Васил Тонов Попов

[illegible]

ЦЕНТРАЛНА ЛАБОРАТОРИЯ ПО
ПАРАЛЕЛНА ОБРАБОТКА НА ИНФОРМАЦИЯТА

Б А Н

Тодор Томов Димов

СМЕСЕН МЕТОД НА КРАЙНИТЕ ЕЛЕМЕНТИ ЗА ЕЛИПТИЧНИ ЗАДАЧИ ОТ ВТОРИ РЕД – КОНЦЕНТРАЦИЯ НА МАСАТА, ЛОКАЛНО СГЪСТЯВАНЕ, МОНТЕ КАРЛО АЛГОРИТМИ

ДИСЕРТАЦИЯ

за присъждане на

образователната и научна степен „Доктор“

Научен ръководител:

проф., дмн Райчо Лазаров

Научен консултант:

доц., д-р Андрей Андреев

София, 1998

Съдържание

Увод

През последните години все по-широко разпространение получава методът на крайните елементи (МКЕ) за решаване на гранични задачи за диференциални уравнения, възникващи при моделиране на широк кръг задачи на механиката, физиката и инженерната практика. По същество МКЕ представлява обобщение на метода Релей-Ритц-Гальоркин, в който изходната диференциална задача се формулира като еквивалентна на нея вариационна задача и приближеното решение се търси като линейна комбинация на предварително зададени пробни функции.

Голяма част от усилията на специалистите, работещи в областта на МКЕ са насочени към подходящ избор на пробните функции. Най-често те са начасти полиноми върху всеки краен елемент и носителят им е съсредоточен в околност само на един възел. Това дава възможност за разработване на ефективни методи за решаване на системата алгебрични уравнения, до която води методът.

Най-простите двумерни и тримерни крайни елементи са въведени от Курант [1] през 1943 година. Основите на математическата теория на метода се полагат от Курант и Хилберт [1]. Усъвършенстването на изчислителната техника доведе до бързо развитие на сферите на приложение на метода, което се съпътства с усъвършенстване и прецизиране на математическите му основи

със средствата на числения анализ.

През последните 20 години като водещ дял на МКЕ се наложи смесеният метод на крайните елементи. Смесеният метод за първи път в литературата е въведен от Херман [1] през 1967 година за задачата за плочата. Първите основни резултати в смесения метод са получени от Оден [1], [2], Реди [1], Оден и Реди [1], Бреци и Равиар [1], Сиарле и Равиар [2], Джонсън [1], Миоши [1] и др. През 1977 година Равиар и Тома [1] въвеждат пространствата, наречени по-късно на тяхно име, което дава силен тласък в развитието на теорията на смесения метод на крайните елементи и в практическото му използване. Същността на метода се състои в замяна на изходното диференциално уравнение със система от диференциални уравнения от по-нисък ред, въвеждайки нови неизвестни. Прост пример, илюстриращ този факт, е задачата за дифузия в порести среди:

$$\left| \begin{array}{ll} -div(a(x)\underline{\nabla} p) = f(x)\,, & x \in \Omega\,; \\ \dfrac{\partial p}{\partial \underline{\nu}} = 0\,, & x \in \partial\Omega\,, \end{array} \right.$$

където $\underline{\nu}$ е външният единичен нормален вектор към границата $\partial\Omega$.

Чрез въвеждане на ново неизвестно $\underline{u}$ заменяме уравнението със следната еквивалентна система:

$$\left| \begin{array}{ll} a(x)\underline{\nabla} p = \underline{u}\,, & x \in \Omega\,; \\ -div\underline{u} = f\,, & x \in \Omega\,. \end{array} \right.$$

Тази замяна дава възможност за независима апроксимация на налягането p и скоростта $\underline{u}$ и води до намаляване на изискванията за гладкост на функциите от крайномерното апроксимиращо пространство. Друго важно предимство на метода е, че пресмятането на p и $a\underline{\nabla} p$ е с една и съща степен на точност, за разлика от стандартния МКЕ, при който оценката за $a(x)\underline{\nabla} p$ е с един порядък по-ниска от тази за p.

Друг съвременен подход за решаване на широк клас задачи е методът Мон-

те Карло. По същество това е метод, който използва моделирането на случайни величини или полета за решаване на математически проблеми. Обикновено за всяка задача се подбира случайна величина или поле, такива, че тяхното математическо очакване да съвпада с решението или с линеен функционал от решението. Методът се състои в компютърна симулация на съответната случайна величина или поле и приближено пресмятане на математическото очакване. За първи път в литературата този проблем е описан от Метрополис, Улам [1]. Бързото развитие на алгоритмите през последните години се дължи преди всичко на отличните възможности за симулация на случайни величини, които дават съвременните мощни паралелни компютри.

0.1 Основни означения, дефиниции и базови резултати

В този раздел представяме някои общоприети означения и помощни твърдения, които ще използваме в по-нататъшния анализ. Въвеждаме някои стандартни означения:

Нека Ω е ограничена област в $\mathbb{R}^n$ с граница $\partial\Omega$, която е на части от C^1. За всяко естествено число m означаваме с $H^m(\Omega)$. Соболевото пространство от ред m на скаларнозначната функция w върху Ω, което се дефинира рекурсивно по следния начин:

$$H^0(\Omega) = L^2(\Omega)$$

$$H^m(\Omega) = \{w \in H^{m-1}(\Omega) : \forall \alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}^n,$$

$$\sum_{i=1}^{n} \alpha_i = m, \quad \partial^\alpha w = \frac{\partial^m w}{\partial x_1^{\alpha_1} \dots \partial x_n^{\alpha_n}} \in L^2(\Omega)\Big\}, \quad \forall m \geq 1 .$$

Съответните норма $\|\cdot\|_m$ и полунорма $|\cdot|_m$ дефинираме чрез:

$$\|w\|_{0,\Omega} = \left[\int_\Omega w^2(x)dx\right]^{1/2};$$

$$|w|_{m,\Omega} = \left[\sum_{|\alpha|=m} \|\partial^\alpha w\|^2_{0,\Omega}\right]^{1/2}, \forall m \geq 1;$$

$$\|w\|_{m,\Omega} = \left[\|w\|^2_{m-1,\Omega} + |w|^2_{m,\Omega}\right]^{1/2}, \quad \forall m \geq 1.$$

Ще разглеждаме също и Соболевото пространство $W^{m,p}(\Omega)$, което за всяко цяло число $m \geq 0$ и всяко число p, удовлетворяващо условията $1 \leq p \leq \infty$, се състои от такива функции $v \in L^p(\Omega)$, за които всички частни производни $\partial^\alpha w$ (в смисъл на обобщени функции) при $|\alpha| \leq m$ принадлежат на пространството $L^p(\Omega)$, снабдено със съответните полунорми:

$$|w|_{m,p,\Omega} = \left[\sum_{|\alpha|=m} \int_\Omega |\partial^\alpha w(x)|^p dx\right]^{1/p}, \ 1 \leq p < \infty,$$

$$|w|_{m,\infty,\Omega} = \max_{|\alpha|=m} \left[ess.\sup_{x\in\Omega} |\partial^\alpha w(x)|\right].$$

и норми:

$$\|w\|_{m,p,\Omega} = \left[\sum_{|\alpha|\leq m} \int_\Omega |\partial^\alpha w(x)|^p dx\right]^{1/p}, \ 1 \leq p < \infty,$$

$$\|w\|_{m,\infty,\Omega} = \max_{|\alpha|\leq m} \left[ess.\sup_{x\in\Omega} |\partial^\alpha w(x)|\right].$$

Отбелязваме, че

$$W^{m,2}(\Omega) = H^m(\Omega)$$

и са в сила равенствата

$$|\cdot|_{m,2,\Omega} = |\cdot|_{m,\Omega}\,, \quad \|\cdot\|_{m,2,\Omega} = \|\cdot\|_{m,\Omega}$$

Пространството $(H^m(\Omega))^n$ от векторнозначни функции

$$(H^m(\Omega))^n = \{\underline{v} = (v_1, v_2, \ldots, v_n) : v_i \in H^m(\Omega),\ \ i = 1, 2, \ldots, n\}$$

е снабдено със следните полунорма

$$|\underline{v}|_{m,\Omega} = \left[\sum_{i=1}^{n} |v_i|^2_{m,\Omega}\right]^{1/2}$$

и съответно норма

$$\|\underline{v}\|_{m,\Omega} = \left[\sum_{i=1}^{n} \|v_i\|^2_{m,\Omega}\right]^{1/2}. \tag{0.1}$$

Нека $H_0^m(\Omega)$ е множеството от функции, принадлежащи на $H^m(\Omega)$, които се анулират почти навсякъде върху границата $\partial\Omega$. За функцията $\varphi \in L^2(\Omega)$ използваме също и пространството $H^{-m}(\Omega)$, което е дуално на $H_0^m(\Omega)$ и е снабдено с нормата

$$\|\varphi\|_{-m,\Omega} = \sup_{0 \not\equiv \psi \in H^m(\Omega)} \frac{|(\varphi, \psi)|}{\|\psi\|_{m,\Omega}}, \qquad m > 0,$$

където $(\varphi, \psi) = \int_\Omega \varphi(x)\psi(x)dx$ е обичайното скаларно произведение в $L^2(\Omega)$.

Означаваме с $\underline{H}(div; \Omega)$ пространството, дефинирано чрез:

$$\underline{H}(div; \Omega) = \left\{\underline{v} \equiv (v_1, v_2, \ldots, v_n) \in (L^2(\Omega))^n : div\underline{v} = \sum_{i=1}^{n} \frac{\partial v_i}{\partial x_i} \in L^2(\Omega)\right\}$$

и снабдено с нормата

$$\|\underline{v}\|_{\underline{H}(div;\Omega)} = \left(\|\underline{v}\|^2_{0,\Omega} + \|div\underline{v}\|^2_{0,\Omega}\right)^{1/2}. \tag{0.2}$$

Ще използваме следните означения за полиноми на две променливи с реални коефициенти:

$$P(n) \equiv P_n(x_1, x_2) = \left\{ p(x_1, x_2) : p(x_1, x_2) = \sum_{0 \le i+j \le n} a_{ij} x_1^i x_2^j \right\}; \tag{0.3}$$

$$Q(n, m) \equiv Q_{n,m}(x_1, x_2) = \left\{ q(x_1, x_2) : q(x_1, x_2) = \sum_{i=0}^{n} \sum_{j=0}^{m} a_{ij} x_1^i x_2^j \right\}. \tag{0.4}$$

Казваме, че две отворени подмножества Ω и $\hat{\Omega}$, принадлежащи на $\mathbb{R}^2$, са афинно-еквивалентни, ако съществува обратимо афинно изображение

$$F : \hat{\Omega} \to \Omega,$$

такова, че

$$\hat{x} \to x = F(\hat{x}) = B\hat{x} + \underline{b}, \tag{0.5}$$

където

$$B \in \mathbb{R}^2 \times \mathbb{R}^2, \quad \underline{b} \in \mathbb{R}^2.$$

При това са в сила следните съотношения:

$$det(B) = \frac{meas(\Omega)}{meas(\hat{\Omega})} \tag{0.6}$$

и

$$\|B\| \le \frac{h}{\hat{\rho}}, \qquad \|B^{-1}\| \le \frac{\hat{h}}{\rho}, \tag{0.7}$$

където параметрите $h, \hat{h}, \rho$ и $\hat{\rho}$ означават съответно:

$$h = diam(\Omega), \qquad \hat{h} = diam(\hat{\Omega}); \tag{0.8}$$

$$\rho = sup\{diam(S); \quad S \text{ е вписано кълбо в } \Omega\}\,,$$
$$\hat{\rho} = sup\{diam(\hat{S}); \quad \hat{S} \text{ е вписано кълбо в } \hat{\Omega}\}\,. \tag{0.9}$$

Без ограничение на общността можем да считаме, че Якобианът на изображението F_Ω $det(B_\Omega)$ е положителен. За всяка скаларнозначна функция w в Ω полагаме:

$$w(x) = \hat{w}(\hat{x}), \text{ където } x = F_\Omega(\hat{x}), \ \forall \hat{x} \in \hat{\Omega}. \tag{0.10}$$

Навсякъде в по-нататъшния анализ ще означаваме със C положителна константа, която в общия случай не е една и съща в различните неравенства.

Ще се нуждаем от някои основни неравенства (виж Сиарле [1]):

Лема 0.1.1 *Нека Ω и $\hat{\Omega}$ са две афинно-еквивалентни отворени подмножества в $\mathbb{R}^n$. Ако функцията w принадлежи на пространството $W^{m,p}(\Omega)$ за някакво цяло число $m \geq 0$ и някакво число $p \in [1, \infty]$, то функцията $\hat{w} = w \circ F$ принадлежи на пространството $W^{m,p}(\hat{\Omega})$ и съществува такава константа $C = C(m, n)$, че:*

$$\forall w \in W^{m,p}(\Omega) \qquad |\hat{w}|_{m,p,\hat{\Omega}} \leq C\|B\|^m |det(B)|^{-1/p} |w|_{m,p,\Omega}\,. \tag{0.11}$$

Аналогично:

$$\forall \hat{w} \in W^{m,p}(\hat{\Omega}) \qquad |w|_{m,p,\Omega} \leq C\|B^{-1}\|^m |det(B)|^{1/p} |\hat{w}|_{m,p,\hat{\Omega}}\,. \tag{0.12}$$

Следващите помощни твърдения са тясно свързани с някои свойства на полиномите, дефинирани в (0.3) и (0.4).

Лема 0.1.2 *Нека $\Omega \subset \mathbb{R}^n$ е ограничена област. Тогава съществува такава константа C, че за всяка функция $w \in W^{r+1,p}(\Omega)$ е в сила:*

$$\inf_{p \in P_r(\Omega)} \|w + p\|_{r+1,p,\Omega} \leq C(\Omega) |w|_{r+1,p,\Omega}\,.$$

В съответствие с дефинираните в (0.4) полиноми разглеждаме следния Лагранжев интерполант на скаларнозначна функция $w(x_1, x_2)$:

(0.13) $$w^I \in Q(r,r) : \ w^I \equiv w \ \text{в} \ (r+1) \times (r+1) \ \text{точки}$$

За векторнозначна функция $\underline{v} = (v_1, v_2)$ дефинираме съответния интерполант $\underline{v}^I = (v_1^I, v_2^I)$ по следния начин:

(0.14) $$v_1^I \in Q(r+1,r) : \ v_1^I \equiv v_1 \ \text{в} \ (r+2) \times (r+1) \ \text{точки}$$

(0.15) $$v_2^I \in Q(r,r+1) : \ v_2^I \equiv v_2 \ \text{в} \ (r+1) \times (r+2) \ \text{точки}$$

Следват някои основни свойства на интерполантите (0.13), (0.14) и (0.15), дефинирани в правоъгълна област $\hat{\Omega} \subset \mathbb{R}^2$.

Лема 0.1.3 *Съществува константа $C(\hat{\Omega})$, такава, че за всяко афинно еквивалентно множество Ω на $\hat{\Omega}$ и за всяка скаларнозначна функция $w \in H^{r+1}(\Omega)$ е изпълнено неравенството:*

$$\|w - w^I\|_{0,\Omega} \leq C(\hat{\Omega}) h^{r+1} |w|_{r+1,\Omega} ,$$

където параметърът h е определен в (0.8).

Доказателство. От неравенство (0.12) при $m = 0$, $p = 2$, и тъй като $\forall w \in P_r \ \ w = w^I$ следва:

$$\begin{aligned} \|w - w^I\|_{0,\Omega} &\leq C|det(B)|^{1/2} \|\hat{w} - \hat{w}^I\|_{0,\hat{\Omega}} \\ &\leq C|det(B)|^{1/2} . \inf_{\hat{p} \in P_r(\hat{\Omega})} \|\hat{w} + \hat{p}\|_{r+1,\hat{\Omega}} . \end{aligned}$$

Прилагайки Лема (0.1.2) и неравенство (0.11) при $m = r+1$ и $p = 2$ получаваме:

$$\begin{aligned} \|w - w^I\|_{0,\Omega} &\leq C(\hat{\Omega})|det(B)|^{1/2} |\hat{w}|_{r+1,\hat{\Omega}} \\ &\leq C(\hat{\Omega}) \|B\|^{r+1} |w|_{r+1,\Omega} . \end{aligned}$$

За да завършим доказателството, достатъчно е да приложим първото неравенство на (0.7).

Аналогични твърдения са в сила и за векторнозначни функции:

Лема 0.1.4 *Съществува константа* $C(\hat{\Omega})$, *такава, че за всяко афинно еквивалентно множество* Ω *на* $\hat{\Omega}$ *и за всяка векторнозначна функция* $\underline{v} \in (H^{r+1}(\Omega))^2$ *е изпълнено неравенството:*

$$\|\underline{v} - \underline{v}^I\|_{0,\Omega} \leq C(\hat{\Omega}) h^{r+1} |\underline{v}|_{r+1,\Omega} ,$$

където параметърът h *е определен в (0.8).*

Доказателство. За доказателството на лемата е достатъчно да приложим два пъти Лема 0.1.3, тъй като по дефиниция (виж 0.1) е в сила:

$$\begin{aligned} \|\underline{v} - \underline{v}^I\|_{0,\Omega} &= \left(\sum_{i=1}^{2} \|v_i - v_i^I\|_{0,\Omega}^2 \right)^{1/2} \\ &\leq \|v_1 - v_1^I\|_{0,\Omega} + \|v_2 - v_2^I\|_{0,\Omega} . \end{aligned} \tag{0.16}$$

Лема 0.1.5 *Съществува константа* $C(\hat{\Omega})$, *такава, че за всяко афинно еквивалентно множество* Ω *и за всяка векторнозначна функция* $\underline{v} \in (H^{r+2}(\Omega))^2$ *е изпълнено неравенството:*

$$\|div(\underline{v} - \underline{v}^I)\|_{0,\Omega} \leq C(\hat{\Omega}) h^{r+1} |div\underline{v}|_{r+1,\Omega} ,$$

където параметърът h *е определен в (0.8).*

Доказателство. За доказателството на лемата достатъчно е да приложим Лема 0.1.3, имайки предвид, че $div\underline{v} \in Q(r,r)$.

Най-характерният аспект на метода на крайните елементи се състои в осъществяването на триангулация τ_h на множеството $\overline{\Omega}$, зависеща от параметъра h. Множеството $\overline{\Omega}$ се разделя на краен брой подмножества T, които

се наричат крайни елементи, такива, че

$$\overline{\Omega} = \bigcup_{T \in \tau_h} T \tag{0.17}$$

Следващата лема дава необходимото и достатъчно условие за принадлежност на една векторнозначна функция $\underline{v}$ към пространството $\underline{H}(div;\Omega)$.

Лема 0.1.6 *Функцията* $\underline{v} \in (L^2(\Omega))^2$, *чиято рестрикция* $\underline{v}_{|T} \in \underline{H}(div;T)$ *за всяко* $T \in \tau_h$ *принадлежи на* $\underline{H}(div;\Omega)$ *тогава и само тогава, когато за всяка обща страна* $T' = T_1 \cap T_2$ *на* $T_1, T_2 \in \tau_h$ *е изпълнено* $\underline{v}_{|T_1}.\underline{\nu_1} = \underline{v}_{|T_2}.\underline{\nu_2}$ *върху* T', *където* $\underline{\nu_i}$ *е единичният външен нормален вектор на* T_i, $i = 1,2$ *по общата им страна* T', *т.е. когато нормалната компонента на* $\underline{v}$ *е непрекъсната по границата между елементите* T *от* τ_h.

Доказателство. Изискването функцията $\underline{v} \in (L^2(\Omega))^2$ да принадлежи на $\underline{H}(div;\Omega)$ е $div\underline{v} \in L^2(\Omega)$, т.е.

$$(div\underline{v}, \varphi) = -(\underline{v}, \underline{\nabla}\varphi), \ \ \forall \varphi \in C_0^\infty(\Omega).$$

където $\underline{\nabla}\varphi$ означава градиента на скаларнозначната функция φ.

Разглеждаме елементите T_1 и T_2, които имат обща страна T'. За всяка функция φ, чийто носител е във вътрешността на тяхното обединение, използвайки формулата на Грийн за всеки елемент, получаваме:

$$(div\underline{v}, \varphi) = \int_{\partial T_1} \left(\underline{v}_{|T_1}.\underline{\nu}_{T_1}\right) \varphi ds + \int_{\partial T_2} \left(\underline{v}_{|T_2}.\underline{\nu}_{T_2}\right) \varphi ds - (\underline{v}.\underline{\nabla}\varphi),$$

което е еквивалентно на

$$(\underline{v}_{|T_1}.\underline{\nu}_{T_1})_{|T'} = (\underline{v}_{|T_2}.\underline{\nu}_{T_2})_{|T'}.$$

Следствие 0.1.1 *В случая на правоъгълни крайни елементи изискването* $\underline{v} = (v_1, v_2) \in \underline{H}(div;\Omega)$ *е еквивалентно с изискването за непрекъснатост на функциите* v_i *съответно по променливите* x_i, $i = 1,2$.

Казваме, че триангулацията τ_h (0.17) е регулярна в смисъл на Сиарле [1], ако съществува такава константа C, че за всеки краен елемент $T \in \tau_h$ параметрите h_T (0.8) и ρ_T (0.9), свързани с него, удовлетворяват условието:

$$\frac{h_T}{\rho_T} \leq C \tag{0.18}$$

при произволно малко

$$h = \max_{T \in \tau_h} h_T. \tag{0.19}$$

Във всички по-нататъшни оценки на грешката степенният показател на параметъра h, дефиниран в (0.19), в десните части на съответните неравенства ще играе основна роля при определяне скоростта на сходимост на разглеждания метод.

0.2 Основни резултати в дисертацията

Най-важните показатели за качеството на даден дискретен метод са:

1) Скоростта на сходимост на приближеното към точното решение;

2) Възможността за решаване на съответната алгебрична задача с минимален обем изчислителна работа.

В настоящата дисертация тези две направления са изследвани в тяхната взаимовръзка, т.е. първоначалната задача се свежда до алгебрична задача със значително по-малка размерност от стандартната, без обаче да се понижава степента на сходимост. В случаите на локално сгъстяване се отчитат физическите особености на задачата, която се моделира (скоростта $a\underline{\nabla}p$ е твърде различна в съответни подобласти на разглежданата област).

Основните направления се изследват чрез две важни явления в метода на крайните елементи, респективно в смесения метод, а именно локално сгъстяване на мрежата и концентрация на матрицата на масата.

Предлага се и един нов Монте Карло подход за обръщане на матрицата на масата, който има самостоятелно значение при обръщане на матрици от общ вид.

Оптимални по порядък оценки (в смисъл на Дъглас и Робъртс [1]) в смесения метод на крайните елементи за елиптични гранични задачи от втори ред са получени от Равиар и Тома [1], Бреци [1], Фолк и Осборн [1] и Дъглас и Робъртс [1], които най-общо имат следния вид:

$$\|\underline{u} - \underline{u_h}\|_{\underline{H}(div;\Omega)} + \|p - p_h\|_{0,\Omega} \leq Ch^{r+1} \left(\|\underline{u}\|_{r+2,\Omega} + \|p\|_{r+1,\Omega} \right).$$

където h е дефиниран в (0.19), p_h и $\underline{u_h}$ са приближените стойности съответно на p и $\underline{u} = a(x)\underline{\nabla} p$, нормите са обичайните в Соболевите пространства, r е индексът на апроксимиращите пространства, дефинирани в (1.6) а C означава (както и в цялата дисертация) положителна константа, независеща от h.

Локалното сгъстяване на мрежата е добре познат прийом, използван успешно за практически цели – мрежата се сгъстява в подобласти, в които решението се изменя по-бързо. Доскоро обаче се използваше само стандартният тип локално сгъстяване, който не допускаше връх на някакъв краен елемент да лежи върху страна на съседен елемент.

Подходът на локално сгъстяване, използвайки концепцията на „подчинени" възли е въведен в смесения метод на крайните елементи през 1990 година от Юинг, Лазаров, Ръсел и Василевски [1], използвайки аналогии от стандартния метод на крайните елементи и крайните разлики (виж Брамбъл, Юинг, Пашек и Шатс [1], Мак Кормик [1], Банк и Дюпон [1], Юинг, Лазаров и Василевски [1]).

Необходимостта от прилагането на този подход възниква от моделирането на реални физически процеси, в които решението проявява локални свойства.

Отразяването на тези свойства изисква относително гъста мрежа, водещо до увеличаване размера на алгебричната задача, което се явява излишно в тези подобласти, в които решението се изменя относително бавно.

В подобластите, където съответните Соболеви норми на решението, участващи в оценката са относително малки, се въвежда „груба" мрежа с параметър h_c, а в тези с относително големи норми – „фина" мрежа с параметър $h_f = \frac{1}{n} h_c$, което дава възможност за балансиране на членовете в дяснята част на оценките на грешката между точното и приближеното решение.

Важно преимущество на тази концепция е възможността лесно да се модифицират и прилагат вече готови пакети програми за решаване на конкретни практически задачи.

Другият основен аспект в оптимизиране на пресмятанията в метода на крайните елементи е явлението „концентрация на масата", което води до привеждане на алгебричната задача в по-прост вид. Отдавна физиците и инженерите са забелязали възможността дадена механична схема или конструкция да се разглежда като тяло, в което масата е концентрирана в определени възли.

Чермак и Зламал [1] са установили, че по този начин се постигат по-устойчиви схеми и по-прости алгебрични системи. Т.Миоши [2] изследва този ефект за редица приложни задачи.

Конструирането на тази процедура обаче трябва да се съпътства със строг математически анализ, за да не се допусне нарушаване на точността (описано като контрапример в монографията на Стренг и Фикс [1]).

В общия случай доказателството, че концентрацията на матрицата на масата в смесения метод не нарушава съществуването, единствеността и сходимостта на решението на дискретната задача, не винаги е възможно. Тогава възниква въпросът за нейното обръщане, с цел намаляване размерността на задачата.

В дисертацията са разгледани въпросите за концентрация на масата, два различни подхода за получаване на оценки на грешката при регулярно локално сгъстяване на мрежата, приложение към задачата за приближено намиране на собствени стойности в смесения метод на крайните елементи и един нов Монте Карло подход за обръщане на матрици.

В Глава 1 се разглежда смесен метод на крайните елементи с концентрация на матрицата на масата за решаване на двумерна моделна задача на Дирихле върху правоъгълна мрежа. Параграфи 1.1 и 1.2 имат въвеждащ характер.

Разгледана е хомогенна задача на Дирихле

$$(0.20)\qquad \left|\begin{array}{ll} -div(a(x)\underline{\nabla} p) = f(x)\,, & x \in \Omega\,; \\ p = 0\,, & x \in \partial\Omega\,, \end{array}\right.$$

където $\underline{\nabla} w$ означава градиента на скаларнозначната функция w, $div\,\underline{v}$ означава дивергенцията на векторнозначната функция $\underline{v}$, а $a(x) = diag\,(a_1(x), a_2(x))$ е диагонална матрица, чиито елементи удовлетворяват изискванията $a_i(x) \geq a_0 > 0$, $i = 1, 2$.

Изходната диференциална задача се формулира като еквивалентна на нея вариационна задача: търси се двойката $(\underline{u}, p) \in \underline{V} \times W$ като решение на системата:

$$(0.21)\qquad \left|\begin{array}{ll} a(\underline{u}, \underline{v}) + b(\underline{v}, p) = 0, & \forall \underline{v} \in \underline{V}\,; \\ b(\underline{u}, w) = -(f, w), & \forall w \in W\,, \end{array}\right.$$

където $\underline{V} = \underline{H}(div; \Omega)$ и $W = L^2(\Omega)$ са характерните пространства, свързани със смесения метод, $a(\underline{u}, \underline{v})$ и $b(\underline{u}, w)$ са билинейни форми, дефинирани в (1.3), а $(\cdot, \cdot)$ означава обичайното скаларно произведение в $L^2(\Omega)$.

Параграф 1.3 има спомагателен характер. В него се въвеждат крайномерните пространства $\underline{V_h}$ и W_h, определят се степените на свобода (в случая

възловия базис) и чрез Лема 1.3.2 се показва, че тези пространства удовлетворяват добре известните условия на Бабушка - Бреци.

В § 1.4 се въвеждат специални квадратури формули от тип Гаус - Лобато за пресмятане на интегралите в алгебричната система, които съчетани с подбрания в § 1.3 интерполационен базис дават концентрация на матрицата на масата. Абстрактна оценка на грешката при използване на числено интегриране в смесения метод на крайните елементи се съдържа в монографията на Робъртс и Тома [1] и е от следния вид:

(0.22)

$$\|\underline{u} - \underline{u}_h^*\|_{\underline{V}} + \|p - p_h^*\|_W$$

$$\leq C \left\{ \inf_{\underline{v}_h \in \underline{V}_h} \|\underline{u} - \underline{v}_h\|_{\underline{V}} + \sup_{\underline{t}_h \in \underline{V}_h} \frac{a(\underline{v}_h, \underline{t}_h) - a_h(\underline{v}_h, \underline{t}_h)}{\|\underline{t}_h\|_{\underline{V}}} + \inf_{w_h \in W_h} \|p - w_h\|_W \right\}.$$

където двойката $(\underline{u}_h^*, p_h^*) \in \underline{V}_h, \times W_h$ е приближеното решение, получено по смесения метод, използвайки числено интегриране, а $a_h(\underline{u}_h, \underline{t}_h)$ е апроксимация на билинейната форма $a(\underline{u}_h, \underline{t}_h)$.

Целта на Лема 1.4.1 е да се докаже, че конструираните пространства и подбраните квадратурни формули изпълняват условията, изискващи се при прилагане на оценката (0.22), т.е. съществуват константи C_1 и C_2, такива, че

$$a_h(\underline{u}_h, \underline{v}_h) \leq C_1 \|\underline{u}_h\|_{\underline{V}} \|\underline{v}_h\|_{\underline{V}}, \quad \forall \underline{v}_h, \ \underline{u}_h \in \underline{V}_h;$$

$$a_h(\underline{v}_h, \underline{v}_h) \geq C_2 \|\underline{v}_h\|_{\underline{V}}^2, \quad \forall \underline{v}_h \in \underline{V}_h^*,$$

където

$$\underline{V}_h^* = \left\{ \underline{v}_h \in \underline{V}_h \ : \ (div \underline{v}_h, w_h) = 0, \ \forall w_h \in W_h \right\}.$$

Лема 1.4.5 дава оценка на основния член в дясната част на неравенство

(0.22) в случая $\underline{v_h} = \underline{u}^I$, т.е.

$$\sup_{\underline{t_h} \in \underline{V_h}} \frac{a(\underline{u}^I, \underline{t_h}) - a_h(\underline{u}^I, \underline{t_h})}{\|\underline{t_h}\|_{\underline{V}}} \leq C_9 h^{r+1} \left(\|\underline{u}\|_{r+2,\Omega} + \|p\|_{r+1,\Omega} \right).$$

където $\underline{u}^I$ е интерполантът на $\underline{u}$ по съответните възли, дефиниран за $\underline{u} = \underline{v}$ в (1.12).

Основният резултат в тази глава се получава на базата на (0.22), Лема 1.4.1 и Лема 1.4.5 и се състои в следното: Нека $(\underline{u}, p)$ е решение на (0.20) и $(\underline{u_h}^*, p_h^*)$ е решение на (1.34), получено чрез прилагане на числено интегриране с помощта на квадратурните формули (1.33). Предполагаме, че $p \in H^{r+3}(\Omega)$. Тогава съществува константа C такава, че

$$\|\underline{u} - \underline{u_h}^*\|_{\underline{H}(div;\Omega)} + \|p - p_h^*\|_{0,\Omega} \leq C h^{r+1} \left(\|\underline{u}\|_{r+2,\Omega} + \|p\|_{r+1,\Omega} \right).$$

Полученият резултат показва, че използването на замяната на интегралите с квадратурни формули не влошава сходимостта на метода, т.е. оценката на грешката остава оптимална по порядък.

Забележка 0.2.1 *В цялата дисертация ще разбираме понятието оптимална по порядък оценка на грешката в смисъла на Дъглас и Робертс.*

Другият основен ефект от така избрания тип числено интегриране е илюстриран в § 1.5. След диагонализирането на матриците M_i, $i = 1, 2$ в системата на метода

$$BY = \begin{pmatrix} M_1 & 0 & N_1 \\ 0 & M_2 & N_2 \\ N_1^T & N_2^T & 0 \end{pmatrix} \begin{pmatrix} U_1 \\ U_2 \\ P \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -F \end{pmatrix}, \tag{0.23}$$

където $M_i \in \mathbb{R}^{n \times n}$ и $N_i \in \mathbb{R}^{n \times m}$, $i = 1, 2$ са матрици, а $U_1, U_2 \in \mathbb{R}^n$ и $P, F \in \mathbb{R}^m$ – вектори. Тя се редуцира в еквивалентна на нея система

$$\left(N_1^T M_1^{-1} N_1 + N_2^T M_2^{-1} N_2 \right) P = F,$$

със значително по-малка размерност (в случая $U_i = -M_i^{-1}N_iP,\ i = 1,2$). В това се състои практическата стойност на изложените в тази глава теоретични резултати.

В Глава 2 се разглежда смесения метод на крайните елементи върху правоъгълна мрежа с използване на „подчинени" възли за решаване на задачата на Дирихле върху правоъгълна мрежа с регулярно локално сгъстяване. § 2.1 има обзорен характер, а в § 2.2 се дефинира изходната задача на Дирихле:

$$\text{(0.24)}\qquad \begin{aligned} &\text{(а)}\quad -div(a(x)\underline{\nabla} p) = f, \quad \text{в } \Omega; \\ &\text{(б)}\quad p = -g, \qquad\qquad\quad \text{върху } \partial\Omega, \end{aligned}$$

където с $\partial\Omega$ е означена границата на $\Omega \in \mathbb{R}^2$, а $a(x) = diag\,(a_1(x), a_2(x))$ е диагонална матрица, чиито елементи удовлетворяват изискванията $a_i(x) \geq a_0 > 0,\ i = 1,2$.

Еквивалентната вариационна формулировка е следната:

$$\text{(0.25)}\qquad \begin{aligned} &\text{(а)}\quad a(\underline{u}, \underline{v}) - b(\underline{v}, p) = \langle g, \underline{v}.\underline{\nu}\rangle \quad \forall \underline{v} \in \underline{V}\,; \\ &\text{(б)}\quad b(\underline{u}, w) = (f, w), \qquad\qquad\quad \forall w \in W\,, \end{aligned}$$

където $\underline{\nu}$ е единичният външен нормален вектор към $\partial\Omega$. Скаларното произведение в $(L^2(\Omega))^2$ е означено с $(\cdot,\cdot)$, а в $L^2(\partial\Omega)$ – с $\langle\cdot,\cdot\rangle$.

Описание на апроксимацията чрез смесения метод върху съставната мрежа се дава в § 2.3. Разделянето на областта Ω се извършва на два етапа:

1) Прави се разделяне на „груби" крайни елементи с характерен размер h_c;

2) Част от тези елементи се разделят повторно на n части по всяко едно направление, така че от всеки „груб" елемент се получават n^2 „фини" елементи с характерен размер $h_f = \frac{1}{n}h_c$.

Означаваме с Ω_1 и Ω_2 подобластите на Ω, състоящи се съответно от „фини" и „груби" крайни елементи. Комбинирайки (наслагвайки) по подхо-

дящ начин пространствата на Равиар - Тома съответно върху грубата и фина мрежа получаваме пространствата от същия вид върху съставната мрежа.

