P_{i+1}}) - u(\widehat{P_{i-1}}))^2] \\ B_i =& \frac{1}{8}[2(u(\widehat{P_{i-1}}) - u(\widehat{P_{i+1}}))^2 + (u(\widehat{P_i}) - u(P_{i-1}))^2 + (u(\widehat{P_i}) - u(P_{i+1}))^2] \\ C_i =& \frac{1}{24}[3(u(P_{i-1}) - u(\widehat{P_i}))^2 + 3(u(P_{i+1}) - u(\widehat{P_i}))^2 \\ &+ 2(u(P_{i-1}) - u(\widehat{P_i}))(u(P_{i+1}) - u(\widehat{P_i})) + 8(u(\widehat{P_{i-1}}) - u(\widehat{P_{i+1}}))^2],\end{aligned}$$

then we have the equalities $\|I_2u - I_1u\|_T^2 = -\sum_{i=1}^{3} r_iA_i$, $\|I_1u\|_T^2 = -\sum_{i=1}^{3} r_iB_i$, $\|I_2u\|_T^2 = -\sum_{i=1}^{3} r_iC_i$. For $i = 1, 2, 3$, we can obtain the following relations by using the Cauchy inequality

$$\begin{aligned}
A_i =& \frac{1}{24}[(u(P_{i-1}) - u(\widehat{P_i}))^2 + (u(P_{i+1}) - u(\widehat{P_i}))^2 + 2(u(\widehat{P_{i-1}}) - u(\widehat{P_{i+1}}))^2 \\
&+ 2(u(P_{i-1}) - u(\widehat{P_i}))(u(\widehat{P_{i-1}}) - u(\widehat{P_{i+1}})) + 2(u(P_{i+1}) - u(\widehat{P_i}))(u(\widehat{P_{i-1}}) - u(\widehat{P_{i+1}}))] \\
\le& \frac{1}{12}[2(u(\widehat{P_{i-1}}) - u(\widehat{P_{i+1}}))^2 + (u(\widehat{P_i}) - u(P_{i-1}))^2 + (u(\widehat{P_i}) - u(P_{i+1}))^2] = \frac{2}{3}B_i
\end{aligned}$$

On the other hand, it is obvious that $r_i \le 0$, hence we can derive the following inequality

$$\|I_2u - I_1u\|_T^2 = -\sum_{i=1}^{3} r_iA_i \le -\sum_{i=1}^{3} r_i\frac{2}{3}B_i = \frac{2}{3}\|I_1u\|_T^2$$

That is the first part of Theorem 2.1. Similarly, we have

$$\begin{aligned}
C_i =& \frac{1}{24}[(u(P_{i-1}) - u(P_{i+1}))^2 + 2(u(P_{i-1}) - u(\widehat{P_i}))^2 + 2(u(P_{i+1}) - u(\widehat{P_i}))^2 \\
&+ 4(u(P_{i-1}) - u(\widehat{P_i}))(u(P_{i+1}) - u(\widehat{P_i})) + 8(u(\widehat{P_{i-1}}) - u(\widehat{P_{i+1}}))^2] \\
=& \frac{1}{24}[(u(P_{i-1}) - u(P_{i+1}))^2 + 8(u(\widehat{P_{i-1}}) - u(\widehat{P_{i+1}}))^2 + 2(u(P_{i-1}) + u(P_{i+1}) - 2u(\widehat{P_i}))^2] \\
=& \frac{1}{24}\Big[\frac{1}{2}(u(P_{i-1}) - u(P_{i+1}))^2 + \frac{1}{2}(u(P_{i+1}) - u(P_{i-1}))^2 \\
&+ (u(P_{i-1}) + u(P_{i+1}) - 2u(\widehat{P_i}) + 2u(\widehat{P_{i-1}}) - 2u(\widehat{P_{i+1}}))^2 \\
&+ (u(P_{i+1}) + u(P_{i-1}) - 2u(\widehat{P_i}) + 2u(\widehat{P_{i+1}}) - 2u(\widehat{P_{i-1}}))^2\Big] \\
=& \frac{1}{24}\Big[\frac{3}{2}(u(P_{i-1}) - u(P_{i+1}))^2 + 4(u(P_{i-1}) - u(\widehat{P_i}) + u(\widehat{P_{i-1}}) - u(\widehat{P_{i+1}}))^2 \\
&+ 4(u(P_{i+1}) - u(P_{i-1}))(u(P_{i-1}) - u(\widehat{P_i}) + u(\widehat{P_{i-1}}) - u(\widehat{P_{i+1}})) \\
&\frac{3}{2}(u(P_{i-1}) - u(P_{i+1}))^2 + 4(u(P_{i+1}) - u(\widehat{P_i}) + u(\widehat{P_{i+1}}) - u(\widehat{P_{i-1}}))^2 \\
&+ 4(u(P_{i-1}) - u(P_{i+1}))(u(P_{i+1}) - u(\widehat{P_i}) + u(\widehat{P_{i+1}}) - u(\widehat{P_{i-1}}))\Big] \\
=& \frac{1}{24}\Big[\Big(\sqrt{\frac{3}{2}}(u(P_{i+1}) - u(P_{i-1})) + 2\sqrt{\frac{2}{3}}(u(P_{i-1}) - u(\widehat{P_i}) + u(\widehat{P_{i-1}}) - u(\widehat{P_{i+1}}))\Big)^2 \\
&+ \Big(\sqrt{\frac{3}{2}}(u(P_{i-1}) - u(P_{i+1})) + 2\sqrt{\frac{2}{3}}(u(P_{i+1}) - u(\widehat{P_i}) + u(\widehat{P_{i+1}}) - u(\widehat{P_{i-1}}))\Big)^2 \\
&+ \frac{4}{3}(u(P_{i-1}) - u(\widehat{P_i}) + u(\widehat{P_{i-1}}) - u(\widehat{P_{i+1}}))^2 \\
&+ \frac{4}{3}(u(P_{i+1}) - u(\widehat{P_i}) + u(\widehat{P_{i+1}}) - u(\widehat{P_{i-1}}))^2\Big] \\
\ge& \frac{4}{3}A_i.
\end{aligned}$$

