The Exponentially Fitted Finite Volume Method Revisited

Song Wang
The University of Western Australia

mailto:swang@maths.uwa.edu.au
Outline

● The continuous problem (linear convection-diffusion equation with a singular perturbation parameter $\varepsilon$).
Outline

- The continuous problem (linear convection-diffusion equation with a singular perturbation parameter $\varepsilon$).
- Formulation of the finite volume method (FVM)
Outline

- The continuous problem (linear convection-diffusion equation with a singular perturbation parameter $\varepsilon$).
- Formulation of the finite volume method (FVM)
- A priori estimates for the solution and derivatives
Outline

- The continuous problem (linear convection-diffusion equation with a singular perturbation parameter $\varepsilon$).
- Formulation of the finite volume method (FVM)
- A priori estimates for the solution and derivatives
- Re-formulation of the FVM (based on early work by Miller and Wang)
Outline

- The continuous problem (linear convection-diffusion equation with a singular perturbation parameter \( \varepsilon \)).
- Formulation of the finite volume method (FVM)
- A priori estimates for the solution and derivatives
- Re-formulation of the FVM (based on early work by Miller and Wang)
- Coercivity analysis – in a discrete energy norm
Outline

- The continuous problem (linear convection-diffusion equation with a singular perturbation parameter \( \varepsilon \)).
- Formulation of the finite volume method (FVM)
- A priori estimates for the solution and derivatives
- Re-formulation of the FVM (based on early work by Miller and Wang)
- Coercivity analysis – in a discrete energy norm
- Error estimates – almost uniform convergence \( O \left( h^{1/2} + h \left| \ln \varepsilon / \ln h \right|^{1/2} \right) \).
Consider stationary, linear, convection-diffusion problems of the form

\[- \nabla \cdot f + Gu = F \quad \text{in} \quad \Omega := (0, 1)^2, \quad (1)\]

\[f = \varepsilon \nabla u - au, \quad (2)\]

\[u|_{\partial \Omega} = 0, \quad (3)\]
Consider stationary, linear, convection-diffusion problems of the form

\[- \nabla \cdot f + Gu = F \quad \text{in} \quad \Omega := (0, 1)^2, \tag{1}\]

\[f = \varepsilon \nabla u - au, \tag{2}\]

\[u|_{\partial \Omega} = 0, \tag{3}\]

where

\(\partial \Omega\) denotes the boundary of \(\Omega\),

\(\varepsilon > 0\) is a positive parameter,

\(a = (a_1, a_2)\) is a known vector-valued function \((|a| \gg \varepsilon)\)

and

\(F\) is given function.
We assume that \( a \in (C^1(\bar{\Omega}))^2 \), \( G \in C(\Omega) \cap H^1(\Omega) \) and \( F \in L^2(\Omega) \).
We assume that $a \in (C^1(\bar{\Omega}))^2$, $G \in C(\Omega) \cap H^1(\Omega)$ and $F \in L^2(\Omega)$.

Also, $a$ satisfies
\[
\frac{1}{2} \nabla \cdot a + G \geq 0 \quad \text{in} \quad \Omega
\] (4)
and
\[
a_1 \geq \alpha_1 > 0, \quad a_2 \geq \alpha_2 > 0 \quad \text{in} \quad \Omega.
\] (5)
We assume that \( \mathbf{a} \in (C^1(\bar{\Omega}))^2 \), \( G \in C(\Omega) \cap H^1(\Omega) \) and \( F \in L^2(\Omega) \).

Also, \( \mathbf{a} \) satisfies
\[
\frac{1}{2} \nabla \cdot \mathbf{a} + G \geq 0 \quad \text{in} \quad \Omega
\]
and
\[
a_1 \geq \alpha_1 > 0, \quad a_2 \geq \alpha_2 > 0 \quad \text{in} \quad \Omega.
\]

In this case, the solution to the problem has two exponential boundary layers of width \( O(\varepsilon) \) at \( x = 1 \) and \( y = 1 \).
The variational problem

Problem 2.1: Find $u \in H_0^1(\Omega)$ such that for all $v \in H_0^1(\Omega)$

$$A(u, v) = (F, v),$$

where $A(\cdot, \cdot)$ is a bilinear form on $(H_0^1(\Omega))^2$ defined by

$$A(u, v) = (\varepsilon \nabla u - au, \nabla v).$$

It is easy to show that

$$A(u, u) = \|u\|_{\varepsilon}^2, \quad \forall u \in H_0^1(\Omega),$$

where

$$\|v\|_{\varepsilon}^2 = (\varepsilon \nabla v, \nabla v) + ((\frac{1}{2} \nabla \cdot a + G)v, v).$$

is a norm on $H_0^1(\Omega)$. 
Formulation of the FVM

Given a set of mesh vertices $X_h$ on $\bar{\Omega}$, let $T_h$ be mesh consisting of triangles and rectangles satisfying the circumcircle of each element contains no other vertices in $X_h$.

