NASA Contractor Report 181902

ICASE Report No. 89–56

# ICASE

## OPTIMAL ERROR ANALYSIS OF SPECTRAL METHODS WITH EMPHASIS ON NON-CONSTANT COEFFICIENTS AND DEFORMED GEOMETRIES

**Yvon Maday**
**Einar M. Rønquist**

Contract No. NAS1–18605
September 1989

Institute for Computer Applications in Science and Engineering
NASA Langley Research Center
Hampton, Virginia 23665–5225

Operated by the Universities Space Research Association



National Aeronautics and Space Administration

**Langley Research Center**
Hampton, Virginia 23665-5225

# OPTIMAL ERROR ANALYSIS OF SPECTRAL METHODS WITH EMPHASIS ON NON-CONSTANT COEFFICIENTS AND DEFORMED GEOMETRIES

*Yvon Maday*[1] [2] [3]

and

*Einar M. Rønquist*[3]

## Abstract

In this paper we present the numerical analysis of spectral methods when non-constant coefficients appear in the equation, either due to the original statement of the equations or to take into account the deformed geometry. A particular attention is devoted to the optimality of the discretization even for low values of the discretization parameter. The effect of some "overintegration" is also addressed, in order to possibly improve the accuracy of the discretization.

[1]This research was supported in part by the National Aeronautics and Space Administration under NASA Contract No. NAS1-18605 while the first author was in residence at the Institute for Computer Applications in Science and Engineering (ICASE), NASA Langley Research Center, Hampton, VA 23665.

[2]Laboratoire C.N.R.S. d'Analyse Numérique de l'Université Pierre et Marie Curie, Tour 55-65, 5eme Étage, 4, Place Jussieu, 75252 Paris Cedex 05, France.

[3]Department of Mechanical Engineering, M.I.T., 77, Massachusetts Avenue, Cambridge, MA 02139, USA

## 1. INTRODUCTION.

Spectral methods were introduced around 20 years ago by S. A. ORSZAG in order to provide a high order accuracy for the numerical simulation of partial differential equations. The original idea was to use truncated Fourier series to approximate the (smooth) solution when the problem was supposed to be provided with periodic boundary conditions. In order to tackel problems with more general boundary conditions (Dirichlet or Neumann type), the set of (algebraic) polynomials replaced the set of truncated series, but the characterization of the unique discrete function that would provide the numerical solution was still achieved following the original strategy. This enters in the context of the collocation method where the numerical solution is chosen so as to satisfy the original partial differential equation at some suitably chosen collocation points. Of course the choice of the set of collocation points is of fundamental importance for the accuracy of the method and the first remark is to notice that the number of collocation points must be equal to the dimension of the space of approximation. Otherwise, the problem could, in general, be overspecified.

When other authors arrived to the analysis of the problem, ([1], [2], [3], [4] among others) they realized that the collocation method could be interpreted as a variational problem with numerical integration. In fact, it is quite common now to realize that the spectral methods are very close to the finite element method in its $p$ or $h-p$ version, where convergence is achieved by increasing the order of the polynomial degree and not by diminishing the size of the elements. It is in this framework that the domain of application of the spectral methods has been generalized. The plain spectral method suffers from being constrained to very simple domains: they are limited to be slightly deformed squares (in 2-D) or cubes (in 3-D). The idea to couple domain decomposition techniques to the spectral discretization developed rapidly in order to cope with this initial drawback ([5], [6]). However, by starting from the strong formulation of the equations, this can only produce the Schwarz algorithm ([6], [7]) or a strong coupling between the elements ([5]), where the solution (in the case of second order elliptic problems) is searched as a global $C^1$ function that is piecewise polynomial. This results in some drawbacks. First, in the Schwarz algorithm, there is an increase of the work due to the double computation over the overlapping region (recall that this one has to be large enough in order to achieve a good convergence rate of the algorithm [7], [8]). Second there is a lack of optimality of the approximation in the strong coupling formulation — both from the numerical analysis point of view and from the algorithmic point of view — a consequence of an overconstraint problem due to the $C^1$ matching.

In this context, understanding of the similarity between the collocation method and a variational formulation used with consistent numerical quadrature brings a lot of flexibility. Indeed, the coupling required in the nonoverlapping decomposition of the domain is weaker (only $C^0$) and allows for constructing an optimal method. This has lead to the spectral element method [9], [10], [11] and more recently to the mortar element method [12], [13], [14]. The variational method involves integrals that can be computed (with or without numerical quadratures) separately over each subdomain. The spectral element method uses consistent quadrature formulas and in this respect (but not only this one) conserves the spirit of former spectral discretization (the other points being the use of tensorial basis and tensorial evaluations of the residuals). It has been shown ([10], [15]) that the method can still be interpreted as a collocation method within each element wheras a suitable equation is satisfied at the interface of each subdomain. There are still fundamental differences between the spectral and the finite element methods that are important as regards the numerical implementation of the method, however the general philosophy is the same. The point that we want to address in this paper is related to one of the differences between the $p$ or $h-p$ version of the finite element method and the spectral methods, more precisely to one possible drawback of the spectral method. Indeed, as we have said, the spectral method can be interpreted in a finite element framework when a particular numerical integration formula is used. Derived from the collocation method, the numerical integration formula is constrained to be based on a certain number of points. This number is not related to the fact that some quantity must be well (or exactely) computed, but has to be equal to the dimension of the discrete space. Using a vocabulary that is standard in the finite element context, the points of the numerical quadrature formula have to be unisolvent with respect to the discrete space. This is a restriction with respect to the fact that in finite element methods, people use more general integration formulas based on accuracy considerations. In this sense, there is often an overintegration in the finite element statement. The current spectral methods use these "*consistent*" (with respect to the numerical discretization) quadrature rules with a priori no lack of accuracy. As shown in the appendix, the possibility to use an overintegration in the context of

spectral methods is however possible, but is not of importance in most interesting cases as demonstrated in this paper.

The situations where overintegration could be of importance are those where non-constant coefficients are involved, either because they are present in the original formulation of the equations or because they are introduced when a deformed geometry is considered and a mapping to a domain of reference is used. Indeed, it is in these cases that the numerical integration based on $N+1$ Gauss-Lobatto points is not accurate enough. We analyze in details these two situations and point out the cases where overintegration is required. The analysis is illustrated with many numerical experiments. This allows to strengthen our theoretical analysis since the exact value of some constants could be of importance and are not exactely taken into account in the theoretical analysis.

The analysis of the effect of overintegration has never been addressed in the spectral context. Previous analysis of Legendre spectral approximations of problems with non-constant coefficients [10] or with deformed geometries [7] could in this sense be misleeding as the results presented there were proving a spectral type convergence (i.e. faster than any algebraic rate if the solution is analytic) but were not optimal. The basic ingredient to get optimality is a new result of [16] and deals with the fact that the interpolation at Gauss-type points is optimal.