Hence $\|I_2u - I_1u\|_T^2 = -\sum_{i=1}^{3} r_iA_i \leq -\sum_{i=1}^{3} r_i\frac{3}{4}C_i = \frac{3}{4}\|I_2u\|_T^2$.
This completes the proof of the second inequality. With similar method, we can also prove the following conclusions.

Theorem 2.2. $\|I_1u\|_T \leq \sqrt{\frac{3}{2}}\|I_2u\|_T$ and $\|I_2u\|_T \leq \sqrt{\frac{4}{3}}\|I_1u\|_T$

3. Linear Elliptic Problem

Let Ω be a bounded, convex polygonal domain in R^2 with the boundary $\partial\Omega$. We consider the elliptic model problem

$$\begin{cases} \text{Find } u \in H_0^1(\Omega), \text{ such that} \\ a(u,v) = (f,v), \quad \forall\ v \in H_0^1(\Omega) \end{cases} \tag{3.1}$$

where $a(u,v) = \int_\Omega \nabla u \nabla v dxdy$, $(f,v) = \int_\Omega fvdxdy$ and f sufficiently smooth.

To define a finite element method, we need a partition of Ω into elements E, for example triangles in this paper. Let the triangulation $\mathcal{T}_2$ be regular with maximum diameter $2h$. Another triangulation $\mathcal{T}_1$ can be obtained by connecting the middle points of edges for each triangle of $\mathcal{T}_2$. Let V_i $(i = 1, 2)$ be the finite element space composed of piecewise polynomials of order i defined on $\mathcal{T}_i$. Thus $V_i \subset H^1(\Omega)$, and denote $V_i^0 = V_i \bigcap H_0^1(\Omega)$. Define $I_i : C^0 \longrightarrow V_i$, i.e., piecewise interpolation operator of order i on the all vertices of triangles of $\mathcal{T}_1$.

This paper formulates the following Petrov-Galerkin approximation model

$$\begin{cases} \text{Find } u_h \in V_2^0 \text{ such that} \\ a(u_h,v) = (f,v), \ \forall v \in V_1^0 \end{cases} \tag{3.2}$$

Theorem 3.1. *Let u be the exact solution of* (3.1)*, then there exists exactly one finite element solution u_h for* (3.2) *and there hold the estimates*

$$\|u - u_h\|_a \leq C\|u - I_2u\|_a \tag{3.3}$$

$$\|u - u_h\| \leq C \cdot h\|u - I_2u\|_a \tag{3.4}$$

where $\|u\|^2 = \int_\Omega u^2dxdy$ *and* $\|u\|_a^2 = \int_\Omega |\nabla u|^2dxdy$.