Първо разглеждаме пространството на Равиар и Тома $\underline{V_c^r}$, асоциирано с триангулацията τ_c на областта Ω.

Нека $\underline{\overset{\circ}{V}{}_f^r}(\Omega_1)$ е отново пространство на Равиар и Тома, асоциирано обаче с триангулацията τ_f на подобластта Ω_1, такова, че нормалната компонента на функцията $\underline{v} \in \underline{\overset{\circ}{V}{}_f^r}(\Omega_1)$ върху границата Γ на Ω_1 е нула. Тогава разглеждаме следните пространства на Равиар и Тома върху съставната мрежа τ_h, както следва:

(0.26)

$$\text{(а)} \quad \underline{V_h^r} = \underline{V_c^r} + \underline{\overset{\circ}{V}{}_f^r}(\Omega_1);$$

$$\text{(б)} \quad W_h^r = \{w \in L^2(\Omega), w(x_1, x_2) \in Q(r, r) \text{ за всяко } T \in \tau_h\}$$

Използвайки пространствата 0.26 дефинираме съответната вариационна формулировка. Смесения метод на крайните елементи за задачата (2.1) се определя чрез намиране на двойката $\left(\underline{u_h}, p_h\right) \in \underline{V_h^r} \times W_h^r$ такава, че:

(0.27)

$$\text{(а)} \qquad a(\underline{u_h}, \underline{v_h}) - b(\underline{v_h}, p_h) = \langle g, \underline{\nu}.\underline{v_h}\rangle, \quad \underline{v_h} \in \underline{V_h^r};$$

$$\text{(б)} \qquad b(\underline{u_h}, w_h) = (f, w_h). \qquad\qquad w_h \in W_h^r.$$

В Лема 2.3.1 се доказва, че така дефинираните пространства удовлетворяват условията на Бабушка - Бреци, което е важна стъпка в оценката на грешката на метода.

В § 2.4 се доказват първо локални оценки на грешката, като интерполантът $\underline{u}^I$ на $\underline{u}$ се дефинира по специален начин върху елементите от подобластите $\Omega_1 \backslash I_f$, Ω_2 и интерфейса I_f на Ω_1. Те са представени в Теорема 2.4.1. В нейното доказателство, освен интерполационна техника, съществено се използва и Лема 2.4.1. При наличие на съответните изисквания на теоремата

съществуват константи C, независещи от h_f и h_c, такава, че:

(а) $\|w - w^I\|_{0,T} \leq Ch_T^{r+1}\|w\|_{r+1,T}$, $T \in \tau_h, T \in \Omega$;

(б) $\|\underline{v} - \underline{v}^I\|_{0,T} \leq Ch_T^{r+1}\|\underline{v}\|_{r+1,T}$, $T \in \tau_h, T \in \Omega\backslash I_f$;

(в) $\|div(\underline{v} - \underline{v}^I)\|_{0,T} \leq Ch_T^{r+1}\|\underline{v}\|_{r+2,T}$, $T \in \tau_h, T \in \Omega\backslash I_f$;

(г) $\|\underline{v} - \underline{v}^I\|_{0,T} \leq C\left(h_f^{r+1}\|\underline{v}\|_{r+1,T} + h_f h_c^{r+1}\|\underline{v}\|_{r+1,\infty,\gamma}\right)$, $T \in \tau_h$, $T \in I_f$;

(д) $\|div(\underline{v} - \underline{v}^I)\|_{0,T} \leq C\left(h_f^{r+1}\|\underline{v}\|_{r+2,T} + h_f h_c^{r}\|\underline{v}\|_{r+1,\infty,\gamma}\right)$, $T \in \tau_h$, $T \in I_f$,

където γ е страната на произволен елемент $T_c \in \Omega_2$, който граничи с подобластта Ω_1, т.е.

$$\gamma = \bigcup_{T_f \in I_f} (T_f \cap T_c)$$

В доказателството на Теорема 2.4.1 се демонстрира една добра характеризация на локалната оценка върху елементите от интерфейса I_f, тъй като разликата между нея и тази за елементите от $\Omega_1\backslash I_f$ се представя в явен вид.

Резултатите от Лема 2.3.1 и Теорема 2.4.1 дават възможност да се докаже основната теорема за оценка на грешката. При съответните изисквания за глаткост на решението задачата (2.6) има единствено решение $(\underline{u_h}, p_h) \in (\underline{V_h} \times W_h)$ и съществува константа C, независеща от h_c и h_f, такава, че:

$$\begin{aligned}\|\underline{u} - \underline{u_h}\|_{\underline{H}(div;\Omega)} + \|p - p_h\|_{0,\Omega} \leq C\Big(&h_f^{r+1}\|p\|_{r+1,\Omega_1} + h_c^{r+1}\|p\|_{r+1,\Omega_2} \\ &+h_f^{r+1}\|\underline{u}\|_{r+2,\Omega_1} + h_c^{r+1}\|\underline{u}\|_{r+2,\Omega_2} \\ &+h_c^{r+1}n^{-1/2}\|\underline{u}\|_{r+1,\infty,\Gamma}\Big),\end{aligned}$$

където $\Gamma = \overline{\Omega}_1 \bigcap \overline{\Omega}_2$.

Характерното за този тип оценки е, че ако притежаваме предварителна информация за големината на нормите на решението (например скоростта на неговото изменение) в различни подобласти ние оптимизираме оценката,

т.е. там където нормите са „големи“ размерът h_f е „малък“, а там където нормите са „малки“ h_c е „голям“.

Известно е, че в основата на оценките за приближено намиране на собствени стойности чрез смесения метод на крайните елементи стои съответната оценка за така наречената „задача на първоизточника“ (виж например монографията на Бабушка и Осборн [1]).

Представено е и едно приложение на гореописаната техника за оценка на грешката при задачата за собствени стойности. Получена е оценка на грешката между точните и приближените собствени стойности, която е от следният вид:

Нека $(\lambda, (\underline{u}, p))$ и $(\lambda_h, (\underline{u_h}, p_h))$ са решения съответно на изходната и приближената задача за собствени стойности по смесения метод. Тогава съществува константа C, независеща от h_f и h_c, такава, че:

$$| \lambda - \lambda_h | \leq C \left[C_1(r) h_f^{2(r+1)} + C_2(r,n) h_c^{2(r+1)} \right] ,$$

където:

$$C_1(r) = \|\underline{u}\|_{r+2,\Omega_1} + \|p\|_{r+1,\Omega_1} ,$$

$$C_2(r,n) = \|\underline{u}\|_{r+2,\Omega_2} + \|p\|_{r+1,\Omega_2} + n^{-1/2} \|\underline{u}\|_{r+1,\infty,\Gamma}$$

а вътрешната граница е $\Gamma = \overline{\Omega}_1 \bigcap \overline{\Omega}_2$.

В Глава 3 се доказва оценка на грешката в смесения метод на крайните елементи за елиптична гранична задача върху мрежи с регулярно локално сгъстяване. За разлика от изследванията в Глава 2 тук елиптичният оператор е в най-общ вид (не се изисква симетричност и положителна-определеност). Концепцията за построяване на апроксимиращите пространства се основава на друг принцип. Съществената разлика от тази на Глава 2 е тази, че тук степените на свобода са определени интеграли и анализът се извършва на

базата на подходящо дефинирани $L^2(\Omega)$-проектори. Подобна техника за една проста моделна задача е демонстрирана в работата на Юинг и Уонг [1].

В § 3.1 и § 3.2 се прави обзор на резултатите на други автори до момента и се дефинират изходната и еквивалентната на нея вариационна задача:

Нека $\Omega \subset \mathbb{R}^2$ е ограничена област с граница $\partial\Omega$ и задачата на Дирихле (0.28)

$$\text{(а)} \quad Lp = -div\,(a(x)\underline{\nabla} p + \underline{b}(x)p) + c(x)p = f(x)\,, \quad x \in \Omega\,;$$

$$\text{(б)} \quad p = -g(x)\,, \quad x \in \partial\Omega\,,$$

притежава единствено решение за $\{f, g\} \in L^2(\Omega) \times H^{3/2}(\partial\Omega)$, и че

$$\|p\|_{2,\Omega} \leq C\left(\|f\|_{0,\Omega} + \|g\|_{3/2;\partial\Omega}\right)\,.$$

Слабата формулировка на (0.28) се представя по следния начин: търсим двойка $(\underline{u}, p) \in \underline{V} \times W$, такава, че

$$\text{(а)} \quad (\alpha\underline{u}, \underline{v}) - (div\underline{v}, p) + (\underline{\beta}p, \underline{v}) = \langle g,\, \underline{v}\,.\,\underline{\nu}\rangle\,, \qquad \underline{v} \in \underline{V}\,;$$

$$\text{(б)} \quad (div\underline{u}, w) + (cp, w) = (f, w)\,, \qquad w \in W\,,$$

където

$$\alpha(x) = a(x)^{-1}\,, \qquad \underline{\beta}(x) = \alpha(x)\underline{b}(x)\,.$$

Съществуване, единственост и оптимална оценка на грешката на решението $(\underline{u_h}, p_h)$ на смесения метод на крайните елементи за пространства на Равиар-Тома с индекс r за елиптични гранични задачи от втори ред без локално сгъстяване е получена от Дъглас и Робъртс [1]. За регулярни крайни елементи представените оценки в L^2-норма са от следния вид:

$$\text{(а)} \quad \|p - p_h\|_{0,\Omega} \leq \begin{cases} Ch\|p\|_{2,\Omega}, \text{ ако } r = 0; \\ Ch^k\|p\|_{k,\Omega}, \text{ ако } r \geq 1 \text{ и } 2 \leq k \leq r+1; \end{cases}$$

$$\text{(б)} \quad \|\underline{u} - \underline{u_h}\|_{0,\Omega} \leq Ch^k\|p\|_{k+1,\Omega}, \text{ ако } 1 \leq k \leq r+1;$$

$$\text{(в)} \quad \|div(\underline{u} - \underline{u_h})\|_{0,\Omega} \leq Ch^k\|p\|_{k+2,\Omega}, \text{ ако } 0 \leq k \leq r+1.$$

където Ω е подобласт на $\mathbb{R}^2$ или $\mathbb{R}^3$ и $a(x) = diag(a_1(x), a_2(x))$ е диагонална матрица, чиито елементи удовлетворявят условията $a_i(x) \geq a_0 > 0$, $i = 1, 2$.

Аналогичната гранична задача на Нойман също се включва в нашите общи разглеждания.

Апроксимацията чрез смесения метод на крайните елементи на задачата (0.30) с елементи на Равиар - Тома води (виж например Робъртс, Тома [1]) до следната система от линейни алгебрични уравнения:

$$BY \equiv \begin{pmatrix} M & N \\ N^* & 0 \end{pmatrix} \begin{pmatrix} U \\ P \end{pmatrix} = \begin{pmatrix} G \\ F \end{pmatrix}, \tag{0.31}$$

където съответните матрици и вектори са от следния вид:

$$B \in \mathbb{R}^{(n_1+n_2)\times(n_1+n_2)},\ M \in \mathbb{R}^{n_1\times n_1},\ N \in \mathbb{R}^{n_1\times n_2},\ N^* \in \mathbb{R}^{n_2\times n_1},$$

$$Y \in \mathbb{R}^{n_1+n_2},\ U, G \in \mathbb{R}^{n_1},\ P, F \in \mathbb{R}^{n_2}.$$

Матрицата B е обратима, но не е положително - определена. Затова директното решаване на тази система в общия случай е затруднено (виж например Робъртс, Тома [1]).

За приближеното решаване на системата (0.31) се използват различни итерационни методи (метод на спрегнатия градиент, наказателен метод, ускорен метод на Лагранж и др.)

От друга страна матрицата M е симетрична и положително - определена. Следователно, теоретически нейното обръщане е винаги възможно. Тогава, ако успеем да пресметнем M^{-1} и положим

$$U = M^{-1}G - M^{-1}NP$$

получаваме системата:

$$KP = H, \tag{0.32}$$

където

$$K = N^* M^{-1} N$$

и

$$H = N^* M^{-1} G - F.$$

По този начин редуцираме линейната алгебрична система (0.31) с размерност $n = (n_1 + n_2)$ до n_2-размерната система (0.32), където

$$n_2 < \frac{1}{3} n.$$

Задачата за обръщане на матрици заслужава отделно разглеждане, тъй като тя се явява основна или помощна подзадача и на други важни математически задачи. Така например в случая, когато се нуждаем от груба оценка на обратната матрица (Колотилина, Йерьомин [1]), която се използва при конструиране на преобуслователи за ускоряване на различни итерационни методи за решаване на линейни алгебрични системи.

В Глава 4 са представени два ефективни Монте Карло алгоритъма за обръщане на матрици, основаващи се на два нови подхода към разглежданата задача.

Разглежданите алгоритми се базират на нов подход (виж Т. Димов [3] и И. Димов, Т. Димов, Гюров [1]) и притежават същата изчислителна сложност, като алгоритмите, описани от Джон Холтън, но те са по-ефективни, защото комбинират съответно различни стоп критерии и релаксационни параметри.

Добре известно е, че числените методи Монте Карло дават статистическа оценка за решението на дадена задача, използвайки някаква случайна величина, чието математическо очакване съвпада с търсеното решение. Алгоритмите, базиращи се на тези методи притежават някои очевидни предимства. Едно от най-съществените е, че те са „присъщо паралелни“. Те притежават висока

ефективност, когато се използват паралелни компютри. Този факт е показан в работите на Метрополис, Улам [1], Димов, Тонев [1], [2] и др. Монте Карло алгоритмите са също достатъчно ефективни, когато разглежданата задача е с свръх голяма размерност или е с „неопределена“ (или „обща“) структура, която не се вмества в алгоритмите, основаващите се на традиционните числени методи.

Едно от най-важните предимства на тези алгоритми е, че те дават възможност да се пресметне определен линеен функционал от решението, и в частност - само една негова компонента, без да се пресмятат останалите му компоненти.

Съществуват два класа алгоритми, основаващи се на числените методи Монте Карло - директни и итерационни.

Директните методи притежават само вероятностна грешка.

Итерационните Монте Карло алгоритми притежават два типа грешки - систематични и вероятностни. В конкретния случай те се наричат съответно грешка от прекъсване и вероятна грешка. Систематичната грешка зависи от броя на итерациите в използвания итерационен метод, докато вероятностната грешка зависи от стохастичната природа на методите Монте Карло.

Първият от разглежданите алгоритми е помощен и пресмята произволна компонента u_0 на решението $\underline{u}$ на линейната алгебрична система. Този алгоритъм е разгледан отделно, тъй като някои от неговите стъпки се използват в описанието на следващите два алгоритъма.

Вторият алгоритъм пресмята приближението $\hat{C}$ на обратната матрица $C = A^{-1}$. Той се базира на специален избор на релаксационния параметър γ, който се контролира чрез апостериорен критерий за всеки стълб на следната резидуална матрица

$$E^c = A\hat{C} - I\,.$$

Този избор на получаване на резудуална матрица (умножавайки матрицата $\hat{C}$ от ляво с матрицата A) дава в най-пълна степен зависимостта на точността, с която пресмятаме стълбовете на $\hat{C}$, от стълбовете на E^c.

Алгоритъмът дава възможност различните вектор–стълбове на матрицата $\hat{C} = (\hat{C}_1, \ldots, \hat{C}_m)$ да се пресмятат, използвайки различни стойности на релаксационния параметър $\gamma = \gamma_p$, $p = 1, 2, \ldots, l$. Стойностите на γ_p се избират така, че да минимизират съответните Евклидови норми на следните вектор–стълбове:

$$E_j^c = A\hat{C}_j - I_j, \quad j = 1, 2, \ldots, m.$$

Фактически, минимизирането на Евклидовата норма на вектор–стълба E_j^c върху някакво предварително зададено множество

$$\gamma = \{\gamma_1, \gamma_2, \ldots, \gamma_l\} \tag{0.33}$$

подобрява (намалява) нормата на Фробениус на резидуалната матрица

$$E^c = (E_1^c, \ldots, E_m^c)$$

и води до по-добра апроксимация на обратната матрица $\hat{C}$ стълб по стълб.

Отбелязваме, че пресмятането на различните стълбове може да се реализира паралелно и независимо един от друг. Като второ ниво на паралелизъм, в зависимост от броя на процесорите във всяка конкретна компютърна конфигурация, могат да се задават съответни стойности на параметъра l, който определя броя на елементите от множеството γ, дефинирано в (0.33).

Основната идея на представения по-горе алгоритъм използва детерминистичен подход, който не зависи от неговата статистическа природа и може да бъде прилагана и при другите традиционни алгоритми. Това негово качество му определя самостоятелна роля в областта на алгоритмите за обръщане на матрици.

Третият алгоритъмът съществено изисква Монте Карло подход и няма детерминистичен аналог. Разликата между втория и третия алгоритъм е, че последният не може да бъде приложен за традиционните (детерминистични) итерационни методи. Тези методи дават възможност да се пресмятат паралелно различните стълбове на обратната матрица, но те не позволяват пресмятане на всеки един от нейните елементи независимо от другите елементи.

Едно от съществените предимства на Монте Карло алгоритмите се състои именно във възможносста за пресмятане на елементите на обратната матрица невзаисимо един от друг. Това свойство дава възможност да се прилагат различни итерационни подходи за намиране на матрицата $\hat{C}$, използвайки априорна информация за редовете на дадената матрица A (например някакви отношения между техните елементи). Отбелязваме, че предварителната ифомация за свойствата на редовете на матриците, възникващи при решаване на конкретни задачи е винаги достъпна, а понякога и единствено възможна.

За всеки ред A_i на матрицата A (когато $|a_{ii}| \neq 0$), въвеждаме съответните параметри K_i, както следва:

$$K_i = \left(\sum_{\substack{j=1 \\ j \neq i}}^{m} |a_{ij}| \right) \Big/ |a_{ii}|, \quad i = 1, 2, \ldots m. \tag{0.34}$$

Казваме, че даден ред A_i на матрицата A е съответно добре обусловен, умерено обусловен и лошо обусловен, когато съответстващият му параметър (0.34) изпълнява следните изисквания:

$$K_i < 1, \quad K_i = 1, \quad \text{и} \quad K_i > 1.$$

Нашите числени експерименти показват, че итерационните Монте Карло алгоритми притежават толкова по-добра сходимост при пресмятане на реда $\hat{C}_i$, колкото е по-малък параметърът K_i, съответстващ на реда A_i, защо-

то първоначалното намаляване на теглата на състоянията на Марковската верига W е гарантирано.

Този факт дава априорна информация за използване на различни стоп критерии ε_i $(i = 1, \ldots, m)$ за пресмятане на съответстващия му ред $\hat{C}_i$ на матрицата $\hat{C}$. Практически това означава, че ние използваме по-голяма разлика между две Монте Карло итерации за прекъсване на итерационния процес, когато той е по-бързо сходящ за сметка на на съответната по-малка разлика в случая на по-бавна сходимост.

Броят на скоковете във верига на Марков (т.е. броят на итерациите) могат също да бъдат контролирани така, че да се получи добър баланс между статистическата и систематичната грешки. Задачата за балансиране на тези грешки е много съществена при използването на Монте Карло алгоритмите. Очевидно, за да получим добри резултати, трябва статистическата (в случая вероятната) грешка r_n да бъде приблизително равна на съответната систематична r_k, т.е.

$$r_n = O(r_k).$$

Разгледана е и задачата за балансиране на грешките. За да запазим баланса на грешките (систематична и стохастична), ние също използваме различен брой реализации n_i $(i = 1, \ldots, m)$ на съответните случайни величини, които са пропорционални на броя на итерациите. Тази процедура наричаме *фин стоп критерий*. Използвайки я, определяме понятието *Рафиниран Монте Карло алгоритъм*.

Така ние намаляваме нормата на Фробениус на следната резидуална матрица:

$$E^r = \hat{C}A - I\,,$$

и следователно подобряваме точността на пресмятане на матрицата $\hat{C}$ ред по

ред, без да увеличаваме изчислителната сложност на алгоритъма в сравнение с алгоритмите, използващи стандартния *груб стоп критерий*.

В тази глава са представени и числени резултати, които демонстрират възможностите на предложените нови Монте Карло алгоритми. В качеството на конкретен пример е разгледана следната система линейни алгебрични уравнения, възникваща след апроксимация на двумерна хомогенна задача на Дирихле (0.30) чрез смесения метод на крайните елементи с правоъгълници, аналогично на Глава 1:

$$BY = \begin{pmatrix} M_1 & 0 & N_1 \\ 0 & M_2 & N_2 \\ N_1^T & N_2^T & 0 \end{pmatrix} \begin{pmatrix} U_1 \\ U_2 \\ P \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -F \end{pmatrix},$$

където $M_i \equiv diag(A_{i1}, A_{i2}, \ldots A_{is})$ са $t \times t$ блочно диагонални матрици, N_i са $t \times t_1$ матрици $(t_1 < t)$, $U_i \in I\!R^t$ и $P, F \in I\!R^{t_1}$, $i = 1, 2$.

Основните резултати от дисертацията са публикувани в работите на автора [1], [2], [3], в съвместните работи с А.Б. Андреев [1], [2], както и в съвместната работа с И.Т. Димов и Т.В. Гюров [1].

Резултатите, включени в дисертацията, са докладвани на семинара на лаборатория „Паралелни алгоритми", ЦЛПОИ – БАН, на семинара на секция „Изчислителна математика", ИМИ – БАН и на семинара на групата по паралелни алгоритми на проекта "APACHE" – IMAG, Гренобъл, Франция. Част от публикуваните резултати са представени на специализирани международни конференции, както следва:

International Conference on Constructive Theory of Functions, Варна, 1991 г.;

Aplication of Mathematics in Engineering, Варна, 1991 г.;

International Workshop on Parallel Algorithms, София, 1992 г.;

International Conference on Scientific Computation & Mathematical Modeling, Созопол, 1993 г.;

Third International Conference on Numerical Methods & Applications – $\mathcal{O}(h^3)$, София, 1994 г.;

First Workshop on Numerical Analysis & Applications, Русе, 1996 г.;

Fourth International Conference on Numerical Methods & Applications – $\mathcal{O}(h^4)$, София, 1998 г..

Част от резултатите на автора са включени в отчета на съвместния проект между IMAG – Grenoble и ЦЛПОИ – БАН, София, финансиран от министерството на образованието на Франция и в отчетите по договори МУ-МСА-3/94 (ръководител), ММ-449/94 (член), И-501/95 (член) с Националния Фонд „Научни изследвания“ при МОНТ.

Глава 1

Смесен метод на крайните елементи с концентрация на масата

1.1 Въведение

Ще разгледаме смесения метод на крайните елементи върху правоъгълна мрежа за решаване на хомогенната задача на Дирихле за уравнение от втори ред. Оптимални оценки на грешката за тази задача са дадени от Равиар и Тома [1], Бреци [1] и Фолк и Осборн [1]. Изчислителните аспекти на разглеждания тип апроксимация са изследвани от Юинг и Уилър [1], Юинг, Коеби, Гонзалес и Уилър [1], Василевски и Лазаров [1], като са представени многобройни числени експерименти. Оценки от тип свръхсходимост в Гаусовите точки са установени от Наката, Уейсър и Уилър [1] и по Гаусовите линии от Юинг, Лазаров и Уонг [1]. В смесения метод на крайните елементи абстрактна оценка за ефекта от числено интегриране е дадена от Робъртс и Тома [1]. Ефектът

от прилагането на специални квадратурни формули за пресмятане на интегралите в смесения метод за пространства на Равиар и Тома от най-нисък ред $r = 0$ е разгледан от ван Ноен[1].

Параграфи 2 и 3 имат спомагателен характер. След постановката на задачата се дефинират стандартните функционални пространства, както и пространствата на Равиар и Тома.

Диагонализирането на матрицата на масата (концентрация на масата) води до по-ефективен алгоритъм за решаване на системата алгебрични уравнения, която се получава при прилагане на метода. Много автори са разглеждали това явление в стандартния метод на крайните елементи (виж например Банержи и Осборн [1], Чен и Томе [1], Андреев, Касчиева и Ванмаеле [1]).

В §4 се прилагат подходящи квадратурни формули за пресмятане на интегралите в системата на смесения метод, които водят до диагонализация на матрицата на масата. Доказва се, че тази процедура запазва оптималния порядък на сходимост на приближеното към точното решение. В §5 се разглежда матричната задача и се показва ефективността на алгоритъма - значително намаляване на размерността на задачата.

1.2 Постановка на задачата

Нека $\Omega \subset \mathbb{R}^2$ е ограничена област с граница $\partial\Omega$. Разглеждаме хомогенната задача на Дирихле:

$$
(1.1) \qquad \left| \begin{array}{ll} -div(a(x)\underline{\nabla} p) = f(x), & x \in \Omega; \\ p = 0, & x \in \partial\Omega, \end{array} \right.
$$

където $\underline{\nabla} w$ означава градиента на скаларнозначната функция w, $div\,\underline{v}$ означава дивергенцията на векторнозначната функция $\underline{v}$, а $a(x) = diag\,(a_1(x), a_2(x))$

е диагонална матрица, чиито елементи удовлетворяват изискванията $a_i(x) \geq a_0 > 0,\ i = 1,2$.

От съображение за простота вземаме областта Ω да бъде единичния квадрат $(0,1)^2$. Резултатите, които получаваме, могат да бъдат продължени за по-общи области, използвайки подхода на Дъглас и Робъртс [1].

За задачата (1.1) полагаме

$$\underline{u} \equiv (u_1, u_2) = a(x)\underline{\nabla} p, \quad \alpha_i(x) = a_i(x)^{-1}, \quad i = 1,2.$$

Разглеждаме пространствата $\underline{V} = \underline{H}(div;\Omega)$ и $W = L^2(\Omega)$, снабдени с нормите:

$$\|\underline{v}\|_{\underline{V}} \equiv \|\underline{v}\|_{\underline{H}(div;\Omega)} = \left[\|\underline{v}\|^2_{0,\Omega} + \|div\underline{v}\|^2_{0,\Omega}\right]^{1/2}$$

и съответно

$$\|w\|_W \equiv \|w\|_{L^2(\Omega)} = \|w\|_{0,\Omega}.$$

Вариационната формулировка на задачата (1.1) се дава чрез двойката $(\underline{u}, p) \in \underline{V} \times W$, като решение на системата:

$$\left|\begin{array}{ll} a(\underline{u},\underline{v}) + b(\underline{v},p) = 0, & \forall \underline{v} \in \underline{V}; \\ b(\underline{u},w) = -(f,w), & \forall w \in W, \end{array}\right. \tag{1.2}$$

където

$$a(\underline{u},\underline{v}) = (\alpha_1 u_1, v_1) + (\alpha_2 u_2, v_2), \quad b(\underline{u},w) = (div\underline{u}, w), \tag{1.3}$$

а $(\cdot,\cdot)$ означава скаларното произведение в $L^2(\Omega)$.

1.3 Апроксимация върху правоъгълна мрежа

Нека D_1 и D_2 са разделяния на интервала $I = [0,1]$, както следва:

$$D_1 = \{0 = x_{1,0} < x_{1,1} < \ldots < x_{1,N} = 1\},$$

$$D_2 = \{0 = x_{2,0} < x_{2,1} < \ldots < x_{2,M} = 1\}.$$

Означаваме

$$h_i = x_{1,i} - x_{1,i-1}, \quad i = 1, \ldots, N,$$

$$h_j' = x_{2,j} - x_{2,j-1}, \quad j = 1, \ldots, M$$

и нека

$$h = \max_{i,j}\{h_i, h_j'\}.$$

Така разделянето на $\bar{\Omega} = [0,1]^2$ се получава като Декартово произведение на D_1 и D_2.

Това разделяне (триангулация) τ_h се дефинира, както следва:

$$\tau_h = \Big\{T \equiv T_{ij} \subset \Omega : T_{ij} = [x_{1,i-1}, x_{1,i}] \times [x_{2,j-1}, x_{2,j}], \quad i = 1, \ldots, N,\ j = 1, \ldots, M\Big\}. \tag{1.4}$$

Предполагаме, че триангулацията τ_h е регулярна (виж (0.18)) и съществуват константи C_1 и C_2 такива, че

$$C_1 h^2 \leq meas(T) \leq C_2 h^2, \quad \forall T \in \tau_h.$$

Полагаме:

$$M_k^n(D_1) = \Big\{p \in C^k(I) : p_{|[x_{1,i-1}, x_{1,i}]} \in P_n,\ i = 1, \ldots, N\Big\};$$

$$M_k^n(D_2) = \Big\{p \in C^k(I) : p_{|[x_{2,j-1}, x_{2,j}]} \in P_n,\ j = 1, \ldots, M\Big\},$$

където $k \geq -1$ е цяло число. Случаят $k = -1$ означава, че съответно $p(x_1)\,(p(x_2))$ могат да имат прекъсвания в точките $x_{1,1}, \ldots, x_{1,N-1}$ $(x_{2,1}, \ldots, x_{2,M-1})$.

Дефинираме пространствата на Равиар и Тома (виж Равиар и Тома [1]) с

индекс r $(r = 0, 1, 2, \dots)$ $\underline{V_h} \equiv \underline{V_h^r}$ и $W_h \equiv W_h^r$, по следния начин:

$$
\begin{aligned}
V_{1,h}^r &= M_0^{r+1}(D_1) \otimes M_{-1}^r(D_2), \\
V_{2,h}^r &= M_{-1}^r(D_1) \otimes M_0^{r+1}(D_2), \\
\underline{V_h^r} &= V_{1,h}^r \times V_{2,h}^r, \\
W_h^r &= M_{-1}^r(D_1) \otimes M_{-1}^r(D_2).
\end{aligned} \tag{1.5}
$$

Така дефинираните пространства са на части полиноми върху всеки краен елемент:

$$
\begin{aligned}
\underline{V_h} &= \left\{ \underline{v_h} \in \underline{V} : \underline{v_h} \in Q(r+1, r) \times Q(r, r+1), \ \forall T \in \tau_h \right\}; \\
W_h &= \left\{ w_h \in W : w_h \in Q(r, r), \ \forall T \in \tau_h \right\}.
\end{aligned} \tag{1.6}
$$

Освен това функциите $v_{i,h} \in V_{i,h}^r$ са непрекъснати съответно по x_i, $i = 1, 2$ и следователно (виж Следствие 0.1.1) $\underline{v_h} \in \underline{H}(div; \Omega)$.

Ще въведем базис от възли за тези пространства, използвайки основния краен елемент $\hat{T} = [-1, 1]^2$. Разглеждаме Гаусовите и Лобатовите точки в интервала [-1,1], както следва:

$$
\begin{aligned}
\hat{G} &= \{ \hat{g}_i : L_{r+1}(\hat{g}_i) = 0, \quad i = 1, \dots, r+1 \}; \\
\hat{L} &= \left\{ \hat{l}_j : L'_{r+1}(\hat{l}_j) = 0, \quad j = 1, \dots, r, \ \hat{l}_0 = -1, \ \hat{l}_{r+1} = 1 \right\},
\end{aligned}
$$

където $L_{r+1}(\xi)$ е полиномът на Льожандър от степен $r+1$, а $L'_{r+1}(\xi)$ е неговата производна. Например ако $r = 0$ то $\hat{l}_0 = -1$, $\hat{l}_1 = 1$, $\hat{g}_1 = 0$; ако $r = 1$, то $\hat{l}_0 = -1$, $\hat{l}_1 = 0$, $\hat{l}_2 = 1$, $\hat{g}_1 = -\sqrt{3}/3$, $\hat{g}_2 = \sqrt{3}/3$, както е показано на Фигура 1.1.

Разглеждаме основния елемент $\hat{T} = [-1, 1]^2$ и дефинираме пространствата, $\underline{\hat{V}}$ и $\hat{W}$, върху $\hat{T}$:

$$
\begin{aligned}
\underline{\hat{V}}(\hat{x}_1, \hat{x}_2) &= \left\{ \underline{\hat{v}} \in \underline{\hat{V}} : \underline{\hat{v}} \in Q(r+1, r) \times Q(r, r+1) \text{ върху } \hat{T} \right\}; \\
\hat{W}(\hat{x}_1, \hat{x}_2) &= \left\{ \hat{w} \in \hat{W} : \hat{w} \in Q(r, r) \text{ върху } \hat{T} \right\}.
\end{aligned} \tag{1.7}
$$

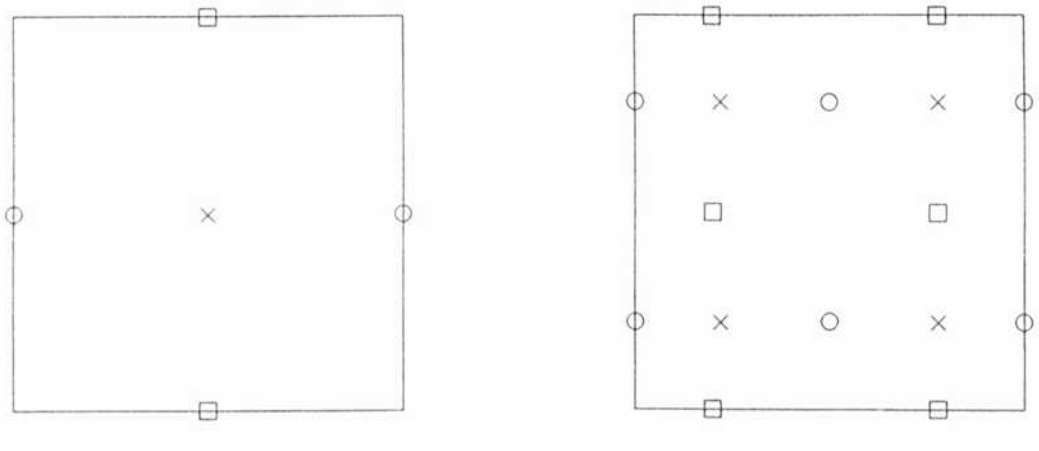
а) $r = 0$ б) $r = 1$

Фигура 1.1: Гаусови и Лобатови точки в интервала $[-1, 1]$.

Тогава функцията $\underline{\hat{u}} \equiv (\hat{u}_1, \hat{u}_2) \in \underline{\hat{V}}$ се определя по единствен начин чрез стойностите в точките $\hat{S}_1 \times \hat{S}_2$, а функцията $w \in W_h$ – в точките $\hat{S}_0$, където

$$
\begin{aligned}
\hat{S}_0 &= \{(\hat{g}_i, \hat{g}_j), \quad i, j = 1, \ldots, r+1\}; \\
\hat{S}_1 &= \left\{(\hat{l}_j, \hat{g}_i), \quad j = 0, \ldots, r+1, \ i = 1, \ldots, r+1\right\}; \\
\hat{S}_2 &= \left\{(\hat{g}_i, \hat{l}_j), \quad i = 1, \ldots, r+1, \ j = 0, \ldots, r+1\right\}.
\end{aligned} \tag{1.8}
$$

Случаите $r = 0$ и $r = 1$ са илюстрирани на Фигура 1.2. Възлите $\times, \circ$ и $\square$ съответстват на $\hat{S}_0$, $\hat{S}_1$ и $\hat{S}_2$.



а) $r = 0$ б) $r = 1$

Фигура 1.2: Базис от възли в основния елемент $\hat{T} = [-1, 1]^2$.