This mesh is called a Delaunay mesh.

Dirichlet tessellation associated with a node $x_i$:

$$d_i = \{ x \in \Omega : |x - x_i| < |x - x_j|, x_i \in X_h, j \neq i \}$$

All $d_i$ form a mesh dual to $T_h$. 
Figure 1: An example of Delaunay triangulation and Dirichlet tessellation.
Now consider a particular tile \( d_i \) (cf. the figure).

**Figure 2: A typical local structure of the meshes.**

Integrating the equation by parts over \( d_i \) we have

\[
- \int_{\partial d_i} f \cdot n \, ds + \int_{d_i} Gudx = \int_{d_i} f \, dx,
\]

\( n \) — the unit outward vector normal to \( \partial d_i \).
But $\partial d_i = \bigcup_{j \in I_i} l_{ij}$ we have

$$-\sum_{j \in I_i} \int_{l_{ij}} f \cdot e_{ij} ds + \int_{d_i} Gudx = \int_{d_i} f dx$$

where $e_{ij}$ is the direction from $x_i$ to $x_j$. On $l_{ij}$ we look for a constant approximation to the flux projection $f \cdot e_{ij}$ by solving

$$\varepsilon \frac{du}{de_{i,j}} - a_{ij} u = f_{ij}$$

$$u(x_i) = u_i, \quad u(x_j) = u_j$$

where $a_{ij}$ is a constant approximation to $a \cdot e_{ij}$. The analytical solution is

$$f \cdot e_{ij} \approx f_{ij} = \frac{\varepsilon}{|e_{ij}|} \left( B(\sigma_{ij})u_j - B(-\sigma_{ij})u_i \right)$$

where $\sigma_{i,j} = a_{ij}|e_{ij}|/\varepsilon$, $|\cdot|$ denotes the measure and $B(\cdot)$ is
the Bernoulli function defined by

\[ B(x) = \begin{cases} 
  \frac{x}{(e^x - 1)}, & x \neq 0, \\
  1, & x = 0. 
\end{cases} \]  

(9)

We also use approximations

\[ \int_{d_i} Gudx \approx G_i u_i |d_i|, \quad \int_{d_i} Fdx \approx F_i |d_i|. \]

Thus the discretised equation is

\[- \sum_{j \in I_i} \frac{\varepsilon |l_{ij}|}{|e_{ij}|} \left( B(\sigma_{ij}) u_j - B(-\sigma_{ij}) u_i \right) + G_i u_i |d_i| = F_i |d_i|\]
or in matrix form

\[ Au = b \]

with

\[
a_{ii} = \sum_{j \in I_i} \frac{\varepsilon |l_{ij}|}{|e_{ij}|} B(-\sigma_{ij}) + G_i |d_i|,
\]

\[
a_{ij} = -\frac{\varepsilon |l_{ij}|}{|e_{ij}|} B(\sigma_{ij}), \quad j \in I_i.
\]

All undefined entries are zeros.

It has been shown \( A \) is an \( M \)-matrix if \( G \geq 0 \).
Estimates on the solution and its derivatives

We divide the solution region $\Omega$ into two parts $\Omega_1$ and $\Omega_2$ given respectively by

\[
\Omega_1 = (0, 1 - \delta_1) \times (0, 1 - \delta_2),
\]

\[
\Omega_2 = (1 - \delta_1, 1) \times (0, 1) \cup (0, 1 - \delta_1) \times (1 - \delta_2, 1),
\]

with $\delta_1, \delta_2 \in (0, 1)$ (cf. Figure 3). Obviously $\overline{\Omega}_1 \cup \overline{\Omega}_2 = \overline{\Omega}$.

The region $\Omega_2$ consists of three subregions $\Omega_2^{(1)}, \Omega_2^{(2)}$ and $\Omega_2^{(3)}$ defined respectively by

\[
\Omega_2^{(1)} = (1 - \delta_1, 1) \times (0, 1 - \delta_2),
\]

\[
\Omega_2^{(2)} = (0, 1 - \delta_1) \times (1 - \delta_2, 1),
\]

\[
\Omega_2^{(3)} = (1 - \delta_1, 1) \times (1 - \delta_2, 1).
\]
Figure 3: Subdomains $\Omega_1$ and $\Omega_2 = \Omega_2^{(1)} \cup \Omega_2^{(2)} \cup \Omega_2^{(3)}$, $\Gamma = \Gamma_1 \cup \Gamma_2$. 
The choice of the transition parameters $\delta_1$ and $\delta_2$ is rather arbitrary, but it is required that $\Omega_2$ contains the boundary layers and $\delta_1, \delta_2 = O(\varepsilon)$. One choice is