Proof. The variational problem (3.2) is equivalent to the following one

$$\begin{cases} \text{Find } u_h \in V_2^0 \text{ such that} \\ a(u_h,I_1v) = (f,I_1v), \ \forall v \in V_2^0 \end{cases} \tag{3.5}$$

From (2.12) of Theorem 2.1 we have

$$\begin{aligned} a(v,I_1v) =& a(v,v) - a(v,v - I_1v) \geq \|v\|_a^2 - \|v\|_a\|v - I_1v\|_a \\ \geq& \|v\|_a^2 - \sqrt{\frac{3}{4}}\|v\|_a^2 = \Big(1 - \sqrt{\frac{3}{4}}\Big)\|v\|_a^2, \quad \forall v \in V_2^0 \end{aligned} \tag{3.6}$$

This means that the bilinear functional $a(v, I_1 v)$ defined on V_2^0 is positive definite. Obviously $a(v, I_1 v)$ is a continuous and $(f, I_1 v)$ is continuous linear form on V_2^0. Hence there exists only one solution u_h for the variational problem (3.2) by Lax-Milgram Lemma. On the other hand, it can be seen from (3.1) and (3.5) that

$$a(u - u_h, I_1 v) = 0, \ \forall v \in V_2^0$$

With the above equality and (3.6) and Theorem 2.2 the following estimates can be achieved

$$\begin{aligned} \|I_2 u - u_h\|_a^2 \le & C \cdot a(I_2 u - u_h, I_1(I_2 u - u_h)) = C \cdot a(I_2 u - u, I_1(I_2 u - u_h)) \\ \le & C \cdot \|I_2 u - u\|_a \cdot \|I_2 u - u_h\|_a \end{aligned}$$

That completes the proof for (3.3).

At last by using Nitsche's technique (3.4) can be proved. This completes the proof of Theorem 2.1.

The iterated defect correction scheme established in [1] and [2] is

$$\begin{cases} \text{Find } u_{i+1}^h \in V_1^0 \\ a(u_{i+1}^h, v) = a(u_i^h, v) - \{a(I_2 u_i^h, v) - (f, v)\}, \ \forall v \in V_1^0 \end{cases} \tag{3.7}$$

In this scheme u_0^h can be taken as the following linear finite element solution

$$\begin{cases} \text{Find } u_0^h \in V_1^0 \\ a(u_0^h, v) = (f, v), \quad \forall v \in V_1^0 \end{cases} \tag{3.8}$$

It can be seen that the coefficient matrices in the linear systems (3.7) and (3.8) are same. After calculating u_0^h by (3.8) the first iterated solution u_1^h can be obtained from (3.7). Many researches[2−5] are dedicated to estimating errors of u_1^h as well as higher order interpolation of u_1^h approximating to the exact solution. It is the goal of this paper to prove that for any chosen initial approximation solution $u_0^h \in V_1^0$ the quadratic interpolation $I_2 u_{i+1}^h$ of the iterated solution u_{i+1}^h of (3.7) converges geometrically to the Petrov-Galerkin approximation solution u_h of (3.2).

Theorem 3.2. $\forall u_0^h \in V_1^0$, *the iterative scheme* (3.7) *is convergent and the following estimation is valid.*

$$\|I_2 u_i^h - u_h\|_a \le \frac{1}{1 - \sqrt{\dfrac{3}{4}}} \Big(\sqrt{\frac{2}{3}}\Big)^i \|u_1^h - u_0^h\|_a \tag{3.9}$$

Proof. By (3.7), for any $v \in V_1^0$,

$$a(u_i^h, v) = a(u_{i-1}^h, v) - \{a(I_2 u_{i-1}^h, v) - (f, v)\} \tag{3.10}$$

Then, from (3.7) and (3.10),

$$a(u_{i+1}^h - u_i^h, v) = a(u_i^h - u_{i-1}^h, v) - a(I_2 u_i^h - I_2 u_{i-1}^h, v)$$

$$= a((I - I_2)(u_i^h - u_{i-1}^h), v)$$

Taking $v = u_{i+1}^h - u_i^h$, by using (2.11), it results in

$$\begin{aligned}\|u_{i+1}^h - u_i^h\|_a^2 &\leq \|(I - I_2)(u_i^h - u_{i-1}^h)\|_a \|u_{i+1}^h - u_i^h\|_a \\ &\leq \sqrt{\frac{2}{3}}\|u_i^h - u_{i-1}^h\|_a \cdot \|u_{i+1}^h - u_i^h\|_a\end{aligned}$$