За всеки елемент $T \in \tau_h$, дефинирано в (1.4), съществува единствено обратимо афинно изображение $F_T : \hat{T} \to T$ в съответствие с (0.5), такова, че

$$\hat{x} \to x = F_T(\hat{x}) = B_T\hat{x} + \underline{b}_T, \tag{1.9}$$

където, във връзка с избора ни на основния краен елемент $\hat{T}$,

$$det(B_T) = \frac{1}{4}meas(T).$$

Изображението F_T въвежда мрежа от възли

$$S_i = S_i(T) = F(\hat{S}_i), \qquad i = 0, 1, 2 \tag{1.10}$$

върху всеки краен елемент $T \in \tau_h$. Тогава степените на свобода във всеки краен елемент, асоциран с пространствата, дефинирани в (1.5) и (1.6), са съответните стойности на функциите в точките от S_0, S_1 и S_2.

За всяка скаларнозначна функция w, дефинирана в T, аналогично на (0.10), полагаме:

$$w(x) = \hat{w}(\hat{x}), \quad \text{където } x = F_T(\hat{x}), \quad \forall \hat{x} \in \hat{T}. \tag{1.11}$$

Дефинираме функциите

$$w^I \in W_h^r, \quad v_1^I \in V_{1,h}^r, \quad v_2^I \in V_{2,h}^r, \tag{1.12}$$

които върху всеки краен елемент $T \in \tau_h$ се явяват интерполанти в съответствие с (0.13), (0.14) и (0.15) съответно в точките S_0, S_1 и S_2.

Апроксимацията по смесения метод на крайните елементи на (1.1) се дефинира по следния начин: търсим двойката $(\underline{u}_h, p_h) \in \underline{V_h} \times W_h$, такава, че

$$\left| \begin{array}{ll} a(\underline{u}_h, \underline{v}_h) + b(\underline{v}_h, p_h) = 0, & \forall \underline{v}_h \in \underline{V_h}; \\ b(\underline{u}_h, w_h) = -(f, w_h), & \forall w_h \in W_h. \end{array} \right. \tag{1.13}$$

Условията за устойчивост на задачата (1.13) и единственост на нейното решение са установени от Бабушка [1] и Бреци [1]. Това са добре известните условия на Бабушка-Бреци, които представяме в следното твърдение:

Лема 1.3.1 *Допускаме, че от*

$$\begin{cases} \underline{v_h} \in \underline{V_h} \\ \forall w_h \in W_h, \qquad b(\underline{v_h}, w_h) = 0 \end{cases}$$

следва $div\,\underline{v_h} = 0$ *и съществува независима от* h *константа* $\beta > 0$, *такава, че*

$$\inf_{w_h \in W_h} \sup_{\underline{v_h} \in \underline{V_h}} \frac{b(\underline{v_h}, w_h)}{\|p_h\|_{0,\Omega} \|\underline{v_h}\|_{\underline{H}(div;\Omega)}} \geq \beta.$$

Тогава задачата (1.13) има единствено решение $(\underline{u_h}, p_h) \in \underline{V_h} \times W_h$ *и съществува константа* C, *независеща от* h, *такава, че:*

(1.14)

$$\|\underline{u} - \underline{u_h}\|_{\underline{H}(div;\Omega)} + \|p - p_h\|_{0,\Omega} \leq C \left\{ \inf_{\underline{v_h} \in \underline{V_h}} \|\underline{u} - \underline{v_h}\|_{\underline{H}(div;\Omega)} + \inf_{w_h \in W_h} \|p - w_h\|_{0,\Omega} \right\}.$$

По-късно Фолк и Осборн [1] показват, че условията на Лема 1.3.1 са в сила, ако са изпълнени следните съотношения:

(1.15)

$$(i) \quad \forall w_h \in W_h^r, \ \exists \underline{v_h} \in \underline{V_h^r}, \text{ такова, че } div\,\underline{v_h} = w_h;$$

$$(ii) \quad \|\underline{v_h}\|_{\underline{H}(div;\Omega)} = \left[\|\underline{v_h}\|_{0,\Omega}^2 + \|div\,\underline{v_h}\|_{0,\Omega}^2 \right]^{1/2} \leq C \|w_h\|_{0,\Omega}.$$

За да приложим абстрактната оценка на грешката (1.14), първо се нуждаем от следното помощно твърдение:

Лема 1.3.2 *Гореописаните пространства* W_h^r *и* $\underline{V_h^r}$ *удовлетворяват условията на Бабушка-Бреци.*

Доказателство. Ще докажем, че са изпълнени условията (1.15). За произволна функция $w_h \in W_h$ ще конструираме фукция $\underline{v_h} \in \underline{V_h}$, такава, че $\underline{v_h} \in \underline{H}(div;\Omega)$ и $div\underline{v_h} = w_h$. Тъй като $\underline{v_h} = (v_{1,h}, v_{2,h})$, в съответствие със Следствие (0.1.1), принадлежността на $\underline{v_h}$ към $\underline{H}(div;\Omega)$, съблюдаваме като държим сметка за непрекъснатостта на компонентата $v_{i,h}$ по съответното направление x_i, $i = 1,2$.

Първо, конструираме функцията $v_{1,h}(x_1,x_2)$ по рекурсивен начин, разглеждайки хоризонтална лента от елементи T_{i,j_0} при фиксирано j_0 и $i = 1,2,\ldots,N$.

За всяко $w_h(x_1,x_2) \in W_h$ дефинираме следната помощна функция $\varphi_1(x_1,x_2)$:

$$\varphi_1(x_1,x_2) = \frac{1}{2}\int_0^{x_1} w(s,x_2)ds; \tag{1.16}$$

С помощта на (1.16) конструираме функцията $v_{1,h}$, както следва:

$$v_{1,h}(x_1,x_2)_{|T_{1,j_0}} = \varphi_1(x_1,x_2)_{|T_{1,j_0}}; \tag{1.17}$$

$$v_{1,h}(x_1,x_2)_{|T_{i,j_0}} = \varphi_1(x_1,x_2)_{|T_{i,j_0}} + \psi_1(x_2)_{|T_{i,j_0}}, \quad i = 2,3,\ldots,N, \tag{1.18}$$

където

$$\psi_1(x_2)_{|T_{i,j_0}} = v_1(x_{1,i-1},x_2)_{|T_{i-1,j_0}} - \varphi_1(x_{1,i-1},x_2)_{|T_{i,j_0}}.$$

Ролята на функцията $\psi_1(x_2)$ е да подсигури непрекъснатост на функцията $v_{1,h}$ по границата на крайните елементи. Същевременно тя е функция само на променливата x_2 и затова нейната производната по x_1 е равна на 0, което е от съществено значение при установяване на условието $(1.15i)$.

По същия начин конструираме функцията $v_{1,h}$ върху всички ленти от крайни елементи $(j_0 = 1,2,\ldots,M)$.

Целият анализ, извършен за първата компонента $v_{1,h}$ на функцията $\underline{v_h}$, важи в пълна сила и за втората компонента $v_{2,h}$ с единствената разлика,

че в разглеждането на вертикални ленти от крайни елементи, направлението x_1 е заменено с направлението x_2 и помощната функция ψ_2 зависи само от променливата x_1 и

$$\varphi_2(x_1, x_2) = \frac{1}{2}\int_0^{x_2} w(x_1, s)ds; \tag{1.19}$$

Така конструираната функция $\underline{v_h} = (v_{1,h}, v_{2,h})$ удовлетворява съотношенията:

$$\underline{v_h}(x_1, x_2)_{|T_{i,j}} \in Q(r+1, r) \times Q(r, r+1); \tag{1.20}$$

$$\underline{v_h}(x_1, x_2) \in \underline{H}(div; \Omega); \tag{1.21}$$

$$div\underline{v_h}(x_1, x_2) = \frac{\partial v_1}{\partial x_1} + \frac{\partial v_2}{\partial x_2} = w(x_1, x_2), \tag{1.22}$$

което означава, че изискването $1.15(i)$ е изпълнено.

Удовлетворяването на изискването $1.15(ii)$ установяме чрез непосредствена проверка. За да облекчим означенията, ще извършим пресмятанията, използвайки функцията $\underline{\varphi} \equiv (\varphi_1, \varphi_2)$, дефинирана чрез (1.16) и (1.19), тъй като функцията $\underline{v_h}$ се конструира рекурсивно с помощта на φ. Прилагайки неравенството на Коши-Шварц, получаваме:

(1.23)

$$\begin{aligned}
&\|\varphi\|^2_{0,\Omega} \\
&= \int_\Omega \varphi_1^2(x_1, x_2)dx_1dx_2 + \int_\Omega \varphi_2^2(x_1, x_2)dx_1dx_2 \\
&= \int_\Omega \left(\frac{1}{2}\int_0^{x_1} w(s, x_2)ds\right)^2 dx_1dx_2 + \int_\Omega \left(\frac{1}{2}\int_0^{x_2} w(x_1, s)ds\right)^2 dx_1dx_2 \\
&= \int_0^1\int_0^1 \left(\frac{1}{2}\int_0^{x_1} 1.w(s, x_2)ds\right)^2 dx_1dx_2 + \int_0^1\int_0^1 \left(\frac{1}{2}\int_0^{x_2} 1.w(x_1, s)ds\right)^2 dx_1dx_2 \\
&\leq \frac{1}{4}\left[\int_0^1\int_0^1 \left(\int_0^1 ds\int_0^1 w^2(s, x_2)ds\right)dx_1dx_2 + \int_0^1\int_0^1 \left(\int_0^1 ds\int_0^1 w^2(x_1, s)ds\right)dx_1dx_2\right] \\
&= \frac{1}{4}\left(\int_0^1 \|w\|^2_{0,\Omega}dx_1 + \int_0^1 \|w\|^2_{0,\Omega}dx_2\right) \\
&= \tfrac{1}{2}\|w\|^2_{0,\Omega}.
\end{aligned}$$

Следователно:

$$\|\underline{v}\|^2_{0,\Omega} \leq C\|w\|^2_{0,\Omega}\,. \tag{1.24}$$

От условие $1.15(i)$ следва, че

$$\|div\underline{v}\|^2_{0,\Omega} = \|w\|^2_{0,\Omega}\,. \tag{1.25}$$

По дефиниция:

$$\|\underline{v}\|^2_{\underline{H}(div,\Omega} = \|\underline{v}\|^2_{0,\Omega} + \|div\underline{v}\|^2_{0,\Omega}\,. \tag{1.26}$$

Съотношенията (1.24), (1.25) и (1.26) показват валидността на условие $1.15(ii)$.

Това завършва доказателството на лемата.

Следващото твърдение дава оптимална оценка на грешката за задачата (1.13).

Теорема 1.3.1 *Допускаме, че решението на задачата (1.2)* $(\underline{u}, p) \in (H^{r+2}(\Omega))^2 \times H^{r+3}(\Omega)$ *за някакво цяло число* $r \geq 0$. *Нека пространствата* $\underline{V_h}$ *и* W_h *са асоциирани с регулярно разделяне* τ_h *на областта* Ω. *Тогава задачата (1.13) има единствено решение* $(\underline{u_h}, p_h) \in (\underline{V_h} \times W_h)$ *и съществува константа* C, *независеща от* h, *такава, че:*

$$\|\underline{u} - \underline{u_h}\|_{\underline{H}(div;\Omega)} + \|p - p_h\|_{0,\Omega} \leq Ch^{r+1}\left(\|\underline{u}\|_{r+2,\Omega} + \|p\|_{r+1,\Omega}\right). \tag{1.27}$$

Доказателство. Използвайки Лема 0.1.3, Лема 0.1.4 и Лема 0.1.5 в случая $\hat{\Omega} \equiv \hat{T}$, $\underline{v} = \underline{u}$ и стандартното сумиране на нормите по крайните елементи, получаваме:

$$\|p - w^I\|_{0,\Omega} \leq Ch^{r+1}\|p\|_{r+1,\Omega}\,. \tag{1.28}$$

и

$$\|\underline{u} - \underline{v^I}\|_{\underline{H}(div;\Omega)} \leq Ch^{r+1}\|\underline{u}\|_{r+2,\Omega}. \tag{1.29}$$

Прилагането на Лема 1.3.2 и неравенства (1.28) и (1.29) , съответно за $\underline{v_h} = \underline{u^I} \in \underline{V_h}$ и $w_h = p^I \in W_h$, в Лема 1.3.1 завършва доказателството на теоремата.

1.4 Числено интегриране с концентрация на матрицата на масата

Съотношенията (0.5) и (0.10) ни дават следното съответствие между интегралите върху произволен елемент T и базисния $\hat{T}$.

$$\int_T \varphi(x)dx = det(B_T)\int_{\hat{T}} \hat{\varphi}(\hat{x})d\hat{x}\,.$$

С други думи, пресмятането на интеграла $\int_T \varphi(x)dx$ е равносилно на пресмятането на интеграла $\int_{\hat{T}} \hat{\varphi}(x)d\hat{x}$. Квадратурната схема върху базисния елемент $\hat{T}$ се състои в замяна на интеграла $\int_{\hat{T}} \hat{\varphi}(x)d\hat{x}$ с крайна сума от вида $\sum_{l=1}^{L} \hat{\omega}_l\hat{\varphi}(\hat{b}_l)$. Този процес на апроксимация ще бележим символично с

$$\int_{\hat{T}} \hat{\varphi}(x)d\hat{x} \sim \sum_{l=1}^{L} \hat{\omega}_l\hat{\varphi}(\hat{b}_l)$$

Квадратурните формули, които се използват за апроксимиране на интегралите в базисния краен елемент индуцират съответни квадратурни формули в произволен краен елемент:

$$\int_T \varphi(x)dx \sim \sum_{l=1}^{L} \omega_{l,T}\varphi(b_{l,T})$$

с тегла $\omega_{l,T}$ и с възли $b_{l,T}$, определени с равенствата

$$\omega_{l,T} = det(B_T)\hat{\omega}_l \quad \text{и} \quad b_{l,T} = F_T(\hat{b}_l)\,, \qquad 1 \leq l \leq L.$$

Въвеждаме функционалите на грешките за съответните квадратури:

$$E_T(\varphi) = \int_T \varphi(x)dx - \sum_{l=1}^{L} \omega_{l,T}\varphi(b_{l,T}),$$

$$\hat{E}(\hat{\varphi}) = \int_{\hat{T}} \hat{\varphi}(\hat{x})d\hat{x} - \sum_{l=1}^{L} \hat{\omega}_l\hat{\varphi}(\hat{b}_l),$$

които са свързани със следното съотношение:

$$E_T(\varphi) = det(B_T)\,\hat{E}(\hat{\varphi}) \tag{1.30}$$

Ще заменим билинейната форма $a(\cdot,\cdot)$ в (1.2) с $a_h(\cdot,\cdot)$, използвайки числено интегриране. Тъй като интегрирането се извършва отделно по всеки краен елемент, първо ще го дефинираме в основния елемент $\hat{T}$. Разглеждаме билинейната форма

(1.31)
$$\hat{a}(\underline{\hat{u}},\underline{\hat{v}}) = \hat{a}_1(\hat{u}_1,\hat{v}_1) + \hat{a}_2(\hat{u}_2,\hat{v}_2) = \int_{\hat{T}} \hat{\alpha}_1\hat{u}_1\hat{v}_1\,d\hat{x} + \int_{\hat{T}} \hat{\alpha}_2\hat{u}_2\hat{v}_2\,d\hat{x}$$

и съответстващата и апроксимация с квадратури

$$\hat{a}_h(\underline{\hat{u}},\underline{\hat{v}}) = \hat{a}_{h_1}(\hat{u}_1,\hat{v}_1) + \hat{a}_{h_2}(\hat{u}_2,\hat{v}_2), \tag{1.32}$$

където

$$\hat{a}_{h_1}(\hat{u}_1,\hat{v}_1) = \sum_{j=0}^{r+1}\sum_{i=1}^{r+1} \hat{\omega}_{ij}\hat{\alpha}_1\hat{u}_1\hat{v}_1(\hat{l}_j,\hat{g}_i),\ (\hat{l}_j,\hat{g}_i)\in\hat{S}_1,\ \hat{\omega}_{ij}>0,$$

$$\hat{a}_{h_2}(\hat{u}_2,\hat{v}_2) = \sum_{i=1}^{r+1}\sum_{j=0}^{r+1} \hat{\omega}_{ij}\hat{\alpha}_2\hat{u}_2\hat{v}_2(\hat{g}_i,\hat{l}_j),\ (\hat{g}_i,\hat{l}_j)\in\hat{S}_2,\ \hat{\omega}_{ij}>0.$$

Така конструираните квадратурни формули се явяват декартово произведение на Гаусови и Лобатови квадратури и следователно (виж Мисовских [1] или Николски [1]) са точни за полиноми от $Q(2r+1,2r+1)$.

Използвайки (1.32), дефинираме $a_h(\cdot,\cdot)$ в съответствие с трансформациите на възлите и функциите, дефинирани в (1.9), (1.10) и (1.11):

(1.33)
$$a_{h_1}(u_1,v_1)=\sum_{j=0}^{r+1}\sum_{i=1}^{r+1}\omega_{ij}\alpha_1 u_1 v_1(l_j,g_i),\ (l_j,g_i)\in S_1,\ \omega_{ij}>0\,;$$
$$a_{h_2}(u_2,v_2)=\sum_{i=1}^{r+1}\sum_{j=0}^{r+1}\omega_{ij}\alpha_2 u_2 v_2(g_i,l_j),\ (g_i,l_j)\in S_2,\ \omega_{ij}>0\,.$$

Забележка 1.4.1 *Да отбележим, че* $\underline{v_h}=(v_{1,h},v_{2,h})$ *и* $(\cdot,\cdot)_h$ *е апроксимация на скаларно произведение, използвайки квадратурни формули в съответствие с (0.10), и (1.32) води до диагонална матрица на масата (концентрация на масата), тъй като възлите на квадтатурните формули съвпадат с възлите на съответните апроксимационни пространства.*

Представяме задачата (1.1), както следва: търсим $(\underline{u_h^*},p_h^*)\in\underline{V_h}\times W_h$, удовлетворяваща условията

(1.34)
$$\left|\begin{array}{ll} a_h(\underline{u_h^*},\underline{v_h})+b(\underline{v_h},p_h^*)=0, & \forall\underline{v_h}\in\underline{V_h};\\ b(\underline{u_h^*},w_h)=-(f,w_h), & \forall w_h\in W_h\,.\end{array}\right.$$

За доказателството на основния резултат в тази глава (Теорема 1.4.1) ще използваме абстрактната оценка на грешката за задачата (1.34), представена от Робъртс и Тома [1] (Теорема 11.2.). Очевидно, при отсъствие на числено интегриране на билинейната форма $b(\cdot,\cdot)$ и на интегралите в дясната част на второто уравнение, достигаме до следното твърдение:

Теорема 1.4.1 *Нека* $a(\cdot,\cdot)$ *и* $b(\cdot,\cdot)$, *дефинирани в (1.3), непрекъснати билинейни форми съответно върху* $\underline{V}\times\underline{V}$ *и* $\underline{V}\times W$. *Тогава, ако съществуват константи* C_1 *и* C_2 *за задачата (1.34), такива, че*

(1.35)
$$a_h(\underline{u_h},\underline{v_h})\le C_1\|\underline{u_h}\|_{\underline{V}}\|\underline{v_h}\|_{\underline{V}},\quad\forall\underline{v_h},\ \underline{u_h}\in\underline{V_h}\,;$$

(1.36) $$a_h(\underline{v_h}, \underline{v_h}) \geq C_2 \|\underline{v_h}\|^2_{\underline{V}}, \quad \forall \underline{v_h} \in \underline{V_h^*},$$

където

$$\underline{V_h^*} = \left\{ \underline{v_h} \in \underline{V_h} \, : \, (div \underline{v_h}, w_h) = 0, \ \forall w_h \in W_h \right\},$$

то е в сила оценката

(1.37) $$\|\underline{u} - \underline{u_h^*}\|_{\underline{V}} + \|p - p_h^*\|_W \leq$$

$$C \left\{ \inf_{\underline{v_h} \in \underline{V_h}} \|\underline{u} - \underline{v_h}\|_{\underline{V}} + \sup_{\underline{t_h} \in \underline{V_h}} \frac{a(\underline{v_h}, \underline{t_h}) - a_h(\underline{v_h}, \underline{t_h})}{\|\underline{t_h}\|_{\underline{V}}} + \inf_{w_h \in W_h} \|p - w_h\|_W \right\}.$$

Нашата цел ще бъде с помощта на няколко леми да установим наличието на необходимите свойства на използваните квадратури, които се изискват в цитираната теорема, и да получим съответна оценка на грешката за допълнителния член в дясната страна на неравенството, възникващ в резултат на численото интегриране.

Лема 1.4.1 *Съществуват константи* C_1 *и* C_2 *, удовлетворяващи изискванията (1.35) и (1.36).*

Доказателство. Нека $a : \underline{V} \times \underline{V} \to \mathbb{R}$ е ограничена билинейна форма, дефинирана чрез (1.3).

Първо доказваме, че изображението

$$\hat{v}_1 \in Q(r+1, r) \to [a_{h_1}(\hat{v}_1, \hat{v}_1)]^{1/2}$$

е норма в $Q(r+1, r)$. Заедно с другите очевидни свойства на нормата трябва да е изпълнено и свойството

$$[a_{h_1}(\hat{v}_1, \hat{v}_1)]^{1/2} = 0 \qquad \text{точно когато } \hat{v}_1 = 0.$$

Достатъчно е да отбележим, че $\hat{v}_1 = 0$ в $(r+2)\times(r+1)$ точки, от което следва, че $\hat{v}_1$ се анулира тъждествено върху $\hat{T}$. Следователно мрежата от точките на Лобато - Гаус е $Q(r+1,r)$ - унисолвентно множество върху $\hat{T}$. От еквивалентността на нормите в крайномерни пространства следва съществуването на константи $\hat{C}_1,\ \hat{C}_2 > 0$, такива, че

$$\hat{C}_1 \hat{a}_{h_1}(\hat{v}_1, \hat{v}_1) \leq \|\hat{v}_1\|^2_{0,\hat{T}} \leq \hat{C}_2 \hat{a}_{h_1}(\hat{v}_1, \hat{v}_1)$$

Използвайки, че

$$a_{h_1,T}(v_1, v_1) \ = \ (det(B_T))\, \hat{a}_{h_1}(\hat{v}_1, \hat{v}_1)$$

и

$$\|w\|_{0,T} \ = \ (det(B_T))^{1/2}\, \|\hat{w}\|_{0,\hat{T}}, \quad \forall w \in L^2(T)\,,$$

получаваме

$$C_3\, a_{h_1,T}(v_1, v_1) \ \leq \ \|v_1\|^2_{0,T} \ \leq \ C_4\, a_{h_1,T}(v_1, v_1)\,.$$

Напълно аналогични разсъждения дават съответната релация за втората компонента v_2 и квадратурната формула $a_{h_2,T}(v_2, v_2)$:

$$C_5 a_{h_2,T}(v_2, v_2) \leq \ \|v_2\|^2_{0,T} \leq C_6 a_{h_2,T}(v_2, v_2)\,.$$

Използвайки следните очевидни равенства за всяко $\underline{v_h} \in (L^2(\Omega))^2$

$$\|\underline{v_h}\|^2_{0,T} \ = \ \|v_{1,h}\|^2_{0,T} + \|v_{2,h}\|^2_{0,T}\,;$$

$$\|v_{i,h}\|^2_{0,\Omega} \ = \sum_{T\in\tau_h} \|v_{i,h}\|^2_{0,T}, \qquad i = 1,2\,;$$

$$a_h(\underline{v_h}, \underline{v_h}) \ = \sum_{T\in\tau_h} a_{h,T}(\underline{v_h}, \underline{v_h}),$$

за всяко $\underline{v_h} \in \underline{V_h}$ получаваме

$$C_7 a_h(\underline{v_h}, \underline{v_h}) \leq \|\underline{v_h}\|^2_{0,\Omega} \leq C_8 a_h(\underline{v_h}, \underline{v_h}) . \tag{1.38}$$

Да отбележим, че

$$a_h(\underline{u_h}, \underline{v_h}) \leq (a_h(\underline{u_h}, \underline{u_h}))^{1/2} (a_h(\underline{v_h}, \underline{v_h}))^{1/2} . \tag{1.39}$$

Тъй като за всяко $\underline{v_h} \in \underline{V_h}$ съществува $w_h \in W_h$, такова, че $div\underline{v_h} = w_h$, то за всяко $\underline{v_h} \in \underline{V_h}^*$ $\|div\underline{v_h}\|_{0,\Omega} = 0$. Следователно

$$\|\underline{v_h}\|_{0,\Omega} = \|\underline{v_h}\|_{\underline{V}} \tag{1.40}$$

Така неравенствата (1.35) и (1.36) следват от (1.38), (1.39) и (1.40).

Важен инструмент в по-нататъшния анализ е лемата на Брамбъл - Хилберт (виж например Брамбъл - Хилберт [1] или Сиарле [1]):

Лема 1.4.2 *(Брамбъл - Хилберт) Нека Ω е отворено подмножество на $\mathbb{R}^n$ с непрекъсната по Липшиц граница. Нека, освен това, за някое цяло число $k \geq 0$ и някое число $p \in [0, \infty]$ f е непрекъснат линеен функционал върху пространството $W^{k+1,p}(\Omega)$, притежаващ свойството:*

$$\forall p \in P_k(\Omega) \qquad f(p) = 0 . \tag{1.41}$$

Тогава съществува константа $C(\Omega)$, такава, че

$$\forall w \in W^{k+1,p}(\Omega) \ |f(w)| \leq C(\Omega)\|f\|^*_{k+1,p,\Omega}|w|_{k+1,p,\Omega} ,$$

където $\|\cdot\|^*_{k+1,p,\Omega}$ е норма в двойственото на $W^{k+1,p}(\Omega)$ пространство.

Ще се нуждаем също и от следния помощен резултат (виж Сиарле [1]):

Лема 1.4.3 *Дадени са функциите* $\varphi \in W^{m,q}(\Omega)$ *и* $w \in W^{m,\infty}(\Omega)$. *Тогава функцията* φw *принадлежи на пространството* $W^{m,q}(\Omega)$ *и*

$$|\varphi w|_{m,q,\Omega} \leq C \sum_{j=0}^{m} |\varphi|_{m-j,q,\Omega} |w|_{j,\infty,\Omega}\,, \tag{1.42}$$

където константата C *не зависи от множеството* Ω.

Лема 1.4.4 *Нека за някакво цяло число* $r \geq 0$ *квадратурната схема е точна за полиноми от* $Q(2r+1, 2r+1)$ *т.е.*

$$\forall \varphi \in Q(2r+1, 2r+1)\,, \qquad E(\varphi) = 0\,. \tag{1.43}$$

Тогава съществува константа C *(независеща от* h *и* $T \in \tau_h$*), такава, че*

$$\forall a \in W^{r+1,\infty}(T), \quad p \in Q(r+1, r)(T), \quad q \in Q(r+1, r)(T),$$

$$|E_T(apq)| \leq C h_T^{r+1} \|p\|_{r+1,T} \|q\|_{0,T}\,. \tag{1.44}$$

Доказателство. От съотношението (1.30) знаем, че

$$E_T(apq) = det(B_T) \hat{E}(\hat{a}\hat{p}\hat{q})$$

За дадената функция $\hat{q} \in Q_{r+1,r}(T)$ и функция $\hat{\varphi} \in W^{r+1,\infty}(\hat{T})$ $(W^{r+1,\infty}(\hat{T}) \hookrightarrow C^0(\hat{T})$, (т.к. $r \geq 0)$.

$$\left|\hat{E}(\hat{\varphi}\hat{q})\right| = \left|\int_T \hat{\varphi}\hat{q} d\hat{x} - \sum_{l=1}^{h} \hat{\omega}_l (\hat{\varphi}\hat{\omega})(b_l)\right| \leq \hat{C} |\hat{\varphi}\hat{q}|_{0,\infty,\hat{T}} \leq \hat{C} |\hat{\varphi}|_{0,\infty,\hat{T}} |\hat{q}|_{0,\infty,\hat{T}}.$$

Тъй като $|\hat{\varphi}|_{0,\infty,\hat{T}} \leq \|\hat{\varphi}\|_{r+1,\infty,\hat{T}}$ и от еквивалентността на нормите в крайномерното пространство $Q_{r+1,r}(T)$, получаваме

$$\hat{f}(\hat{\varphi}) = |\hat{E}(\hat{\varphi}\hat{q})| \leq C \|\hat{\varphi}\|_{r+1,\infty,\hat{T}} |\hat{q}|_{0,\hat{T}}\,.$$

За даденото $\hat{q}$ линейният функционал

$$\hat{\varphi} \in W^{r+1,\infty}(\hat{T}) \to \hat{E}(\hat{\varphi}\hat{q})$$

е непрекъснат при норма $\leq C|\hat{q}|_{0,T}$ и се анулира по силата на предположение (1.43) за полиноми

$$\hat{\varphi} \in P_r(\hat{T}).$$

Следователно, използвайки лемата на Брамбъл - Хилберт, получаваме, че съществува константа $\hat{C}$, такава, че

$$\forall \varphi \in W^{r+1,\infty}(\hat{T}),\ \forall q \in Q_{r+1,r}(\hat{T}) \quad |\hat{E}(\hat{\varphi}\hat{q})| \leq \hat{C}|\hat{\varphi}|_{r+1,\infty,\hat{T}}|\hat{q}|_{0,\hat{T}}\,.$$

Използваме, че $\hat{\varphi} = \hat{a}\hat{p}$ при $a \in W^{r+1,\infty}(\hat{T})$. От Лема 1.4.3 и отчитайки, че $|\hat{p}|_{r+1,\infty,\hat{T}} = 0$, получаваме

$$|\hat{\varphi}|_{r+1,\infty,\hat{T}} = |\hat{a}\hat{p}|_{r+1,\infty,\hat{T}} \leq \hat{C}\sum_{j=0}^{r}|\hat{a}|_{r+1-j,\infty,\hat{T}}|\hat{p}|_{j,\infty,\hat{T}} \leq \hat{C}\sum_{j=0}^{r}|\hat{a}|_{r+1-j,\infty,\hat{T}}|\hat{p}|_{j,\hat{T}}\,,$$

където в последното неравенство отново използваме енвивалентност на нормите в крайномерното пространство $Q(r+1,r)(\hat{T})$. Следователно получаваме

$$\forall a \in W^{r+1,\infty}(T)\,, \quad \forall p \in Q(r+1,r)(T)\,, \quad \forall q \in Q(r+1,r)(T)$$

$$\left|\hat{E}(\hat{a}\hat{p}\hat{q})\right| \leq \left(\sum_{j=0}^{r}|\hat{a}|_{r+1-j,\infty,\hat{T}}|\hat{p}|_{j,\hat{T}}\right)|\hat{q}|_{0,\hat{T}}\,. \tag{1.45}$$

Тогава достатъчно е да използваме Лема 0.1.1 и неравенство (0.7):

$$\begin{aligned}
&|\hat{a}|_{r+1-j,\infty,\hat{T}} \leq \hat{C}h_T^{r+1-j}|a|_{r+1-j,\infty,T}\,, && 0 \leq j \leq r\,;\\
&|\hat{p}|_{j,\hat{T}} \leq \hat{C}h_T^{j}(det(B_T))^{-1/2}|p|_{j,T}\,, && 0 \leq j \leq r\,;\\
&|\hat{q}|_{0,\hat{T}} \leq \hat{C}(det(B_T))^{-1/2}|q|_{0,T}\,,
\end{aligned} \tag{1.46}$$

заедно със съотношенията (1.30),(1.45). По този начин получаваме

$$\forall a \in W^{r+1,\infty}(T), \quad \forall p \in Q(r+1,r)(T), \quad q \in Q(r+1,r)(T)\,,$$

$$|E(apq)| \leq C h_T^{r+1} \|a\|_{r+1,\infty,T} \|p\|_{r,T} \|q\|_{0,T}\,. \tag{1.47}$$

Неравенство (1.47), имайки предвид, че константата може да зависи от функцията $a(x)$, завършва доказателството на лемата.

Забележка 1.4.2 *Оценката, получена в Лема 1.4.4, е в сила, ако заменим условието (1.44) със следното условие:*

$$\forall a \in W^{r+1,\infty}(T)\,, \quad \forall p \in Q(r,r+1)(T)\,, \quad \forall q \in Q(r,r+1)(T)\,,$$

което използваме в оценките за вторите компоненти на съответните векторнозначни функции.

За да приложим абстрактната оценка (1.37), се нуждаем и от следната помощна оценка:

Лема 1.4.5 *Нека* $(\underline{u}, p) \in \underline{V} \times W$ *е решение на (1.1),* $p \in H^{r+3}(\Omega)$ *и* $\underline{u}^I$ *е интерполантът на* $\underline{u}$*, дефиниран в (1.12) в съответствие с (0.10). Тогава съществува константа* C_9*, такава, че*

$$\sup_{\underline{t_h} \in \underline{V_h}} \frac{a(\underline{u}^I, \underline{t_h}) - a_h(\underline{u}^I, \underline{t_h})}{\|\underline{t_h}\|_{\underline{V}}} \leq C_9\, h^{r+1} \left(\|\underline{u}\|_{r+2,\Omega} + \|p\|_{r+1,\Omega} \right). \tag{1.48}$$

Доказателство. Първо оценяваме числителя на лявата страна на (1.48):

$$|a(\underline{u}^I, \underline{t_h}) - a_h(\underline{u}^I, \underline{t_h})| \leq \sum_{T \in \tau_h} |E_{1,T}(\alpha_1 u_1^I t_{1,h})| + |E_{2,T}(\alpha_2 u_2^I t_{2,h})|\,,$$

където

$$E_{i,T}(\alpha_i u_i^I t_{i,h}) = a_i(u_i^I, t_{i,h}) - a_{i,h}(u_i^I, t_{i,h}), \qquad i = 1, 2\,. \tag{1.49}$$

Използвайки Лема 1.4.3 за $i = 1$ и Забележка 1.4.2 за $i = 2$, получаваме следните неравенства:

$$|E_{i,T}(\alpha_i u_i^I t_{i,h})|_{0,T} \le C_i h^{r+1} \|u_i^I\|_{r+1,T} \|t_{i,h}\|_{0,T}, \qquad i = 1, 2\,, \tag{1.50}$$

където коефициентите C_i не зависят от параметъра h, а зависят само от α_i, които по условие притежават необходимите свойства за ограниченост.

Прилагайки неравенството на триъгълника и Теорема 0.1.3, получаваме следната оценка:

$$\|u_i^I\|_{r+1,T} \le \|u_i^I - u_i\|_{r+1,T} + \|u_i\|_{r+1,T} \le C\|u_i\|_{r+2,T}, \qquad i = 1, 2\,. \tag{1.51}$$

Така за числителя в лявата страна на неравенство (1.48) получаваме:

$$\begin{aligned}
&\left|a(\underline{u}^I, \underline{t}_h) - a_h(\underline{u}^I, \underline{t}_h)\right| \\
&\le Ch^{r+1} \left(\sum_{T \in \tau_h} \|u_1\|_{r+2,T} \|t_{1,h}\|_{0,T} + \sum_{T \in \tau_h} \|u_2\|_{r+2,T} \|t_{2,h}\|_{0,T} \right) \\
&\le Ch^{r+1} \left[\left(\sum_{T \in \tau_h} \|u_1\|_{r+2,T}^2 \right)^{1/2} \left(\sum_{T \in \tau_h} \|t_{1,h}\|_{0,T}^2 \right)^{1/2} \right. \\
&\quad \left. + \left(\sum_{T \in \tau_h} \|u_2\|_{r+2,T}^2 \right)^{1/2} \left(\sum_{T \in \tau_h} \|t_{2,h}\|_{0,T}^2 \right)^{1/2} \right] \\
&= Ch^{r+1} \left(\|u_1\|_{r+2,\Omega} \|t_{1,h}\|_{0,\Omega} + \|u_2\|_{r+2,\Omega} \|t_{2,h}\|_{0,\Omega} \right) \\
&= Ch^{r+1} \left(\|u_1\|_{r+2,\Omega} + \|u_2\|_{r+2,\Omega}\| \right) \left(\|t_{1,h}\|_{0,\Omega} + \|t_{2,h}\|_{0,\Omega} \right) \\
&= Ch^{r+1} \left(\|\underline{u}\|_{r+2,\Omega} + \|\underline{t}_h\|_{0,\Omega} \right) \\
&\le Ch^{r+1} \left(\|\underline{u}\|_{r+2,\Omega} + \|\underline{t}_h\|_V \right)\,.
\end{aligned} \tag{1.52}$$

Доказателството на лемата следва от веригата неравенства (1.52).