$$
\delta_1 = \frac{\beta}{\alpha_1} \varepsilon \ln(1/\varepsilon) \quad \text{and} \quad \delta_2 = \frac{\beta}{\alpha_2} \varepsilon \ln(1/\varepsilon), \quad (10)
$$

where $\beta \geq 1$ is a positive constant.
Assumption 2.1: \( u \) has the representation

\[ u = U_1 + U_2 + U_3 + U_4, \]

such that

\[ ||U_1||_{k,\infty,\Omega} \leq C \quad \text{for} \quad k = 0, 1, 2, \]

and \( U_2, U_3 \) and \( U_4 \) satisfy

\[
\left| \frac{\partial^{i+j}U_2}{\partial x^i \partial y^j} \right| \leq C \varepsilon^{-i} \exp \left( -\frac{\alpha_1 (1-x)}{\varepsilon} \right),
\]

\[
\left| \frac{\partial^{i+j}U_3}{\partial x^i \partial y^j} \right| \leq C \varepsilon^{-j} \exp \left( -\frac{\alpha_2 (1-y)}{\varepsilon} \right),
\]

\[
\left| \frac{\partial^{i+j}U_4}{\partial x^i \partial y^j} \right| \leq C \varepsilon^{-(i+j)} \exp \left( -\frac{\alpha_1 (1-x)}{\varepsilon} \right) \exp \left( -\frac{\alpha_2 (1-y)}{\varepsilon} \right),
\]

for \( 0 \leq i + j \leq 2 \) and some positive constant \( C \).
Theorem 2.1: If $\beta \geq 3/2$, then the solution $u$ to Problem 2.1 satisfies

$$||u||_{i,\Omega_1} \leq C \quad i = 0, 1, 2,$$  \hspace{1cm} (12)

for some positive constant $C$, independent of $u$. Furthermore, if $\beta \geq 2$, then

$$||u||_{i,\infty,\Omega_1} \leq C \quad i = 0, 1, 2.$$  \hspace{1cm} (13)
Proof: Let $C$ be a generic positive constant independent of $u$ and $\varepsilon$. We only show that $\|u\|_{2, \Omega_1} \leq C$.

From the definition of $\Omega_1$ and (11) we have

$$\|\partial^2 U_2 / \partial x^2\|_{0, \Omega_1}^2 \leq C\varepsilon^{-4} \int_{\Omega_1} \exp \left( -\frac{2\alpha_1}{\varepsilon} (1 - x) \right) dx$$

$$= \frac{C(1 - \delta_2)}{\varepsilon^4} e^{-2\alpha_1/\varepsilon} \frac{e^{2\alpha_1 x/\varepsilon}}{2\alpha_1/\varepsilon} \Bigg|_{1-\delta_1}^0$$

$$= \frac{C(1 - \delta_2)}{2\alpha_1 \varepsilon^3} \left( e^{-2\alpha_1 \delta_1/\varepsilon} - e^{-2\alpha_1/\varepsilon} \right)$$

$$= C \frac{1 - \delta_2}{2\alpha_1 \varepsilon^3} \left( e^{-2\beta \ln \frac{1}{\varepsilon}} - e^{-2\alpha_1/\varepsilon} \right)$$

$$= \frac{C(1 - \delta_2)}{2\alpha_1} \left( \varepsilon^{2\beta - 3} - \varepsilon^{-3} e^{-2\alpha_1/\varepsilon} \right).$$

Note that $\varepsilon^{-3} e^{-2\alpha_1/\varepsilon}$ is uniformly bounded above for all $0 < \varepsilon \leq O(1)$ since $\alpha_1 > 0$. Thus, if $\beta \geq 3/2$ we have from
the above inequality

\[ \| \partial^2 U_2 / \partial x^2 \|_{0, \Omega_1}^2 \leq C. \]

The proof of boundedness for the first and second seminorms of all other terms are analogous to this case. Therefore we have shown (12).
the above inequality
\[ \| \partial^2 U_2 / \partial x^2 \|_{0, \Omega_1}^2 \leq C. \]

The proof of boundedness for the first and second seminorms of all other terms are analogous to this case. Therefore we have shown (12).