Hence we have

$$\|u_{i+1}^h - u_i^h\|_a \leq \sqrt{\frac{2}{3}}\|u_i^h - u_{i-1}^h\|_a \leq \Big(\sqrt{\frac{2}{3}}\Big)^i \|u_1^h - u_0^h\|_a$$

Substituting (3.2) into (3.7) results in $a(u_{i+1}^h - u_i^h, v) = a(u_h - I_2 u_i^h, v)$, $\forall v \in V_1^0$. Let $P : H_0^1(\Omega) \longrightarrow V_1^0$ denote the Ritz projection operator, then

$$\begin{aligned}&u_{i+1}^h - u_i^h = P(u_h - I_2 u_i^h) \\ &u_h - I_2 u_i^h = u_h - I_2 u_i^h - P(u_h - I_2 u_i^h) + u_{i+1}^h - u_i^h\end{aligned}$$

It can be seen, from the property of the orthogonal projection operator and (2.12), that

$$\begin{aligned}\|u_h - I_2 u_i^h\|_a &\leq \|u_h - I_2 u_i^h - P(u_h - I_2 u_i^h)\|_a + \|u_{i+1}^h - u_i^h\|_a \\ &\leq \|u_h - I_2 u_i^h - I_1(u_h - I_2 u_i^h)\|_a + \|u_{i+1}^h - u_i^h\|_a \\ &\leq \sqrt{\frac{3}{4}}\|u_h - I_2 u_i^h\|_a + \|u_{i+1}^h - u_i^h\|_a\end{aligned}$$

Finally one can write the following inequality

$$\|u_h - I_2 u_i^h\|_a \leq \frac{1}{1 - \sqrt{\frac{3}{4}}}\|u_{i+1}^h - u_i^h\|_a \leq \frac{1}{1 - \sqrt{\frac{3}{4}}}\Big(\sqrt{\frac{2}{3}}\Big)^i \|u_1^h - u_0^h\|_a$$

This completes the proof.

4. The semi-linear elliptic boundary problem

Consider the semi-linear elliptic boundary problem

$$\begin{cases} -\triangle u = f(z, u), & \text{in } \Omega \\ u = 0, & \text{on } \partial\Omega \end{cases} \tag{4.1}$$

where $z = (x, y)$, $f(z, u)$ and $f_u(z, u)$ are respectively continuous on the domain $\Omega \times (-\infty, \infty)$, and for each $(z, u) \in \Omega \times (-\infty, \infty)$

$$|f_u(z, u)| \leq \lambda < \Lambda \equiv \inf_{w \in W_{1,2}^0(\Omega)} \frac{a(w, w)}{(w, w)} \tag{4.2}$$

The weak form of (4.1) is

$$\begin{cases} \text{Find } u \in H_0^1(\Omega), \text{ such that} \\ a(u,v) = (f(z,u),v), \quad \forall v \in H_0^1(\Omega) \end{cases} \tag{4.3}$$

Similar to (3.2), define corresponding Petrov-Galerkin approximation as

$$\begin{cases} \text{Find } u_h \in V_2^0, \text{ such that} \\ a(u_h,v) = (f(z,u_h),v), \quad \forall v \in V_1^0 \end{cases} \tag{4.3}$$

and the iterated defect correction scheme of (4.3) as ([1])

$$\begin{cases} \text{Find } u_{i+1}^h \in V_1^0, \text{ such that} \\ a(u_{i+1}^h,v) = a(u_i^h,v) - \{a(I_2u_i^h,v) - (f(z,I_2u_i^h),v)\}, \quad \forall v \in V_1^0 \end{cases} \tag{4.4}$$

Assumed (4.2) satisfied, then there exists only one solution u, respectively, for (4.1) and (4.3). In this section we will use the mean value formula in the following form many times

$$f(z,w) - f(z,v) = f_u(z, \theta w + (1-\theta)v)(w-v)$$

where $\theta \in (0,1)$ is a function of z and w, v. For simplicity, abbreviate the above formula by

$$f(z,w) - f(z,v) = f_u(w-v)$$

and $f(z,u)$ by $f(u)$.