Доказаните по-горе твърдения дават възможност да представим основната оценка в тази глава.

Теорема 1.4.2 *Нека* $(\underline{u}, p)$ *е решение на (1.1) и* $(\underline{u_h}^*, p_h^*)$ *е решение на (1.34), получено чрез прилагане на числено интегриране с помощта на квадратурните формули (1.33). Предполагаме, че* $p \in H^{r+3}(\Omega)$. *Тогава съществува константа* C *такава, че*

$$\|\underline{u} - \underline{u_h}^*\|_{\underline{H}(div;\Omega)} + \|p - p_h^*\|_{0,\Omega} \leq C\, h^{r+1} \left(\|\underline{u}\|_{r+2,\Omega} + \|p\|_{r+1,\Omega}\right).$$

Доказателство. За доказателството на теремата е достатъчно да и приложим Лема 1.4.1, неравенства (1.28) и (1.29), съответно за $\underline{v_h} = \underline{u}^I \in \underline{V_h}$ и $w_h = p^I \in W_h$, и Лема 1.4.5 в Теорема 1.4.1.

Теорема 1.4.2 показва, че използването на гореописаните квадратурни формули, водещи до диагонализиране матрицата на масата, не променя порядъка на грешката на метода.

1.5 Съответстващата матрична задача

Въвеждаме векторни означения U_1, U_2, P за апроксимиране на неизвестните стойности $u_{1,h}^*, u_{2,h}^*, p_h^*$ съответно и V_1, V_2, W за $v_{1,h}, v_{2,h}, p_h$, както следва:

$$\begin{aligned} a_h(\underline{u_h}^*, \underline{v_h}) &= \sum_{i=1}^{2} U_i^T M_i V_i\,; \\ b(\underline{v_h}, p_h^*) &= \sum_{i=1}^{2} V_i^T N_i P\,; \\ b(\underline{u_h}^*, w_h) &= \sum_{i=1}^{2} W^T N_i^T U_i\,. \end{aligned}$$

Тогава системата алгебрични уравнения за определяне на приближеното решение, представено в матрична форма, е следното:

$$BY = \begin{pmatrix} M_1 & 0 & N_1 \\ 0 & M_2 & N_2 \\ N_1^T & N_2^T & 0 \end{pmatrix} \begin{pmatrix} U_1 \\ U_2 \\ P \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -F \end{pmatrix}, \tag{1.53}$$

където $M_i \in \mathbb{R}^{n\times n}$ и $N_i \in \mathbb{R}^{n\times m}$, $i = 1,2$ са матрици, а $U_1, U_2 \in \mathbb{R}^n$ и $P, F \in \mathbb{R}^m$ – вектори.

В стандартния случай на получаване на системата (1.53) матрицата B е обратима, но не е положително-определена. Този факт я прави трудна за преобуславяне, което води до съществени затруднения при прилагане на съответните итерационни процедури (виж Василевски, Лазаров [1]). Използването на квадратурни формули, чиито възли съвпадат с възлите на базисните функции (те са единици в един от възлите и нули във всички останали), води до диагонализация на матриците $M_i\,(i = 1,2)$. Това дава възможност за директно пресмятане на техните обратни $M_i^{-1}\,(i = 1,2)$ и задачата (1.53) автоматично се редуцира до задача със значително по-малка размерност:

$$KP = \left(N_1^T M_1^{-1} N_1 + N_2^T M_2^{-1} N_2\right) P = F\,,$$

където $K \in \mathbb{R}^{m\times m}$, $m < n$. В този случай $U_i = -M_i^{-1} N_i P$, $i = 1,2$.

Забележка 1.5.1 *В случая на елементи на Равиар-Тома от най-нисък ред, т.е.* $r = 0$, *това е стандартната 5-точкова апроксимация върху клетъчно-центрирана мрежа (виж Юинг, Лазаров и Василевски [1]).*

Забележка 1.5.2 *Ръсел и Уилър [1] показват, че в случая* $r = 0$ *стандартния петточков блочно-центриран метод на краайните разлики е еквивалентен на прилагането на подходящи квадратурни формули в смесения метод.*

Глава 2

Оценка на грешката върху правоъгълна мрежа с използване на „подчинени“ възли

2.1 Въведение

Концепцията за „подчинени“ възли се изучава от Брамбъл, Юинг, Пашек и Шатс [1] и Дрия и Видлунд [1] за стандартния метод на крайни елементи и крайните разлики. Тя е описана за смесения метод на крайните елементи от Юинг, Лазаров, Ръсел и Василевски [1]. Там са дадени ефективни алгоритми за решаване на недефинитната система алгебрични уравнения и са представени числени резултати.

Идеята на тази концепция се състои в разделяне на разглежданата област на подобласти Ω_1 и Ω_2, състоящи се съответно от „фини“ и „груби“ крайни

елементи, с цел да се отразят някой локални свойства на решението. Резултатите от числените експерименти за пространствата на Равиар и Тома с индекс r показват, че грешката на метода в Ω_1 и Ω_2 е пропорционална на h_f^{r+1} и h_c^{r+1}, където $h_c = nh_f$. Изключение правят само тези елементи от Ω_1, които имат общи страни с елементите от Ω_2 (елементите от интерфейса на Ω_1).

Целта на тази глава е получаването на оптимални (неподобряеми по порядък) оценки за смесения метод на крайните елементи върху правоъгълна мрежа с регулярно локално сгъстяване, като максимално се прецизира членът, в който участват елементите от интерфейса. Неподобряемостта на оценката следва от факта, че разликата между локалните оценки върху елементите от интерфейса и вътрешността на Ω_1 е намерена в явен вид.

В §3 въвеждаме специфичното разделяне на областта и конструираме апроксимиращите пространства върху съставната мрежа. В Лема 2.3.1 показваме, че тези пространства удовлетворяват условията на Бабушка-Бреци (виж Бабушка [1], Бреци [1] и Фолк и Осборн [1]).

Основния резултат в тази глава – оценка на грешката (Теорема 2.4.2) доказваме в §4, базирайки се на Лема 2.4.1, в която получаваме локални оценки, използвайки интерполационната техника. Представени са и някои приложения на получения резултат за задачата за приближено намиране на собствените стойности, използвайки локално сгъстяване на мрежата.

2.2 Постановка на задачата

Разглеждаме задачата на Дирихле

$$\begin{aligned} &\text{(а)} \quad -div(a(x_1,x_2)\underline{\nabla} p) = f, \quad \text{в } \Omega; \\ &\text{(б)} \quad p = -g, \qquad\qquad\qquad\quad \text{върху } \partial\Omega, \end{aligned} \tag{2.1}$$

където с $\partial\Omega$ е означена границата на $\Omega \in \mathbb{R}^2$, а $a(x) = diag\,(a_1(x), a_2(x))$ е диагонална матрица, чиито елементи удовлетворяват изискванията $a_i(x) \geq a_0 > 0$, $i = 1, 2$. От съображение за простота вземаме областта Ω да бъде квадратът $(0,1)^2$.

Полагаме

$$\underline{u} = a(x_1, x_2)\underline{\nabla} p, \quad \alpha_i(x_1, x_2) = a_i(x_1, x_2)^{-1}, \qquad i = 1, 2.$$

Нека

$$\underline{V} = \underline{H}(div; \Omega) = \{\underline{u} \equiv (u_1, u_2) \in (L^2(\Omega))^2 : div\underline{u} \in L^2(\Omega)\},$$

$$W = L^2(\Omega).$$

За $\underline{u}, \underline{v} \in \underline{V}$ и $p, w \in W$ дефинираме билинейните форми:

$$(2.2) \qquad \begin{aligned} &\text{(а)} \quad a(\underline{u}, \underline{v}) = (\alpha_1 u_1, v_1) + (\alpha_2 u_2, v_2); \\ &\text{(б)} \quad b(\underline{u}, w) = (div\underline{u}, w). \end{aligned}$$

Тогава задачата (2.1) е еквивалентна с решаването на задачата за седлова точка, дадена чрез

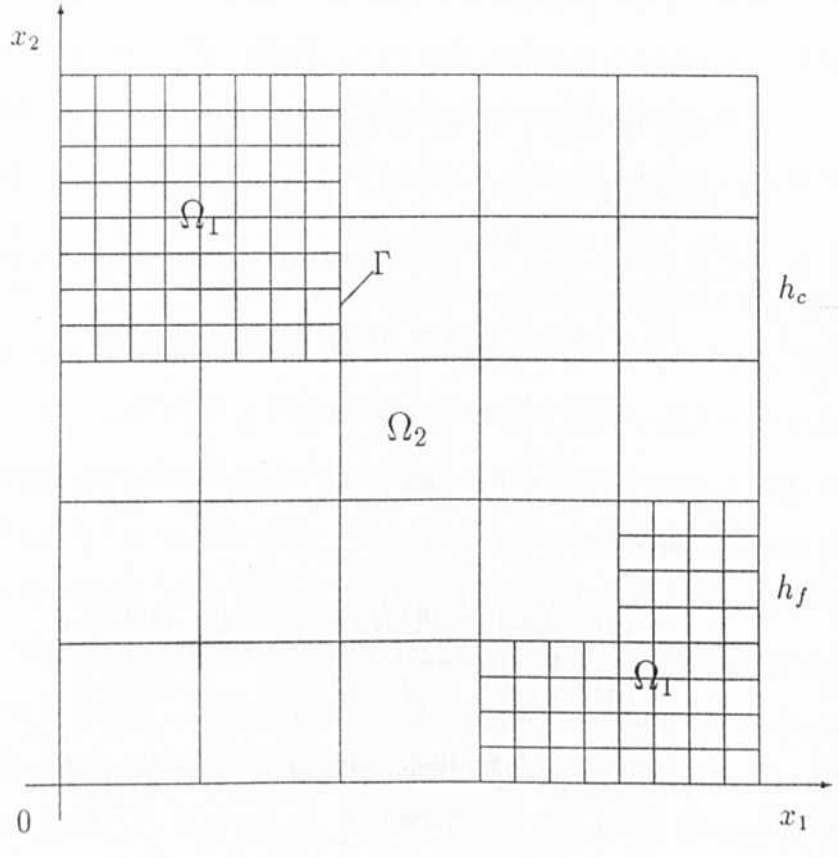
$$(2.3) \qquad \begin{aligned} &\text{(а)} \quad a(\underline{u}, \underline{v}) - b(\underline{v}, p) = \langle g, \underline{v}.\underline{\nu} \rangle \quad \forall \underline{v} \in \underline{V}; \\ &\text{(б)} \quad b(\underline{u}, w) = (f, w), \qquad\qquad\quad \forall w \in W, \end{aligned}$$

където $\underline{\nu}$ е единичният външен нормален вектор към $\partial\Omega$. Скаларното произведение в $(L^2(\Omega))^2$ е означено с $(\cdot, \cdot)$, а в $L^2(\partial\Omega)$ – с $\langle \cdot, \cdot \rangle$.

2.3 Апроксимация върху съставна мрежа с „подчинени“ възли.

Аналогично на (1.4), първо въвеждаме регулярно (в смисъл на (0.18)) грубо разделяне на областта Ω, означено с τ_c, с характерен параметър h_c. Нека

Ω_1 е предварително зададена подобласт на Ω, в която решението се изменя по-бързо от останалата част на Ω. Нашата цел е да получим по-прецизни оценки на грешката в Ω_1, благодарение на по-малкия параметър h_f. За тази цел повторно разделяме елементите от Ω_1, въвеждайки фина мрежа, означена с τ_f, както е показано на Фигура 2.1.



Фигура 2.1: Разделяне на областта с регулярно локално сгъстяване

Предполагаме, че това рафиниране (сгъстяване) е равномерно и $h_c = nh_f$ (виж например Юинг, Лазаров, Ръсел и Василевски [1]). По този начин получаваме три характерни подобласти на областта Ω, а именно Ω_1, Ω_2 и елементите от интерфейса, които образуват съответната подобласт $I_f \in \Omega_1$.

Съвкупността от „грубото“ разделяне τ_c и „финото“ τ_f обрауват съставната мрежа τ_h. Накратко, в сила са следните отношения:

$$
\begin{aligned}
&\text{(а)} \quad \Omega_2 = \Omega \backslash \Omega_1; \\
&\text{(б)} \quad I_f = \left\{ T \in \tau_f \subset \Omega_1 : T \bigcap \overline{\Omega}_2 \neq \emptyset \right\}; \\
&\text{(в)} \quad \tau_h = \tau_c \bigcup \tau_f .
\end{aligned} \tag{2.4}
$$

Разглеждаме пространството на Равиар и Тома $\underline{V_c^r}$, дефинирано в (1.5) и (1.6), асоциирано с триангулацията τ_c на областта Ω.

Нека $\underline{\overset{\circ}{V}{}_f^r}(\Omega_1)$ е пространство на Равиар и Тома, асоциирано с триангулацията τ_f на подобластта Ω_1, такова, че нормалната компонента на функцията $\underline{v} \in \underline{\overset{\circ}{V}{}_f^r}(\Omega_1)$ върху границата Γ на Ω_1 е нула. Дефинираме пространствата на Равиар и Тома върху съставната мрежа τ_h, както следва:

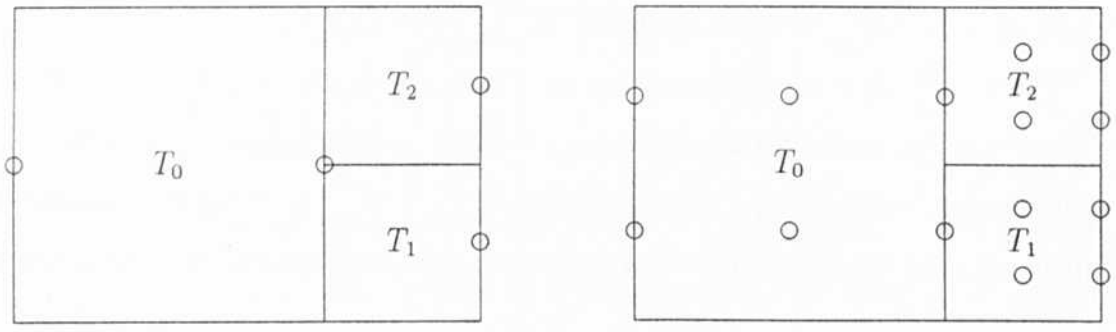
$$
\begin{aligned}
&\text{(а)} \quad \underline{V_h^r} = \underline{V_c^r} + \underline{\overset{\circ}{V}{}_f^r}(\Omega_1); \\
&\text{(б)} \quad W_h^r = \{ w \in L^2(\Omega), w(x_1, x_2) \in Q(r,r) \text{ за всяко} T \in \tau_h \}
\end{aligned} \tag{2.5}
$$

За крайните елементи $T \in \Omega_2 \bigcup \Omega_1 \backslash I_f$ въвеждаме мрежа от възли, дефинирана чрез (1.8), (1.9) и (1.10).

Забележка 2.3.1 *Изискването възлите в крайните елементи да са декартово произведение на Гаусови и Лобатови точки не е задължително. Важното в случая е те да възстановяват полиномите от съответните пространства по единствен начин.*

По различен начин конструираме възлите на елементите T_f от интерфейса I_f. Пример за разположението на възлите за компонентата $v_{1,h}$ е показан на Фигура 2.2. Част от възлите на елементите $T_f \in I_f$ не лежат върху тях, а върху общата им страна $\gamma = T_f \bigcap T_0$ и съвпадат с възлите на елемента

$T_0 \in \Omega_2$. Тези възли, от друга страна, са общи за всички елементи от интерфейса на Ω_1, такива, че $T_f \in I_f : T_f \bigcap T_0 \neq \emptyset$. Практически това означава наличието на фиктивни възли върху елементите $T_f \in I_f$, които се оказват „подчинени" на възлите върху γ. Този избор на възлите е в пряка връзка с изискването за непрекъснатост на първата компонента v_1 по направлението x_1.



а) $r = 0$ б) $r = 1$

Фигура 2.2: Базис от възли за елементите $T_0 \in \Omega_2$ и $T_1, T_2 \in I_f$ за компонентата v_1 при $n = 2$.

Аналогично дефинираме базисните възли за v_2 в случая, когато имаме преход от „груби" към „фини" крайни елементи по направлението x_2.

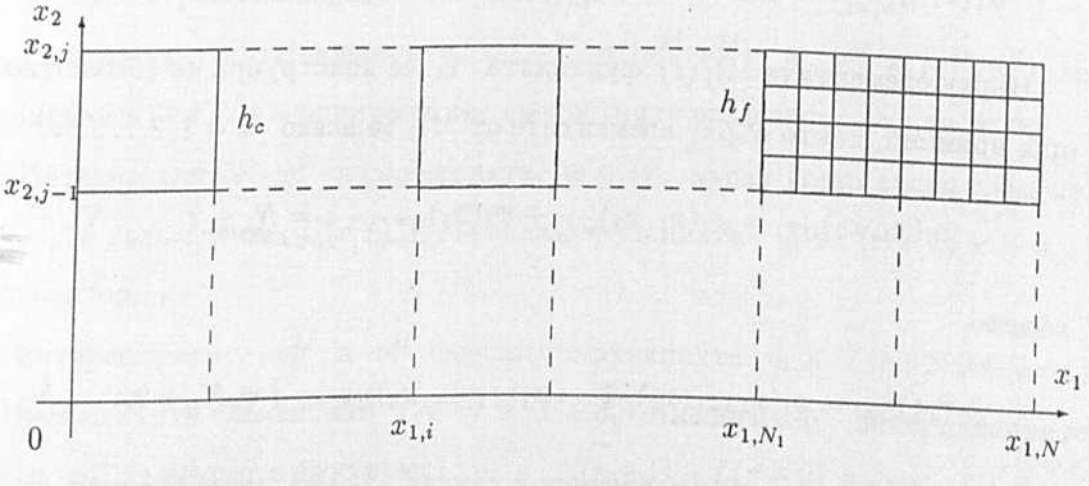
Формулировката на смесения метод на крайните елементи за задачата (2.1) се дефинира чрез определяне на двойката $\left(\underline{u_h}, p_h\right) \in \underline{V_h^r} \times W_h^r$ такава, че:

$$
\begin{aligned}
&\text{(а)} \qquad a(\underline{u_h}, \underline{v_h}) - b(\underline{v_h}, p_h) = \langle g, \underline{\nu}.\underline{v_h} \rangle, \quad \underline{v_h} \in \underline{V_h^r}; \\
&\text{(б)} \qquad b(\underline{u_h}, w_h) = (f, w). \qquad\qquad\qquad\quad w_h \in W_h^r.
\end{aligned}
\tag{2.6}
$$

За да използваме абстрактната оценка (1.14), трябва да установим, че са изпълнени условията на Лема 1.3.1.

Лема 2.3.1 *Пространствата* $\underline{V_h^r}$ *и* W_h^r, *дефинирани в (2.5), върху съставното разделяне на областта* Ω, *удовлетворяват условията на Бабушка - Бреци.*

Доказателство. Имайки предвид Следствие 0.1.1, принадлежността на функцията $\underline{v_h}$ към пространството $\underline{H}(div;\Omega)$ изисква непрекъснатост на компонентите $v_{i,h}$ по съответните направления x_i, $i = 1,2$. Анализът на това условие ще извършим, започвайки от функцията v_1. Аналогично на Лема 1.3.2 разглеждаме хоризонтална лента от елементи $T_{i,j}$ $T_{i,j} = [x_{1,i-1}, x_{1,i}] \times [x_{2,j-1}, x_{2,j}] \in [0,1] \times [x_{2,j-1}, x_{2,j}]$ $i = 1, \ldots, N$ при фиксирано j, показани на Фигура 2.3.



Фигура 2.3: Лента от елементи, принадлежащи на Ω при $n = 4$.

В този случай, движейки се от ляво на дясно, след елемента $T_{N_1,j}$ се осъществява преминаване от „груба“ към „фина“ мрежа.

За всяко $w_h(x_1, x_2) \in W_h^r$, посредством помощната функция φ_1, дефинирана в (1.16), конструираме рекурсивно функцията $v_{1,h}$, използвайки (1.17) за $i = 1, 2, \ldots, N_1$.

Разглеждаме останалите елементи от лентата:

$$T_{i,j}^k \in [x_{1,N_1+1}, x_{1,N}] \times [x_{2,j-1}, x_{2,j}], \qquad k = 1, 2, \ldots, n$$

При преминаване от „груба“ към „фина“ мрежа, от ляво на дясно, процедираме по аналогичен начин. Разликата е единствено при елементите от интерфейса I_f, които имат общи страни с елемента $T_{N_1,j} \in \Omega_2$. За всичките n коригиращата функция $\psi_1(x_2)$ е една и съща. В този случай

$$v_1(x_1, x_2)_{|T_{N_1+1,j}^k} = \varphi_1(x_1, x_2)_{|T_{N_1+1,j}^k} + \psi_1(x_2)_{|T_{N_1+1,j}^k}, \quad k = 1, 2, \ldots, n,$$

където

$$\psi_1(x_2)_{|T_{N_1+1,j}^k} = v_1(x_{1,N}, x_2)_{|T_{N_1,j}^k} - \varphi_1(x_{1,N_1}, x_2)_{|T_{N_1,j}^k}, \quad k = 1, 2, \ldots, n.$$

Върху елементите $\Omega_1 \backslash I_f$ функцията v_1 се конструира на абсолютно същия принцип, както върху елементите от Ω_2 за всяко $k = 1, 2, \ldots, n$.

$$v_1(x_1, x_2)_{|T_{i,j}^k} = \varphi_1(x_1, x_2)_{|T_{i,j}^k} + \psi_1(x_2)_{|T_{i,j}^k}, \quad i = N_1 + 2, \ldots, N,$$

където

$$\psi_1(x_2)_{|T_{i,j}^k} = v_1(x_{1,i-1}, x_2)_{|T_{i,j}^k} - \varphi_1(x_{1,i-1}, x_2)_{|T_{i,j}^k}, \quad i = N_1 + 2, \ldots, N.$$

Анализът е напълно аналогичен в случая, когато преминаването от „груба“ към „фина“ мрежа си извършва от дясно на ляво.

Конструирането на втората компонента $v_{2,h}$ се извършва по същия начин, използвайки помощната функция ψ_2, дефинирана в (1.19), и взаимната замяна на променливите x_1 и x_2.

Така конструираната функция $\underline{v_h} = (v_{1,h}, v_{2,h})$ удовлетворява съотношенията (1.20), което означава, че изискването (i) в (1.15) е изпълнено.

Удовлетворяването на изискването (ii) в (1.15) също се установява чрез непосредствена проверка, описана в Лема 1.3.1, което завършва доказателството на лемата.

Забележка 2.3.2 *Техниката на доказателството изисква преминаването от „груба“ към „фина“ мрежа да се извършва еднократно в съответните направления, както е показано на Фигура 2.1.*

2.4 Оценки на грешката

След като вече сме показали валидността на условията (i) и (ii), за да приложим абстрактната оценка, се нуждаем от оценки на грешката на членовете от дясната страна на неравенство (1.14).

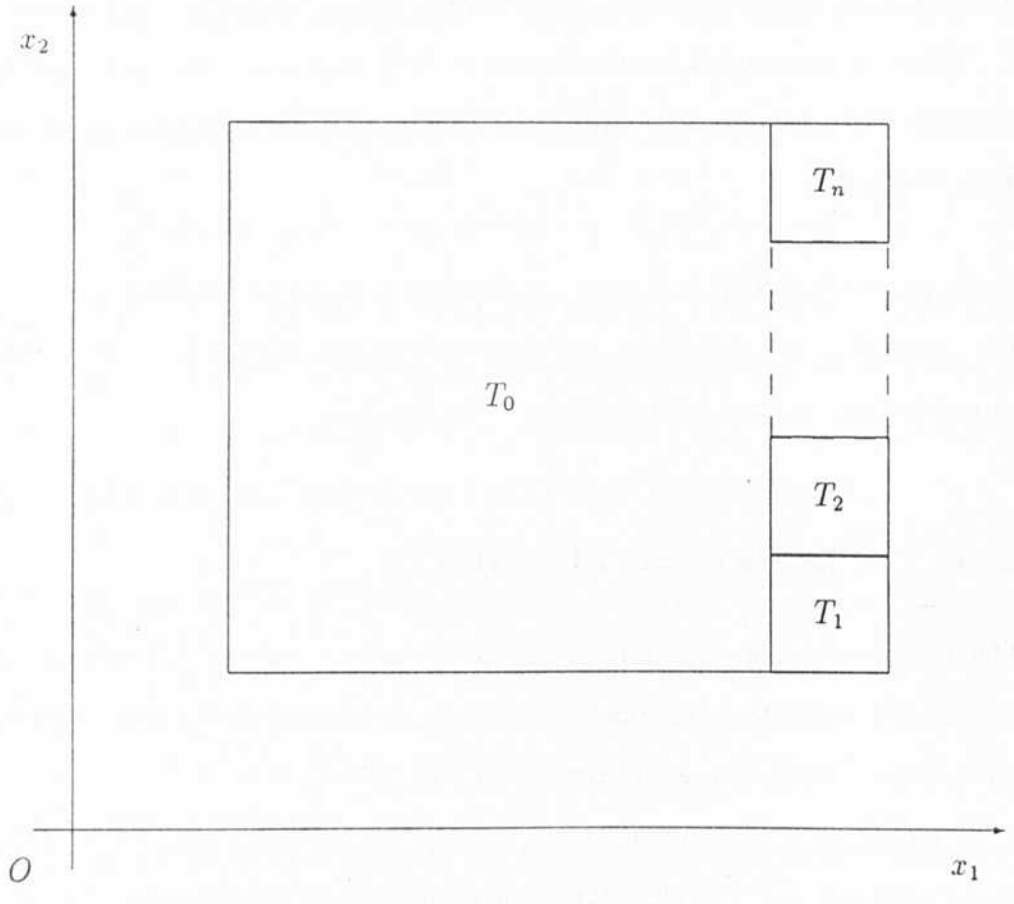
Разглеждаме интерполантите w^I, $\underline{v}^{fI} = (v_1^{fI}, v_2^{fI})$ и $\underline{v}^{cI} = (v_1^{cI}, v_2^{cI})$, дефинирани в (1.12) и асоциирани със съответните разделяния τ_h, τ_f и τ_c.

Интерполантът w^I на функцията $w \in W$ върху всеки краен елемент $T \in \tau_h$ е полином от $Q(r,r)$ и съвпада с функцията w в $(r+1) \times (r+1)$ Гаусови точки.

Интерполантите v_1^{fI} и v_1^{cI} съвпадат с функцията v_1 в $L(r+2) \times G(r+1)$ съответно върху елементите $T \in \tau_f$ и $T \in \tau_c$. Аналогично – интерполантите v_2^{fI} и v_2^{cI} съвпадат с функцията v_2 в $G(r+1) \times L(r+2)$ върху $T \in \tau_f$ и $T \in \tau_c$.

Ще разгледаме отделно двата случая за компонентите v_i, $(i = 1,2)$ на векторнозначната функция $\underline{v} \in \underline{V}$. Първо се спираме на преминаването от „груба“ към „фина“ мрежа за компонентата v_1, в направлението x_1, показан

на Фигура 2.4. В този случай $T_0 \in \Omega_2,\ T_i \in I_f \in \Omega_1,\ i = 1, 2, \ldots, n$.



Фигура 2.4: Преминаване от „груба“ към „фина“ мрежа за компонентата v_1.

За елементите от интерфейса $T_i,\ i = 1, 2, \ldots, n$ дефинираме „псевдоинтерполант“ v_1^{pI} със следните изисквания:

(2.7)

(а) $v_1^{pI} \in Q(r+1, r)$ върху $T_i \bigcap T_0, \qquad i = 0, 1, \ldots, n;$

(б) $v_1^{pI}{}_{|T_i} \equiv v_1^{cI}{}_{|T_0}$ върху $T_i \bigcap T_0, \qquad i = 1, 2, \ldots, n;$

(в) $v_1^{pI} \equiv v_1^{fI}$ във възлите на $T_i \backslash (T_i \bigcap T_0), \quad i = 1, 2, \ldots, n.$

Забележка 2.4.1 *Псевдоинтерполантът* $v_1^{pI}(x_1, x_2)$ *съвпада с функцията* $v_1(x_1, x_2)$ *във възлите, показани на Фигура 2.2.*

Това ни дава възможност да дефинираме интерполанта $v_1^I \in V_h^{r,1}$ във всички подобласти на областта Ω.

$$v_1^I = \begin{cases} v_1^{fI} & \text{в} \quad \Omega_1 \backslash I_f \\ v_1^{cI} & \text{в} \quad \Omega_2 \\ v_1^{pI} & \text{в} \quad I_f \end{cases} \tag{2.8}$$

Интерполанта на $v_2^I \in V_h^{r,2}$ на втората компонента v_2 дефинираме по аналогичен начин, имайки предвид преминаването от „груба“ към „фина“ мрежа за компонентата v_2.

Очевидно, така дефинираните интерполанти w^I и $\underline{v}^I = (v_1^I, v_2^I)$ принадлежат съответно на пространствата W_h^r и $\underline{V_h^r}$, дефинирани в (2.5).

За доказателството на локалните интерполационни резултати ще се нуждаем и от следното помощно твърдение:

Лема 2.4.1 *Разглеждаме алгебричен полином на една променлива* $p_m(x) \in P_m(x)$, *дефиниран в интервала* $[0,1]$. *Нека параметърът* $\delta = 1/m$ *и стойностите на* $p_m(x)$ *в точките* $k_i = i\delta$, $(i = 0, 1, 2, \ldots, m)$ *са следните:*

$$p_m(k_0) = \varepsilon \neq 0, \quad p_m(k_i) = 0, \quad i = 1, 2, \ldots, m.$$

Тогава производната на $p_m(x)$ *достига максимума на модула си в левия край на интервала* $[0,1]$, *т.е.*

$$|p_m'(x)| \leq |p_m'(k_0)|, \qquad x \in [0,1].$$

Доказателство. Полиномът $p_m(x) \in P_m(x)$ се определя по единствен начиин от стойностите си в разглежданите $m+1$ точки:

$$p_m(x) = Cp(x)\,,$$

където

$$C = \frac{\varepsilon}{(-1)^m\, m!\, \delta^m}\,, \quad p(x) = (x-\delta)(x-2\delta)\ldots(x-m\delta)\,.$$

От съображение за простота ще разглеждаме полинома $p(x)$, тъй като максимумът на модула на неговата производна се достига в същата точка, както този на $p_m(x)$.

$$\begin{aligned} p'(x) \;&= (x-2\delta)(x-3\delta)\ldots(x-m\delta) \\ &\quad +(x-\delta)(x-3\delta)\ldots(x-m\delta) \\ &\qquad\qquad \cdot \\ &\qquad\qquad \cdot \\ &\qquad\qquad \cdot \\ &\quad +(x-\delta)(x-2\delta)\ldots(x-(m-1)\delta)\,. \end{aligned} \tag{2.9}$$

Очевидно:

$$|p'(k_0)| = \delta^{m-1}\left(\frac{m!}{1} + \frac{m!}{2} + \ldots + \frac{m!}{m}\right) = m!\,\delta^{m-1}\left(1 + \frac{1}{2} + \ldots + \frac{1}{m}\right)\,. \tag{2.10}$$

Нека първо $x = k_i$, $i = 1, 2, \ldots, m$. Тогава само i-тото събираемо в дясната страна на равенство (2.9) е различно от нула и

$$|p'(k_i)| = \delta^{m-1}(i-1)!\,(m-i)! \le \delta^{m-1}(m-1)! \le \delta^{m-1}\frac{m!}{i}$$

и се мажорира от съответното i-то събираемо в дясната страна на равенство (2.10).

Разглеждаме $x \neq k_i$, $i = 1, 2, \ldots, m$. Случаят, когато $x \in (k_0, k_1)$ е очевиден, тъй като всички членове в дясната страна на равенство (2.9) се мажорират от съответните им членове в дясната страна на равенство (2.10).

Нека

$$x \in (k_{i-1}, k_i) \quad i = 2, 3, \ldots, m.$$

Оценяваме поотделно модулите на всеки j член $r_j(x)$ на сумата в дясната страна на равенство (2.9) за $j = 1, 2, \ldots, m$, където

$$r_j(x) = (x - \delta) \ldots (x - (j-1)\delta)(x - (j+1)\delta) \ldots (x - m\delta).$$

Първите $i - 1$ члена се мажорират от съответсващите им $i - 1$ члена на $|r_j(k_i)|$. Следващите $m - i$ члена се мажорират от съответсващите им $i - 1$ члена на $|r_j(k_{i-1})|$. Така, за трите възможни случая, получаваме:

$$|r_j(x)| \leq \begin{cases} \delta^{m-1} \dfrac{(i-1)!\,(m-i+1)!}{j+1-i} & \text{при } i < j; \\ \delta^{m-1}\,(i-1)!\,(m-i+1)! & \text{при } i = j; \\ \delta^{m-1} \dfrac{(i-1)!\,(i-j)!}{j+1-i} & \text{при } i > j. \end{cases}$$

Тогава, за всяко $i = 2, 3, \ldots, m$ е в сила:

$$\begin{aligned} |p'(x)| &\leq \sum_{j=1}^{m} |r_j(x)| \\ &\leq \delta^{m-1}\,(i-1)!\,(m-i+1)! \left(\frac{1}{i-1} + \frac{1}{i-2} + \ldots + \frac{1}{1} + \frac{1}{1} + \frac{1}{2} + \ldots + \frac{1}{m+1-i} \right) \\ &< \delta^{m-1}\,2\,(i-1)!\,(m-i+1)! \left(1 + \frac{1}{2} + \ldots + \frac{1}{m} \right). \end{aligned}$$

Тъй като, за всяко $i = 2, 3, \ldots, m$ и $m > 2$

$$2\,(i-1)!\,(m-i+1)! \leq 2\,(m-1)! \leq m!,$$

то

$$|p'(x)| \leq \delta^{m-1}\, m! \left(1 + \frac{1}{2} + \ldots + \frac{1}{m} \right). \tag{2.11}$$

Отбелязваме, че всички събираеми в равенството (2.9) са с еднакви знаци. Следователно

$$|p'(k_0)| = \sum_{j=1}^{m} |r_j(k_0)| .$$

Така, неравенство (2.11) и равенство (2.9) завършват доказателството на лемата.