For any fixed \( \varepsilon \), the function in the right side of (11) is increasing in \( x \). Thus we have
\[ \| \partial^2 U_2 / \partial x^2 \|_{0, \infty, \Omega_1} \leq C \varepsilon^{-2} e^{-\alpha_1 \delta_1 / \varepsilon} = C \varepsilon^{\beta - 2}. \]
The last term in the above is uniformly bounded for all \( \beta \geq 2 \).
The finite element formulation of the FVM
The finite element formulation of the FVM

$T_h^{(1)}$ — a regular triangulation on $\Omega_1$, i.e.,

$$\min_{e \in E_h} \frac{|e|}{h} \geq \gamma > 0$$
The finite element formulation of the FVM

\( T_h^{(1)} \) — a regular triangulation on \( \Omega_1 \), i.e.,

\[
\min_{e \in E_h} \frac{|e|}{h} \geq \gamma > 0
\]

\( T_h^{(2)} \) — a mesh on \( \Omega_2 \) containing long, thin rectangles.
The finite element formulation of the FVM

\( T_h^{(1)} \) — a regular triangulation on \( \Omega_1 \), i.e.,

\[
\min_{e \in E_h} \frac{|e|}{h} \geq \gamma > 0
\]

\( T_h^{(2)} \) — a mesh on \( \Omega_2 \) containing long, thin rectangles.

\( T_h \) — combination of \( T_h^{(1)} \) and \( T_h^{(2)} \) containing \( N \) nodes and \( J \) edges.
The finite element formulation of the FVM

\( T_h^{(1)} \) — a regular triangulation on \( \Omega_1 \), i.e.,

\[
\min_{e \in E_h} |e|/h \geq \gamma > 0
\]

\( T_h^{(2)} \) — a mesh on \( \Omega_2 \) containing long, thin rectangles.

\( T_h \) — combination of \( T_h^{(1)} \) and \( T_h^{(2)} \) containing \( N \) nodes and \( J \) edges.

Assumption: \( T_h \) is a Delaunay mesh.
The finite element formulation of the FVM

\( T_h^{(1)} \) — a regular triangulation on \( \Omega_1 \), i.e.,

\[
\min_{e \in E_h} \frac{|e|}{h} \geq \gamma > 0
\]

\( T_h^{(2)} \) — a mesh on \( \Omega_2 \) containing long, thin rectangles.

\( T_h \) — combination of \( T_h^{(1)} \) and \( T_h^{(2)} \) containing \( N \) nodes and \( J \) edges.

Assumption: \( T_h \) is a Delaunay mesh.

We also use

\( N' \) and \( J' \) — numbers of nodes and edges in \( T_h \) not on \( \partial \Omega \).

\( X_h \) and \( X'_h \) — sets of nodes in \( T_h \) and in \( T_h \) not on \( \partial \Omega \).

\( E_h \) and \( E'_h \) — sets of edges in \( T_h \) and in \( T_h \) not on \( \partial \Omega \).
Figure 4: A hybrid Delaunay mesh.
Associated with $T_h$, we define two meshes dual to it.

$D_h$ — the Dirichlet tessellation (associated with each node)

$B_h$ — quadrilateral mesh (associated with each edge) by connecting the two end-points of an edge and the circumcentres of the elements sharing the edge.
Associated with $T_h$, we define two meshes dual to it.

$D_h$ — the Dirichlet tessellation (associated with each node)

$B_h$ — quadrilateral mesh (associated with each edge) by connecting the two end-points of an edge and the circumcentres of the elements sharing the edge.

$D'_h$ — subset of $D_h$ excluding those associated with the boundary points.
Figure 5: Elements and edges associated with the node $x_i \in \Omega_2$. 
Finite element subspaces:
Finite element subspaces:

Test space

\[ V_h = \text{span}\{\xi_i\}_{1}^{N'} \] — test space on \( T_h \), where \( \xi_i \) is piecewise constant given by

\[
\xi_i = \begin{cases} 
1 & \text{on } d_i \\
0 & \text{otherwise}
\end{cases}
\]
Trial space

$U_h$ — exponentially fitted trial space constructed below.

For each $e_{i,j} \in E_h$ connecting the two neighbouring nodes $x_i$ and $x_j$, we define an exponential function $\phi_{i,j}$ on $e_{i,j}$ by

$$\frac{d}{de_{i,j}}(\varepsilon \frac{d\phi_{i,j}}{de_{i,j}} - \bar{a}_{i,j} \phi_{i,j}) = 0, \quad (14)$$

$$\phi_{i,j}(x_i) = 1, \quad \phi_{i,j}(x_j) = 0,$$

$e_{i,j}$ — unit vector from $x_i$ to $x_j$

$\bar{a}_{i,j}$ — constant approximation to $a \cdot e_{i,j}$

$$\phi_i = \begin{cases} 
\phi_{i,j} & \text{on } b_{i,j} \text{ if } j \in I_i \\
0 & \text{otherwise},
\end{cases}$$
\( b_{i,j} \) — the element of \( B_h \) containing \( e_{i,j} \),

\[
I_i = \{ j : e_{i,j} \in E_h \}
\]  

(15)

denotes the index set of all neighbour nodes of \( x_i \).

Figure 6: Support of \( \phi_i \) is star-shaped.
We put $U_h = \text{span}\{\phi_i\}_{1}^{N'}$. 
We put $U_h = \text{span}\{\phi_i\}_{1}^{N'}$.