Theorem 4.1. *Assumed* (4.2) *satisfied and* $\gamma_1 = \sqrt{\frac{3}{4}} + \lambda\Lambda^{-1}\sqrt{\frac{3}{2}} < 1$, *then* (4.4) *exists only one solution* u_h *and the following estimates are valid*

$$\|u - u_h\|_a \le C\|I_2u - u\|_a \tag{4.6}$$

$$\|u - u_h\| \le C \cdot h\|I_2u - u\| \tag{4.7}$$

Proof. For any $w, v \in V_2^0$, from (3.6) and (4.2), we obtain

$$\begin{aligned} a(w-v, I_1(w-v)) - (f(w)-f(v), I_1(w-v)) &\ge \Big(1 - \sqrt{\frac{3}{4}}\Big)\|w-v\|_a^2 \\ - \|f(w)-f(v)\|\|I_1(w-v)\| &\ge \Big(1 - \sqrt{\frac{3}{4}}\Big)\|w-v\|_a^2 \\ - \|f_u(w-v)\|\|I_1(w-v)\| &\ge \Big(1 - \sqrt{\frac{3}{4}}\Big)\|w-v\|_a^2 \\ - \lambda\Lambda^{-1}\sqrt{\frac{3}{2}}\|w-v\|_a^2 &= \Big(1 - \sqrt{\frac{3}{4}} - \lambda\Lambda^{-1}\sqrt{\frac{3}{2}}\Big)\|w-v\|_a^2 \\ =&(1-\gamma_1)\|w-v\|_a^2 \end{aligned} \tag{4.8}$$

That is, the strong elliptic condition is satisfied. On the other hand, for any $w_1, w_2, v \in V_2^0$,

$$|a(w_1 - w_2, I_1v) - (f(w_1) - f(w_2), I_1v)| \le C\|w_1 - w_2\|_a \cdot \|I_1v\|_a + \|f_u(w_1 - w_2)\| \cdot \|I_1v\|$$

$$\leq C \cdot \|w_1 - w_2\|_a \cdot \|v\|_a \tag{4.9}$$

That is, the continuity condition is also satisfied. Hence by Lax-Milgram in the nonlinear form (4.4) has one solution.

Also based on (4.3) and (4.4) it can be seen that

$$a(u - u_h, v) - (f(u) - f(u_h), v) = 0, \quad \forall v \in V_1^0 \tag{4.10}$$

It follows from (4.8) and (4.10)

$$\begin{aligned}(1-\gamma_1)\|I_2u - u_h\|_a^2 \leq& a(I_2u - u_h, I_1(I_2u - u_h)) - (f(I_2u) - f(u_h), I_1(I_2u - u_h))\\ =& a(I_2u - u, I_1(I_2u - u_h)) - (f(I_2u) - f(u), I_1(I_2u - u_h))\\ &+ a(u - u_h, I_1(I_2u - u_h)) - (f(u) - f(u_h), I_1(I_2u - u_h))\\ =& a(I_2u - u, I_1(I_2u - u_h)) - (f(I_2u) - f(u_h), I_1(I_2u - u_h))\\ \leq& C\|I_2u - u\|_a \cdot \|I_2u - u_h\|_a + \lambda\|I_2u - u\| \cdot \|I_1(I_2u - u_h)\|\\ \leq& C\|I_2u - u\|_a \cdot \|I_2u - u_h\|_a \end{aligned} \tag{4.11}$$

Thus (4.6) is derived from the above inequality.

By using the mean value theorem, from (4.10) we derive

$$a(u - u_h, v) - (f_u(u - u_h), v) = 0, \quad \forall v \in V_1^0 \tag{4.12}$$

Let $\Psi \equiv \dfrac{1}{\|u - u_h\|}(u - u_h)$ and $\varphi \in H_0^1(\Omega)$ be the exact solution of the following linear problem

$$a(v, \varphi) + (f_u \cdot v, \varphi) = (v, \Psi), \quad \forall v \in H_0^1(\Omega) \tag*{(4.13.)}$$

Then taking $v = u - u_h$ we have, by using of (4.12),

$$\begin{aligned}\|u - u_h\| &= a(u - u_h, \varphi) + (f_u \cdot (u - u_h), \varphi)\\ &= a(u - u_h, \varphi - I_1\varphi) + (f_u \cdot (u - u_h), \varphi - I_1\varphi)\\ &\leq C \cdot h\|u - u_h\|_a \cdot \|\varphi\|_2\end{aligned} \tag{4.14}$$