Използвайки описаните в този параграф интерполанти, представяме следните основните помощни резултати:

Теорема 2.4.1 *Нека* $w^I \in W_h^r$ *и* $\underline{v}^I = (v_1^I, v_2^I) \in \underline{V_h^r}$ *са дефинирани върху съставната мрежа* τ_h *и са интерполанти съответно на функциите* $w \in H^{r+1}(\Omega)$ *и* $\underline{v} = (v_1, v_2) \in (H^{r+3}(\Omega))^2$. *Тогава съществуват константи* C, *независещи от* h_f *и* h_c, *такава, че:*

(2.12)

(*а*) $\|w - w^I\|_{0,T} \leq C h_T^{r+1} \|w\|_{r+1,T}$, $T \in \tau_h, T \in \Omega$;

(*б*) $\|\underline{v} - \underline{v}^I\|_{0,T} \leq C h_T^{r+1} \|\underline{v}\|_{r+1,T}$, $T \in \tau_h, T \in \Omega \backslash I_f$;

(*в*) $\|div(\underline{v} - \underline{v}^I)\|_{0,T} \leq C h_T^{r+1} \|\underline{v}\|_{r+2,T}$, $T \in \tau_h, T \in \Omega \backslash I_f$;

(*г*) $\|\underline{v} - \underline{v}^I\|_{0,T} \leq C \left(h_f^{r+1} \|\underline{v}\|_{r+1,T} + h_f h_c^{r+1} \|\underline{v}\|_{r+1,\infty,\gamma} \right)$, $T \in \tau_h$, $T \in I_f$;

(*д*) $\|div(\underline{v} - \underline{v}^I)\|_{0,T} \leq C \left(h_f^{r+1} \|\underline{v}\|_{r+2,T} + h_f h_c^{r} \|\underline{v}\|_{r+1,\infty,\gamma} \right)$, $T \in \tau_h$, $T \in I_f$,

където γ *е страната на произволен елемент* $T_c \in \Omega_2$, *който граничи с подобластта* Ω_1, *т.е.*

$$\gamma = \bigcup_{T_f \in I_f} (T_f \cap T_c)$$

Доказателство. Доказателствата на първите три неравенства се основават непосредствено на класически резултати.

Неравенство (2.12а) следва от Лема 0.1.3 при $\Omega \equiv T$ и $\hat{\Omega} \equiv \hat{T}$ и тъй като $P(r) \in Q(r, r)$.

Неравенство (2.12б) следва от Лема 0.1.4 при $\Omega \equiv T$ и $\hat{\Omega} \equiv \hat{T}$ и тъй като $P(r) \in Q(r+1, r)$ и $P(r) \in Q(r, r+1)$.

Неравенство (2.12в) следва от Лема 0.1.5 при $\Omega \equiv T$ и $\hat{\Omega} \equiv \hat{T}$.

Тъй като (виж 0.16)

$$\|\underline{v} - \underline{v}^I\|_{0,T} \leq \|v_1 - v_1^I\|_{0,T} + \|v_2 - v_2^I\|_{0,T}\,, \tag{2.13}$$

ще разгледаме отделно двата члена от дясната страна на равенство (2.13). При преминаване от „груба“ към „фина“ мрежа по направлението x_i $(i = 1, 2)$ за елементите $T \in I_f$ съществено значение ще има оценяването на i-тия член в сумата, тъй като оценката на другия е стандартна (виж (2.12б)) и не нарушава порядъка на грешката.

Първо разглеждаме преминаването по направлението x_1 за произволен елемент $T \equiv T_i \in I_f\,, i = 1, 2, \ldots, n$, $T_0 \in \Omega_2$, показано на Фигура 2.4.

$$\|v_1 - v_1^I\|_{0,T} = \|v_1 - v_1^{pI}\|_{0,T} \leq \|v_1 - v_1^{fI}\|_{0,T} + \|v_1^{fI} - v_1^{pI}\|_{0,T}\,. \tag{2.14}$$

Първия член от дясната страна на неравенството оценяваме като (2.12б). За да опростим означенията, при оценяването на втория член, полагаме:

$$\psi(x_1, x_2) = \left(v_1^{fI} - v_1^{pI}\right)_{|T} \in Q(r+1, r)\,. \tag{2.15}$$

Функцията ψ се анулира в $(r+1) \times (r+1)$ точки върху елемента T и е различна от нула в $1 \times (r+1)$ точки, които се намират на границата му $T \bigcap T_0$ с областта Ω_2. Това дава възможност тя да бъде представена в явен вид, който зависи само от параметрите ε_j, $j = 1, 2, \ldots, r+1$. Нека $l_i \in L(r+2)$ и $g_j \in G(r+1)$, $i = 0, 1, \ldots, r+1$; $j = 1, 2, \ldots, r+1$. Тогава

$$\psi(x_1, x_2) = \varphi_0(x_1) \sum_{i=1}^{r+1} \varepsilon_i \varphi_i(x_2)\,,$$

където

$$\varphi_0(x_1) = \frac{(x_1 - l_1)\dots(x_1 - l_{r+1})}{(l_0 - l_1)\dots(l_0 - l_{r+1})}$$

и

$$\varphi_j(x_2) = \frac{(x_2 - g_1)\dots(x_2 - g_{j-1})(x_2 - g_{j+1})\dots(x_2 - g_{r+1})}{(g_j - g_1)\dots(g_j - g_{j-1})(g_j - g_{j+1})\dots(g_j - g_{r+1})},$$
$$j = 1, 2, \dots, r+1.$$

Тъй като, v_1^{pI} съвпада с v_1^{cI} върху границата на елемента $T \in I_f$ с елемента $T_0 \in \Omega_2$, то параметрите ε_j се явяват разликите между „финия" и „грубия" интерполант в точките (l_0, g_j), $j = 1, 2, \dots, r+1$, т.е.

$$\varepsilon_j = (v_1^{fI} - v_1^{cI})_{|(l_0, g_j)} \qquad j = 1, 2, \dots, r+1.$$

Ще покажем, че функцията $\psi(x_1, x_2)$, дефинирана в T, достига максимума на модула си върху страната му T', която лежи на границата γ.

Тъй като $\psi(x_1, x_2) \in Q(r+1, r)$ върху T се анулира в $r+1$ точки от мрежата S_2, дефинирана в (1.10), по направлението x_1, то тя се анулира навсякъде по вертикалните линии $\psi(l_i, x_2)$, $i = 1, 2, \dots, r+1$. Следователно, за нашите цели е достатъчно да разглеждаме поведението на едномерната функция (при фиксирано $x_2 = x_{2,0}$) $\varphi(x_1) = \psi(x_1, x_{2,0})$.

Във връзка със Забележка 2.3.1 можем да считаме, че възлите от мрежата S_1 (1.10) са равноотдалечени. Фактът, че

$$|\varphi(x_1)| \le |\varphi(l_0)|, \tag{2.16}$$

се установява чрез непосредствена проверка.

Прилагайки Лема 2.4.1 при $p_m(x) = \varphi(x_1)$ и $m = r+1$ за интервала (x_{1,N_1}, x_{1,N_1+1}), получаваме аналогичен резултат и за производната на разглежданата функция, т.е.

$$|\varphi'(x_1)| \le |\varphi'(l_0)|. \tag{2.17}$$

Нека ε е най-голямата стойност на модула на функцията $\psi(l_0, x_2)$ върху границата γ. Тогава

$$\|\psi\|_{0,T} \leq \left(h_f^2 \max_T |\psi|^2\right)^{1/2} \leq \varepsilon h_f .$$

Използвайки неравенството на триъгълника, получаваме:

$$\varepsilon = |(v_1^{fI} - v_1^{cI})|_\gamma \leq |(v_1^{fI} - v_1)|_\gamma + |(v_1 - v_1^{cI})|_\gamma \leq 2|(v_1 - v_1^{cI})|_\gamma ,$$

което се явява грешката от Лагранжевата интерполация по $r+1$ точки върху страната γ на елемента T_0. Прилагайки стандартната интерполационна оценка, получаваме:

$$\|\psi\|_{0,T} \leq C h_f h_c^{r+1} \|v_1\|_{r+1,\infty,\gamma} , \tag{2.18}$$

където $\gamma = T_0 \bigcap (T_1 \bigcup T_2 \bigcup \ldots \bigcup T_n)$.

Оценката (2.18) на втория член на неравенство (2.14) и аналогичните разсъждения за елементите $T \in I_f$, при вертикално преминаване от „груба“ към „фина“ мрежа, по променливата x_2 завършват доказателството на оценката (2.12г).

Използвайки, че

$$\|div(\underline{v} - \underline{v}^I)\|_{0,T} \leq \left\| \frac{\partial(v_1 - v_1^I)}{\partial x_1} \right\|_{0,T} + \left\| \frac{\partial(v_2 - v_2^I)}{\partial x_2} \right\|_{0,T}$$

и неравенство (2.17), с аналогични разсъждения получаваме:

$$\|div(\underline{v} - \underline{v}^I)\|_{0,T} \leq C h_f h_c^r \|\underline{v}\|_{r+1,\infty,\gamma} . \tag{2.19}$$

Аналогично доказваме (2.12д), основавайки се на неравенство (2.19), което завършва доказателството на теоремата.

Основният резултат в тази глава се съдържа в следната оценка на грешката на смесения метод на крайните елементи:

Теорема 2.4.2 *Допускаме, че решението на задачата (2.3)* $(\underline{u}, p) \in (H^{r+3}(\Omega))^2 \times H^{r+1}(\Omega)$ *за някакво цяло число* $r \geq 0$. *Нека пространствата* $\underline{V_h}$ *и* W_h, *дефинирани в (2.5), са асоциирани със съставната мрежа* τ_h *на областта* Ω. *Тогава задачата (2.6) има единствено решение* $(\underline{u_h}, p_h) \in (\underline{V_h} \times W_h)$ *и съществува константа* C, *независеща от* h_c *и* h_f, *такава, че:*

(2.20)

$$\begin{aligned}\|\underline{u} - \underline{u_h}\|_{\underline{H}(div;\Omega)} + \|p - p_h\|_{0,\Omega} \leq C\Big(&h_f^{r+1}\|p\|_{r+1,\Omega_1} + h_c^{r+1}\|p\|_{r+1,\Omega_2} \\ &+h_f^{r+1}\|\underline{u}\|_{r+2,\Omega_1} + h_c^{r+1}\|\underline{u}\|_{r+2,\Omega_2} \\ &+h_c^{r+1}n^{-1/2}\|\underline{u}\|_{r+1,\infty,\Gamma}\Big),\end{aligned}$$

където $\Gamma = \overline{\Omega}_1 \bigcap \overline{\Omega}_2$.

Доказателство. Оценката (2.18) е в сила за всички елементи $T_{1,j}, T_{2,j} \ldots, T_{n,j}$, граничащи с елементите $T_{0,j}$, $(j = 1, 2, \ldots, m)$ при преминаването от „груба“ към „фина“ мрежа от ляво на дясно. Ако означим с I_l тази част от интерфейса I_f, която се намира в левия край на подобластта Ω_1, използвайки неравенство (2.18), получаваме:

$$\|\psi\|_{0,I_r} \leq C\sqrt{n}h_f h_c^{r+1}\|v_1\|_{r+1,\infty,\Gamma_l}, \tag{2.21}$$

където $\Gamma_r = T_{0,j} \bigcap (T_{1,j} \bigcup T_{2,j} \bigcup \ldots \bigcup T_{n,j})$, $(j = 1, 2, \ldots, m)$.

Аналогично, за функцията $div\psi$ получаваме:

$$\|div\psi\|_{0,I_r} \leq C\sqrt{n}h_f h_c^{r}\|v_1\|_{r+1,\infty,\Gamma_l}, \tag{2.22}$$

където $\Gamma_l = T_{0,j} \bigcap (T_{1,j} \bigcup T_{2,j} \bigcup \ldots \bigcup T_{n,j})$, $(j = 1, 2, \ldots, m)$.

По същият начин получаваме оценката $\|\psi\|_{0,I_r}$ и $\|div\psi\|_{0,I_r}$ за дясната част от интерфейса I_f, когато преминаването от „груба“ към „фина“ мрежа се извършва от дясно на ляво.

Напълно аналогични оценки са в сила и за долната и горната част от интерфейса I_d и I_u, по отношение на втория член от дясната страна на равенство (2.13), при вертикално преминаване от „груба“ към „фина“ мрежа.

Използвайки стандартното сумиране на нормите по крайните елементи, получаваме:

(2.23)

(а) $\|w - w^I\|_{0,\Omega_1} \leq C\left(h_f^{r+1}\|w\|_{r+1,\Omega_1} + h_c^{r+1}\|w\|_{r+1,\Omega_2}\right)$;

(б) $\|\underline{v} - \underline{v}^I\|_{0,\Omega} \leq C\left(h_f^{r+1}\|\underline{v}\|_{r+1,\Omega_1} + h_c^{r+1}\|\underline{v}\|_{r+1,\Omega_2} + n^{1/2}h_f h_c^{r+1}\|\underline{v}\|_{r+1,\infty,\Gamma}\right)$;

(в) $\|div(\underline{v} - \underline{v}^I)\|_{0,\Omega} \leq C\left(h_f^{r+1}\|\underline{v}\|_{r+2,\Omega_1} + h_c^{r+1}\|\underline{v}\|_{r+2,\Omega_2} + h_c^{r+1}n^{-1/2}\|\underline{u}\|_{r+1,\infty,\Gamma}\right)$,

където $\Gamma = \Omega_2 \bigcap (I_l \bigcup I_r \bigcup I_d \bigcup I_u)$.

Прилагането на Лема 2.3.1 и неравенства (2.23) в Лема 1.3.1 завършва доказателството на теоремата.

Забележка 2.4.2 *Ако разгледаме задачата (2.3) с нулеви гранични условия (т.е. $g(x) = 0$), то тя и съответната и приближена задача (2.6) чрез регулярно локално сгъстяване ще са еднозначно разрешими за всяка функция $f(x) \in L^2(\Omega)$. От това следва, че съществуват разрешаващи оператори $\underline{S}$ и D и съответстващите им дискретни аналози $\underline{S_h}$ и D_h, такива, че*

$$\underline{S} : L_2 \longrightarrow \underline{V} \qquad \underline{S}f = \underline{u};$$

$$\underline{S_h} : L_2 \longrightarrow \underline{V_h} \qquad \underline{S_h}f = \underline{u_h};$$

$$D : L_2 \longrightarrow W \qquad Df = p;$$

$$D_h : L_2 \longrightarrow W_h \qquad D_h f = p_h.$$

Забележка 2.4.2 дава възможност да приложим получените резултати от оценката на грешката за задача за собствени стойности.

Разглеждаме задачата за вибрираща мембрана:

$$(2.24)\qquad \left|\begin{array}{ll} -div(a(x)\underline{\nabla} p) = \lambda p, & p \in \Omega; \\ p = 0, & p \in \partial\Omega. \end{array}\right.$$

След обичайните полагания (виж Бабушка и Осборн [1]) стигаме до следната вариационна задача: Търсим $\lambda \in \mathbb{C}, (\underline{u}, p) \in \underline{V} \times W, (\underline{u}, p) \neq (\underline{0}, 0)$ такива, че

$$(2.25)\qquad \left|\begin{array}{ll} a(\underline{u}, \underline{v}) + b(\underline{v}, p) = 0, & \forall \underline{v} \in \underline{V}\,; \\ b(\underline{u}, w) = -\lambda(p, w), & \forall w \in W. \end{array}\right.$$

Използвайки пространствата $\underline{V_h}$ и W_h, описани в (2.5) дефинираме приближената задача

$$(2.26)\qquad \left|\begin{array}{l} a(\underline{u_h}, \underline{v_h}) + b(\underline{v_h}, p_h) = 0, \quad \forall \underline{v_h} \in \underline{V_h}\,; \\ b(\underline{u_h}, w_h) = -\lambda_h(p_h, w_h), \quad \forall w_h \in W_h. \end{array}\right.$$

За да получим оценка на грешката между точните и приближените собствени стойности, получени съответно от системите (2.25) и (2.26), използваме общата теория за спектрална апроксимация на компактен оператор в Банахово пространство, развита от Осборн [1] (виж също Бабушка, Осборн [1]).

Разрешаващият оператор D и фамилията разрешаващи оператори $\{D_h\}$, дефинирани в Забележка 2.4.2 притежават апроксимационните свойства, следващи от Теорема 2.4.2 за решаване на изходната елиптична задача чрез смесения метод с регулярно локално сгъстяване. Тези оператори са компактни и следователно за тях можем да използваме класически абстрактни оценки (виж Бабушка, Осборн [1], Теореми 7.3 и 7.4).

Нека $(\lambda, (\underline{u}, p))$ и $(\lambda_h, (\underline{u_h}, p_h))$ са решения съответно на (2.25) и (2.26). Ако са изпълнени условията на Теорема 2.4.2, то съществува константа C,

независеща от h_f и h_c, такава, че:

$$| \lambda - \lambda_h | \leq C \left[C_1(r) h_f^{2(r+1)} + C_2(r,n) h_c^{2(r+1)} \right] ,$$

където:

$$C_1(r) = \|\underline{u}\|_{r+2,\Omega_1} + \|p\|_{r+1,\Omega_1} ,$$

$$C_2(r,n) = \|\underline{u}\|_{r+2,\Omega_2} + \|p\|_{r+1,\Omega_2} + n^{-1/2} \|\underline{u}\|_{r+1,\infty,\Gamma}$$

а вътрешната граница е $\Gamma = \overline{\Omega}_1 \bigcap \overline{\Omega}_2$.

Глава 3

Оптимални оценки върху мрежи с локално сгъстяване

3.1 Въведение

Съществуване, единственост и глобална оценка на грешката на решението $(\underline{u}_h, p_h)$ на смесения метод на крайните елементи за пространства на Равиар - Тома от ииндекс r, за елиптични гранични задачи от втори ред, е получена от Дъглас и Робъртс [1]. За регулярни крайни елементи представените оценки в L^2- норма са от следния тип:

(3.1)

$$\text{(а)} \quad \|p - p_h\|_{0,\Omega} \leq \begin{cases} Ch\|p\|_{2,\Omega}, \text{ ако } r = 0; \\ Ch^k\|p\|_{k,\Omega}, \text{ ако } r \geq 1 \text{ и } 2 \leq k \leq r+1; \end{cases}$$

$$\text{(б)} \quad \|\underline{u} - \underline{u}_h\|_{0,\Omega} \leq Ch^k\|p\|_{k+1,\Omega}, \text{ ако } 1 \leq k \leq r+1;$$

$$\text{(в)} \quad \|div(\underline{u} - \underline{u}_h)\|_{0,\Omega} \leq Ch^k\|p\|_{k+2,\Omega}, \text{ ако } 0 \leq k \leq r+1.$$

Ефективни итерационни методи за решаване на системите алгебрични урав-

нения, получени от апроксимацията по смесения метод на крайните елементи с регулярно локално сгъстяване върху съответната съставна мрежа, се предлагат от Матью [1], Юинг, Лазаров, Ръсел и Василивски [1], Василевски и Лазаров [1] и др. В тази глава, използвайки модифициран проектор на Равиар-Тома, доказваме оптимални оценки на грешката за смесения метод на крайните елементи върху мрежа с регулярно локално сгъстяване. Подобна техника за една проста моделна задача е демонстрирана в работата на Юинг и Уонг [1]. В §3 разглеждаме мрежа с регулярно локално сгъстяване и конструираме пространствата на Равиар-Тома. Най-съществената част тук е конструирането на модифициран проектор на Равиар-Тома върху съставната мрежа. След това се представят локални оценки. В §4 доказваме дуални леми, които лежат в основата на по-нататъшния анализ на грешката. Накрая в §5 доказваме основния резултат в тази глава – степента на сходимост на решението, получено по смесения метод на крайните елементи.

3.2 Постановка на задачата

Нека $\Omega \subset \mathbb{R}^2$ е ограничена област с граница $\partial\Omega$. Допускаме, че задачата на Дирихле

(3.2)

$$\text{(а)} \quad Lp = -div\,(a(x)\underline{\nabla} p + \underline{b}(x)p) + c(x)p = f(x), \quad x \in \Omega;$$

$$\text{(б)} \quad p = -g(x), \quad x \in \partial\Omega,$$

притежава единствено решение за $\{f, g\} \in L^2(\Omega) \times H^{3/2}(\partial\Omega)$, и че

$$\|p\|_{2,\Omega} \leq C\left(\|f\|_{0,\Omega} + \|g\|_{3/2;\partial\Omega}\right),$$

където $\underline{\nabla} w$ означава градиента на скаларнозначната функция w, $div\underline{v} = \underline{\nabla}.\underline{v}$ означава дивергенцията на векторнозначната функция $\underline{v}$, и функцията

$a(x) : \overline{\Omega} \to \mathbb{R}$ удовлетворява изискването $a(x) \geq a_0 > 0$. Полагаме

$$\underline{u} \equiv (u_1, u_2) = -\Big(a(x)\underline{\nabla} p + \underline{b}(x)p\Big) \tag{3.3}$$

и

$$\alpha(x) = a(x)^{-1}, \qquad \underline{\beta}(x) = \alpha(x)\underline{b}(x). \tag{3.4}$$

Нека

$$\begin{aligned} &\text{(а)} \quad \underline{V} = \underline{H}(div; \Omega) = \{\underline{u} \in L^2(\Omega)^2 : div\underline{u} \in L^2(\Omega)\}\ ; \\ &\text{(б)} \quad W = L^2(\Omega). \end{aligned} \tag{3.5}$$

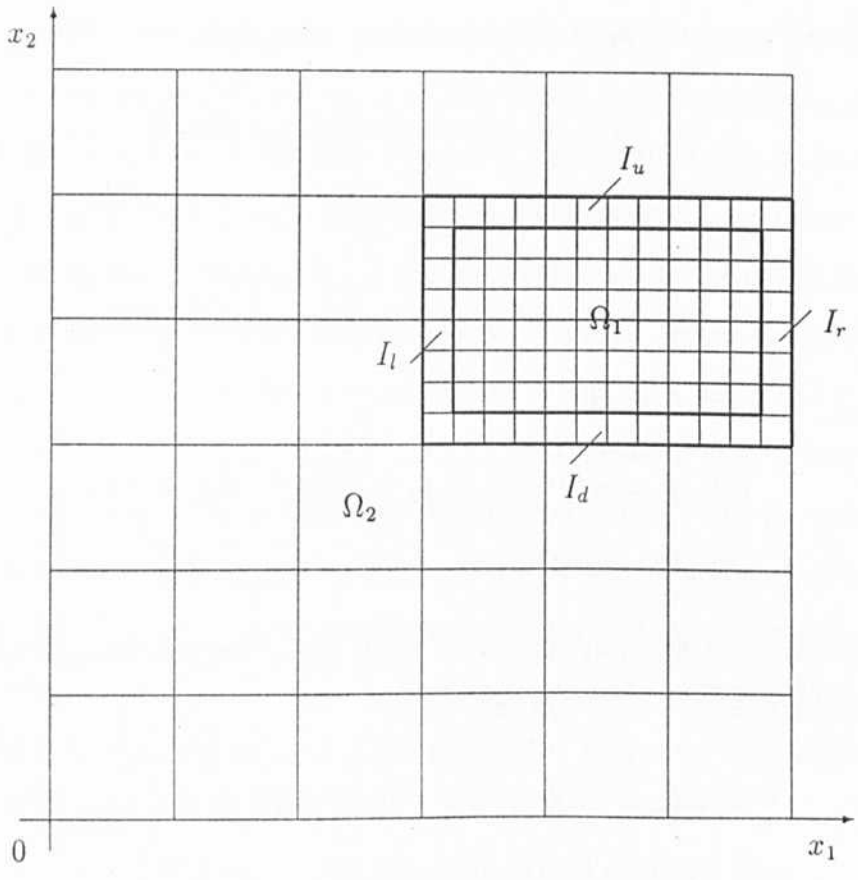
Тогава слабата формулировка на (3.2) се представя по следния начин: търсим двойка $(\underline{u}, p) \in \underline{V} \times W$, такава, че

$$\begin{aligned} &\text{(а)} \quad (\alpha\underline{u}, \underline{v}) - (div\underline{v}, p) + (\underline{\beta}p, \underline{v}) = \langle g, \underline{v}.\underline{\nu}\rangle, \qquad \underline{v} \in \underline{V}; \\ &\text{(б)} \quad (div\underline{u}, w) + (cp, w) = (f, w), \qquad w \in W, \end{aligned} \tag{3.6}$$

където $\underline{\nu}$ е външният единичен нормален вектор към $\partial\Omega$, а скаларните произведения съответно в $L^2(\Omega)^2$ са означени с $(\cdot,\cdot)$, а в $L^2(\partial\Omega)$ с $\langle\cdot,\cdot\rangle$.

3.3 Апроксимация върху мрежа с регулярно локално сгъстяване

Разглеждаме трите подобласти Ω_1, Ω_2 и $I_f \in \Omega_1$ на областта Ω и съставната мрежа (триангулация) τ_h, дефинирани в (2.4). На фигура 3.1 е показан пример за такава триангулация, като с I_l, I_r, I_d и I_u са означени лентите от елементи от I_f, съответно на лявата, дясната, долната и горната части на подобластта Ω_1.



Фигура 3.1: Мрежа с локално сгъстяване

Нека $\underline{V_h^r}$ и W_h^r са пространствата на Равиар и Тома с индекс r, дефинирани в (2.5), върху съставната мрежа τ_h.

Апроксимацията по смесения метод на крайните елементи на задачата (3.2) се дефинира чрез определяне на двойка $\{\underline{u_h}, p_h\} \in \underline{V_h} \times W_h$, такава, че

$$
\begin{aligned}
&\text{(а)} \quad (\alpha \underline{u_h}, \underline{v}) - (div \underline{v}, p_h) + (\underline{\beta} p_h, \underline{v}) = \langle g, \underline{v}.\underline{\nu} \rangle\,, \quad \underline{v} \in \underline{V_h}\,; \\
&\text{(б)} \quad (div \underline{u_h}, w) + (c p_h, w) = (f, w)\,, \qquad\qquad\quad w \in W_h\,.
\end{aligned}
\tag{3.7}
$$

Тъй като анализът се извършва отделно върху всеки елемент, то първо ще разгледаме пространството $\underline{\hat{V}}$, дефинирано в (1.7) върху основния елемент

$\hat{T} = [-1,1]^2$, който е асоцииран с елементите $T \in \tau_h : \ T \in \Omega \backslash I_f$.

Лема 3.3.1 *Степените на свобода на функцията* $\underline{\hat{v}} = (\hat{v}_1, \hat{v}_2) \in \underline{\hat{V}}$ *са определени чрез следните функционали:*

$$
\begin{aligned}
&(a) && \int_{\hat{e}} \underline{\hat{v}}.\underline{\hat{\nu}}\hat{\varphi} d\hat{\gamma}\,, && \hat{\varphi} \in Q(r,r) \ \text{върху} \ \forall \hat{e} \in \hat{T}\,; \\
&(б) && \int_{\hat{T}} \hat{v}_1 \hat{x}_1^l \hat{x}_2^m d\hat{x}\,, && 0 \le l \le r-1, \quad 0 \le m \le r\,; \\
&(в) && \int_{\hat{T}} \hat{v}_2 \hat{x}_1^l \hat{x}_2^m d\hat{x}\,, && 0 \le l \le r, \quad 0 \le m \le r-1\,,
\end{aligned}
\tag{3.8}
$$

където $\hat{e}$ *е страна на елемента* $\hat{T}$, $\underline{\hat{\nu}}$ *е единичният външен нормален вектор към* $\partial\hat{T}$, *а в случая* $r = 0$ *е в сила само условие (а).*

Доказателство. Ще докажем, че функцията $\underline{\hat{v}} \in \underline{\hat{V}}$ се определя по единствен начин от функционалите, дефинирани в (3.8). Размерността на пространството $\underline{\hat{V}}$ е равно на броя на степените на свобода N, т.е.:

$$\dim \underline{\hat{V}} = 2(r+1)(r+2)$$

и

$$N = 4(r+1) + 2r(r+1) = 2(r+1)(r+2)\,.$$

Следователно достатъчно е да докажем, че функцията $\underline{v} \in \underline{\hat{V}}$, която удовлетворява условията

$$
\begin{aligned}
&(i) && \int_{\hat{e}} \underline{\hat{v}}.\underline{\hat{\nu}}\hat{\varphi} d\hat{\gamma} = 0, && \hat{\varphi} \in Q(r,r) \ \text{върху} \ \forall \hat{e} \in \hat{T}\,; \\
&(ii) \\
&\quad \text{(а)} && \int_{\hat{T}} \hat{v}_1 \hat{x}_1^l \hat{x}_2^m d\hat{x} = 0, && 0 \le l \le r-1,\ 0 \le m \le r\,; \\
&\quad \text{(б)} && \int_{\hat{T}} \hat{v}_2 \hat{x}_1^l \hat{x}_2^m d\hat{x} = 0, && 0 \le l \le r,\ 0 \le m \le r-1\,,
\end{aligned}
\tag{3.9}
$$

се анулира тъждествено.

Условието $(3.9(i))$ предполага, че $\underline{\hat{v}}.\underline{\hat{\nu}} = 0$ върху $\partial\hat{T}$. Следователно използвайки $(3.9(ii))$ и прилагайки формулата на Грийн в $\hat{T}$, получаваме за всяко $\hat{\varphi} \in Q(r,r)$:

$$\int_{\hat{T}} \hat{\varphi} div\underline{\hat{v}} d\hat{x} = -\int_{\hat{T}} \underline{\nabla}\hat{\varphi}.\underline{\hat{v}} d\hat{x} + \int_{\partial\hat{T}} \hat{\varphi}\underline{\hat{v}}.\underline{\hat{\nu}} d\hat{\gamma} = 0\,.$$

Тъй като $div\underline{\hat{v}} \in Q(r,r)$, следователно $div\underline{\hat{v}} = 0$. От дефиницията на $\underline{\hat{v}}$ следва, че съществува полином $\hat{w} \in Q(r+1, r+1)$, определен с точност до константа:

$$\underline{\hat{v}} = \underline{curl}\hat{w} = \left[\frac{\partial\hat{w}}{\partial\hat{x}_2}, -\frac{\partial\hat{w}}{\partial\hat{x}_1}\right].$$

Отбелязваме, че

$$\underline{\hat{v}}.\underline{\hat{\nu}} = \frac{\partial\hat{w}}{\partial\hat{x}_2}\hat{\nu}_1 - \frac{\partial\hat{w}}{\partial\hat{x}_1}\hat{\nu}_2 = \frac{\partial\hat{w}}{\partial\hat{x}_1}\hat{\tau}_1 + \frac{\partial\hat{w}}{\partial\hat{x}_2}\hat{\tau}_2 = \frac{\partial\hat{w}}{\partial\hat{\tau}} = 0\,,$$

където $\frac{\partial}{\partial\hat{\tau}}$ означава тангенциалната производна по $\partial\hat{T}$. Така можем да допуснем, че $\hat{w} = 0$ върху $\partial\hat{T}$, което ни дава възможност да положим

$$\hat{w} = \hat{x}_1(1-\hat{x}_1)\hat{x}_2(1-\hat{x}_2)\hat{z}\,, \quad \hat{z} \in Q(r-1, r-1)\,, \quad \text{за } r \geq 1\,;$$
$$\hat{z} = 0\,, \qquad\qquad \text{за } r = 0\,.$$

Използвайки отново $(3.9(ii))$, получаваме за всяко $\underline{\hat{s}} \in Q(r-1,r) \times Q(r,r-1)$

$$0 = \int_{\hat{T}} \underline{\hat{v}}.\underline{\hat{s}} d\hat{x} = \int_{\hat{T}} \underline{curl}\hat{w}.\underline{\hat{s}} d\hat{x} = \int_{\hat{T}} \hat{w} curl\underline{\hat{s}} d\hat{x} =$$
$$= \int_{\hat{T}} \hat{x}_1(1-\hat{x}_1)\hat{x}_2(1-\hat{x}_2)\hat{z} curl\underline{\hat{s}} d\hat{x}\,,$$

където

$$curl\underline{\hat{s}} = \frac{\partial\hat{s}_2}{\partial\hat{x}_1} - \frac{\partial\hat{s}_1}{\partial\hat{x}_2} \in Q(r-1, r-1)\,.$$

Очевидно можем да изберем $\underline{\hat{s}}$, такова, че $\hat{z} = curl\underline{\hat{s}}$, и тогава

$$\int_{\hat{T}} \hat{x}_1(1-\hat{x}_1)\hat{x}_2(1-\hat{x}_2)\hat{z}^2 d\hat{x} = 0\,.$$

Така получаваме $\hat{z} = 0$, следователно $\hat{w} = 0$ и $\underline{\hat{v}} = curl\hat{w} = \underline{0}$. Това завършва доказателството на лемата.

Определянето на степените на свобода ни дава възможност да въведем проектор

$$\underline{\hat{\pi}} : (H^{r+2}(\hat{T}))^2 \longrightarrow \underline{\hat{V}},$$

чието основно свойство се характеризира в следващата теорема.

Теорема 3.3.1 *За всяка функция* $\underline{\hat{v}} \in (H^{r+2}(\hat{T}))^2$, *съществува единствена функция* $\underline{\hat{\pi}}\,\underline{\hat{v}} \in \underline{\hat{V}}$, *такава, че* $div(\underline{\hat{\pi}}\,\underline{\hat{v}})$ *е ортогонален* $L^2(\hat{T})$ *- проектор на* $div\underline{\hat{v}}$ *върху* $\hat{W}$.