The paper is organized as follows:
In section 2, we consider the case of original non-constant coefficients equations and we analyze in which cases the overintegration may improve the accuracy of the approximation to allow for optimal results. An appendix presents some numerical considerations on the implementation that shows that the tensor product evaluations can still be preserved even in the case of overintegrations.
In section 3, we analyze the case where the non-constant coefficients are induced by the treatment of a deformed geometry. We prove that in this situation the overintegration is unnecessary to provide the optimality of the approximation.
In sections 2 and 3 we also present some remarks related to the impact of our analysis as regard the design of the best schemes in the case of deformed geometries.

**Acknowledgements**: The question of the importance of overintegration has been raised during a meeting with I. BABUŠKA. The many discussions with A. T. PATERA have been very enlightening. We thank both of them for their help. This research has been supported in part by NSF under ASC–8806925, ONR and DARPA under contract N00014–85–K 0208. Part of this work has been done while the first author was in residence at ICASE, NASA Langley Research Center, Hampton, VA.

## 2. ANALYSIS OF SPECTRAL APPROXIMATION WITH NON CONSTANT COEFFICIENTS.

### 2.1 — The One Dimensional Case.

Let us consider the following problem: *Find* $u \in H_0^1(-1,1)$, *such that*

$$-(p\ u_x)_x + q\ u = f, \quad \text{in } (-1,1). \tag{1}$$

This problem is supposed to be well posed in $H_0^1(-1,1)$ in the sense that the bilinear form $a$, defined over $(H_0^1(-1,1))^2$ as follows

$$a(u,v) = \int_{-1}^{1} p u_x v_x dx + \int_{-1}^{1} quv dx, \tag{2}$$

is elliptic and continuous over $(H_0^1(-1,1))^2$. In fact, for sake of simplicity, we shall assume that there exists three constants $p_1, p_2$ and $q_2$ such that

$$0 < p_1 \leq p(x) \leq p_2, \quad \forall x \in (-1,1); \tag{3}$$

$$0 \leq q(x) \leq q_2, \quad \forall x \in (-1,1). \tag{4}$$

In addition, the forcing function $f$ is supposed to be at least in $L^2(-1,1)$. We want to approximate this problem by a spectral element method. Our first work will be to design the discrete space that will approximate $H_0^1(-1,1)$. Given a fixed integer $K$, we shall first consider a partition of $(-1,1)$ in $K$ subintervals $I_k$, where

$$I_k = (a_k, a_{k+1}), \tag{5}$$

and the $a_k$ are $(K+1)$ points in (-1,1) such that

$$-1 = a_0 < a_1 < \ldots < a_{K-1} < a_K = 1,$$

(and the lengh $a_{k+1} - a_k$ is supposed to be $\mathcal{O}(K^{-1})$.)

The spaces of discretization are imbedded in

$$Y_N = \{\phi \in L^2(-1,1); \quad \phi_{|I_k} \in I\!P_N(I_k)\}, \tag{6}$$

where $N$ is some integer and $I\!P_N(\Lambda)$ denotes the set of all polynomials of degree $\leq N$ over $\Lambda$. This space will discretize $L^2(-1,1)$. In order to discretize $H_0^1(-1,1)$, let us introduce

$$X_N = Y_N \cap H_0^1(-1,1), \tag{7}$$

(which means that $\phi_{|I_k}(a_{k+1}) = \phi_{|I_{k+1}}(a_{k+1})$ and $\phi$ vanishes at $\pm 1$.)

The second step in the discretization process is to define the discrete problem. As indicated in the introduction, the point of departure is the following continuous variational formulation of the problem (1): *Find* $u \in H_0^1(-1,1)$ *such that*

$$a(u,v) = (f,v), \quad \forall v \in H_0^1(-1,1), \tag{8}$$

where the notation $(.,.)$ stands for the $L^2(-1,1)$–scalar product. Given an integer $M$, $M \geq N$, the discrete problem is: *Find* $u_N \in X_N$ *such that*

$$a_M(u_N, v_N) = (f, v_N)_M, \quad \forall v_N \in X_N. \tag{9}$$

The definition of $a_M$ and $(.,.)_M$ from $a$ and $(.,.)$ will use a composed Gauss-Lobatto quadrature formula. More precisely, this one is based on the data, over each segment $I_k$ of a set of $M+1$ points $\xi_i^k$

$$a_k = \xi_0^k < \xi_1^k < \ldots < \xi_{M-1}^k < \xi_M^k = a_{k+1}, \tag{10}$$

and a set of associated weights $\rho_i^k$ that are such that

$$\sum_{GL} \Phi \equiv \sum_{k=0}^{K} \sum_{i=0}^{M} \Phi(\xi_i^k)\rho_i^k = \int_I \Phi(x)dx, \quad \forall \Phi \in Y_{2M-1}. \tag{11}$$

The definition of $a_M$ and $(.,.)_M$ is now as follows:

$$a_M(u_N, v_N) = \sum_{GL} p u_{Nx} v_{Nx} + \sum_{GL} q u_N v_N, \quad \forall (u_N, v_N) \in X_N^2, \tag{12}$$

$$(\phi, \psi)_M = \sum_{GL} \phi\psi, \quad \forall (\phi, \psi) \in Y_N^2. \tag{13}$$

**Remark 2.1.** The standard spectral element method utilizes a consistent Gauss-Lobatto quadrature rule, i.e. $M = N$. The case $M > N$ corresponds to an overintegration with respect to the standard spectral element method. Such an overintegration might be of interest to introduce in the case of non-constant coefficients. The analysis that follows indicates the situations for which this is the case.

The interesting feature about $(.,.)_M$ is that it defines over $Y_M$ a discrete scalar product that is uniformly equivalent to the $L^2(-1,1)$ one. Indeed, it has been proven in [4] that

$$(\phi, \phi) \leq (\phi, \phi)_M \leq 3(\phi, \phi), \quad \forall \phi \in Y_M. \tag{14}$$

The following Lemma is then completely standard

**Lemma 2.1.** *There exists one and only one solution $u_N$ to problem (9) and there exists a constant $C$ such that*

$$\|u_N\|_{H^1(-1,1)} \leq C\|f\|_{L^\infty(-1,1)}$$

**Proof:** The Lemma follows easily from the constatation that $a_M$ is uniformly continuous and elliptic over $X_N$, i.e. there exists two positive constants $\alpha$ and $\gamma$ such that

$$\alpha\|u_N\|^2_{H^1(-1,1)} \leq a_M(u_N, u_N) \leq \gamma\|u_N\|^2_{H^1(-1,1)} \tag{15}$$

and can be deduced from (3), (4) and (14).

Our further analysis is devoted to the derivation of optimal error bounds for the numerical solution. The following Lemma is also standard and follows from (15).