The projection of the flux of the $U_h$-interpolant $u_I$ of $u$ on $e_{i,j}$ satisfies

$$f_{i,j} := \varepsilon \frac{du_I}{d|e_{i,j}|} - \bar{a}_{i,j} u_I = \frac{\varepsilon}{|e_{i,j}|} \left( B\left( \frac{\bar{a}_{i,j}|e_{i,j}|}{\varepsilon} \right) u_j - B\left( -\frac{\bar{a}_{i,j}|e_{i,j}|}{\varepsilon} \right) u_i \right)$$

on the edge $e_{i,j}$, where $B$ denotes the Bernoulli function defined above.
We put \( U_h = \text{span}\{\phi_i\}_1^{N'} \).

The projection of the flux of the \( U_h \)-interpolant \( u_I \) of \( u \) on \( e_{i,j} \) satisfies

\[
f_{i,j} := \varepsilon \frac{d u_I}{d e_{i,j}} - \bar{a}_{i,j} u_I = \frac{\varepsilon}{|e_{i,j}|} \left( B\left( \frac{\bar{a}_{i,j}|e_{i,j}|}{\varepsilon} \right) u_j - B\left( -\frac{\bar{a}_{i,j}|e_{i,j}|}{\varepsilon} \right) u_i \right)
\]

(16)
on the edge \( e_{i,j} \), where \( B \) denotes the Bernoulli function defined above.

Also, the approximation error in \( f_{i,j} \) in the region \( \Omega_1 \) satisfies

\[
\| f \cdot e_{i,j} - f_{i,j} \|_{\infty,e_{i,j}} \leq C|e_{i,j}| \max\{\varepsilon, |\bar{a}_{i,j}| + |a|_{1,\infty,b_{i,j}}\}.
\]

(17)
where \( C \) is a positive constant independent of \( h, u \) and \( \varepsilon \).
We let
\[ \hat{a}_{b_{i,j}} = \bar{a}_{i,j} e_{i,j} + \bar{a}_{i,j}^\perp l_{i,j}, \]  
(18)
and introduce the mass lumping operator
\[ P : C(\bar{\Omega}) \mapsto \text{span}\{\xi_i\}_1^N \] such that
\[ P(w)(x) = \sum_{i=1}^{N} w(x_i) \xi_i(x), \quad x \in \bar{\Omega}, \]  
(19)
Problem 3.1: Find $u_h \in U_h$ such that

$$a(u_h, v_h) + (P(Gu_h), v_h) = (F, v_h), \quad \forall v_h \in V_h, \quad (20)$$

where $a(\cdot, \cdot)$ is a bilinear form on defined respectively by

$$a(u_h, v_h) = - \sum_{d \in D'_h} \int_{\partial d} (\varepsilon \nabla u_h - \hat{a} u_h) \cdot n v_h|_d ds$$

$n$ — unit outward normal vector of $\partial d$
Let $u_h = \sum_{i=1}^{N'} u_i \phi_i$ and $v_h = \xi_j$. It was shown by Miller & Wang that the above equation reduces to

$$- \sum_{k \in I_j} \int_{l_{j,k}} (\varepsilon \frac{d u_h}{d n} - \bar{a}_{j,k} u_h) \, ds + G_j u_j |d_j| = \int_{d_j} F \, dx$$

for $j = 1, 2, ..., N'$. We've shown that this reduces to

$$- \sum_{j \in I_i} \frac{\varepsilon |l_{ij}|}{|e_{ij}|} \left( B(\sigma_{ij}) u_j - B(-\sigma_{ij}) u_i \right) + G_i u_i |d_i| = \int_{d_j} F \, dx$$

which is the same as that of the FVM except the RHS.
Coercivity of the bilinear form

Observation: \( V_h \) and \( U_h \) are equivalent under the mapping \( P \), i.e., \( P : U_h \leftrightarrow V_h \) is surjective.
Coercivity of the bilinear form

Observation: $V_h$ and $U_h$ are equivalent under the mapping $P$, i.e., $P : U_h \hookrightarrow V_h$ is surjective.

The Bubnov-Galerkin problem corresponding to Problem 3.1:

Problem 3.2: Find $u_h \in U_h$ such that for all

$$b(u_h, v_h) = (F, P(v_h)) \quad \forall v_h \in U_h,$$

(21)

where $b(\cdot, \cdot)$ is a bilinear form on $U_h \times U_h$ defined by

$$b(u_h, v_h) := a(u_h, P(v_h)) + (P(Gu_h), P(v_h)).$$

(22)
For any \( i = 1, 2, \ldots, N' \), integrating the condition on \( a \) by parts over \( d_i \),

\[
\frac{1}{2} \int_{\partial d_i} \mathbf{a} \cdot \mathbf{n} ds + \int_{d_i} G dx \geq 0
\]

Based on the above inequality we make the following assumption.