Доказателство. От Лема 3.3.1 следва, че за всяка функция $\underline{\hat{v}} \in (H^{r+2}(\hat{T}))^2$, съществува единствена функция $\underline{\hat{\pi}}\,\underline{\hat{v}} \in \underline{\hat{V}}$, такава, че

(3.10)

(а) $\int_{\hat{e}} (\underline{\hat{v}} - \underline{\pi}\,\underline{\hat{v}}).\underline{\hat{\nu}}\,\hat{\varphi} d\hat{\gamma} = 0, \qquad \forall \hat{\varphi} \in Q(r,r),$

върху всяка страна $\hat{e}$ на $\hat{T}$;

(б) $\int_{\hat{T}} (\hat{v}_1 - \hat{\pi}\hat{v}_1)\hat{x}_1^l \hat{x}_2^m d\hat{x} = 0, \quad 0 \le l \le r-1, \; 0 \le m \le r;$

(в) $\int_{\hat{T}} (\hat{v}_2 - \hat{\pi}\hat{v}_2)\hat{x}_1^l \hat{x}_2^m d\hat{x} = 0, \quad 0 \le l \le r, \; 0 \le m \le r-1.$

Прилагайки формулата на Грийн в $\hat{T}$, получаваме за всяко $\hat{\varphi} \in Q(r,r)$:

(3.11)

$$\int_{\hat{T}} div(\underline{\hat{v}} - \underline{\hat{\pi}}\,\underline{\hat{v}})\hat{\varphi} d\hat{x} = -\int_{\hat{T}} (\underline{\hat{v}} - \underline{\hat{\pi}}\,\underline{\hat{v}}).\underline{\nabla}\hat{\varphi} d\hat{x} + \int_{\partial\hat{T}} (\underline{\hat{v}} - \underline{\hat{\pi}}\,\underline{\hat{v}}).\underline{\hat{\nu}}\hat{\varphi} d\hat{\gamma}.$$

Първият член в дясната страна на равенство (3.11) се анулира по силата на (3.10б) и (3.10в) тъй като $\underline{\nabla}\hat{\varphi} \in Q(r-1,r) \times Q(r,r-1)$. Анулирането на съответния втори член следва от (3.10а). Следователно

$$\int_{\hat{T}} div(\underline{\hat{\pi}}\,\underline{\hat{v}} - \underline{\hat{v}})\hat{\varphi} d\hat{x} = 0.$$

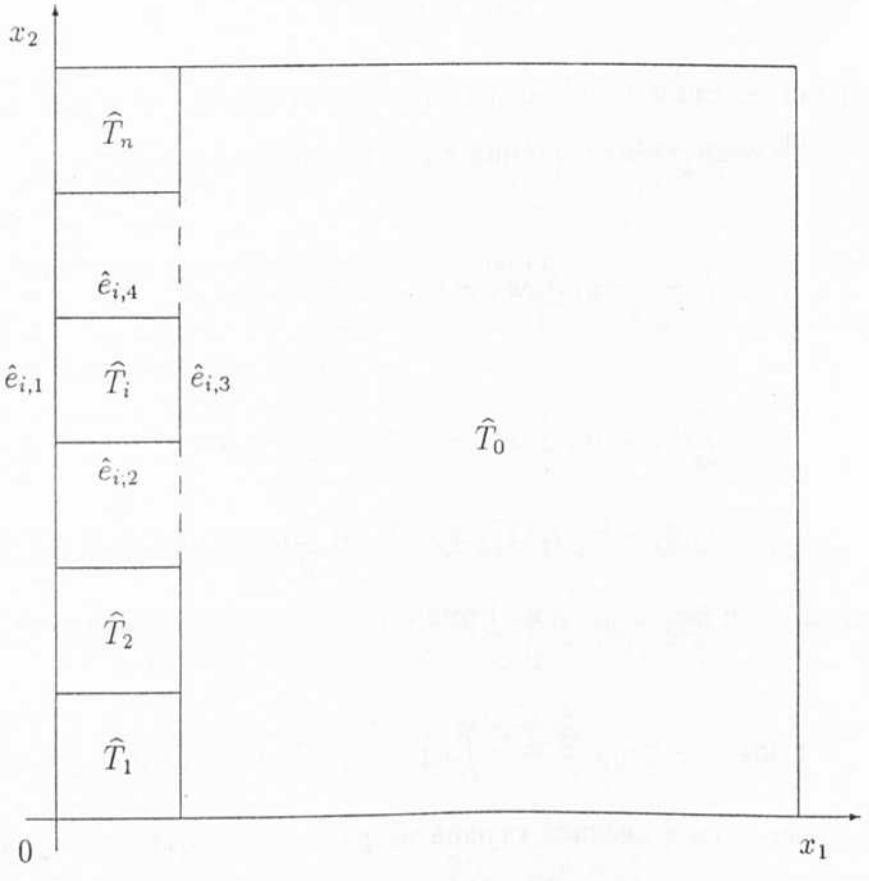
Това завършва доказателството на теоремата.

Във връзка със случая, когато се извършва преход от „груби“ към „фини“ елементи по направлението x_1 от ляво на дясно, въвеждаме следната конфигурация от базисни елементи.

(3.12)

$$\hat{T}_0 = [0,1]^2\,, \quad \hat{T}_i = [0,1/n] \times [(i-1)/n, i/n], \quad i = 1, \ldots, n\,,$$

които са асоциирани с елементите от интерфейса $T_0 \in \tau_c \,:\, T_0 \in \Omega_1$ и съответно $T_i \in \tau_f \,:\, T_i \in I_l\,,\, T_i \in T_0$, както е показано на Фигура 3.2.



Фигура 3.2: Базисна конфигурация от елементи

Забележка 3.3.1 *Базисните конфигурации от елементи, които съответстват на елементите от* I_r, I_d, *и* I_u, *са аналогични.*

Върху елементите $\hat{T}_i$, $i = 1, 2, \ldots, n$ пространствата $\underline{\hat{V}}$ и $\hat{W}$ (1.7) са от същия вид както върху $\hat{T}$. Аналогично на Лема 3.3.1 определяме степените на свобода на $\underline{\hat{v}}$ върху $\hat{T}_i$, $i = 1, 2, \ldots, n$, чрез функционалите:

(3.13)

(а) $\int_{\hat{e}} \underline{\hat{v}}.\underline{\hat{\nu}}\hat{\varphi} d\hat{\gamma}$, $\quad \hat{\varphi} \in Q(r, r)$, върху страните $\hat{e}$,
$\hat{e} = \hat{e}_0, \hat{e}_{1,2}, \hat{e}_{n,4}, \hat{e}_{i,3}$, $i = 1, 2, \ldots, n$;

(б) $\int_{\hat{e}_{i,4}} \hat{v}_2 \hat{x}_1^l d\hat{x}_1$, $\quad 1 \le l \le r$, $i = 1, 2, \ldots, n-1$;

(в) $\int_{\hat{T}_i} \hat{v}_1 \hat{x}_1^l \hat{x}_2^m d\hat{x}$, $\quad 0 \le l \le r-1$, $0 \le m \le r$;

(г) $\int_{\hat{T}_i} \hat{v}_2 \hat{x}_1^l \hat{x}_2^m d\hat{x}$, $\quad 1 \le l \le r$, $0 \le m \le r-1$

и чрез съответно комбиниране на функционалите (3.13а) със следните функционали:

(3.14)

(а) $\int_{\hat{e}_{i,4}} \hat{v}_2 d\hat{x}_1$, $\quad i = 1, 2, \ldots, n-1$;

(б) $\int_{\hat{T}_i} \hat{v}_2 \hat{x}_2^m d\hat{x}$, $\quad 0 \le m \le r-1$,

където $\hat{e}_0$ е лявата страна на елемента $\hat{T}_0$, а разположението на страните $\hat{e}_{i,j}$ на елементите $\hat{T}_i$ е показано на Фигура 3.2.

Степените на свобода, свързани с останалите базисни конфигурации, споменати в Забележка 3.3.1 е напълно аналогичен.

Нашата цел е да въведем модифицирани проектори на Равиар-Тома

$$\underline{\pi_h} \times P_h \; : \; \underline{V} \times W \longrightarrow \underline{V_h} \times W_h \,,$$

които не съвпадат с класическите върху елементите от интерфейса $T \in \tau_f$, $T \in I_f$.

Аналогично на Теорема 3.3.1 първо разглеждаме базисната конфигурация (3.12), съответстваща на елементите от интерфейса $T \in \tau_f$, $T \in I_l$, намиращи се в лявата част на областта Ω_1.

При конструирането на проекторите

$$\underline{\hat{\pi}_i} \; : \; (H^{r+2}(\hat{T}_i))^2 \longrightarrow \underline{\hat{V}_i}, \;\; i = 1, 2, \ldots, n,$$

съблюдаваме следните изисквания:

$$\begin{aligned} &\text{(а)} \qquad \int_{\hat{T}_i} div(\underline{\hat{v}} - \underline{\hat{\pi}_i}\,\underline{\hat{v}})\hat{\varphi} d\hat{x} = 0, \qquad \forall \varphi \in Q(r,r); \\ &\text{(б)} \qquad \int_{\hat{e}} (\underline{\hat{v}} - \underline{\hat{\pi}_i}\,\underline{\hat{v}}).\underline{\hat{\nu}}\hat{\varphi} d\hat{\gamma} = 0, \qquad \forall \varphi \in Q(r,r), \end{aligned} \tag{3.15}$$

където $\hat{e} = \hat{e}_0$, $\hat{e}_{1,2}$, $\hat{e}_{n,4}$, $e_{i,3}$, $i = 1, 2, \ldots, n$.

Условието (3.15а) се определя от изискването $div(\underline{\hat{\pi}_i}\,\underline{\hat{v}})$ да бъде ортогонален $L^2(\hat{T}_i)$-проектор на $div\underline{\hat{v}}$ върху $\hat{W}$, а условието (3.15б), основаващо се на Лема 0.1.6, подсигурява принадлежност на $\underline{\hat{\pi}_i}\,\underline{\hat{v}}$ към $\underline{H}(div; \hat{T})$.

Забележка 3.3.2 *При определянето на степените на свобода (3.13) и (3.14) така, че да възстановят по единствен начин функциите* $\underline{\hat{v}} \in \underline{\hat{V}}$ *върху* $\hat{T}_i$, $i = 1, 2, \ldots, n$, *сме държали сметка проекторите, които ще дефинираме впоследствие чрез тях да изпълняват условията (3.15).*

Обемисти, но несложни от гледна точка на анализа пресмятания ни позволяват да конструираме проекторите $\underline{\hat{\pi}_i}$ върху елементите $\hat{T}_i$, $i = 1, 2, \ldots, n$ (3.12), по следния начин:

(3.16)

(а) $\int_{\hat{e}}(\underline{\hat{v}} - \underline{\hat{\pi}}_i\,\underline{\hat{v}}).\underline{\hat{\nu}}\hat{\varphi}d\hat{\gamma} = 0\,,$ $\hat{\varphi} \in Q(r,r)\,,$ върху страните $\hat{e}\,,$ $\hat{e} = \hat{e}_0, \hat{e}_{1,2}, \hat{e}_{n,4}, \hat{e}_{i,3}\,,\ i = 1, \ldots, n\,;$

(б) $\int_{\hat{e}_{i,4}}(\hat{v}_2 - \hat{\pi}_i\,\hat{v}_2)\hat{x}_1^l d\hat{x}_1 = 0\,,$ $1 \le l \le r\,,\ i = 1, 2, \ldots, n-1\,;$

(в) $\int_{\hat{T}_i}(\hat{v}_1 - \hat{\pi}_i\,\hat{v}_1)\hat{x}_1^l\hat{x}_2^m d\hat{x} = 0\,,$ $0 \le l \le r-1\,,\ 0 \le m \le r\,;$

(г) $\int_{\hat{T}_i}(\hat{v}_2 - \hat{\pi}_i\,\hat{v}_2)\hat{x}_1^l\hat{x}_2^m d\hat{x} = 0\,,$ $1 \le l \le r\,,\ 0 \le m \le r-1\,;$

(д) $\int_{\hat{e}_{i,4}}(\hat{v}_2 - \hat{\pi}_i\,\hat{v}_2)d\hat{x}_1 = F_1(b_m)\,,$ $i = 1, 2, \ldots, n-1\,;$

(е) $\int_{\hat{T}_i}(\hat{v}_2 - \hat{\pi}_i\,\hat{v}_2)\hat{x}_2^m d\hat{x} = F_2(b_m)\,,$ $0 \le m \le r-1\,,$

където функционалите $F_j(b_m),\ j = 1,2$ се конструират с помощта на основните функционали

$$b_m(\hat{v}_1) = \int_{e_0}\hat{v}_1\hat{x}_2^m d\hat{x}_2\,, \qquad m = 0, 1, \ldots, r\,.$$

Забележка 3.3.3 *На практика смесеният метод на крайните елементи се прилага най-често в случаите* $r = 0$ *и* $r = 1$*, които в някакъв смисъл отговарят съответно на линейни и квадратични елементи в класическия метод на крайните елементи. Получаването на конкретният вид на проекторите (3.16) в първия случай е описан подробно в публикацията на автора (виж Димов [2]).*

Поради взаимната връзка между проекторите $\underline{\hat{\pi}}_i\,,\ i = 1, 2, \ldots, n\,,$ можем да дефинираме проектор $\underline{\tilde{\pi}}$ върху ивицата $\hat{I} = \bigcup_{i=1}^n \hat{T}_i$ по следния начин:

(3.17) $$\underline{\tilde{\pi}}\,\underline{\hat{v}}_{|T} = \underline{\hat{\pi}}_i\,\underline{\hat{v}} \qquad i = 1, 2, \ldots, n\,,$$

чието основно свойство е, че запазва полиномите от $Q(r+1,r)\times Q(r,r+1)$ върху съответната ивица.

В по-нататъшния анализ ще се нуждаем и от ортогоналния L^2-проектор $\hat{P}$ на функциите $\hat{w}\in\hat{W}$ върху пространството $div\,\underline{\hat{v}}$, дефинирано чрез (1.7) в основния елемент $\hat{T}=[-1,1]^2$, който е асоцииран с елементите $T\in\tau_h$ т.е.

$$(div\,\underline{\hat{v}},\,\hat{w}-\hat{P}\hat{w})=0\,,\qquad \forall\underline{\hat{v}}\in\underline{\hat{V}}\,. \tag{3.18}$$

В този случай, тъй като не съществуват никакви изисквания за непрекъснатост по границата на крайните елементи, те съвпадат с тези, дефинирани от Равиар и Тома [1] и не зависят от локалното сгъстяване.

Разглеждаме единствената афинна обратима трансформация $F_T:\hat{x}\to x=F_T(\hat{x})$, дефинирана в (1.9), на $\hat{T}$ върху T. За всяка скаларнозначна функция w върху T, дефинираме:

$$w=\hat{w}\circ F_T^{-1}\,,\qquad (\hat{w}=w\circ F_T)\,, \tag{3.19}$$

а за всяка векторнозначна функция $\underline{v}$ върху T:

$$\underline{v}=\frac{1}{J_T}B_T\underline{\hat{v}}\circ F_T^{-1},\qquad \left(\underline{\hat{v}}=J_TB_T^{-1}\underline{v}\circ F_T\right), \tag{3.20}$$

където $J_T=\det(B_T)$.

Върху елементите $T\in\tau_h:\ T\in\Omega\backslash I_f$, асоциирани с основния елемент $\hat{T}$, разглеждаме пространството

$$\underline{V}_T=\left\{\underline{v}\in\underline{H}(div;T),\,\underline{\hat{v}}\in\underline{\hat{V}}\right\}.$$

Върху елементите $T\in\tau_f,\ T\in I_l$, намиращи се в лявата част на областта Ω_1, асоциирани с базисната конфигурация (3.12), разглеждаме пространствата

$$\underline{V}_{T_i}=\left\{\underline{v}\in\underline{H}(div;T_i),\,\underline{\hat{v}}\in\underline{\hat{V}}\right\},\quad i=1,2,\ldots,n\,.$$

Пространствата, които разглеждаме върху елементите от I_r, I_d, и I_u, са подобни.

Така връзката между пространствата $\underline{V_h}$ и $\underline{V_T}$ се осъществява чрез следното съотношение:

$$\underline{v_h}_{|T} \in \underline{V_T}, \text{ за всяко } \underline{v_h} \in \underline{V_h}, \ T \in \tau_h. \tag{3.21}$$

Аналогична е и връзката между пространствата W_h и W_T:

$$w_{h|T} \in W_T, \text{ за всяко } w_h \in W_h, \ T \in \tau_h, \tag{3.22}$$

където

$$W_T = \left\{ w \in L^2(T), \hat{w} \in \hat{W} \right\}.$$

За елементите $T \in \tau_h$, използвайки (3.18) и трансформацията (3.19), дефинираме проектора P_T чрез

$$\widehat{P_T w} = \hat{P}\hat{w}, \qquad \text{за всяко } w \in H^{r+1}(T). \tag{3.23}$$

Аналогично, за елементите $T \in \tau_h : T \in \Omega \backslash I_f$, използвайки (3.10) и трансформацията (3.20), дефинираме проектора $\underline{\pi_T}$ чрез

$$\widehat{\underline{\pi_T}\, \underline{v}} = \hat{\underline{\pi}}\, \hat{\underline{v}}, \qquad \text{за всяко } \underline{v} \in (H^{r+2}(T))^2. \tag{3.24}$$

Накрая, за елементите от ивицата $I = \bigcup_{i=1}^{n} T_i : T_i \in I_f$, $T_i \in T_c \in \Omega_1$, използвайки (3.17) дефинираме проектора $\underline{\pi_I}$ чрез

$$\widehat{\underline{\pi_I}\, \underline{v}} = \tilde{\underline{\pi}}\, \hat{\underline{v}}, \qquad \text{за всяко } \underline{v} \in (H^{r+2}(T))^2. \tag{3.25}$$

В съответствие с (3.23), (3.24), (3.25) и връзките (3.21), (3.22), въвеждаме модифицирани проектори на Равиар - Тома

$$\underline{\pi_h} \times P_h \ : \ \underline{V_h} \times W_h \longrightarrow \underline{V_h} \times W_h. \tag{3.26}$$

притежаващи следните свойства за $L^2(\Omega)$-ортогоналност:

$$
\begin{aligned}
&\text{(а)} \qquad (div\,(\underline{v} - \underline{\pi_h}\,\underline{v}),\, w_h) = 0\,, \qquad \forall w_h \in W_h\,; \\
&\text{(б)} \qquad (div\,\underline{v_h},\, w - P_h w) = 0\,, \qquad \forall \underline{v_h} \in \underline{V_h}\,.
\end{aligned}
\tag{3.27}
$$

Следващата теорема дава основните апроксимационни свойства на въведените по-горе проектори.

Теорема 3.3.2 *Нека* $w \in H^{r+1}(\Omega)$ *и* $\underline{v} \in (H^{r+2}(\Omega))^2$. *Тогава за проекторите (3.26) са в сила следните оценки:*

(3.28)

$$
\begin{aligned}
&(а) \quad \|w - P_h w\|_{-s,\Omega} \le C\left[h_f^{k+s}\|w\|_{k,\Omega_1} + h_c^{k+s}\|w\|_{k,\Omega_2}\right],\ 0 \le k,\, s \le r+1\,; \\
&(б) \quad \|div(\underline{v} - \underline{\pi_h v})\|_{-s,\Omega} \le C\left[h_f^{k+s}\|div\underline{v}\|_{k,\Omega_1} + h_c^{k+s}\|div\underline{v}\|_{k,\Omega_2}\right],\ 0 \le k, s \le r+1\,; \\
&(в) \quad \|\underline{v} - \underline{\pi_h v}\|_{0,\Omega} \le C\left[h_f^{k}\|\underline{v}\|_{k,\Omega_1 \setminus I_f} + h_c^{k}\|\underline{v}\|_{k,\Omega_2} + h_c^{k}\|\underline{v}\|_{k,I_f}\right],\ 1 \le k \le r+1\,.
\end{aligned}
$$

Доказателство. Тъй като техниката на конкретните пресмятания съвпада по същество с тази, използвана в предните глави, тук поясним само някои конкретни различия, породени от конкретния вид на проекторите.

Неравенствата (3.28а) и (3.28б) се получават със стандартна техника (Равиар и Тома [1]) и са аналогични на резултатите от Глава 2, използващи интерполационна техника. Отбелязваме, че апроксимирането чрез L^2-проекторите (3.26), притежаващи свойствата (3.27), е по-качествено от съответното използване на интерполанти (виж Колац [1]), което играе съществена роля в получаването на оценката (3.28б).

Видът на първите два члена в дясната страна на неравенство (3.28в) непосредствено следва от прилагането на стандартната техника в случая на липса на гореописания тип локално сгъстяване.

Произходът на третия член в дясната страна на неравенство (3.28в) изисква някои допълнителни уточнения, които следват от факта, че в този случай не можем да се позовем на аналогични свойства на (3.27) за ортогоналност за елементите от интерфейса I_f. Тук анализът не се прави отделно върху всеки елемент $T \in I_f$, а върху ивицата $I = \bigcup_{i=1}^{n} T_i$, състояща се от елементи $T_i \in I_f \ : \ T_i \in T_0$, където елемента $T_0 \in \tau_c \ : \ T_0 \in \Omega_1$. Локалните оценки върху всяка една ивица се основават на дефинирания в (3.25) проектор, съответстващ на проектора (3.17), който притежава свойството да запазва полиномите от $Q(r+1,r) \times Q(r,r+1)$. Това е фактора, влияещ на порядъка в оценката. Трансформацията (3.20) при $T \equiv T_0$, дава връзката между $I \in T_0$ и базисната конфигурация, дефинирана в (3.12), което обуславя появяването на параметъра h_c в третия член от дясната страна на неравенство (3.28в).

3.4 Дуални леми

Казваме, че областта Ω е $s+2$-регулярна, ако задачата на Дирихле

$$\begin{aligned} &\text{(а)} \quad L^*\varphi = \psi, \qquad x \in \Omega; \\ &\text{(б)} \quad \varphi = 0, \qquad x \in \partial\Omega \end{aligned} \tag{3.29}$$

притежава единствено решение за $\psi \in L^2(\Omega)$ и ако

$$\|\varphi\|_{s+2} \leq C\|\psi\|_s \tag{3.30}$$

за всяко $\psi \in H^s(\Omega)$.

Ако $f \in \underline{V}'$ е дуалното пространство на $\underline{V}$, тогава тя може да бъде представена чрез двойката функции $\{\underline{f_0}, f_1\} \in L^2(\Omega)^2 \times L^2(\Omega)$, такава, че

$$\underline{f}(\underline{v}) = (\underline{f_0}, \underline{v}) + (f_1, div\, \underline{v}) \tag{3.31}$$

Първата дуална лема е следната:

Лема 3.4.1 *Нека индексът* r *на* $\underline{V_h} \times W_h$, *дефинирани в (2.5), е най-малко едно, и нека* $0 \le s \le r-1$. *Допускаме, че* Ω *е* $s+2$*-регулярна. Нека* $\underline{\zeta} \in \underline{V}, \underline{f} = \{\underline{f_0}, 0\} \in \underline{V}'$ *и* $g \in W' = L^2(\Omega)$. *Ако* $z \in W_h$ *удовлетворява релациите:*

$$
\begin{aligned}
&(a) \qquad (\alpha\underline{\zeta}, \underline{v}) - (div\underline{v}, z) + (\underline{\beta} z, \underline{v}) = \underline{f}(\underline{v}), \qquad \underline{v} \in \underline{V_h}; \\
&(б) \qquad (div\underline{\zeta}, w) + (cz, w) = g(w), \qquad\qquad w \in W_h,
\end{aligned}
\tag{3.32}
$$

тогава за достатъчно малко h_c *е в сила:*

$$
\begin{aligned}
\|z\|_{-s,\Omega} \le\ & C\Big[\|\underline{f}\|_{-s-1,\Omega} + \|g\|_{-s-2,\Omega} \\
&+ h_f^{s+2}\|g\|_{0,\Omega_1} + h_c^{s+2}\|g\|_{0,\Omega_2} \\
&+ h_f^{s+1}\|\underline{f}\|_{0,\Omega_1\setminus I_f} + h_c^{s+1}\|\underline{f}\|_{0,\Omega_2} + h_c^{s+1}\|\underline{f}\|_{0,I_f} \\
&+ h_f^{s+1}\|\underline{\zeta}\|_{0,\Omega_1\setminus I_f} + h_c^{s+1}\|\underline{\zeta}\|_{0,\Omega_2} + h_c^{s+1}\|\underline{\zeta}\|_{0,I_f} \\
&+ h_f^{s+2}\|div\,\underline{\zeta}\|_{0,\Omega_1} + h_c^{s+2}\|div\,\underline{\zeta}\|_{0,\Omega_2}\Big].
\end{aligned}
\tag{3.33}
$$

Доказателство. Нека $\psi \in H^s(\Omega)$ и нека $\varphi \in H^{s+2}(\Omega) \bigcap H_0^1(\Omega)$ е решение на (3.29). Комбинирайки (3.32а) при $\underline{v} = \underline{\pi_h}(a\underline{\nabla}\varphi)$ и (3.27а) при $\underline{v} = a\underline{\nabla}\varphi$ и $w_h = z$, получаваме

$$
\begin{aligned}
(z,\psi) =\ & \Big(z,\ -div(a\underline{\nabla}\varphi) + \underline{\beta}\cdot a\underline{\nabla}\varphi + c\varphi\Big) \\
=\ & -\Big(\alpha\underline{\zeta},\ \underline{\pi_h}(a\underline{\nabla}\varphi)\Big) + \underline{f}\Big(\underline{\pi_h}(a\underline{\nabla}\varphi)\Big) \\
& + \Big(\underline{\beta} z,\ a\underline{\nabla}\varphi - \underline{\pi_h}(a\underline{\nabla}\varphi)\Big) + (cz,\ \varphi).
\end{aligned}
$$

Тогава от (3.32б) при $w = \varphi$ следва

$$
(cz,\ \varphi) = (cz + div\,\underline{\zeta},\ \varphi - P_h\varphi) + g(P_h\varphi) - (div\,\underline{\zeta},\ \varphi)
$$

Продължаваме приведението в подходящ вид, използвайки формулата на Грийн и означенията (3.4):

$$
\begin{aligned}
-(div\,\underline{\zeta},\ \varphi) &= (\underline{\zeta},\ \underline{\nabla}\varphi) = (\alpha\underline{\zeta},\ a\underline{\nabla}\varphi) \\
&= \Big(\alpha\underline{\zeta},\ \underline{\pi_h}(a\underline{\nabla}\varphi)\Big) + \Big(\alpha\underline{\zeta},\ a\underline{\nabla}\varphi - \underline{\pi_h}(a\underline{\nabla}\varphi)\Big).
\end{aligned}
$$

Така получаваме следното междинно представяне:

$$
\begin{aligned}
(z,\psi) = \; & \underline{f}\left(\underline{\pi_h}(a\underline{\nabla}\varphi)\right) + g(P_h\varphi) \\
& + \left(\alpha\underline{\zeta} + \underline{\beta}z, a\underline{\nabla}\varphi - \underline{\pi_h}(a\underline{\nabla}\varphi)\right) \\
& + (div\underline{\zeta} + cz, \varphi - P_h\varphi)\,.
\end{aligned} \tag{3.34}
$$

Използвайки Теорема 3.3.2 и (3.30), оценяваме последователно членовете от дясната страна на равенство (3.34):

(3.35)

$$
\begin{aligned}
& |\underline{f}(\underline{\pi_h}(a\underline{\nabla}\varphi))| \le |\underline{f}(a\underline{\nabla}\varphi)| + |\underline{f}(\underline{\pi_h}(a\underline{\nabla}\varphi) - a\underline{\nabla}\varphi))| \\
\le \; & C\left[\|\underline{f}\|_{-s-1,\Omega}\|a\underline{\nabla}\varphi\|_{s+1,\Omega} + \|\underline{f}\|_{0,\Omega}\|\underline{\pi_h}(a\underline{\nabla}\varphi) - a\underline{\nabla}\varphi\|_{0,\Omega}\right] \\
\le \; & C\left[\|\underline{f}\|_{-s-1,\Omega} + h_f^{s+1}\|\underline{f}\|_{0,\Omega_1\setminus I_f} + h_c^{s+1}\|\underline{f}\|_{0,\Omega_2} + h_c^{s+1}\|\underline{f}\|_{0,I_f}\right]\|\psi\|_{0,\Omega}\,.
\end{aligned}
$$

Аналогично:

$$
\begin{aligned}
& |g(P_h(\varphi))| \le |g(\varphi)| + |g(P_h(\varphi)) - g(\varphi)| \\
& \le C\left[\|g\|_{-s-2,\Omega} + h_f^{s+2}\|g\|_{0,\Omega_1} + h_c^{s+2}\|g\|_{0,\Omega_2}\right]\|\psi\|_{0,\Omega}\,;
\end{aligned} \tag{3.36}
$$

(3.37)

$$
\begin{aligned}
& (\alpha\underline{\zeta} + \underline{\beta}z, a\underline{\nabla}\varphi - \underline{\pi_h}(a\underline{\nabla}\varphi)) \\
\le \; & C\left[h_f^{s+1}\|\underline{\zeta}\|_{0,\Omega_1\setminus I_f} + h_c^{s+1}\|\underline{\zeta}\|_{0,\Omega_2} + h_c^{s+1}\|\underline{\zeta}\|_{0,I_f}\right. \\
& + \left. h_f^{s+1}\|z\|_{0,\Omega_1\setminus I_f} + h_c^{s+1}\|z\|_{0,\Omega_2} + h_c^{s+1}\|z\|_{0,I_f}\right]\|\psi\|_{0,\Omega}
\end{aligned}
$$

и

$$
\begin{aligned}
& (div\underline{\zeta} + cz,\, \varphi - P_h\varphi) \\
\le \; & C\left[h_f^{s+2}\|div\underline{\zeta}\|_{0,\Omega_1} + h_c^{s+2}\|div\underline{\zeta}\|_{0,\Omega_2}\right. \\
& + \left. h_f^{s+2}\|z\|_{0,\Omega_1} + h_c^{s+2}\|z\|_{0,\Omega_2}\right]\|\psi\|_{0,\Omega}\,.
\end{aligned} \tag{3.38}
$$

От неравенства (3.35) – (3.38) получаваме следната оценка:

(3.39)

$$\begin{aligned}\|z\|_{-s,\Omega} \leq\ & C\Big[\|\underline{f}\|_{-s-1,\Omega} + \|g\|_{-s-2,\Omega} \\ & + h_f^{s+2}\|g\|_{0,\Omega_1} + h_c^{s+2}\|g\|_{0,\Omega_2} \\ & + h_f^{s+1}\|\underline{f}\|_{0,\Omega_1\setminus I_f} + h_c^{s+1}\|\underline{f}\|_{0,\Omega_2} + h_c^{s+1}\|\underline{f}\|_{0,I_f} \\ & + h_f^{s+1}\|\underline{\zeta}\|_{0,\Omega_1\setminus I_f} + h_c^{s+1}\|\underline{\zeta}\|_{0,\Omega_2} + h_c^{s+1}\|\underline{\zeta}\|_{0,I_f} \\ & + h_f^{s+2}\|div\,\underline{\zeta}\|_{0,\Omega_1} + h_c^{s+2}\|div\,\underline{\zeta}\|_{0,\Omega_2} \\ & + h_f^{s+1}\ \|z\|_{0,\Omega_1\setminus I_f} + h_c^{s+1}\|z\|_{0,\Omega_2} + h_c^{s+1}\|z\|_{0,I_f}\Big]\ .\end{aligned}$$

В случая $s = 0$ и при достатъчно малко h_c, последните три члена в дясната страна на неравенство (3.39) се поглъщат от лявата му част, като променят само константата C, т.е. неравенство (3.33) е в сила. За $0 < s \leq r-1$ заместваме рекурсивно членовете, съдържащи $\|z\|_0$. Стандартната интерполационна теория за пространствата $H(\Omega)'$ (виж Лионс и Мадженес [1]) предполага

$$\|g\|_{-2,\Omega_1} \leq C\left[h_f\|g\|_{0,\Omega_1} + h_f^{-s/2}\|g\|_{-s-2,\Omega_1}\right];$$

$$\|g\|_{-2,\Omega_2} \leq C\left[h_c\|g\|_{0,\Omega_2} + h_c^{-s/2}\|g\|_{-s-2,\Omega_2}\right]$$

и води до мажориране на последните три члена в неравенство (3.39) от членовете в дясната страна на неравенство (3.33), което завършва доказателството на лемата.

Лема 3.4.2 *Нека индексът* r *на* $\underline{V_h} \times W_h$, *дефинирани в (2.5), е неотрицателен и* Ω *е* $(r+2)$*-регулярна. Нека* $\underline{\zeta} \in \underline{V}$, $\underline{f} = \{\underline{f_0}, 0\} \in \underline{V'}$ *и* $g \in W' = L^2(\Omega)$. *Ако* z *удовлетворява (3.32), тогава за достатъчно*

малко h_c *е в сила:*

$$
\begin{aligned}
\|z\|_{-r,\Omega} \leq\ & C\ \Big[\|\underline{f}\|_{-r-1,\Omega} + \|g\|_{-r-2,\Omega} \\
& + h_f^{r+1}\|g\|_{0,\Omega_1} + h_c^{r+1}\|g\|_{0,\Omega_2} \\
& + h_f^{r+1}\|\underline{f}\|_{0,\Omega_1\setminus I_f} + h_c^{r+1}\|\underline{f}\|_{0,\Omega_2} + h_c^{r+1}\|\underline{f}\|_{0,I_f} \\
& + h_f^{r+1}\|\underline{\zeta}\|_{0,\Omega_1\setminus I_f} + h_c^{r+1}\|\underline{\zeta}\|_{0,\Omega_2} + h_c^{r+1}\|\underline{\zeta}\|_{0,I_f} \\
& + h_f^{r+1}\|div\underline{\zeta}\|_{0,\Omega_1} + h_c^{r+1}\|div\underline{\zeta}\|_{0,\Omega_2}\Big]\,.
\end{aligned}
\tag{3.40}
$$

Доказателство. Доказателството е напълно аналогично с това на Лема 3.4.1 с тази разлика, че на две места в разсъжденията, ограничението в апроксимацията на $\varphi - P_h\varphi$ (3.28a) редуцира порядъка от $s+2 = r+2$ на $r+1$.