**Lemma 2.2.** *There exists a constant $C$ independent of $N$ such that*

$$\begin{aligned}\|u - u_N\|_{H^1(-1,1)} \leq C[\inf_{w_N \in X_N} \{\|u - w_N\|_{H^1(-1,1)} + \sup_{v_N \in X_N} \frac{a(u, v_N) - a_M(w_N, v_N)}{\|v_N\|_{H^1(-1,1)}}\} \\ + \sup_{v_N \in X_N} \frac{(f, v_N) - (f, v_N)_M}{\|v_N\|_{H^1(-1,1)}}].\end{aligned} \tag{16}$$

The following Lemma proves that the last term in (16) is of the same order as the best $L^2(-1,1)$-fit of $f$ by polynomials of degree $2M - N$. Let us first denote by $C_K^0(-1,1)$ the space

$$C_K^0(-1,1) = \{\psi \in L^2(-1,1), \quad \exists \psi_k \in C^0(I_k), \psi_{|I_k} = \psi_k\} \tag{17}$$

and similarly, for any $\mu$, by $H_K^\mu(-1,1)$ the space

$$H_K^\mu(-1,1) = \{\psi \in L^2(-1,1), \quad \psi_{|I_k} \in H^\mu(I_k)\}, \tag{18}$$

we have

**Lemma 2.3.** *Let $\mu$ be a real number $> 1/2$. For any $\psi \in H^\mu_K(-1,1)$, we have*

$$\sup_{v_N \in X_N} \frac{(\psi, v_N) - (\psi, v_N)_M}{\|v_N\|_{H^1(-1,1)}} \leq C(2M-N)^{-\mu} \sum_{k=1}^{K} \|\psi_{|I_k}\|_{H^\mu(I_k)}, \tag{19}$$

$$\sup_{v_N \in Y_{N-1}} \frac{(\psi, v_N) - (\psi, v_N)_M}{\|v_N\|_{L^2(-1,1)}} \leq C(2M-N)^{-\mu} \sum_{k=1}^{K} \|\psi_{|I_k}\|_{H^\mu(I_k)}. \tag{20}$$

**Proof:** Let us first consider the case where $M = N$. Let us introduce, for any integer $L$, the $L^2(-1,1)$–projection $\psi_L = \Pi_L(\psi)$ of $\psi$ over $\mathbb{P}_L(-1,1)$, we have

$$\begin{aligned}(\psi, v_N) - (\psi, v_N)_N = [(\psi, v_N) - (\psi_N, v_N)] + [(\psi_N, v_N) - (\psi_N, v_N)_N] \\ + [(\psi_N, v_N)_N - (\psi, v_N)_N].\end{aligned} \tag{21}$$

It is an easy matter to note that

$$|(\psi, v_N) - (\psi_N, v_N)| \leq C\|\psi - \psi_N\|_{L^2(-1,1)}\|v_N\|_{L^2(-1,1)}, \tag{22}$$

Besides, let us denote by $\mathcal{I}_N$ the operator of interpolation from the set $C^0_K(-1,1)$ onto $Y_N$ such that

$$\mathcal{I}_N\psi(\xi_i^k) = \psi(\xi_i^k), \quad \forall i = 0, \ldots, N, \ \forall k = 1, \ldots, K. \tag{23}$$

then

$$\begin{aligned}|(\psi_N, v_N)_N - (\psi, v_N)_N| &= |(\psi_N - \mathcal{I}_N\psi, v_N)_N| \\ &\leq (\psi_N - \mathcal{I}_N\psi, \psi_N - \mathcal{I}_N\psi)_N^{1/2}(v_N, v_N)_N^{1/2},\end{aligned}$$

which, thanks to (14), gives

$$\begin{aligned}|(\psi_N, v_N)_N - (\psi, v_N)_N| &\leq C\|\psi_N - \mathcal{I}_N\psi\|_{L^2(-1,1)}\|v_N\|_{L^2(-1,1)} \\ &\leq C(\|\psi - \psi_N\|_{L^2(-1,1)} + \|\psi - \mathcal{I}_N\psi\|_{L^2(-1,1)})\|v_N\|_{L^2(-1,1)}.\end{aligned} \tag{24}$$

It remains to estimate the middle term in (21). Let us first assume that $K = 1$, corresponding to a plain spectral method with no domain decomposition. The term in question can then be written as follows

$$(\psi_N, v_N) - (\psi_N, v_N)_N = (\hat{\psi}_N \hat{v}_N)[(L_N, L_N) - (L_N, L_N)_N]$$

where, for any $j \in \mathbb{N}$

$$\hat{\psi}_j = \frac{(\psi, L_j)}{(L_j, L_j)}, \quad \hat{v}_j = \frac{(v_N, L_j)}{(L_j, L_j)}, \tag{25}$$

are the coefficients of $\psi$ (and thus of $\psi_N$) and $v_N$ respectively in the basis of the Legendre polynomials $(L_j)_{j \in \mathbb{N}}$. It is then an easy matter to note, from (14), that

$$\begin{aligned}|(\psi_N, v_N) - (\psi_N, v_N)_N| &\leq C|\hat{\psi}_N \hat{v}_N|\|L_N\|^2_{L^2(-1,1)} \\ &\leq C|\hat{\psi}_N| \ \|L_N\|_{L^2(-1,1)}|\hat{v}_N| \ \|L_N\|_{L^2(-1,1)}\end{aligned}$$

Also, it follows easily that

$$|\hat{\psi}_N|\|L_N\|_{L^2(-1,1)} \leq \|\psi - \psi_{N-1}\|_{L^2(-1,1)}$$

and

$$|\hat{v}_N|\|L_N\|_{L^2(-1,1)} \leq \|v_N - \Pi_{N-1}(v_N)\|_{L^2(-1,1)}.$$

so that we deduce

$$|(\psi_N, v_N) - (\psi_N, v_N)_N| \leq C\|\psi - \psi_{N-1}\|_{L^2(-1,1)}\|v_N - \Pi_{N-1}(v_N)\|_{L^2(-1,1)}. \tag{26}$$

The multidomain case where $K > 1$ is treated similarly over each subdomain $I_k$ and the error is the same (the notations are in this case just more complicated). Let us recall that we have for any $\rho \geq 0$, and any $L$

$$\|\phi - \Pi_L \phi\|_{L^2(-1,1)} \leq C L^{-\rho} \|\phi\|_{H^\rho(-1,1)}, \quad \forall \phi \in H^\rho(-1,1). \tag{27}$$

we derive by summarizing (21), (22), (24), and (26) that

$$\sup_{v_N \in X_N} \frac{(\psi, v_N) - (\psi, v_N)_N}{\|v_N\|_{H^1(-1,1)}} \leq 2\|\psi - \psi_N\|_{L^2(-1,1)} + \|\psi - \mathcal{I}_N \psi\|_{L^2(-1,1)} + C N^{-1} \|\psi - \psi_{N-1}\|_{L^2(-1,1)}$$

This completes Lemma 2.3 in the case $M = N$, after we recall that for any $\rho > 1/2$, and any $L$, we have [16]

$$\|\phi - \mathcal{I}_L \phi\|_{L^2(-1,1)} \leq C L^{-\rho} \|\phi\|_{H^\rho(-1,1)}, \quad \forall \phi \in H^\rho(-1,1). \tag{28}$$

so that the contribution of $N^{-1}\|\psi - \psi_{N-1}\|_{L^2(-1,1)}$ is much smaller than the other terms.