Let us rewrite equation (4.13) as follows

$$a(v, \varphi) = (v, \Psi) - (f_u \cdot v, \varphi), \quad \forall v \in H_0^1(\Omega).$$

Based on the priori estimate of generalized solution of the boundary problem we have

$$\|\varphi\|_2 \leq C \cdot \|f_u\varphi + \Psi\| \tag{4.15}$$

Moreover with the positive definiteness of $a(\cdot, \cdot)$ and (4.2), it can be derived from (4.13)

$$\|\varphi\|_a \leq C \cdot \|\Psi\|$$

Finally from this inequality and (4.15) we can obtain

$$\|\varphi\|_2 \leq C \cdot \|\Psi\| = C.$$

Subsituting the above inequality into (4.14) results in the final estimate (4.7).

Theorem 4.2. *Suppose that the hypotheses of Theorem* 4.1 *are satisfied and that* $\gamma_2 = \sqrt{\frac{2}{3}} + \lambda\Lambda^{-1}\sqrt{\frac{4}{3}} < 1$, *then the interpolation* $I_2 u_{i+1}^h$ *of the* $(i+1)$*th iterated solution* u_{i+1}^h *converges to the Petrov-Galerkin solution* u_h *of* (4.4) *and the corrected solution obeys the estimate*

$$\|I_2 u_{i+1}^h - u_h\|_a \le \sqrt{2}\cdot\gamma_2^{i+1}\|I_2 u_0^h - u_h\|_a \tag{4.16}$$

Proof. From (4.5) it is obvious that

$$\begin{aligned} a(u_{i+1}^h - u_i^h, v) &= a(u_i^h - u_{i-1}^h, v) - \{a(I_2u_i^h - I_2u_{i-1}^h, v) - (f(I_2u_i^h) - f(I_2u_{i-1}^h), v)\} \\ &\le a(u_i^h - u_{i-1}^h, v) - a(I_2u_i^h - I_2u_{i-1}^h, v) + (f_u'(I_2u_i^h - I_2u_{i-1}^h), v) \\ &\le \|u_i^h - u_{i-1}^h - I_2(u_i^h - u_{i-1}^h)\|_a \cdot \|v\|_a + \lambda\|I_2(u_i^h - u_{i-1}^h)\| \cdot \|v\| \\ &\le \sqrt{\frac{2}{3}}\|u_i^h - u_{i-1}^h\|_a\|v\|_a + \lambda\Lambda^{-1}\|I_2(u_i^h - u_{i-1}^h)\|_a \cdot \|v\|_a \\ &\le \Big(\sqrt{\frac{2}{3}} + \lambda\Lambda^{-1}\sqrt{\frac{4}{3}}\Big)\Big\|u_i^h - u_{i-1}^h\Big\|_a\|v\|_a \end{aligned}$$

Take $v = u_{i+1}^h - u_i^h$, then

$$\|u_{i+1}^h - u_i^h\|_a \le \gamma_2\|u_i^h - u_{i-1}^h\|_a \le \cdots \le \gamma_2^i\|u_1^h - u_0^h\|_a$$

As $\gamma_2 < 1$, it can be seen that $\{u_i^h\}$ is a Cauchy sequence of V_1^0. Hence there exists $\widehat{u} \in V_1^0$ such that $\|u_i^h - \widehat{u}\|_a \longrightarrow 0$, for $i \longrightarrow \infty$. This derives

$$\|I_2u_i^h - I_2\widehat{u}\|_a \le \sqrt{\frac{4}{3}}\|u_i^h - \widehat{u}\|_a \longrightarrow 0, \quad \text{for } i \longrightarrow \infty.$$

Take $i \longrightarrow \infty$ in (4.5), then

$$a(\widehat{u}, v) = a(\widehat{u}, v) - \{a(I_2\widehat{u}, v) - (f(I_2\widehat{u}), v)\} \tag{4.17}$$