Assumption 4.1: Let the mesh \( T_h \) be sufficiently fine such that the inequality

\[
\frac{1}{2} \int_{\partial d_i} \hat{\mathbf{a}} \cdot \mathbf{n} ds + G(x_i)|d_i| \geq 0
\]  

holds for all \( d_i \in D'_h \), where \( \hat{\mathbf{a}} \) is the approximation of \( \mathbf{a} \) defined in (18) and \( x_i \) denotes the mesh node contained in \( d_i \).
Furthermore, since all the mesh lines in $T_h^{(2)}$ are parallel to one of the axes and $a$ satisfies (5), it is obvious that

$$\min_{e_{i,j} \in E_{h}^{(2)}} |\hat{a} \cdot e_{i,j}| = \min_{e_{i,j} \in E_{h}^{(2)}} |\bar{a}_{i,j}| \geq \min\{\alpha_1, \alpha_2\}. \quad (24)$$
We now define some functional.

\[
\| u_h \|_h^2 = \sum_{e_{i,j} \in E'_h} \max\{ |e_{i,j}| |\bar{a}_{i,j}|, \varepsilon \} \left( \frac{u_j - u_i}{|e_{i,j}|} \right)^2 |b_{i,j}|
\]

and

\[
\| u_h \| = \| u_h \|_h + \sum_{i=1}^{N'} u_i^2 \left( \frac{1}{2} \int_{\partial d_i} \hat{a} \cdot n ds + G_i |d_i| \right)
\]

for each \( u_h = \sum_{i=1}^{N'} u_i \phi_i \in U_h \).

Both are norms on \( U_h \) when \( h \) is sufficiently small.

(\( \bar{a}_{i,j} = \hat{a} \cdot e_{i,j} \).)

When \( |\bar{a}_{i,j}| \geq a_0 > 0 \), the norms are \( \varepsilon \)-independent.
Theorem 4.1: Let Assumption 4.1 be fulfilled. Then, for all $u \in U_h$, we have

$$b(u, u) \geq C||u||^2,$$  \hspace{1cm} (25)

where $C$ denotes a generic positive constant independent of $\varepsilon$, $h$ and $u$. 
PROOF.

It has been shown by Miller & Wang

\[
a(u_h, P(u_h)) = - \sum_{d \in D'_h} \int_{\partial d} (\varepsilon \nabla u_h - \hat{a} u_h) \cdot n P(u_h) ds
\]

\[
= \sum_{e_{i,j} \in E'_h} \frac{\sigma_{i,j}}{2} \frac{\rho_{i,j}}{e^{\rho_{i,j}} - 1} (1 + e^{\rho_{i,j}})(u_j - u_i)^2 \frac{2|b_{i,j}|}{|e_{i,j}|}
\]

\[
+ \sum_{e_{i,j} \in E'_h} \frac{\bar{a}_{i,j}}{2} (u_i^2 - u_j^2)|l_{i,j}|
\]

where

\[
l_{i,j} = \partial d_i \cap \partial d_j, \sigma_{i,j} = \varepsilon / |e_{i,j}|, \text{ and} \]

\[
\rho_{i,j} = \frac{\bar{a}_{i,j}}{\sigma_{i,j}}.
\]
Since
\[
\frac{(e^{\rho_{i,j}} + 1)}{(e^{\rho_{i,j}} - 1)} \bar{a}_{i,j} \geq |\bar{a}_{i,j}|
\]

We have, if $|\bar{a}_{i,j}| > \varepsilon$, 
\[
\sum_{e_{i,j} \in E'_h} \frac{\sigma_{i,j}}{2} \frac{\rho_{i,j}}{e^{\rho_{i,j}} - 1} (1 + e^{\rho_{i,j}})(u_j - u_i)^2 \frac{2|b_{i,j}|}{|e_{i,j}|} \\
\geq C \sum_{e_{i,j} \in E'_h} |e_{i,j}| |\bar{a}_{i,j}| \left( \frac{u_j - u_i}{|e_{i,j}|} \right)^2 |b_{i,j}|.
\]
If $|\tilde{a}_{i,j}| < \varepsilon$, we have

$$\sum_{e_{i,j} \in E_h'} \frac{\sigma_{i,j}}{2} \frac{\rho_{i,j}}{e^{\rho_{i,j}} - 1} (1 + e^{\rho_{i,j}})(u_j - u_i)^2 \frac{2|b_{i,j}|}{|e_{i,j}|} \geq C \sum_{e_{i,j} \in E_h'} \varepsilon \left( \frac{u_j - u_i}{|e_{i,j}|} \right)^2 |b_{i,j}|$$
Transforming from a summation over the edges to a summation over the nodes of $X'_h$,