Let us now consider the case where $M > N$ and let us use the same strategy as before, with $\tilde{\psi}_{2M-N}$ in place of $\psi_N$, where $\tilde{\psi}_{2M-N}$ is some good approximation of $\psi$ of degree $\leq 2M - N$ that will not be necessarily its $L^2(-1,1)$–projection and will be precised in the following. In order to prove Lemma 2.3, we must now estimate

$$\begin{aligned}(\psi, v_N) - (\psi, v_N)_M = (\psi, v_N) - (\tilde{\psi}_{2M-N}, v_N) + (\tilde{\psi}_{2M-N}, v_N) - (\tilde{\psi}_{2M-N}, v_N)_M \\ + (\tilde{\psi}_{2M-N}, v_N)_M - (\psi, v_N)_M\end{aligned} \tag{29}$$

The first term and the second term are treated exactely in the same way as previously and we obtain

$$|(\psi, v_N) - (\tilde{\psi}_{2M-N}, v_N)| \leq \|\psi - \tilde{\psi}_{2M-N}\|_{L^2(-1,1)} \|v_N\|_{L^2(-1,1)}, \tag{30}$$

and

$$(\tilde{\psi}_{2M-N}, v_N) - (\tilde{\psi}_{2M-N}, v_N)_M = \hat{\psi}_{2M-N} \hat{v}_N [(L_{2M-N}, L_N) - (L_{2M-N}, L_N)_M]. \tag{31}$$

¿From the basic properties of the Legendre polynomials we recall that that

$$(L_m, L_n) = \frac{2\delta_{m,n}}{2n+1}, \qquad \forall (m,n),$$

and that

$$(n+1)L_{n+1}(x) = (2n+1)xL_n(x) - nL_{n-1}(x).$$

A reiterate use of these two formulas yields

$$\begin{aligned}(L_{2M-N}, L_N)_M &= \frac{4M-2N-1}{2M-N} \frac{N+1}{2N+1} (L_{2M-N-1}, L_{N+1})_M = \dots \\ &\leq (L_M, L_M)_M.\end{aligned}$$

Now using (14), we derive that

$$(L_{2M-N}, L_N)_M \leq 3\|L_M\|^2_{L^2(-1,1)} \leq 3\|L_{2M-N}\|_{L^2(-1,1)} \|L_N\|_{L^2(-1,1)},$$

so that, as before,

$$|(\tilde{\psi}_{2M-N}, v_N) - (\tilde{\psi}_{2M-N}, v_N)_M| \leq C N^{-1} \|\psi - \tilde{\psi}_{2M-N-1}\|_{L^2(-1,1)}. \tag{32}$$

The treatment of the last term in (29) has to be different. Indeed the same proceedure as before would result in a bound of this term by the quantity $\|\psi - \mathcal{I}_M \psi\|_{L^2(-1,1)}$ and thus a loss with respect to the optimality we could expect from the other terms. Besides, the numerical experiments (see fig 2.1) clearly shows a behaviour

like $\|\psi - \psi_{2M-N}\|_{L^2(-1,1)}$. In order to prove this type of convergence, let us go back to the case $K = 1$ and note that (28) yields, for any $\mu > 1/2$

$$\|\mathcal{I}_M \phi\|_{L^2(-1,1)} \leq \|\phi\|_{L^2(-1,1)} + M^{-\mu}\|\phi\|_{H^\mu(-1,1)}, \quad \forall \phi \in H^\mu(-1,1). \tag{33}$$

¿From (14) it now follows that

$$\begin{aligned}(\tilde{\psi}_{2M-N}, v_N)_M - (\psi, v_N)_M &= (\tilde{\psi}_{2M-N} - \psi, v_N)_M \\ &\leq 3\|\mathcal{I}_M(\tilde{\psi}_{2M-N} - \psi)\|_{L^2(-1,1)}\|v_N\|_{L^2(-1,1)} \\ &\leq 3[\|\tilde{\psi}_{2M-N} - \psi\|_{L^2(-1,1)} + M^{-\mu}\|\tilde{\psi}_{2M-N} - \psi\|_{H^\mu(-1,1)}]\|v_N\|_{L^2(-1,1)}.\end{aligned}$$

At this point the correct $\tilde{\psi}_{2M-N}$ will be chosen. We see that it has to be a good approximation of $\psi$ in the $H^\mu(-1,1)$–norm as well. This will be the case of the projection of $\psi$ with respect to the $H^\mu(-1,1)$–norm over $Y_N$. It is proven in [17] that this element satisfies

$$\|\psi - \tilde{\psi}_{2M-N}\|_{H^\nu_K(-1,1)} \leq CN^{\nu-\mu}\|\psi\|_{H^\mu_K(-1,1)}, \quad \forall \nu \leq \mu, \tag{34}$$

and from (29), (30) and (32), the proof of Lemma 2.3 is complete also in the case where $M > N$.

The second inequality (20) is quite trivial since the middle terms in (21) and (29) simply vanish.

We are now in position to state the following

**Theorem 2.4.** *Let us suppose that the solution $u$ to problem (1) belongs to $H^\sigma_K(-1,1)$, that $p$ belongs to $H^\nu_K(-1,1)$, $q$ belongs to $H^\mu_K(-1,1)$ and that $f$ belongs to $H^\rho_K(-1,1)$ and in addition that the four real numbers $\sigma - 1$, $\mu$, $\nu$ and $\rho$ are larger than $1/2$. Then the following error estimate holds*

$$\begin{aligned}\|u - u_N\|_{H^1(-1,1)} \leq & C[N^{1-\sigma}\|u\|_{H^\sigma(-1,1)} + (2M-N)^{-\min(\nu,\mu,\sigma-1,\rho)}[\|p\|_{H^\nu_K(-1,1)}\|u\|_{H^{\min(\nu+1,\sigma)}_K(-1,1)} \\ & + \|q\|_{H^\mu_K(-1,1)}\|u\|_{H^{\min(\mu,\sigma)}(-1,1)} + \|f\|_{H^\rho_K(-1,1)}]].\end{aligned} \tag{35}$$