That is

$$a(I_2\widehat{u}, v) = (f(I_2\widehat{u}), v), \quad \forall v \in V_1^0.$$

Consequently $I_2\widehat{u}$ is a solution of (4.4) and $I_2\widehat{u} = u_h$ with uniqueness. This completes the proof of first part of Theorem 4.2. On the other hand with (4.5)-(4.17) we have

$$\begin{aligned} a(u_{i+1}^h - \widehat{u}, v) &= a(u_i^h - \widehat{u}, v) - \{a(I_2u_i^h - I_2\widehat{u}, v) - (f(I_2u_i^h) - f(I_2\widehat{u}), v)\} \\ &\le a(u_i^h - \widehat{u} - I_2(u_i^h - \widehat{u}), v) + \|f(I_2u_i^h) - f(I_2\widehat{u})\|\ \|v\| \\ &\le \|u_i^h - \widehat{u} - I_2(u_i^h - \widehat{u})\|_a \cdot \|v\|_a + \|f_u'(I_2u_i^h - I_2\widehat{u})\| \cdot \|v\| \\ &\le \sqrt{\frac{2}{3}}\|u_i^h - \widehat{u}\|_a\|v\|_a + \lambda\Lambda^{-1}\|I_2(u_i^h - \widehat{u})\|_a \cdot \|v\|_a \\ &\le \Big(\sqrt{\frac{2}{3}} + \lambda\Lambda^{-1}\sqrt{\frac{4}{3}}\Big)\|u_i^h - \widehat{u}\|_a\|v\|_a \le \gamma_2\|u_i^h - \widehat{u}\|_a\|v\|_a \end{aligned}$$

Write $v = u_{i+1}^h - \widehat{u}$, then

$$\|u_{i+1}^h - \widehat{u}\|_a \leq \gamma_2 \|u_i^h - \widehat{u}\|_a \leq \cdots \leq \gamma_2^{i+1} \|u_0^h - \widehat{u}\|_a$$

and by Theorem 2.2

$$\|I_2 u_{i+1}^h - I_2 \widehat{u}\|_a \leq \sqrt{\frac{4}{3}} \gamma_2^{i+1} \|I_1 I_2 (u_0^h - \widehat{u})\|_a \leq \sqrt{2} \gamma_2^{i+1} \|I_2 u_0^h - I_2 \widehat{u}\|_a$$

This completes the proof.

Remark 1. When Ω is a concave polygonal domain, the all results except for (2.12) of Theorem 3.1 and (4.7) of Theorem 4.1 are valid with some remedy.

Remark 2. Let

$$a(u, v) = \int_\Omega K(x, y) \nabla u(x, y) \nabla v(x, y) dxdy$$

where $K(x, y)$ is continuous and $K(x, y) \geq \delta > \dot{0}$, then it is easily proved that

$$\int_\Omega K[((I_2 u - I_1 u)'_x)^2 + ((I_2 u - I_1 u)'_y)^2] dxdy \leq K_h \int_\Omega K[((I_1 u)'_x)^2 + ((I_1 u)'_y)^2] dxdy$$

and

$$\int_\Omega K[((I_2 u - I_1 u)'_x)^2 + ((I_2 u - I_1 u)'_y)^2] dxdy \leq K^h \int_\Omega K[((I_2 u)'_x)^2 + ((I_2 u)'_y)^2] dxdy$$

where $\lim\limits_{h \to 0} K_h = \dfrac{2}{3}$, $\lim\limits_{h \to 0} K^h = \dfrac{3}{4}$.

Hence while h is sufficiently small, the theorems of this paper are valid for $a(u, v) = \int_\Omega K(x, y) \nabla u(x, y) \nabla v(x, y) dxdy$.

References

[1] R. Frank, J. Hertling, J.P. Monnet, The application of iterated defect correction to variational methods for elliptic boundary value problems, *Computing*, **30** (1983), 121–135.

[2] H. Blum, Asymptotic Error Expansion and Defect Correction in the Finite Element Method, Habilitation-Schrift, Universität Heidelberg 1990.

[3] R. Rannacher, Defect correction techniques in the finite element method, *Metz Days on Numerical Analysis*, Univ. Metz., June 1990.

[4] Q. Lin, Y.D. Yang, Interpolation and correction of finite elements, *Math. in Practice and Theory*, No.3 (1991), 29–33.

[5] Q. Lin, A.H. Zhou, Defect correction for finite element gradient, *Syst. Sci. and Math. Sci.,* **5** : 3 (1992), 278–288.

[6] Y.D. Yang, Correction method of the finite element for quasilinear elliptic boundary value problems, *Math. Numer. Sinica*, **14** (1992), 467–471.

[7] J.W. Barrett, G. Moore, Optimal recovery in the finite element method, Part 2: Defect correction for ordinary differential equations, *IMA J. Numer. Anal.,* **8** (1988), 527–540.

[8] M.A. Krasnosel'sk etc., Approximation Solution of Operator Equations, Moscow Press, 1969 (in Russian).