3.5 Оценка на грешката в $L^2(\Omega)$

Базирайки се на конструкциите и твърденията в предходните параграфи достигаме до основният резултат в тази глава:

Теорема 3.5.1 *Предполагаме, че задачата на Дирихле (3.2) притежава единствено решение* $p \in H^2(\Omega)$ *за всяка двойка* $\{\underline{f}, g\} \in L^2(\Omega) \times H^{3/2}(\partial\Omega)$ *и, че* Ω *е* $r+2$*-регулярна. Тогава, за достатъчно малко* h_c *съществува единствено решение* $\{\underline{v_h}, p_h\} \in \underline{V_h} \times W_h$ *на смесения метод на крайните елементи върху мрежа с регулярно локално сгъстяване даден с (3.7). Освен това при* $p \in H^{r+3}(\Omega)$ *са в сила следните оценки на грешката между*

точното и приближено решение

$$(а)\quad \|p-p_h\|_{0,\Omega} \le C\left[h_f\|p\|_{2,\Omega_1} + h_c\|p\|_{2,\Omega_2}\right], \qquad r=0;$$

$$(б)\quad \|p-p_h\|_{0,\Omega} \le C\left[h_f^k\|p\|_{k,\Omega_1\setminus I_f} + h_c^k\|p\|_{k,\Omega_2} + h_c^k\|p\|_{k,I_f}\right],$$

$$r\ge 1, \quad 2\le k\le r+1;$$

$$(в)\quad \|\underline{u}-\underline{u_h}\|_{0,\Omega} \le C\left[h_f^k\|p\|_{k+1,\Omega_1\setminus I_f} + h_c^k\|p\|_{k+1,\Omega_2} + h_c^k\|p\|_{k+1,I_f}\right],$$

$$1\le k\le r+1;$$

$$(г)\quad \|div(\underline{u}-\underline{u_h})\|_{0,\Omega} \le C\left[h_f^k\|p\|_{k+2,\Omega_1} + h_c^k\|p\|_{k+2,\Omega_2} + h_c^k\|p\|_{k,I_f}\right],$$

$$0\le k\le r+1.$$

Доказателство. Допускаме че (3.7) притежава единствено решение за достатъчно малки h_f и h_c (което ще следва непосредствено от по-нататъшния анализ за сходимост). Полагаме

$$(3.41)\qquad \begin{aligned} &(а)\quad \underline{\zeta}=\underline{u}-\underline{u_h}, \quad \underline{\sigma}=\underline{\pi_h u}-\underline{u_h};\\ &(б)\quad \eta=p-p_h, \quad \tau=P_h p-p_h, \quad \rho=p-P_h p. \end{aligned}$$

Тогава от (3.2) и (3.7) получаваме

$$(3.42)\qquad \begin{aligned} &(а)\quad (\alpha\underline{\xi},\underline{v})-(div\underline{v},\eta)+(\underline{\beta}\eta,\underline{v})=0, \qquad \underline{v}\in\underline{V_h};\\ &(б)\quad (div\underline{\xi},w)+(c\eta,w)=0, \qquad w\in W_h, \end{aligned}$$

което, имайки предвид (3.27б) е еквивалентно на

(3.43)

$$\begin{aligned} &(а)\quad (\alpha\underline{\xi},\underline{v})-(div\underline{v},\tau)+(\underline{\beta}\tau,\underline{v})=-(\underline{\beta}\rho,\underline{v}), \qquad \underline{v}\in\underline{V_h};\\ &(б)\quad (div\underline{\xi},w)+(c\tau,w)=-(c\rho,w), \qquad w\in W_h. \end{aligned}$$

Прилагайки Лема 3.3.1 при $s=0$ и Лема 3.3.2 при $r=0$ към уравненията на грешката (3.43) за достатъчно малки h_f и h_c и Ω 2-регулярна,

получаваме

(3.44)

$$\begin{aligned}\|\tau\|_{0,\Omega} \leq\ & C\left[\|\rho\|_{-1,\Omega} + h_f\|\rho\|_{0,\Omega_1\setminus I_f} + h_c\|\rho\|_{0,\Omega_2} + h_c\|\rho\|_{0,I_f}\right.\\ & + h_f\|\underline{\xi}\|_{0,\Omega_1\setminus I_f} + h_c\|\underline{\xi}\|_{0,\Omega_2} + h_c\|\underline{\xi}\|_{0,I_f}\\ & \left. + h_f^{2-\delta_{r0}}\|div\underline{\xi}\|_{0,\Omega_1} + h_c^{2-\delta_{r0}}\|div\underline{\xi}\|_{0,\Omega_2}\right].\end{aligned}$$

Използвайки апроксимационните свойства на проектора P_h получаваме

$$\begin{aligned}&\text{(а)} && \|\rho\|_{0,\Omega_1\setminus I_f} \leq Ch_f^k\|p\|_{k,\Omega_1\setminus I_f}, && 0 \leq k \leq r+1;\\ &\text{(б)} && \|\rho\|_{0,\Omega_2} \leq Ch_c^k\|p\|_{k,\Omega_2}, && 0 \leq k \leq r+1;\\ &\text{(в)} && \|\rho\|_{0,I_f} \leq Ch_f^k\|p\|_{k,I_f}, && 0 \leq k \leq r+1.\end{aligned} \tag{3.45}$$

От друга страна

$$\|\rho\|_{-1,\Omega} \leq C\left[h_f^{t+1}\|p\|_{t,\Omega_1} + h_c^{t+1}\|p\|_{t,\Omega_2}\right],\ 0 \leq t \leq r+1. \tag{3.46}$$

От (3.44), (3.45) и (3.46) за $t = k$ и тъй като $\eta = \rho + \tau$ за достатъчно малко h_c получаваме

(3.47)

$$\begin{aligned}\|\eta\|_{0,\Omega} \leq\ & C\left[h_f^k\|p\|_{k,\Omega_1} + h_c^k\|p\|_{k,\Omega_2}\right.\\ & + h_f\|\underline{\xi}\|_{0,\Omega_1\setminus I_f} + h_c\|\underline{\xi}\|_{0,\Omega_2} + h_c\|\underline{\xi}\|_{0,I_f}\\ & \left. + h_f^{2-\delta_{r0}}\|div\underline{\xi}\|_{0,\Omega_1} + h_c^{2-\delta_{r0}}\|div\underline{\xi}\|_{0,\Omega_2}\right], \qquad 0 \leq k \leq r+1.\end{aligned}$$

Тъй като от (3.27а), $(div\underline{\sigma}, w) = (div\underline{\xi}, w)$ за $w \in W_h$, от (3.42б) при избора $w = div\underline{\sigma} \in W_h$ следва, че

$$\|div\underline{\sigma}\|_{0,\Omega} \leq C\|\eta\|_{0,\Omega}.$$

Следователно, използвайки означенията (3.41а), получаваме

(3.48)

$$\|div\underline{\xi}\|_{0,\Omega_1} \leq C\left[\|\eta\|_{0,\Omega_1} + h_f^q\|div\underline{u}\|_{q,\Omega_1}\right], \quad 0 \leq q \leq r+1;$$

$$\|div\underline{\xi}\|_{0,\Omega_2} \leq C\left[\|\eta\|_{0,\Omega_2} + h_c^q\|div\underline{u}\|_{q,\Omega_2}\right], \quad 0 \leq q \leq r+1.$$

Ако в (3.42a) изберем тестова функция $\underline{v} = \underline{\sigma}$ тогава

$$(\alpha\underline{\sigma}, \underline{\sigma}) = (div\underline{\sigma}, \eta) - (\underline{\beta}\eta, \underline{\sigma}) - ((\alpha\underline{u} - \underline{\pi_h u}), \underline{\sigma});$$

Следователно

$$\|\underline{\sigma}\|_{0,\Omega} \leq C\left[\|\eta\|_{0,\Omega} + \|\underline{u} - \underline{\pi_h u}\|_{0,\Omega}\right]$$

и тъй като $\underline{\xi} = \underline{\sigma} + \underline{u} - \underline{\pi_h u}$, получаваме

(3.49)
$$\|\underline{\xi}\|_{0,\Omega} \leq C\left[\|\eta\|_{0,\Omega} + \|\underline{u} - \underline{\pi_h u}\|_{0,\Omega}\right]$$

Тогава от неравенство (3.28в) следва:

(3.50)

$$\|\underline{\xi}\|_{0,\Omega} \leq C\left[\|\eta\|_{0,\Omega} + h_f^t\|\underline{u}\|_{t,\Omega_1\setminus I_f} + h_c^t\|\underline{u}\|_{t,\Omega_2} + h_c^t\|\underline{u}\|_{t,I_f}\right], \; 0 \leq t \leq r+1.$$

Ако заместим (3.48) и (3.50) в (3.47), тогава за $0 \leq k,\ q \leq r+1$ и $1 \leq t \leq r+1$ следва:

$$\begin{aligned}\|\eta\|_{0,\Omega} \leq\ & C\left[h_f^k\|p\|_{k,\Omega_1} + h_c^k\|p\|_{k,\Omega_2} + h_c\|\eta\|_{0,\Omega}\right.\\ & + h_f^{t+1}\|\underline{u}\|_{t,\Omega_1\setminus I_f} + h_c^{t+1}\|\underline{u}\|_{t,\Omega_2} + h_c^{t+1}\|\underline{u}\|_{t,I_f}\\ & + h_f^{q+2-\delta_{r0}}\|div\underline{u}\|_{q,\Omega_1} + h_c^{q+2-\delta_{r0}}\|div\underline{u}\|_{q,\Omega_2}\\ & \left. + h_f^{2-\delta_{r0}}\|\eta\|_{0,\Omega_1} + h_c^{2-\delta_{r0}}\|\eta\|_{0,\Omega_2}\right],\end{aligned}$$

т.е. за достатъчно малко h_c получаваме

$$\begin{aligned}\|\eta\|_{0,\Omega} \leq\ & C\left[h_f^k\|p\|_{k,\Omega_1} + h_c^k\|p\|_{k,\Omega_2}\right.\\ & + h_f^{t+1}\|\underline{u}\|_{t,\Omega_1\setminus I_f} + h_c^{t+1}\|\underline{u}\|_{t,\Omega_2} + h_c^{t+1}\|\underline{u}\|_{t,I_f}\\ & \left. + h_f^{q+2-\delta_{r0}}\|div\underline{u}\|_{q,\Omega_1} + h_c^{q+2-\delta_{r0}}\|div\underline{u}\|_{q,\Omega_2}\right].\end{aligned}$$

Тогава при избор на $k = t+1 = q+2-\delta_{r0}$ и тъй като

$$\|\underline{u}\|_{k-1} + \|div\underline{u}\|_{k-2} \leq C\|p\|_k\,,$$

получаваме

$$\|\eta\|_{0,\Omega} \leq C\left[h_f\|p\|_{2,\Omega_1} + h_c\|p\|_{2,\Omega_2}\right], \quad r = 0\,;$$

$$\|\eta\|_{0,\Omega} \leq C\left[h_f^k\|p\|_{k,\Omega_1\setminus I_f} + h_c^k\|p\|_{k,\Omega_2} + h_c^k\|p\|_{k,I_f}\right], \quad r \geq 1\,,\ 2 \leq k \leq r+1.$$

От (3.49) и (3.48) следва, че

$$\|\underline{\xi}\|_{0,\Omega} \leq C\left[h_f^k\|p\|_{k+1,\Omega_1\setminus I_f} + h_c^k\|p\|_{k+1,\Omega_2} + h_c^k\|p\|_{k+1,I_f}\right],$$
$$1 \leq k \leq r+1\,.$$

и

$$\|div\underline{\xi}\|_{0,\Omega} \leq C\left[h_f^k\|p\|_{k+2,\Omega_1} + h_c^k\|p\|_{k+2,\Omega_2} + h_c^k\|p\|_{k,I_f}\right],$$
$$0 \leq k \leq r+1\,.$$

Преди да завършим анализа на грешката ще покажем съществуване и единственост на решението на (3.7). Тъй като системата (3.7) е линейна, то достатъчно е да се докаже единственост. Нека функциите $\underline{f}$ и g се анулират. Тогава при избора $w = div\underline{u_h}$ в (3.7б) получаваме

$$\|div\underline{u_h}\|_{0,\Omega} \leq C\|p_h\|_{0,\Omega}\,. \tag{3.51}$$

От Лема 3.4.1 или Лема 3.4.2 следва, че

$$\begin{aligned}\|p_h\|_{0,\Omega} \leq\ & C\left[h_f\|\underline{u_h}\|_{0,\Omega_1\setminus I_f} + h_c\|\underline{u_h}\|_{0,\Omega_2} + h_c\|\underline{u_h}\|_{0,I_f}\right.\\ & \left.+h_f\|div\,\underline{u_h}\|_{0,\Omega_1} + h_c\|div\,\underline{u_h}\|_{0,\Omega_2}\right].\end{aligned} \tag{3.52}$$

От (3.51) и (3.52) следва, че за достатъчно малки h_f и h_c е изпълнено

$$\|p_h\|_{0,\Omega} \leq C\left[h_f\|\underline{u_h}\|_{0,\Omega_1\setminus I_f} + h_c\|\underline{u_h}\|_{0,\Omega_2} + h_c\|\underline{u_h}\|_{0,I_f}\right]. \tag{3.53}$$

Ако в (3.7a) изберем тестова функция $\underline{v} = \underline{u_h}$, тогава

$$\|\underline{u_h}\|_{0,\Omega} \leq C\|p_h\|_{0,\Omega}\,. \tag{3.54}$$

Неравенствата (3.53) и (3.54) показват, че за достатъчно малко h_f и h_c, $\underline{u_h} = \underline{0}$ и $p_h = 0$, т.е. системата (3.7) притежава единствено решение.

Това завършва доказателството на теоремата.

Глава 4

Нов Монте Карло подход за обръщане на матрици, възникващи в смесения метод на крайните елементи

4.1 Въведение

В тази глава отделяме особено внимание на една важна подзадача, свързана с практическото прилагане на смесения метод на крайните елементи за решаване на елиптични гранични задачи от втори ред.

Разглеждаме следната елиптична гранична задача на Дирихле от втори ред:

$$
\begin{aligned}
-div(a(x)\underline{\nabla} p) &= f(x), \qquad \text{в } \Omega\,; \\
p &= -g, \qquad \text{върху } \partial\Omega\,,
\end{aligned}
\tag{4.1}
$$

където Ω е подобласт на $I\!R^2$ или $I\!R^3$ с граница $\partial\Omega$.

Забележка 4.1.1 *Аналогичната гранична задача на Нойман също се включва в нашите общи разглеждания.*

Апроксимацията чрез смесения метод на крайните елементи на задачата (4.1) с елементи на Равиар-Тома води (виж например Робъртс, Тома [1]) до следната система от линейни алгебрични уравнения:

$$BY \equiv \begin{pmatrix} M & N \\ N^* & 0 \end{pmatrix} \begin{pmatrix} U \\ P \end{pmatrix} = \begin{pmatrix} G \\ F \end{pmatrix}, \tag{4.2}$$

където съответните матрици и вектори са от следния вид:

$$B \in \mathbb{R}^{(n_1+n_2)\times(n_1+n_2)},\ M \in \mathbb{R}^{n_1\times n_1},\ N \in \mathbb{R}^{n_1\times n_2},\ N^* \in \mathbb{R}^{n_2\times n_1},$$

$$Y \in \mathbb{R}^{n_1+n_2},\ U, G \in \mathbb{R}^{n_1},\ P, F \in \mathbb{R}^{n_2}.$$

Матрицата B е обратима, но не е положително-определена. Затова директното решаване на тази система в общия случай представлява определена трудност (виж например Робъртс, Тома [1]).

За приближеното решаване на системата (4.2) се използват различни итерационни методи (метод на спрегнатия градиент, наказателен метод, ускорен метод на Лагранж и др.)

От друга страна матрицата M е симетрична и положително-определена. Следователно, теоретически нейното обръщане е винаги възможно. Тогава, ако успеем да пресметнем M^{-1} и положим

$$U = M^{-1}G - M^{-1}NP$$

получаваме системата:

$$KP = H, \tag{4.3}$$

където

$$K = N^*M^{-1}N$$

и

$$H = N^* M^{-1} G - F.$$

Така редуцираме линейната алгебрична система (4.2) с размерност $n = (n_1 + n_2)$ до n_2-размерната система (4.3), където

$$n_2 < \frac{1}{3} n.$$

Задачата за обръщане на матрици заслужава отделно разглеждане, тъй като тя се явява основна или помощна подзадача и на други важни математически задачи. Така например в случая, когато се нуждаем от груба оценка на обратната матрица (Колотилина, Йерьомин [1]), която се използва при конструиране на преобуслователи за ускоряване на различни итерационни методи за решаване на линейни алгебрични системи.

В тази глава представяме два ефективни Монте Карло алгоритъма за обръщане на матрици, основаващи се на два нови подхода към разглежданата задача.

Основен параметър за ефективността на всеки алгоритъм е неговата изчислителна сложност или времето за което той се реализира. Някои теоретични оценки за сложността на различни класически и статистически алгоритми са описани от Холтън [1]. Задачата за обръщане на дадена матрица е представена като задача за решаване на система от линейни системи алгебрични уравнения, представена в следния обобщен вид:

$$AV = B, \tag{4.4}$$

където $(m \times m)$ матрицата A и $(m \times m)$ матрицата B са известни, докато търсената в този случай е $(m \times m)$ матрицата V.

Задачата за обръщане на матрицата A се състои в пресмятането на матрицата V в случая, когато дясната страна на системата (4.4) е единичната

матрица, т.е. $B = I$.

Съществуват редица класически числени методи за решаване на $(m \times m \times m)$ системата (4.4). В детерминистичните случаи директните методи, като метод на Гаус и метод на Жордан, изискват следния брой стъпки S_D:

$$S_D(m) = O(m^3),$$

От друга страна итерационните методи, като метод на Якоби, метод на Гаус - Зайдел и методите основаващи се на различни релаксационни техники, използващи k на брой итерации изискват

$$S_I(m,k) = O(m^3 k)$$

стъпки.

Алгоритмите за пресмятане на обратни матрици, основаващи се на директните методи са удобни за паралелизиране, но те изискват твърде голям брой процесори p. Известно е, че процедурата на Цанки за паралелно обръщане на квадратна матрица използва $O(log_2^2 m)$ стъпки (Цанки [1]). Тя обаче изисква $p = m^4$ процесори. Отбелязваме, че този резултат има само теоретично значение, тъй като изискваят брой процесори е нереалистичен при решаване на задачи с големи размерности.

Фактически, следвайки принципа на декомпозицията в алгоритъма, представен от Хаефил и Кунг [1], може да се представи следната по-обща оценка за изчислителната сложност на един директен алгоритъм с размерност m върху p процесора

$$S_D(p,m) = O\left(\frac{m^5}{p} log_2^2 m\right). \tag{4.5}$$

Оценката (4.5) е неподобряема по порядък.

Известно е (Холтън [1]), че Монте Карло алгоритмите за пресмятане на компонентите на U изискват средно

$$S_{MC}(m,k,n) = O(m^2kn)$$

стъпки, включващи n случайни блуждаения със средна дължина k на веригата на Марков. В сравнение с другите итерационни алгоритми тук е налице замяна на параметъра m с n. От това следва, че колкото е по-малко отношението $\frac{n}{m}$, толкова той е по-ефективен от класическите итерационни алгоритми. Тъй като параметърът n контролира вероятната грешка, ефективността на Монте Карло алгоритмите нараства, ако предварително зададената вероятна грешка, с която искаме да обърнем матрицата е по-голяма.

Ако разполагаме със система от p процесора, тогава Монте Карло техниката изисква

$$S_{MC_p}(m,k,n,p) = O\left(\frac{m^2}{p}kn\right) \tag{4.6}$$

стъпки.

Съпоставянето на оценките (4.5) и (4.6) показва, че при многопроцесорна реализация за матрици алгоритмите от типа Монте Карло могат да бъдат предпочетени пред класическите алгоритми.

Алгоритмите, които се разглеждат в тази глава се базират на нов подход (виж Т. Димов [3] и И. Димов, Т. Димов, Гюров [1]) и притежават същата изчислителна сложност, като алгоритмите, описани от Джон Холтън, но те са по-ефективни, защото комбинират различни съответно стоп критерии и релаксационни параметри.

4.2 Постановка на задачата

Разглеждаме задачата за числено пресмятане на обратната матрица A^{-1} на дадена квадратна матрица A.

За нашите цели ще използваме следната система от линейни алгебрични уравнения:

$$A\underline{u} = \underline{b}, \tag{4.7}$$

където

$$A = \{a_{ij}\}_{i,j=1}^{m} \in I\!R^{m\times m}; \quad \underline{b}, \underline{u} \in I\!R^{m}.$$

Задачата за обръщане на матрицата A е еквивалентна на решаването m пъти на задачата (4.7), т.е.

$$AC_j = I_j, \quad j = 1, \ldots, m$$

където

$$I_j \equiv (0, \ldots, 0, \underbrace{1}_{j}, 0, \ldots, 0)^T$$

и

$$C_j \equiv (c_{1j}, c_{2j}, \ldots, c_{mj})^T$$

е j-тия стълб на обратната матрица $C = A^{-1}$.

Добре известно е, че числените методи Монте Карло дават статистическа оценка за решението на дадена задача, използвайки някаква случайна величина, чието математическо очакване съвпада с търсеното решение. Алгоритмите, базиращи се на тези методи притежават някои очевидни предимства. Едно от най-съществените е, че те са „присъщо паралелни“. Те притежават висока ефективност, когато се използват паралелни компютри. Този факт е показан в работите на Метрополис, Улам [1], Димов, Тонев [1], [2], Мегсон,

Александров, Димов [1] и др. Монте Карло алгоритмите са също достатъчно ефективни, когато разглежданата задача е с свръх голяма размерност или е с „неопределена“ (или „обща“) структура, която не се вмества в алгоритмите, основаващите се на традиционните числени методи.

Едно от най-важните предимства на тези алгоритми е, че те дават възможност да се пресметне определен линеен функционал от решението, и в частност – само една негова компонента, без да се пресмятат останалите му компоненти.

Съществуват два класа алгоритми, основаващи се на числените методи Монте Карло – директни и итерационни.

Директните методи притежават само вероятностна грешка.

Итерационните Монте Карло алгоритми притежават два типа грешки – систематични и вероятностни. В конкретния случай те се наричат съответно грешка от прекъсване и вероятна грешка. Систематичната грешка зависи от броя на итерациите в използвания итерационен метод, докато вероятностната грешка зависи от стохастичната природа на методите Монте Карло.

Класическите примери, които дават Уестлейк [1] и Къртис [1] показват, че алгоритмите, основаващи на методите Монте Карло са предпочитани в случаите, когато се решават задачи за разредени матрици с голяма размерност.

Разглеждаме линейната алгебрична система (4.7). Дефинираме итерация от *ред* i като функция от следния вид:

$$\underline{u}^{(k+1)} = F_k(A, \underline{b}, \underline{u}^{(k)}, \underline{u}^{(k-1)}, \ldots, \underline{u}^{(k-i+1)}),$$

където $\underline{u}^{(k)}$ е m-тата компонента на вектора, получен от k-тата итерация. Очевидно:

$$\underline{u}^{(k)} \to \underline{u} = A^{-1}\underline{b} \text{ когато } k \to \infty.$$

Методът се нарича *стационарен* ако $F_k = F$ за всяко k, т.е. F_k е независима от k.

Итерационният процес се нарича *линеен* ако F_k е линейна функция на $\underline{u}^{(k)}, \ldots, \underline{u}^{(k-i+1)}$. Ние ще разглеждаме *стационарни линейни итерационни Монте Карло* алгоритми.

Разглеждаме *метод на Якоби* с релаксационен параметър γ. Ще използваме матрицата $L = \{l_{ij}\}_{ij=1}^{m}$, която се получава от следното очевидно отношение

$$L = I - DA,$$

където D е диагонална матрица $D = diag(d_1, \ldots, d_m)$ и

$$d_i = \frac{\gamma}{a_{ii}}, \ \gamma \in (0,1] \ i = 1, \ldots, m.$$

Системата (4.7) може да се представи в следния вид:

$$\underline{u} = L\underline{u} + \underline{f}, \tag{4.8}$$

където

$$\underline{f} = D\,\underline{b}.$$

Да допуснем, че матрицата A е с преобладаващ главен диагонал. Фактически, това изискване е твърде силно и както ще покажем по-късно, представените алгоритми работят за по-широк клас матрици. Очевидно, ако матрицата A е с преобладаващ главен диагонал, тогава елементите на матрицата L удовлетворяват следното условие:

$$\sum_{j=1}^{m} |l_{ij}| < 1 \qquad i = 1, \ldots, m. \tag{4.9}$$

Да разгледаме *стационарен линеен итерационен процес от първи ред* за системата (4.8)

$$\underline{u}^{(k)} = L\underline{u}^{(k-1)} + \underline{f}, \ k = 1, 2, \ldots. \tag{4.10}$$

Фактически, (4.10) дефинира следния *ред на Нойман* :

$$\underline{u}^{(k)} = \underline{f} + L\underline{f} + \cdots + L^{k-1}\underline{f} + L^k\underline{u}^{(0)}, \qquad k > 0\,.$$

От (4.8) и (4.10) може да се получи грешката от прекъсване. Ако $\underline{u}^{(0)} = \underline{f}$, тогава

$$\underline{u}^{(k)} - \underline{u} = L^k(\underline{f} - \underline{u}).$$

Добре известно е, че съотношението (4.9) е достатъчно условие за сходимост на реда на Нойман, т.е.

$$\underline{u} = \lim_{k\to\infty} \underline{u}^{(k)}.$$

Очевидно, всеки итерационен алгоритъм използва краен брой итерации k. Практически параметърът k не е зададен предварително, а се определя от разликата между две итерации. В нашите алгоритми определяме броя на Монте Карло итерациите $\underline{u}^{(q)}$, $1 \le q \le k$, използвайки практическото правило разликата между две стохастични итерации да бъде по-малка от предварително зададен достатъчно малък параметър ε.

4.3 Дискретни процеси на Марков

Разглеждаме задачата за приближено пресмятане на линейния функционал $V(\underline{u})$ от решението $\underline{u}$ на системата (4.8):

$$V(\underline{u}) \equiv (\underline{v}, \underline{u}) = \sum_{i=1}^{m} v_i u_i\,, \tag{4.11}$$

където $\underline{v} \in I\!R^m$ е даден вектор.

Конструираме случайна величина $X[\underline{v}]$, чието математическо очакване съвпада с търсеното решение (4.11), т.е.

$$EX[\underline{v}] = V(\underline{u}),$$

използвайки дискретни процеси на Марков с крайно множество от състояния.

Тогава изчислителната задача се свежда до пресмятане на определен брой независими реализации на $X[\underline{v}]$, чието използване дава подходяща статистическа оценка на $V(\underline{u})$. Отбелязваме, че природата на всяка реализация на $X[\underline{v}]$ е процес на Марков. Ще разглеждаме само *Дискретни процеси на Марков с крайно множество от състояния*, така наречените *Крайни дискретни вериги на Марков.*

Дефиниция 4.3.1 *Крайна дискретна верига на Марков* S *се дефинира като крайно множество от състояния* $s_1, s_2, ..., s_m$. *Във всяко дискретно състояние от време* $t = 0, 1, \ldots, k, \ldots$ *веригата* S *се намира в едно от следните състояния* $s_{t_0}, s_{t_1}, \ldots, s_{t_k}, \ldots$, *които удовлетворяват условието на Марков:*

$$P(s_{t_q} = \alpha_q | s_{t_{q-1}}, s_{t_{q-2}}, \ldots, s_{t_0}) = P(s_{t_q} = \alpha_q | s_{t_{q-1}}), \quad q = 0, 1, \ldots, \quad \alpha_q \in \{s_1, \ldots, s_m\}.$$

Всяко от състоянията s_i са асоциирани с множество от условни вероятности p_{ij}, такива, че p_{ij} е вероятността системата, която във времето t се е намирала в състояние s_i, да премине в състояние s_j в съответното $(t+1)$ време, т.е.

$$P(s_{t_{k+1}} = s_j | s_{t_k} = s_i) = p_{ij}.$$

Така, p_{ij} е вероятността за прехода s_i към s_j. Множеството от условни вероятности p_{ij} дефинира матрица на преходната плътност $P = \{p_{ij}\}_{i,j=1}^{m}$, която се определя от вероятностите на дадената верига S.

Дефиниция 4.3.2 *Дадено състояние се нарича поглъщащо, ако съответната верига се прекъсва в него с вероятност единица.*

В общия случай, итерационните Монте Карло алгоритми могат да се дефинират като *вериги на Марков с поглъщане*:

$$S_k = s_{t_0} \to s_{t_1} \to s_{t_2} \to \cdots \to s_{t_k}\,, \tag{4.12}$$

където s_{t_q}, $(q = 1, \ldots, k)$ е едно от състоянията с поглъщане (абсорбиращите състояния). Така определяме стойностите на някаква функция $F(S) = X\,[\underline{v}]$, която зависи от редицата (4.12). Функцията $F(S)$ е случайна променлива. След като $F(S)$ е пресметната, тогава системата се стартира от нейното начално състояние s_{t_0} и преходите започват отново. Осъществяват се n на брой различни независими реализации на веригата на Марков, започвайки от състоянието s_{t_0} до абсорбиращите състояния.

Средната стойност

$$\frac{1}{n}\sum_{T} F(S) \tag{4.13}$$

е пресметната на базата на реализираните редици от преходи (4.12). Изразът (4.13) апроксимира $E\{F(S)\}$, което е търсеният линеен функционал от решението.

Ние ще се интересуваме и от *изчислителната сложност* на съответния алгоритъм.

Дефиниция 4.3.3 *Изчислителната сложност на алгоритъма се дефинира чрез*

$$n\,E(k)\,t_0\,,$$

където $E(k)$ *е математическото очакване на броя на блуждаенията в редицата (4.12) и* t_0 *означава времето, необходимо за реализирането на един преход.*

Фактически, дефиницията на изчислителна сложност се използва за получаване на теоретични оценки, защото тя дава оценка само за зависимостта на математическото очакване от броя на преходите k. Практически, за всяка реализация на даден Монте Карло алгоритъм трябва да се определи предварително броя на преходите във всички реализации на съответната верига на Марков, които означаваме с R:

$$R = \sum_{i=1}^{n} k_i \,, \tag{4.14}$$

където k_i са броят на преходите в i-тата реализация на веригата на Марков.

4.4 Итерационни Монте Карло методи

Да разгледаме следната верига на Марков:

$$S = s_0 \to s_1 \to s_2 \to \cdots \to s_k \to \ldots \tag{4.15}$$

с m състояния $\{1, 2, \ldots, m\}$. Нека

$$P(s_0 = i) = p_i, \quad P(s_q = j | s_{q-1} = i) = p_{ij},$$

където p_i (началната вероятност) е вероятността веригата да стартира от състояние i и p_{ij} е преходната вероятност за случайния процес да заеме състояние j, след като се е намирал в състояние i. Вероятностите p_{ij} формират матрицата на преход $P = \{p_{ij}\}_{ij=1}^{m}$. Очевидно, p_i и p_{ij} трябва да бъдат неотрицателни и

$$\sum_{i=1}^{m} p_i = 1 \,, \quad \sum_{j=1}^{m} p_{ij} = 1 \,, \qquad \text{за всяко } i = 1, 2, \ldots, m.$$

Дефинираме следната теглова функция W_q, за веригата на Марков (4.15) с m състояния, използвайки рекурсивната формула:

$$W_0 = 1; \qquad W_q = W_{q-1} \frac{l_{s_{q-1} s_q}}{p_{s_{q-1} s_q}}, \qquad q = 1, 2, \ldots,$$

където редицата от състояния $s_0, s_1, s_2, \ldots$ е дадена случайна траектория с начална вероятност p_{s_0} и преходни вероятности $p_{s_{q-1}s_q}$.

Фактически

$$W_q = \frac{l_{s_0 s_1} l_{s_1 s_2} \ldots l_{s_{q-1} s_q}}{p_{s_0 s_1} p_{s_1 s_2} \cdots p_{s_{q-1} s_q}}. \tag{4.16}$$

Дефинираме следната случайна величина:

$$X[\underline{v}] = \frac{v_{s_0}}{p_{s_0}} \sum_{q=0}^{\infty} W_q f_{s_q}. \tag{4.17}$$

В сила е следното основно твърдение, което е обобщение на съответното, представено от Спанер и Джилбърт [1]:

Теорема 4.4.1 *Нека $\underline{u}$ е решение на системата (4.8) и елементите l_{ij} на матрицата L удовлетворяват условието (4.9). Тогава математическото очакване на случайната величина $X[\underline{v}]$ съвпада със стойността на линейния функционал $V(\underline{u})$, дефиниран чрез (4.11), т.е.*

$$EX[\underline{v}] = (\underline{v}, \underline{u}).$$

Доказателство: За дадена редица от състояния $\alpha_0 \to \alpha_1 \to \cdots \to \alpha_q \to \cdots$, разглеждаме верига на Марков, дефинирана чрез (4.15).

Тогава имаме

$$P\{s_0 = \alpha_0, \cdots, s_q = \alpha_q\} = p_{\alpha_0} p_{\alpha_0 \alpha_1} \cdots p_{\alpha_{q-1} \alpha_q},$$

където

$$\alpha_i \in \{1, 2, \ldots, m\} \quad i = 1, 2, \ldots, q.$$

Използвайки (4.16), получаваме

$$E\{\frac{v_{s_0}}{p_{s_0}} W_q f_{s_q}\} = \sum_{\alpha_0, \cdots, \alpha_q = 1}^{m} \frac{v_{\alpha_0}}{p_{\alpha_0}} \frac{l_{\alpha_0 \alpha_1} l_{\alpha_1 \alpha_2} \ldots l_{\alpha_{q-1} \alpha_q}}{p_{\alpha_0 \alpha_1} p_{\alpha_1 \alpha_2} \cdots p_{\alpha_{q-1} \alpha_q}} f_{\alpha_q} p_{\alpha_0} p_{\alpha_0 \alpha_1} p_{\alpha_1 \alpha_2} \cdots p_{\alpha_{q-1} \alpha_q}$$

$$= \sum_{\alpha_0=1}^{m} v_{\alpha_0} \sum_{\alpha_1=1}^{m} \cdots \sum_{\alpha_{q-1}=1}^{m} l_{\alpha_0\alpha_1} l_{\alpha_1\alpha_2} \ldots l_{\alpha_{q-2}\alpha_{q-1}} \sum_{\alpha_q=1}^{m} l_{\alpha_{q-1}\alpha_q} f_{\alpha_q}$$

$$= \sum_{\alpha_0=1}^{m} v_{\alpha_0} (L^q f)_{\alpha_0} = (\underline{v}, L^q \underline{f}).$$

Следователно

$$EX\,[\underline{v}] = \sum_{q=0}^{\infty} E\{\frac{v_{s_0}}{p_{s_0}} W_q f_{s_q}\}$$

$$= \sum_{q=0}^{\infty} (\underline{v}, L^q \underline{f}) = (\underline{v}, \underline{u}).$$

Това завършва доказателството на теоремата.

Очевидно, ако $\underline{u}^{(k)}$ е k-тото итерационно решение (4.10) с начално условие $\underline{u}^{(0)} = \underline{f}$, тогава математическото очакване на случайната величина

$$X_k\,[\underline{v}] = \frac{v_{s_0}}{p_{s_0}} \sum_{q=0}^{k} W_q f_{s_q} \tag{4.18}$$

е равна на линейния функционал $V(\underline{u}^{(k)}) = (\underline{v}, \underline{u}^{(k)})$, т.е.

$$EX_k\,[\underline{v}] = (\underline{v}, \underline{u}^{(k)}).$$

Параметърът k се избира от следното изискване за прекъсване на веригата на Марков

$$|W_q| < \varepsilon. \tag{4.19}$$

Като апроксимация на линейния функционал (4.11) получаваме следната средна стойност на случайната величина $X\,[\underline{v}]$:

$$\overline{X}_n\,[\underline{v}] = \frac{1}{n} \sum_{i=1}^{n} \{X[\underline{v}]\}_i,$$

където $\{X\,[\underline{v}]\}_i$ е i-тата независима реализация на случайната величина (4.17). Вероятната грешка се дефинира като стойност r_n (виж например Собол [1] или Ермаков, Михайлов [1]), за която следното условие

$$P\{|\overline{X}_n - EX\,[\underline{v}]\,| < r_n\} \approx \frac{1}{2} \approx P\{|\overline{X}_n - EX[\underline{v}]| > r_n\}\,.$$

е удовлетворено.

Лесно се установява (следвайки Собол [1]), че в разглеждания случай

$$r_n \approx 0.6745\sqrt{VarX\,[\underline{v}]\,/n}\,,$$

където

$$VarX = E(X^2) - (EX)^2\,.$$

Подходящ избор на вектора на началните вероятности $p = \{p_i\}_{i=1}^m$ е

$$p_i = \frac{|v_i|}{\sum_{j=1}^m |v_j|},$$

и съответно на матрицата на преходните вероятности $P = \{p_{ij}\}_{i,j=1}^m$

$$p_{ij} = \frac{|l_{ij}|}{\sum_{j=1}^m |l_{ij}|}\ i = 1, 2, \ldots, m\,.$$

Така дефинираният избор води до така наречения в работата на Мегсон, Александров, Димов [1] *почти оптимален алгоритъм Монте Карло.*

Забележка 4.4.1 *Очевидно, ако използваме специален избор на линейния функционал* $V(\underline{u})$, *който да съответства на вектора*

$$\underline{v} = I_{i_0} \equiv (0, \ldots, 0, \underbrace{1}_{i_0}, 0, \ldots, 0)\,,$$

(където единицата е на i_0*-то място), тогава получаваме* i_0*-та компонента на решението.*

Забележка 4.4.2 *Грешката от прекъсване се получава от замяната на случайната величина (4.17) със съответстващата и, дефинирана в (4.18). Вероятната грешка възниква от факта, че за апроксимиране на линейния функционал (4.11), използваме средната стойност на случайната величина (4.18).*

4.5 Итерационни Монте Карло алгоритми

В Монте Карло алгоритмите, които ще представим ще се нуждаем от следните норми (виж например Акселсон [1]) на вектори $\underline{u} \in \mathbb{R}^m$ и матрици $A \in \mathbb{R}^{m \times m}$:

$$\|\underline{u}\|_2 = \left(\sum_{i=1}^{m} |u_i|^2 \right)^{1/2} \quad \text{(Евклидова норма)} \tag{4.20}$$

и

$$\| A \|_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{m} a_{ij}^2 \right)^{1/2} \quad \text{(норма на Фробениус)}. \tag{4.21}$$

Първият алгоритъм е помощен и пресмята произволна компонента u_0 на решението $\underline{u}$ на линейната алгебрична система (4.7). Алгоритъмът 4.5.1 се разглежда отделно, тъй като някои от неговите стъпки се използват в описанието на следващите два алгоритъма.