**Proof:** It is a direct application of Lemma 2.3 to derive

$$\sup_{v_N \in X_N} \frac{(f, v_N) - (f, v_N)_M}{\|v_N\|_{H^1(-1,1)}} \leq (2M-N)^{-\rho}\|f\|_{H^\rho_K(-1,1)}. \tag{36}$$

Let us choose $w_N$ as being the best fit of $u$ in the $H^1(-1,1)$–norm. It follows from [17] that

$$\|u - w_N\|_{H^1(-1,1)} \leq CN^{1-\sigma}\|u\|_{H^\sigma_K(-1,1)}. \tag{37}$$

If we now let $\tilde{w}_N$ be any polynomial in $X_N$, it follows from (3) and (4) that,

$$\begin{aligned}|a_M(w_N, v_N) - a_M(\tilde{w}_N, v_N)| &\leq C(p_2, q_2)\|w_N - \tilde{w}_N\|_{H^1(-1,1)}\|v_N\|_{H^1(-1,1)} \\ &\leq C(p_2, q_2)[\|u - w_N\|_{H^1(-1,1)} + \|u - \tilde{w}_N\|_{H^1(-1,1)}]\|v_N\|_{H^1(-1,1)}\end{aligned}$$

so that

$$\begin{aligned}\sup_{v_N \in X_N} \frac{a(u, v_N) - a_M(w_N, v_N)}{\|v_N\|_{H^1(-1,1)}} \leq & \sup_{v_N \in X_N} \frac{a(\tilde{w}_N, v_N) - a_M(\tilde{w}_N, v_N)}{\|v_N\|_{H^1(-1,1)}} \\ & + [\|u - w_N\|_{H^1(-1,1)} + \|u - \tilde{w}_N\|_{H^1(-1,1)}].\end{aligned} \tag{38}$$

The interest of this decomposition is that we can now choose $\tilde{w}_N$, still close to $u$ but such that the higher norms (i.e. $H^\sigma_K(-1,1)$) is uniformly bounded. This will be the case if we choose $\tilde{w}_N$ as being the projection of $u$ with respect to the $H^\sigma(-1,1)$–norm over $X_N$. It is proven in [17] that this element satisfies

$$\|u - \tilde{w}_N\|_{H^\gamma_K(-1,1)} \leq CN^{\gamma-\sigma}\|u\|_{H^\sigma_K(-1,1)}, \quad \forall \gamma \leq \sigma. \tag{39}$$

¿From the definition of $a$ and $a_M$ we have

$$a(\tilde{w}_N, v_N) - a_M(\tilde{w}_N, v_N) = (p\tilde{w}_{Nx}, v_{Nx}) - (p\tilde{w}_{Nx}, v_{Nx})_M \tag{40}$$
$$+ (q\tilde{w}_N, v_N) - (q\tilde{w}_N, v_N)_M. \tag{41}$$

Using Lemma 2.3, first with $\psi = p\tilde{w}_{Nx}$ and then with $\psi = q\tilde{w}_N$, we derive that

$$\begin{aligned}
|a(\tilde{w}_N, v_N) - a_M(\tilde{w}_N, v_N)| \leq\ & C[(2M-N)^{-min(\nu,\sigma-1)}\|p\tilde{w}_{Nx}\|_{H_K^{min(\nu,\sigma-1)}(-1,1)} \\
& + (2M-N)^{-min(\mu,\sigma)}\|q\tilde{w}_N\|_{H_K^{min(\mu,\sigma)}(-1,1)}]\|v_N\|_{H^1(-1,1)} \\
\leq\ & C[(2M-N)^{-min(\nu,\sigma-1)}\|pu_x\|_{H_K^{min(\nu,\sigma-1)}(-1,1)} \\
& + (2M-N)^{-min(\mu,\sigma)}\|qu\|_{H_K^{min(\mu,\sigma)}(-1,1)}]\|v_N\|_{H^1(-1,1)} \\
& + C[(2M-N)^{-min(\nu,\sigma-1)}\|p(u-\tilde{w}_N)_x\|_{H_K^{min(\nu,\sigma-1)}(-1,1)} \\
& + (2M-N)^{-min(\mu,\sigma)}\|q(u-\tilde{w}_N)\|_{H_K^{min(\mu,\sigma)}(-1,1)}]\|v_N\|_{H^1(-1,1)}
\end{aligned}$$

Let us recall that $min(\nu, \mu, \sigma - 1) > 1/2$ so that $H_K^{min(\nu,\sigma-1)}(-1,1)$ and $H_K^{min(\mu,\sigma)}(-1,1)$ are algebras. This proves that

$$\begin{aligned}
|a(\tilde{w}_N, v_N) - a_M(\tilde{w}_N, v_N)| \leq\ & C[(2M-N)^{-min(\nu,\sigma-1)}\|pu_x\|_{H_K^{min(\nu,\sigma-1)}(-1,1)} \\
& + (2M-N)^{-min(\mu,\sigma)}\|qu\|_{H_K^{min(\mu,\sigma)}(-1,1)}]\|v_N\|_{H^1(-1,1)} \\
& + C[(2M-N)^{-min(\nu,\sigma-1)}\|p\|_{H_K^{\nu}(-1,1)}\|(u-\tilde{w}_N)_x\|_{H_K^{min(\nu,\sigma-1)}(-1,1)} \\
& + (2M-N)^{-min(\mu,\sigma)}\|q\|_{H_K^{\mu}(-1,1)}\|(u-\tilde{w}_N)\|_{H_K^{min(\mu,\sigma)}(-1,1)}]\|v_N\|_{H^1(-1,1)}
\end{aligned}$$

which together with (39) and the fact that $M \geq N$ (so that $2M - N \geq N$) gives

$$\begin{aligned}
|a(\tilde{w}_N, v_N) - a_M(\tilde{w}_N, v_N)| \leq\ & C[(2M-N)^{-min(\nu,\sigma-1)}\|pu_x\|_{H_K^{min(\nu,\sigma-1)}(-1,1)} \\
& + (2M-N)^{-min(\mu,\sigma)}\|qu\|_{H_K^{min(\mu,\sigma)}(-1,1)}]\|v_N\|_{H^1(-1,1)} \\
& + [N^{1-\sigma}\|p\|_{H_K^{\nu}(-1,1)}\|u\|_{H_K^{\sigma}(-1,1)} \\
& + N^{-\sigma}\|q\|_{H_K^{\mu}(-1,1)}\|u\|_{H_K^{\sigma}(-1,1)}]\|v_N\|_{H^1(-1,1)}
\end{aligned}$$

Theorem 2.4 now follows from the fact that $H^r(-1,1)$ is an algebra when $r > 1/2$.