$$
\sum_{e_{i,j} \in E'_h} \frac{\bar{a}_{i,j}}{2} (u_i^2 - u_j^2) |l_{i,j}| = \frac{1}{2} \sum_{i=1}^{N'} u_i^2 \sum_{j \in I_i} \bar{a}_{i,j} |l_{i,j}|
$$

$$
= \frac{1}{2} \sum_{i=1}^{N'} u_i^2 \int_{\partial d_i} \hat{a} \cdot n ds
$$

Therefore,

$$
b(u_h, u_h) = a(u_h, P(u_h)) + (P(Gu_h), P(u_h))
$$

$$
\geq C ||u_h||_h^2 + \sum_{i=1}^{N'} u_i^2 \left( \frac{1}{2} \int_{\partial d_i} \hat{a} \cdot n ds + G_i |d_i| \right) = C ||u_h||^2.
$$
Convergence
Lemma 5.1: Let Assumptions 2.1 be fulfilled. If $\beta \geq 2$ in (10) and $M_1 = M_2 = M$, a positive integer, then, for any element edge $e_{i,j} \in E_h^{(2)}$, the set of edges in $T_h^{(2)}$, there exists a positive integer $C$, independent of $h$, $u$ and $\epsilon$, such that

$$\int_{l_{i,j}} |f \cdot e_{i,j} - f_{i,j}| ds \leq \begin{cases} 
C |l_{i,j}| hK_1, & e_{i,j} \subset \bar{\Omega}_2^{(1)} \cup \bar{\Omega}_2^{(2)}, \\
C |l_{i,j}| \frac{1}{M} \ln(\frac{1}{\epsilon}), & e_{i,j} \subset \Omega_2^{(3)}, 
\end{cases}$$

where $f$ and $f_{i,j}$ are defined in (2) and (16) respectively and

$$K_1 = \max \left\{ 1, h^{-1} \epsilon^{\beta/2M}, h^{-1} \epsilon \ln \frac{1}{\epsilon} \right\}. \quad (26)$$
Theorem 5.1: Let Assumptions 2.1 and 4.1 be fulfilled. If $\beta \geq 3$ in (10) and $M_1 = M_2 = M$, a positive integer, then there exists a positive integer $C$, independent of $h$, $u$ and $\varepsilon$, such that

$$||u_I - u_h|| \leq Ch^{1/2} \left(1 + h^{1/2}(K_1^{1/2} + K_2)\right),$$

(27)

where $u_I$ and $u_h$ denote respectively the $U_h$-interpolation of the solution $u$ to Problem 2.1 and the solution to Problem 3.1, $K_1$ is defined in (26) and $K_2$ is defined as

$$K_2 = \max\{M^{1/2}K_1, M^{-1/2}\sqrt{\varepsilon \ln \frac{1}{\varepsilon}}\}.$$

(28)
PROOF. Sketch only.

$C$ — a generic positive constant, independent of $h$, $\varepsilon$ and $u$.

For any $v_h \in U_h$, multiplying the continuous PDE by $P(v_h)$ and integrating by parts,

$$- \sum_{d \in D'_h} \int_{\partial d} f \cdot n P(v_h) ds + (Gu, P(v_h)) = (F, P(v_h))$$

From this and the FE problem

$$a(u_h - u_I, P(v_h)) + (P(Gu_h) - P(Gu_I), P(v_h))$$

$$= - \sum_{d \in D'_h} \int_{\partial d} f \cdot n P(v_h) ds - a(u_I, P(v_h))$$

$$+ (Gu - P(Gu_I), P(v_h))$$
So,

\[ |b(u_h - u_I, v_h)| \leq -\sum_{d \in D'_h} \int_{\partial d} f \cdot nP(v_h)ds - a(u_I, P(v_h)) + (Gu - P(Gu_I), P(v_h)) =: R_1 + R_2 \]
For $R_1$, it can be shown that

$$R_1 \leq \left| \sum_{e_{i,j} \in E'_h} (v_i - v_j) \int_{l_{i,j}} (\mathbf{f} \cdot e_{i,j} - f_{i,j}) \, ds \right| .$$

We split $E'_h$ into two parts: $E_h^{(1)}$ and $E_h^{(2)}$, and so $R_1 = R_1^{(1)} + R_1^{(2)}$. It is easy to show that

$$R_1^{(1)} \leq C h^{1/2} \left[ \sum_{e_{i,j} \in E_h^{(1)}} \max\{|e_{i,j}|, \bar{a}_{i,j}, \varepsilon\} \left( \frac{v_j - v_i}{|e_{i,j}|} \right)^2 |b_{i,j}| \right]^{1/2}.$$
Using Lemma 5.1,

\[
R_1^{(2)} \leq ChK_2 \left[ \sum_{e_{i,j} \in E_h^{(2)}} |e_{i,j}| \left( \frac{v_j - v_i}{e_{i,j}} \right)^2 |b_{i,j}| \right]^{1/2}
\]