Алгоритъм 4.5.1

1. **Input** *начални данни: матрица* A, *вектор* $\underline{b}$, *константа* ε, *релаксационен параметър* $\gamma \in (0, 1]$ *и цяло число* n.

2. *Предварителни пресмятания:*

   2.1. **Compute** *матрицата* L:

$$\{l_{ij}\}_{i,j=1}^{m} = \begin{cases} 1 - \gamma & \text{когато } i = j; \\ -\gamma \frac{a_{ij}}{a_{ii}} & \text{когато } i \neq j. \end{cases}$$

   2.2. **Compute** *вектора* $\underline{f}$:

$$f_i = \gamma \frac{b_i}{a_{ii}}, \quad i = 1, 2, \ldots, m;$$

2.3. Compute *вектора lsum :*

$$lsum(i) = \sum_{j=1}^{m} |l_{ij}| \quad за \quad i = 1, 2, \ldots, m ;$$

3. Основни пресмятания

3.1. Реализиране на една траектория:

3.1.1. Set *начални стойности* $X := 0$, $W := 1$;

3.1.2. Calculate $X := X + W f_{i_0}$;

3.1.3. Generate *равномерно разпределено случайно число* $\xi \in (0, 1)$;

3.1.4. Set $j := 1$;

3.1.5. If $\xi < \sum_{k=1}^{j} p_{i_0 k}$ then

3.1.5.1. Calculate $W := W \times sign(l_{i_0 j}) \times lsum(i_0)$;

3.1.5.2. Calculate $X := X + W f_j$ *(един скок в траекторията);*

3.1.5.3. If $|W| < \varepsilon$ then go to *стъпка 3.1.6 ;*

3.1.5.4. Update *индексът* $i_0 := j$ *и* go to *стъпка 3.1.3 ;*

else

3.1.5.5. Update $j := j + 1$ *и* go to *стъпка 3.1.5.*

3.1.6. End *на текторията.*

3.2. Calculate *средната стойност, базирана на* n *независими траектории:*

3.2.1. Do n *пъти стъпка 3.1;*

3.2.2. Calculate $\overline{X}_n$ *и* $u_{i_0} := \overline{X}_n$.

4. End *на Алгоритъма 4.5.1.*

В Алгоритъм 4.5.1, за намиране на i_0-вата компонента на решението на задачата (4.7) използваме следния линеен функционал

$$V(\underline{u}) = (\underline{v}, \underline{u}),$$

където $\underline{v} = I_{i_0} = (0, 0, \ldots, \underbrace{1}_{i_0}, 0, \ldots, 0)$.

Алгоритъмът 4.5.2 пресмята приближението $\hat{C}$ на обратната матрица $C = A^{-1}$. Той се базира на специален избор на релаксационния параметър γ, който се контролира чрез апостериорен критерий за всеки стълб на следната резидуална матрица

$$E^c = A\hat{C} - I. \tag{4.22}$$

Този избор на получаване на резудуална матрица (умножавайки матрицата $\hat{C}$ от ляво с матрицата A) дава в най-пълна степен зависимостта на точността, с която пресмятаме стълбовете на $\hat{C}$ от стълбовете на E^c.

Алгоритъмът дава възможност различните вектор–стълбове на матрицата $\hat{C} = (\hat{C}_1, \ldots, \hat{C}_m)$ да се пресмятат, използвайки различни стойности на релаксационния параметър $\gamma = \gamma_p$, $p = 1, 2, \ldots, l$. Стойностите на γ_p се избират така, че да минимизират съответните Евклидови норми на следните вектор–стълбове:

$$E_j^c = A\hat{C}_j - I_j, \ j = 1, 2, \ldots, m.$$

Фактически, минимизирането на Евклидовата норма (4.20) на вектор–стълба E_j^c върху някакво предварително зададено множество

$$\gamma = \{\gamma_1, \gamma_2, \ldots, \gamma_l\} \tag{4.23}$$

подобрява (намалява) нормата на Фробениус (4.21) на резидуалната матрица

$$E^c = (E_1^c, \ldots, E_m^c)$$

и води до по-добра апроксимация на обратната матрица $\hat{C}$ стълб по стълб.

Отбелязваме, че пресмятането на различните стълбове може да се реализира паралелно и независимо един от друг. Като второ ниво на паралелизъм, в зависимост от броя на процесорите във всяка конкретна компютърна конфигурация, могат да се задават съответни стойности на параметъра l, който определя броя на елементите от множеството γ, дефинирано в (4.23).

Алгоритъм 4.5.2

1. **Input** *начални данни: елементите на матрица* T, *константа* ε, *достатъчно голяма стойност на параметъра* $E.norm$, *цели числа* n *и* l *и множество от релаксационни параметри* $\gamma_1, \gamma_2, \ldots, \gamma_l$, *където* $\gamma_i \in (0,1]$, $i = 1, \ldots, l$.

2. **Compute** *асамблираната матрица* A *и съответния параметър за нейната размерност* m.

3. **For** $j_0 = 1$ **to** m **do**

Calculate *елементите на* j_0*-вия вектор-стълб на матрицата* $\hat{C}$:

3.1. **For** $k = 1$ **to** l **do**

3.1.1. **For** $i_0 = 1$ **to** m **do**

Apply *стъпка 2. и стъпка 3. на Алгоритъма 4.5.1 за* A, ε, n, $\gamma = \gamma_k$, *и дясна част* $\underline{b} = I_{j_0} = (0, \ldots, 0, \underbrace{1}_{j_0}, 0, \ldots, 0)^T$ *за намиране на вектора-стълб* $\hat{C}_{j_0}^{(k)} = (\hat{c}_{1j_0}^{(k)}, \ldots, \hat{c}_{mj_0}^{(k)})^T$;

3.1.2. **Compute** *Евклидовата норма на резидуалния вектор-стълб* $E_{j_0}^{c(k)}$:

$$\left\| E_{j_0}^{(k)} \right\|_2 = \left(\sum_{j=1}^{m} \left(\sum_{i=1}^{m} a_{ji} \hat{c}_{ij_0}^{(k)} - \delta_{jj_0} \right)^2 \right)^{1/2};$$

3.1.3. **If** $\left\|E_{j_0}^{c(k)}\right\|_2 < E.norm$ **then** $\hat{C}_{j_0} := \hat{C}_{j_0}^{(k)}$ *и* $E.norm := \left\|E_{j_0}^{c(k)}\right\|_2$.

4. **End** *на Алгоритъма 4.5.2.*

Основната идея на представения по-горе алгоритъм използва детерминистичен подход, който не зависи от неговата статистическа природа и може да бъде прилагана и при другите традиционни алгоритми. Това негово качество му определя самостоятелна роля в областта на алгоритмите за обръщане на матрици.

Алгоритъмът 4.5.3, обаче съществено изисква Монте Карло подход и няма детерминистичен аналог. Разликата между двата алгоритъма е, че Алгоритъмът 4.5.3 не може да бъде приложен за традиционните (детерминистични) итерационни методи. Тези методи дават възможност да се пресмятат паралелно различните стълбове на обратната матрица, но те не позволяват пресмятане на всеки един от нейните елементи независимо от другите елементи.

Едно от съществените предимства на Монте Карло алгоритмите се състои именно във възможносста за пресмятане на елементите на обратната матрица независимо един от друг. Това свойство дава възможност да се прилагат различни итерационни подходи за намиране на матрицата $\hat{C}$, използвайки априорна информация за редовете на дадената матрица A (например, някакви отношения между техните елементи). Отбелязваме че предварителната информация за свойствата на редовете на матриците, възникващи при решаване на конкретни задачи е винаги достъпна, а понякога и единствено възможна.

За всеки ред A_i на матрицата A (когато $|a_{ii}| \neq 0$), въвеждаме съответните параметри K_i, както следва:

$$K_i = \left(\sum_{\substack{j=1 \\ j\neq i}}^{m} |a_{ij}|\right) \Big/ |a_{ii}|, \quad i = 1, 2, \ldots m. \tag{4.24}$$

Дефиниция 4.5.1 *Казваме, че даден ред* A_i *на матрицата* A *е съответно добре обусловен, умерено обусловен и лошо обусловен, когато съответстващият му параметър (4.24) изпълнява следните изисквания:*

$$K_i < 1, \quad K_i = 1, \quad и \quad K_i > 1.$$

Нашите числени експерименти показват, че итерационните Монте Карло алгоритми притежават толкова по-добра сходимост при пресмятане на реда $\hat{C}_i$, колкото е по-малък параметърът K_i, съответстващ на реда A_i, защото първоначалното намаляване на W е гарантирано.

Този факт дава априорна информация за използване на различни стоп критерии ε_i $(i = 1, \ldots, m)$ за пресмятане на съответстващия му ред $\hat{C}_i$ на матрицата $\hat{C}$. Практически това означава, че ние използваме по-голяма разлика между две Монте Карло итерации за прекъсване на итерационния процес, когато той е по-бързо сходящ за сметка на на съответната по-малка разлика в случая на по-бавна сходимост.

Броят на скоковете във верига на Марков (т.е. броят на итерациите) могат също да бъдат контролирани така, че да се получи добър баланс между статистическата и систематичната грешки. Задачата за балансиране на тези грешки е много съществена при използването на Монте Карло алгоритмите. Очевидно, за да получим добри резултати, трябва статистическата (в случая вероятната) грешка r_n да бъде приблизително равна на съответната систематична r_k, т.е.

$$r_n = O(r_k).$$

Задачата за балансиране на грешките е тясно свързана със задачата за намиране на оптимално (от изчислителна гледна точка) отношение между броя на реализациите n на случайната величина и на математическото очакване $E(k)$ от броя на на стъпките във всяка случайна траектория. Това

балансиране ни дава възможност да подобрим точността на алгоритъма, без да увеличаваме съответната изчислителната сложност, която се контролира от параметъра R, дефиниран в (4.14), чрез избора на различни дължини на реализации на съответната верига на Марков. Практически, ние избираме абсорбиращите състояния на случайната траектория, използвайки неравенство (4.19).

За да запазим баланса на грешките (систематична и стохастична), ние също използваме различен брой реализации n_i $(i = 1, \ldots, m)$ на съответните случайни величини, които са пропорционални на броя на итерациите. Тази процедура наричаме *фин стоп критерий*. Използвайки я, определяме понятието *Рафиниран Монте Карло алгоритъм*.

Така ние намаляваме нормата на Фробениус на следната резидуална матрица:

$$E^r = \hat{C}A - I\,, \tag{4.25}$$

и следователно подобряваме точността на пресмятане на матрицата $\hat{C}$ ред по ред, без да увеличаваме изчислителната сложност на алгоритъма в сравнение с алгоритмите, използващи стандартния *груб стоп критерий*.

Алгоритъм 4.5.3

1. **Input** *начални данни: елементна матрица* T, *константа* ε, *константата* $\gamma = 1$, *цяло число* n *и достатъчно голяма стойност* $E.norm$.

2. **Compute** *асамблираната матрица* A *и съответния параметър за нейната размерност* m.

3. **Compute** *коефициентите* δ_i *и вектора* $\mathcal{E} = (\varepsilon_1, \ldots, \varepsilon_m)$:

$$\varepsilon_i = \varepsilon\delta_i, \quad i = 1, \ldots, m;$$

4. **For** $i_0 = 1$ **to** m **do**

Calculate *елементите на* i_0*-я вектор-ред на матрицата* $\hat{C}$ *:*

4.1. **For** $j_0 = 1$ **to** m **do**

Apply *стъпка 2. и стъпка 3. на Алгоритъма 4.5.1 за* A, $\varepsilon = \varepsilon_{i_0}$, n, $\gamma = 1$, *и дясна част - векторът* $b = I_{j_0} = (0, \ldots, 0, \underbrace{1}_{j_0}, 0, \ldots, 0)^T$ *за намиране на вектора-ред* $\hat{C}_{i_0}^{(k)} = (\hat{c}_{i_0 1}^{(k)}, \ldots, \hat{c}_{i_0 m}^{(k)})$ *;*

4.2. **Calculate** *Евклидовата норма на резидуалния вектор-ред* $E_{i_0}^{r(k)}$.

4.3. **If** $\left\| E_{i_0}^{r(k)} \right\|_2 < E.norm$, **then** $\hat{C}_{i_0} := \hat{C}_{i_0}^{(k)}$ *и* $E.norm := \left\| E_{i_0}^{r(k)} \right\|_2$.

5. **End** *на Алгоритъма 4.5.3.*

4.6 Числени резултати и коментари

В качеството на конкретен пример разглеждаме следната система линейни алгебрични уравнения, възникваща след апроксимация на двумерна хомогенна задача на Дирихле (4.1) чрез смесения метод на крайните елементи с правоъгълници, аналогично на Глава 1:

$$BY = \begin{pmatrix} M_1 & 0 & N_1 \\ 0 & M_2 & N_2 \\ N_1^T & N_2^T & 0 \end{pmatrix} \begin{pmatrix} U_1 \\ U_2 \\ P \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -F \end{pmatrix}, \tag{4.26}$$

където $M_i \equiv diag(A_{i1}, A_{i2}, \ldots A_{is})$ са $t \times t$ блочно диагонални матрици, N_i са $t \times t_1$ матрици $(t_1 < t)$, $U_i \in I\!R^t$ и $P, F \in I\!R^{t_1}$, $i = 1, 2$.

Ако се пресметнат $M_i^{-1} (i = 1, 2)$, тогава системата (4.26) се достига до системата

$$KP = F,$$

където

$$K = N_1^T M_1^{-1} N_1 + N_2^T M_2^{-1} N_2.$$

По този начин получаваме t_1-размерна линейна алгебрична система.

За числените ни експерименти използваме пространства на Равиар–Тома с индекс $r = 1$ и равномерно разделяне на областта $\Omega = [0,1]^2$ с $N_e = n_e^2$ елементи. В този случай $t_1 = 4n_e^2$, $dim A_{ij} = 4n_e + 2$ и следователно $t = n_e(4n_e + 2)$. Отбелязваме, че в конкретната реализация матриците, участващи в системата (4.26) се генерират програмно от съответните елементни метрици.

Очевидно, задачата за намиране на M^{-1} се редуцира до задачата за пресмятане на матриците A_{ij}^{-1}. В общия случай, матриците A_{ij}^{-1} не са със строго преобладаващ главен диагонал, но техните собствени стойности се намират в единичния кръг.

От съображения за простота, в по-нататъшните разглеждания, ще използваме означенията $A = A_{ij}$ и $m = 4n_e + 2$.

Представени са числени експерименти в случая $n_e = 4$ т.е. $t = 18$. В този конкретен пример стойностите на параметъра K_i (4.24) са следните:

$$K_i = 1\,, \quad i = 1, 2, 5, 6, 9, 10, 13, 14, 17, 18$$

и

$$K_i = 3/7\,, \quad i = 3, 4, 7, 8, 11, 12, 15, 16\,.$$

Пресметнати са нормите на Фробениус на съответните резидуални матрици E^c и E^r. Някои от числените резултати са показани в Таблици 4.1 – 4.3 и на Фигура 4.1.

Таблица 4.1 представя относителната норма на Фробениус на матриците E^c по отношение на $\hat{C}$ (Алгоритъм 4.5.2), получени при различни стойности на параметъра γ, в случая $\varepsilon = 0.01$. Резултатите показват подобряването, което се получава при комбиниране на стълбовете на обратната

γ	0.7	0.8	0.9	1.0	Комбинирани γ
$n = 200$	0.00317	0.00325	0.00335	0.00245	0.00217
$n = 250$	0.00300	0.00274	0.00286	0.00262	0.00210

Таблица 4.1: Относителна норма на Фробениус за различни γ.

γ	0.7	0.8	0.9	1.0
$n = 200$	4	3	3	8
$n = 250$	2	6	4	6

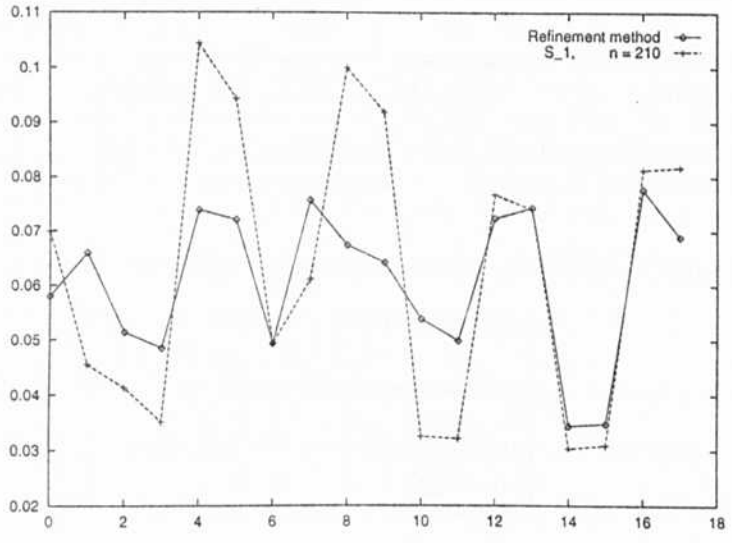
Таблица 4.2: Брой на стълбовете на $\hat{C}$, получени при съответното γ.

матрица, получени при използване на най-подходящия (от дадените) релаксационен параметър. Броят на тези стълбове при съответното ε е представен в Таблица 4.2.

	Изчислителна сложност	Норма на Фробениус
Рафиниран подход	1.00000	1.00000
S_1 $n = 210$	1.04390	1.09164
S_2 $n = 220$	1.09599	1.07406
S_3 $n = 230$	1.14489	1.06554
S_4 $n = 240$	1.19322	1.04587

Таблица 4.3: Отношения между рафинирания и S_i алгоритмите.

Сравнени са изчислителните сложности и нормите на Фробениус на резидулните матрици E^r, получени съответно чрез рафинирания и четири стандартни алгоритъма. В този случай рафинираният подход използва парамет-



Фигура 4.1: Евклидова норма на E^r ред по ред.

рите

$$n = 100\,, \qquad \varepsilon = 0.05$$

за добре обусловените редове и

$$n = 300\,, \qquad \varepsilon = 0.01$$

за умерено обусловените редове на матрицата A. Представени са съответните n на брой реализации и стоп критериите $\varepsilon = 0.01$ за алгоритмите S_i, $i = 1, 2, 3, 4$. Резлтатите илюстрират, че рафинираният алгоритъм (Алгоритъм 4.5.3) в сравнение със стандартните алгоритми, от една страна подобрява (намалява) нормата на Фробениус на резидуалната матрица, а от друга – притежава по-малка изчислителна сложност.

Във Фигура 4.1 се сравняват Евлидовите норми по редове на резидуалната матрица E^r (18 на брой) на рафинирания алгоритъм с един от стандартните (в случая с S_1 алгоритъма). На графиката ясно се вижда, че рафинираният

алгоритъм притежава по-регулярни свойства, по отношение на Евлидовите норми по редове. Очевидно, този факт в най-пълна степен показва качеството на пресмятане на съответната обратна матрица.

Алгоритъм 4.5.3 дава по-добри резултати, защото той използва прецизно съществените свойства на първоначално зададената матрица A.

Заключение

Съществени резултати в дисертацията са:

1. Конструирани са подходящи квадратурни формули за пресмятане на част от интегралите в смесения метод на крайните елементи върху правоъгълна мрежа, които водят до диагонална матрица на масата (концентрация на масата). Доказано е, че това числено интегриране, което намалява изчислителната сложност на задачата повече от три пъти, не води до повишаване на порядъка на грешката и той остава оптимален.

2. Доказана е оптимална по порядък оценка на грешката в смесения метод на крайните елементи с регулярно локално сгъстяване за решаване на двумерни гранични задачи за коерцитивен елиптичен оператор, като е използван интерполационен подход. Получените теоретични оценки дават пълна характеризация на добре известни ефекти, получени преди това при числената реализация на съответния подход.

3. Предложен е модифициран проектор на Равиар - Тома в смесения метод на крайните елементи с регулярно локално сгъстяване за решаване на общи елиптични гранични задачи в $I\!R^2$. Доказаната сходимост е оптимална по порядък и съгласува в най-пълна степен членовете в дясната част на оценките на грешката.

4. Представен е нов подход, използващ метода Монте Карло, който е подходящ за решаване на дискретната задача, възникваща в смесения метод на крайните елементи. На базата на него са предложени алгоритми за обръщане на матрици, които редуцират съществено съответната линейна алгебрична система.

5. Предложени са два нови Монте Карло алгоритъма за обръщане на матрици. Първият се основава на използването на различни релаксационни параметри при формиране на съответните стълбове на обратната матрица и се контролира автоматично посредством подходящо избран апостериорен критерий. Вторият използва лесно достъпна априорна информация за едно специфично (в случая съществено) свойство на началната матрица, влияещо на сходимостта на алгоритъма и постига по-прецизен от стандартния баланс на грешките на метода. Тези алгоритми са реализирани числено и изследвани върху моделна задача, възникваща в смесения метод на крайните елементи.

Литература

Акселсон (O. Axelsson)

[1] *Iterative Solution Methods*, **Cambridge University Press**, Cambridge, 1994.

Андреев, Димов (A.B. Andreev, T.T. Dimov)

[1] *Mixed finite element method for eigenvalue problems with regular local refinement*, in **Scientific Computation and Mathematical Modeling, Dates Publishing**, Sofia, 1993, pp. 51-53.

[2] *Lumped Mass Approximation for the Mixed Finite Element Method*, in: **Advances in Numerical Methods and Applications** - $O(h^3)$, Proceedings of the Third International Conference on Numerical Methods and Applications, Sofia, 21-26 August 1994, **World Scientific**, New Jersey, London, Hong Kong, pp. 11–18.

Андреев, Касчиева, Ванмаеле (A.B. Andreev, V.A. Kascieva, M. Vanmaele)

[1] *Some Results in Lumped Mass Finite Element Apprximation of Eigenvalue Problems using Numerical Quadrature Formulas*, **J.Comp. Appl.Math. 43**, 1992, pp. 291-311.

Бабушка (I. Babuška)

[1] *Error bounds for the finite element method*, **Numer. Math., 16**, 1971, 322-333.

Бабушка, Осборн (I. Babuška, J. Osborn)

[1] *Handbook of Numerical Analysis*, Vol. 2., 1991 (**North-Holland**).

Банержи, Осборн (U. Banerjee, J.E. Osborn)

[1] *Estimation of the Effect of Numerical Integration in Finite Element Eigenvalue Approximation*, **Math. of Comp. 56**, 1990, pp. 735-762.

Банк, Дюпон (R. Bank, T. Dupont)

[1] *Analysis of a two-level scheme for solving finite element eqations*, Report CNA-159, **Center for Numerical Analysis, The University of Texas at Austin**, 1980.

Брамбъл, Хилберт (J.H. Bramble, S. Hilbert)

[1] *Bounds for yhe class of linear functionals with application to Hermite interpolation*, **Numer. Math. v. 16**, 1971, N4, pp. 362- 369.

Брамбъл, Юинг, Пашек, Шатс (J.H. Bramble, R.E. Ewing, J.E. Pasciak, A.H. Shatz)

[1] *A preconditioning technique for efficient solution of problems with local grid refinement*, **Comp. Math. Appl. Mech. Eng., 67**, 1988, 149-159.

Бреци (F. Brezzi)

[1] *On the existence, uniqueness and approximation of saddle point problems arising from Lagrangian multipliers*, **RAIRO, Anal. Numer., 2**, 1974, 129-151.

Бреци, Равиар (F. Brezzi, P.A. Raviart)

[1] *Mixed finite element methods for 4 th order problems.* To appear.

ван Ноен (R.R.P. van Nooyen)

[1] *An improved accuracy version of the mixed finite-element method for a second-order elliptic equation*, **Journal of Computational and Applied Mathematics 47**, 1993, pp. 11-33, North-Holland.

Василевски, Лазаров (P.S. Vassilevski, R.D. Lazarov)

[1] *Preconditioning Saddle-Point Problems Arising from Mixed Finite Element Discretization of Elliptic Equations*, **UCLA Comp. and Appl. Math.**, Report, 1992.

Джонсън (C. Johnson)

[1] *On the convergence of a mixed finite element method for plate bending problems*, **Numer. Math. 21**, 1973, pp. 43-62.

Димов, Тонев (I. Dimov, O. Tonev)

[1] *Monte Carlo algorithms: performance analysis for some computer architectures.* **J. Computational and Applied Mathematics, vol. 48** (1993) pp. 253–277.

[2] *Random walk on distant mesh points Monte Carlo methods*, **J. Statistical Physics, vol. 70(5/6)** (1993), pp. 1333–1342.

И. Димов, Т. Димов, Гюров (I. Dimov, T. Dimov, T. Gurov)

[1] *A New Iterative Monte Carlo Approach for Inverse Matrix Problem.* **Journal of Computational and Applied Mathematics, vol. 92** (1998) pp. 15–35.

Димов (T.T. Dimov)

[1] *Error estimates for mixed finite element method on rectangular grid with regular local refinement, using "slave" nodes*, **Aplication of Mathematics in Engineering**, Varna 1991 pp 51-53.

[2] *Error estimates of the lowest order mixed finite element method on grids with regular local refinement*, **Advances in Constructive Theory of Functions**, Proceedings of the International Conference on Constructive Theory of Functions, Varna, May 28 – June 31, 1991, Publishing House of the Bulgarian Academy of Sciences, Sofia, 1992, pp. 105–114.

[3] *Efficient Monte Carlo Algorithms for Inverting Matrices Arising in Mixed Finite Element Approximation.* 4^{th} International Conference on Numerical Methods and Applications, Sofia 1998, Submited to publish.

Дрия, Видлунд (M. Dryja, O.B. Widlund)

[1] *On the optimality of an additive iterative refinement method*, Proceedings Coper Mountain Multigrid Conference, April 1989.

Дъглас, Робъртс (J. Douglas, Jr., J.E. Roberts)

[1] *Global estimates for mixed methods for second order elliptic problems*, **Math of Comp., Vol.44**, 1985, 39-52.

Ермаков, Михайлов (С.М. Ермаков, Г.А. Михайлов)

[1] *Статистическое моделирование*, **Наука**, Москва, 1982.

Колац (Л. Коллатц)

[1] *Функциональный анализ и вычислительная математика*, **Мир**, Москва, 1969.

Колотилина, Йерьомин (L.Yu. Kolotilina, A. Yu. Yeremin)

[1] *Factorized Sparse Approximate Inverse Preconditionings I. Theory*, **SIAM J. Matrix Anal. Appl., vol. 14**, (1993), No 1, pp. 45–58.

Курант (R. Courant)

[1] *Variational methods for the solution of problems of equilibrium and vibrations*, Bull. Amer. Math. Soc. 1943, 49, 1-23.

Курант, Хилберт

[1] *Методы мат. физики*, Т. 1 и 2, **Гостехиздат**, 1951.

Къртис (J.H Curtiss)

[1] *Monte Karlo methods for the iterations of linear operators*, **J. Statist. Phis. 70 (5/6)**, 1993.

Лионс, Мадженес (J.-L. Lions, E. Magenes)

[1] *Non - Homogeneous Boundary Value Problems and Applications*, **Springer - Verlag, Vol. 1.**, Berlin, 1970.

Матьо (T.P. Mathew)

[1] *Domain Decomposition and Iterative Refinement Methods for Mixed Finite Element Discretizations of Elliptic Problems*, Technical report, Computer Science Department, New York University.

Мак Кормик (S. McCormick)

[1] *Fast adaptive composite grid (FAC) methods: Theory for the variational case*, **Computing Suppl., 5**, 1984, pp. 115–121.

Миоши (T. Miyoshi)

[1] *A finite element method for the solutions of fourth order partial differential equations*, **Kumamoto J. Math. 9**, 1973, pp. 87–116.

[2] *Lumped mass approximation to the nonlinear bending of elastic plates*, **Kimamoto J. of Science Math., v. 12,** *N*2, pp. 29–34, 1977.

Мегсон, Александров, Димов (G. Megson, V. Aleksandrov, I. Dimov)

[1] *Systolic Matrix Inversion Using a Monte Carlo Method*, **J. of Parallel Algorithms and Appl., vol. 3**, No 3/4 (1994), pp. 311–330.

Метрополис, Улам (N. Metropolis, S. Ulam)

[1] *The Monte Carlo Method*, **J. of Amer. Statist. Assoc., 44**, (1949), pp. 335–341.

Мисовских (И.П. Мысовских)

[1] *Интерполяционные кубатурные формулы*, **Наука**, 1981, 336 с.

Наката, Уейсър, Уилър (M. Nakata, A. Weiser, M.F. Wheeler)

[1] *Some superconvergence results for mixed finite element methods for elliptic problems on rectangular domains*, in **The Mathematics of Finite Elements and Applications** *V*, J.R.Whiteman, ed., Academic Press, London, 1985.

Николски (С.М. Никольский)

[1] *Квадратурные формулы*, **Наука**, Москва 1974, 224 с.

Оден (J.D. Oden)

[1] *Generalized conjugate functions for mixed finite element approximations of boundary-value problems*, **The Mathematical Foundations of the Finite Element Method (A.K.Aziz Editor), Academic Press**, New-York, 1973, pp. 629–670.

[2] *Some contributions to the mathematical theory of mixed finite element approximations*, **Tokyo Seminar on Finite Elements**, Tokyo, 1973.

Оден, Реди (J.T. Oden, J.N. Reddy)

[1] *On mixed element approximations*, Texas Institute for Computational Mechanics, The University of Texas at Austin, 1974.

Осборн (J.E. Osborn)

[1] *Spectral approximation for compact operators*, **Math. Comp. 26**, 1975, pp. 712–725.

Парлет (B. Parlet)

[1] *The Symetric eigenvalue problem*, **Prentice-Hall**, N.J. 1980.

Равиар, Тома (P.A. Raviart, J.E. Thomas)

[1] *A mixed finite element method for 2nd order elliptic problems*, Math. Asp. of the FEM, **Lecture Notes in Math., Vol. 606, Springer Verlag**, Berlin, 293-315.

Реди (J.N. Reddy)

[1] *A Mathematical Theory of Complementary-Dual Variational Principles and Mixed Finite Element Approximations of Linear Boundary-Value Problems in Continuous Mechanics*, Ph. D. Dissertation, The University of Alabama in Huntsville, 1973.

Робъртс, Тома (J.E. Roberts, J.M. Thomas)

[1] *Mixed and Hybrid Methods*, Handlook of Numerical Analysis, **North Holland**, 1991.

Ръсел, Уилър (T.F. Russell, M.F. Wheeler)

[1] *Finite element and finite difference methods for continuous flows in porous media*, **Mathematics of Reservoir Simulation, SIAM Publications**, Philadelphia 1984.

Сиарле (P.G. Ciarlet)

[1] *Finite Element Method for Elliptic Problems*, **North-Holland**, Amsterdam 1978.

Сиарле, Равиар (P.G. Ciarlet, P.A. Raviart)

[1] *General Lagrange and Hermite interpolation in R^n with applications to finite element methods*, **Arch. Rat. Mech. Anal. 46**, 1972, pp. 177–199.

[2] *A mixed finite element method for the biharmonic equation*, **Mathematical Aspects of Finite Elements in Partial Differential Equations, Academic Press**, New-York, 1974, pp. 125–145.

Собол (И.М. Соболь)

[1] *Численные методы Монте-Карло*, **Наука**, Москва, 1973.

Спанер, Джирбърд (J. Spanier, E.M. Gelberd)

[1] *Monte Carlo Principles and Neutron Transport problems*, **Addison-Wesley Publishing Company, Reading, Ma**, 1969.

Стренг, Фикс (G. Strang, G. Fix)

[1] *An Analysis of the Finite Element Method*, **Prentice-Hall, Inc.,Englewood Cliffs**, N.J. 1973.

Томе (V. Thomeé)

[1] *Galerkin Finite Element Methods for Parabolic Problems*, **Lecture Notes in Mathematics, Vol. 1054** Springer-Verlag, Berlin, 1984 pp. 205–220.

Уестлейк (J.R. Westlake)

[1] *A Handbook of Numerical matrix Inversion and Solution of Linear Equations*, **John Wiley & Sons, inc.**, New York, London, Sydney, 1968.

Уонг (J. Wang)

[1] *Analysis of the mixed finite element method for grids with local refinement*, **EORI, University of Wyoming**, Preprint, 1990.

Фолк, Осборн (R.S. Falk, J.E. Osborn)

[1] *Error estimates for mixed methods*, **RAIRO Anal. Numer., 14**, 1980, 249-277.

Хаефил, Кунг (Hyafil L., H.T. Kung)

[1] *Parallel algorithms for solving triangular linear systems with small parallelism*, Dept. of Comp. Sci., Carnegie-Mellon Univ., Dec., 1974.

Холтън (J.H. Halton)

[1] *Sequential Monte Carlo Techniques for the Solution of Linear Systems*, **TR 92-033**, University of North Carolina at Chapel Hill, Department of Computer Science, 46 pp., 1992.

Херман (L.R. Hermann) [1] *Finite element bending analysis for plates*, **J. Eng. Mech. Div. ASCE 94**, (1967), pp. 13–25.

Цанки (L. Csanky)

[1] *Fast parallel matrix inversion algorithms*, **SIAM J. Comput. vol. 5**, (1976), No 4, pp. 618–623.

Чен, Томе (C.M.Chen, V.Thomeé)

[1] *The Lumped Mass Finite Element Method for a Parabolic Problem*, **J. Austral. Math. Soc. Ser.** *B* **26**, 1985, pp. 329–354.

Чермак, Зламал (L. Cermak, M. Zlamal)

[1] *Transformation of dependant variables and the finite element solution of nonlinear evolution equations*, **Intern. J. for Numer. Meth. in Engin., v. 15**, 1980, *N*1, pp. 31–40.

Юинг, Лазаров, Василевски (R.E.Ewing, R.D.Lazarov, P.S.Vassilevski)

[1] *Local refinement techniques for elliptic problems in cell-centered grids*, Department of Mathematics, University of Wyoming, Preprint # 1988-16 (1988), 75 pages.

Юинг, Лазаров, Ръсел, Василевски (R.E. Ewing, R.D. Lazarov, T.F. Russel, P.S. Vassilevski)

[1] *Local refinement via domain decomposition techniques for mixed finite element methods with rectangular Raviart-Thomas elements*, **Domain Decomposition Methods for PDE's, SIAM**, 1990, 92-114.

Юинг, Лазаров, Уонг (R.E. Ewing, R.D. Lazarov, J. Wang)

[1] *Superconvergence of the Velocity alongg the Gauss Lines in Mixed Finite Element Methods*, **SIAM J. Numer. Anal. 28**, 1991, pp. 1015–1029.

Юинг, Уонг (R.E. Ewing, J. Wang)

[1] *Analysis of mixed finite element methods on locally refined grids*, **J. Numer. Math. 63, No.2**, 1992, pp. 183–194.

Юинг, Уилър (R.E. Ewing, M.F. Wheeler)

[1] *Computational aspects of mixed finite element methods*, in Numerical Methods for Scientific Computing, R.S.Stepleman, ed., **North-Holland, New York**, 1983, pp. 163–172.

Юинг, Коебби, Гонзалез, Уилър (R.E. Ewing, J.V. Koebbi, R. Gonzalez, M.F. Wheeler)

[1] *Computing accurate velocities for fluid flow in porous media*, **Proceedings of TICOM Conference**, Austin, Texas (1983).