**Corollary 2.5.** *Let us suppose that the hypothesis of theorem 2.4 are fullfilled and in addition let us suppose that $M = N$. Then the following error estimate holds*

$$\begin{aligned}
\|u - u_N\|_{H^1(-1,1)} \leq & C[N^{min(\nu,\mu,\sigma-1,\rho)}[\|p\|_{H_K^{\nu}(-1,1)}\|u\|_{H_K^{min(\nu+1,\sigma)}(-1,1)} \\
& + \|q\|_{H_K^{\mu}(-1,1)}\|u\|_{H^{min(\mu,\sigma)}(-1,1)} + \|f\|_{H_K^{s}(-1,1)}]
\end{aligned}$$

**Remark 2.2** — Interpretation of the results: In the practical computations, the more common situation is the one where the regularity of the solution is limited by the regularity of the non-constant data. More precisely, and especially if we consider the extension to the 2-D case that will be treated in the forthcoming section, we shall have $\sigma - 1 \leq min(\nu, \mu, \rho)$. In this case, the estimates given in corollary 2.5 prove that the optimality of the scheme is achieved with no use of overintegration. This statement is confirmed by the numerical experiments shown in Figure 2.1. In the case where the regularity of the solution is better than the regularity of the non-constant data, we expect, from theorem 2.4 that some amount of overintegration is necessary to recover the optimality of the discretization. The numerical experiments of Figure 2.1 also confirm this statement with exactly the rate theoretically proven. It is important to note that the possibility

to include some amount of overintegration only is true in the case where the variational formulation is the starting point of the discretization. The pure collocation method is incompatible with the concept of overintegration.

2.2 — The Two Dimensional Case.

Let us suppose here that the problem is set on a square $\Omega = (-1,1)^2$ and that no domain decomposition is used. The extension of our results to the spectral element case is very simple and it is only for sake of limitation of the notations that we restrict our analysis to this particular case. Let us consider now the following test problem: *Find* $u \in H_0^1(\Omega)$ *such that*

$$\nabla \cdot \mathbf{A} \cdot \nabla u = f \tag{42}$$

where $f$ is a given force (supposed to be in $L^2(\Omega)$), $\mathbf{A} = (a_{ij})$ is a $2 \times 2$ matrix with non-constant coefficients and we assume, for sake of simplicity, that it is bounded, symmetric, positive and non-degenerate, i.e. such that there exists 2 real numbers $p_1 > 0$ and $p_2 \geq 0$ such that

$$\mathbf{AV} \cdot \mathbf{V} \geq p_1(V_1^2 + V_2^2), \quad \forall \mathbf{V} = (V_1, V_2) \in I\!R^2 \tag{43}$$

$$|a_{ij}| \leq p_2, \quad \forall i,j = 1,2. \tag{44}$$

The variational formulation of this problem : *Find* $u \in H_0^1(\Omega)$ *such that*

$$a(u,v) = ((f,v)), \quad \forall v \in H_0^1(\Omega) \tag{45}$$

will be used as a starting point to define the discrete problem. Here

$$((\phi,\psi)) = \int\int_\Omega (\phi\psi)(\mathbf{x})d\mathbf{x}, \quad \forall (\phi,\psi) \in (L^2(\Omega))^2,$$

$$a(u,v) = ((\mathbf{A}\nabla u, \nabla v)), \quad \forall (u,v) \in (H^1(\Omega))^2.$$

The space of discretization, $X_N$, consists of all polynomials of degree $\leq N$ in each variable ($I\!P_N(\Omega)$) that vanish at the boundary

$$X_N = I\!P_N(\Omega) \cap H_0^1(\Omega)$$

The quadrature formula is based on $(M+1)^2$ points derived by a tensor product of the one dimensional Gauss-Lobatto formula, and consists of the evaluation $\sum_{GL,1}\sum_{GL,2}$ where the index 1 and 2 refer to the first and second spatial direction respectively. The discrete problem is then: *Find* $u_N \in X_N$ *such that*

$$a_M(u_N, v_N) = ((f, v_N))_M, \quad \forall v_N \in X_N \tag{46}$$

where we have introduced the following discrete scalar product over $I\!P_N(\Omega)$

$$((\phi,\psi))_M = \sum_{GL,1}\sum_{GL,2} \phi\psi, \quad \forall (\phi,\psi) \in (I\!P_N(\Omega))^2 \tag{47}$$

and

$$a_M(u_N, v_N) = ((\mathbf{A}\nabla u_N, \nabla v_N))_M.$$

It is standard to note that, from (43) and (44), this problem has a unique solution and that in the case where $M = N$ it is equivalent to a collocation problem based on the $(N+1)^2$ Gauss-Lobatto points. The analysis of the error will be based on Lemma 2.3 and on the following complement of this Lemma

**Lemma 2.6.** *Let $\mu$ be a real number $> 1/2$. For any $\psi \in H_K^\mu(-1,1)$, we have*

$$\sup_{v_N \in Y_N} \frac{(\psi, v_N) - (\psi, v_N)_M}{\|v_N\|_{L^2(-1,1)}} \leq C(2M-N-1)^{-\mu} \sum_{k=1}^{K} \|\psi_{|I_k}\|_{H^\mu(I_k)}. \tag{48}$$

**Proof:** This proof is derived from (29) by taking $\tilde{\psi}_{2M-N-1}$ instead of $\tilde{\psi}_{2M-N}$, which has the effect to cancel the middle difference in the right hand side of (29).

**Remark 2.3:** It is an easy matter to note that (48) cannot be improved to yield a bound by the best fit by polynomials of degree $2M-N$ as it was in Lemma 2.3.

We are now in position to give the following error estimate

**Theorem 2.7.** *Let us suppose that the solution $u$ to problem (42) belongs to $H^\sigma(\Omega)$, that the coefficients $(a_{ij})$ of the matrix $\mathbf{A}$ belong to $H^\mu(\Omega)$ and that $f$ belongs to $H^\rho(\Omega)$ and let us suppose in addition that the real numbers $\mu, \sigma - 1, \rho$ are larger than 1. Then the following error estimate holds*

$$\begin{aligned} \|u - u_N\|_{H^1(\Omega)} \leq C[N^{1-\sigma}\|u\|_{H^\sigma(\Omega)} + (2M-N-1)^{-\min(\mu,\sigma-1)} \sum_{i,j=1}^{2} \|a_{ij}u\|_{H^{\min(\mu,\sigma-1)}(\Omega)} \\ + (2M-N)^{-\rho}\|f\|_{H^\rho(\Omega)}]. \end{aligned} \tag{49}$$

**Proof:** The proof is derived from Lemma 2.2 (with obvious changes in the notations) following exactly the same lines as in the previous subsection. The only difference relies on the fact that Lemma 2.6 has to be used in the analysis of the quantity

$$\sup_{v_N \in X_N} \frac{a(u, v_N) - a_M(w_M, v_N)}{\|v_N\|_{H^1(\Omega)}}. \tag{50}$$