Similarly, it can be shown that

\[
R_2 \leq Ch \left( 1 + \varepsilon \ln \frac{1}{\varepsilon} + \varepsilon^{\beta/2M} \right) ||v_h||_\infty.
\]
Combining all these terms,

\[ |b(u_h - u_I, v_h)| \leq C \left[ h^{1/2} (1 + h^{1/2} K_2) ||v_h|| + h^2 K_1 ||v_h||_{\infty} \right], \]

Choosing \( v_h = u_h - u_I \) and using (25) we obtain

\[ ||u_h - u_I||^2 \leq C \left[ h^{1/2} (1 + h^{1/2} K_2) ||u_h - u_I|| + h^2 K_1 \right]. \]
Combining all these terms,

\[ |b(u_h - u_I, v_h)| \leq C \left[ h^{1/2}(1 + h^{1/2} K_2) ||v_h|| + h^2 K_1 ||v_h||_{\infty} \right], \]

Choosing \( v_h = u_h - u_I \) and using (25) we obtain

\[ ||u_h - u_I||^2 \leq C \left[ h^{1/2}(1 + h^{1/2} K_2) ||u_h - u_I|| + h^2 K_1 \right]. \]

This is of the form

\[ y^2 \leq C(1 + h^{1/2} K_2) h^{1/2} y + CK_1 h^2 \]

or

\[ \left( y - \frac{1}{2} C(1 + h^{1/2} K_2) h^{1/2} \right)^2 \leq CK_1 h^2 + \frac{(C(1 + h^{1/2} K_2))^2}{4} h. \]
The above reduces to

\[
y \leq \sqrt{CK_1 h^2 + \frac{(C(1 + h^{1/2}K_2))^2}{4} h + \frac{CK_2}{2} h^{1/2}}
\leq Ch^{1/2} \left[ 1 + h^{1/2} (K_2 + K_1^{1/2}) \right].
\]

Replacing \( y \) with \( ||u_h - u_I|| \) we obtain

\[
||u_h - u_I|| \leq Ch^{1/2} \left[ 1 + h^{1/2} (K_2 + K_1^{1/2}) \right].
\]

This completes the proof of the theorem. \( \square \)
Corollary: Let the Assumptions in Theorem 5.1 be fulfilled. If $\varepsilon, h, M$ and $\beta$ are such that $h^{-1}\varepsilon^{\beta/2M} \leq O(1)$ and $h^{-1}\varepsilon \ln(1/\varepsilon) \leq O(1)$, then we have

$$\|u_I - u_h\| \leq C h^{1/2} (1 + h^{1/2} M^{1/2}).$$  \hspace{1cm} (29)$$

Furthermore, if choose $h^{-1}\varepsilon^{\beta/2M} = O(1)$, then

$$\|u_I - u_h\| \leq C h^{1/2} \left(1 + h^{1/2} \sqrt{\ln \frac{\varepsilon}{\ln h}}\right).$$  \hspace{1cm} (30)$$

Note $h^{-1}\varepsilon^{\beta/2M} = 1 \Rightarrow M = \left\lceil \frac{\beta \ln \varepsilon}{2 \ln h} \right\rceil$.

($\beta$ can be chosen to be 2 according to Theorem 2.1 and Theorem 5.1.)
Remark: When $M = 1$ and $\beta = 2$, we have uniform convergence of order $O(h^{1/2})$. This is true because all the mesh points are outside the layers.

In this case the method does not resolve the layers.
Conclusions

• When $\varepsilon << h$, $\|u_I - u_h\|$ converges to zero at the rate of $h^{1/2}$ almost uniformly in $\varepsilon$. 
Conclusions

- When $\varepsilon << h$, $||u_I - u_h||$ converges to zero at the rate of $h^{1/2}$ almost uniformly in $\varepsilon$.

- A mixture of triangular and rectangular elements. (or a triangular mesh)
Conclusions

- When $\varepsilon \ll h$, $||u_I - u_h||$ converges to zero at the rate of $h^{1/2}$ almost uniformly in $\varepsilon$.

- A mixture of triangular and rectangular elements. (or a triangular mesh)

- Neuman boundary conditions can be included (as in the paper by Miller & Wang)
Conclusions

- When $\varepsilon << h$, $||u_I - u_h||$ converges to zero at the rate of $h^{1/2}$ almost uniformly in $\varepsilon$.

- A mixture of triangular and rectangular elements. (or a triangular mesh)

- Neuman boundary conditions can be included (as in the paper by Miller & Wang)

- Possibility of extending to singularly perturbed problems with non-rectangular, polygonal regions. (the a priori estimates corresponding to those in Assumption 2.1 need to be established on non-rectangular polygonal regions.)