Here $w_N$ is a polynomial of degree $N$ (and not $N-1$) suitably chosen to be an optimal approximation of $u$ in the norm $H^\sigma(\Omega)$, say, and in any other lower norm from the bidimensional equivalent to (39) (see [16]). As in (38) we can write

$$\begin{aligned} \sup_{v_N \in X_N} \frac{a(u, v_N) - a_M(w_N, v_N)}{\|v_N\|_{H^1(\Omega)}} \leq \sup_{v_N \in X_N} \frac{a(w_N, v_N) - a_M(w_N, v_N)}{\|v_N\|_{H^1(\Omega)}} \\ + C\|u - w_N\|_{H^1(\Omega)}. \end{aligned}$$

The first term on the left-hand side is a sum of four terms that are similar. The first one reads a follows

$$((a_{11}w_{Nx}, v_{Nx})) - ((a_{11}w_{Nx}, v_{Nx}))_M$$

and can be bounded from Lemma 2.6 as

$$((a_{11}w_{Nx}, v_{Nx})) - ((a_{11}w_{Nx}, v_{Nx}))_M \leq C(2M-N-1)^{-\min(\mu,\sigma-1)}\|a_{11}u_x\|_{H^{\min(\mu,\sigma-1)}(\Omega)}\|v_x\|_{L^2(\Omega)}$$

The reason why (20) is not sufficient to analyze this term is that in the second direction, $v_{Nx}$ is of degree $\leq N$.

It follows easily from the analysis in the one dimensional case that we can prove

$$\sup_{v_N \in X_N} \frac{(f, v_N) - (f, v_N)_M}{\|v_N\|_{H^1(-1,1)}} \leq C(2M-N)^{-\rho}\|f\|_{H^\rho(\Omega)}, \tag{51}$$

which allows to derive Theorem 2.7.

**Corollary 2.8.** *Let us suppose that the hypothesis of theorem 2.7 are fullfilled and let us suppose in addition that $M = N$. Then the following error estimate holds*

$$\|u - u_N\|_{H^1(\Omega)} \leq CN^{-min(\nu,\sigma-1,\rho)}\{\|\mathbf{A}\|_{H^\nu(\Omega)}\|u\|_{H^\sigma(\Omega)} + \|f\|_{H^\rho(\Omega)}\}.$$

**Remark 2.4** — Interpretation of the results: As pointed out in the previous subsection, there is no need to overintegrate in the general case where the solution has the same regularity as the non-constant coefficients. In 2 dimensions this conclusion still holds, however more weakly. Indeed, we have only been able to prove that the plain spectral or spectral element method provides the same accuracy as the best fit by polynomials of degree one less. An overintegration with just one more point (M=N+1) is sufficient to recover the optimal accuracy. Although the results are asymptotically the same, this remark could suggest the use an overintegration. However, the associated increase in work makes it preferable to use a standard spectral element method with polynomial degree of N+1 (instead of N) since this is less expensive than an overintegration with M=N+1 (see appendix). It is only in the case where the spectral method would require much higher values of N that overintegration, coupled with the preservation of the tensor product formula, could be of interest. Finally, note that the numerical experiments again confirm the theoretical statements with an even less important difference between the cases $M = N + 1$ and $M = N$.

## 3. ANALYSIS OF SPECTRAL APPROXIMATION IN DEFORMED GEOMETRIES.

Our purpose is now to extend the previous analysis in the case where the non-constant coefficients are induced from the treatment of a non-rectangular geometry. Since the point here is not to analyze the behaviour of the approximation when original non-constant coefficients are present in the equation, we shall restrict our analysis to the simple problem of a Poisson equation in a deformed domain $\Omega$ which reads as follows: *Find* $u \in H_0^1(\Omega)$ *such that*

$$\iint_\Omega \nabla u \nabla v dx dy = \iint_\Omega f v dx dy \quad \forall v \in H_0^1(\Omega). \tag{52}$$

In order to explain an important issue related to the error analysis, let us first consider the simple case where the geometry is rectilinear but not rectangular.

### 3.1 — Rectilinear Geometries.

Let us suppose that the domain $\Omega$ is the trapezoid with vertices A=(0,0), B=(4,0), C=(0,1), and D=(4-4/a,1), where $a$ is a real number $> 1$. It is an easy matter to check that the transformation $F : (r,s) \mapsto (x,y)$ with

$$\begin{aligned} x &= r(4 - \frac{4}{a}s), \\ y &= s, \end{aligned} \tag{53}$$

is a one to one maping from the square $\hat{\Omega} = (0,1)^2$ onto the domain $\Omega$. We can now define a natural correspondance between functions $\phi$ defined over $\Omega$ and functions $\hat{\phi}$ defined over $\hat{\Omega}$ as follows

$$\hat{\phi}(r,s) = \phi(F(r,s)).$$

This will allow to set the discrete problem. Indeed as we have indicated in the introduction, the spectral metod has to be used on square domains. The other domains have to be mapped on such a reference square in order to allow for a spectral computation (the spectral element method allows to deal with more complex geometries, however, the elements must still be mapped to such a reference square). The variational formulation transferred to the domain $\hat{\Omega}$ is the following: *Find* $\hat{u} \in H_0^1(\hat{\Omega})$ *such that*

$$\iint_{\hat{\Omega}} \frac{1}{J(r,s)}[\frac{\partial \hat{u}}{\partial r}\frac{\partial \hat{v}}{\partial r}\omega_1(r,s) + [\frac{\partial \hat{u}}{\partial r}\frac{\partial \hat{v}}{\partial s} + \frac{\partial \hat{u}}{\partial s}\frac{\partial \hat{v}}{\partial r}]\omega_2(r,s) + \frac{\partial \hat{u}}{\partial r}\frac{\partial \hat{v}}{\partial r}\omega_3(r,s)]drds = \iint_{\hat{\Omega}} \hat{f}\hat{v}J(r,s)drds, \quad \forall \hat{v} \in H_0^1(\hat{\Omega}). \tag{54}$$

The coefficients $J$ and $\omega_i$, $i = 1,2,3$, are respectively the Jacobian of the transformation $F$ and some smooth geometric factor given by

$$J = \frac{4}{a}(a - s), \quad \omega_1 = (1 + \frac{16r^2}{a^2}), \quad \omega_2 = (4 - \frac{4s}{a})(\frac{4r}{a}), \quad \omega_3 = (4 - \frac{4s}{a})^2. \tag{55}$$

The discretization of the problem can now proceed and consists in the following: *Find* $\hat{u}_N \in I\!P_N(\hat{\Omega}) \cap H_0^1(\hat{\Omega})$ *such that*

$$a_M(\hat{u}_N, \hat{v}_N) = ((J\hat{f}, \hat{v}_N))_M, \quad \forall \hat{v} \in I\!P_N(\hat{\Omega}) \tag{56}$$

where $((.,.))_M$ has been defined in (47) and

$$\begin{aligned} a_M(\hat{u}_N, \hat{v}_N) = &((\frac{\omega_1}{J}\frac{\partial \hat{u}_N}{\partial r}, \frac{\partial \hat{v}_N}{\partial r}))_M + ((\frac{\omega_2}{J}\frac{\partial \hat{u}_N}{\partial r}, \frac{\partial \hat{v}_N}{\partial s}))_M \\ &+ ((\frac{\omega_2}{J}\frac{\partial \hat{u}_N}{\partial s}, \frac{\partial \hat{v}_N}{\partial r}))_M + ((\frac{\omega_3}{J}\frac{\partial \hat{u}_N}{\partial s}, \frac{\partial \hat{v}_N}{\partial s}))_M. \end{aligned}$$

The coefficients in the right-hand side of (56) are extremely regular. We can deduce from the previous analysis that no overintegration is required to compute accurately this term in order to achieve optimal

approximation. On the contrary, the coefficients that arise in $a_M$ on the left-hand side, although analytic, may give rise to problem since the factor $\omega_i/J$ has a pole where $J$ vanishes, i.e. at $s = a$. This pole is not in $\hat{\Omega}$ but can be very close to it. The similarity of this problem with the one treated in the second section (Figure 2.2) seems to imply that, when $a$ is sufficently close to 1, an overintegration is required to compute correctly $a_M$ in order to get an optimal aproximation. However, surprisingly, the experiments performed in Figure 3.1 shows that no overintegration is required, even with those values of $a$ that in the one-dimensional case induced a big error associated with the consistency error (50). The interpretation of this fact can be easily understood, a posteriori, as the pole of $1/J$ is compensated by the fact that the other part of the integrand vanishes and makes the term

$$\frac{\omega_1}{J}\frac{\partial \hat{u}}{\partial r} + \frac{\omega_2}{J}\frac{\partial \hat{u}}{\partial r} + \dots$$

more regular than it appears. In order to be more precise, let us state

**Lemma 3.1.** *Let us assume that the solution $u$ to problem (52) belongs to $H^\sigma(\Omega)$ for some real number $\sigma > 2$. Then we have*

$$\sup_{\hat{v}_N \in \mathbb{P}_N(\hat{\Omega})} \frac{a(\hat{u},\hat{v}_N) - a_M(\hat{w}_N,\hat{v}_N)}{\|\hat{v}_N\|_{H^1(\hat{\Omega})}} \leq C[\|\hat{u}-\hat{w}_N\|_{H^1(\hat{\Omega})} + N^{-1}\|\hat{u}-\hat{w}_N\|_{H^2(\hat{\Omega})} + (2M-N-1)^{1-\sigma}[\|\widehat{\frac{\partial u}{\partial x}}\|^2_{H^{\sigma-1}(\hat{\Omega})} + \|\widehat{\frac{\partial u}{\partial y}}\|^2_{H^{\sigma-1}(\hat{\Omega})}]. \tag{57}$$

**Proof:** Let us go back to the original equation (52) and remark that

$$\iint_\Omega \nabla u \nabla v dx dy = \iint_{\hat{\Omega}} [\widehat{\frac{\partial u}{\partial x}}(r,s)(\frac{\partial \hat{v}_N}{\partial r})(r,s) + \widehat{\frac{\partial u}{\partial y}}(r,s)[(\frac{\partial \hat{v}_N}{\partial r}(r,s)(\frac{4r}{a}) + \frac{\partial \hat{v}_N}{\partial s}(r,s)(4-\frac{4}{a}s)]]drds$$

and that a simple change of variable formula yields

$$\begin{aligned}
((\widehat{\frac{\partial u}{\partial x}}, \frac{\partial \hat{v}_N}{\partial r}))_M &= ((\frac{1}{J}\frac{\partial \hat{u}}{\partial r}, \frac{\partial \hat{v}_N}{\partial r}))_M \\
(((\frac{4r}{a})\widehat{\frac{\partial u}{\partial y}}, \frac{\partial \hat{v}_N}{\partial r}))_M &= (((\frac{\omega_1 - 1}{J})\frac{\partial \hat{u}}{\partial r}, \frac{\partial \hat{v}_N}{\partial r}))_M + ((\frac{\omega_2}{J}\frac{\partial \hat{u}}{\partial s}, \frac{\partial \hat{v}_N}{\partial r}))_M \\
(((4-\frac{4}{a}s)\widehat{\frac{\partial u}{\partial y}}, \frac{\partial \hat{v}_N}{\partial s}))_M &= ((\frac{\omega_2}{J}\frac{\partial \hat{u}}{\partial r}, \frac{\partial \hat{v}_N}{\partial s}))_M + ((\frac{\omega_3}{J}\frac{\partial \hat{u}}{\partial s}, \frac{\partial \hat{v}_N}{\partial s}))_M
\end{aligned}$$

so that

$$a(u,v_N) - a_M(\hat{w}_N, v_N) = A_1 + B_1 + A_2 + B_{2,1} + B_{2,2} + A_3 + B_{3,1} + B_{3,2} \tag{58}$$

where

$$\begin{aligned}
A_1 &= \iint_{\hat{\Omega}} \widehat{\frac{\partial u}{\partial x}}(r,s)(\frac{\partial \hat{v}_N}{\partial r})(r,s)drds - ((\widehat{\frac{\partial u}{\partial x}}, \frac{\partial \hat{v}_N}{\partial r}))_M \\
B_1 &= ((\frac{1}{J}\frac{\partial \hat{u}}{\partial r}, \frac{\partial \hat{v}_N}{\partial r}))_M - ((\frac{1}{J}\frac{\partial \hat{w}_N}{\partial r}, \frac{\partial \hat{v}_N}{\partial r}))_M \\
A_2 &= \iint_{\hat{\Omega}} \widehat{\frac{\partial u}{\partial y}}(r,s)(\frac{\partial \hat{v}_N}{\partial r}(r,s)(\frac{4r}{a}))drds - (((-\frac{4r}{a})\widehat{\frac{\partial u}{\partial y}}, \frac{\partial \hat{v}_N}{\partial r}))_M \\
B_{2,1} &= (((\frac{\omega_1 - 1}{J})\frac{\partial \hat{u}}{\partial r}, \frac{\partial \hat{v}_N}{\partial r}))_M - (((\frac{\omega_1 - 1}{J})\frac{\partial \hat{w}_N}{\partial r}, \frac{\partial \hat{v}_N}{\partial r}))_M \\
B_{2,2} &= (((\frac{\omega_2}{J})\frac{\partial \hat{u}}{\partial s}, \frac{\partial \hat{v}_N}{\partial r}))_M - (((\frac{\omega_2}{J})\frac{\partial \hat{w}_N}{\partial s}, \frac{\partial \hat{v}_N}{\partial r}))_M
\end{aligned}$$

$$A_3 = \iint_{\hat\Omega} \widehat{\frac{\partial u}{\partial y}}(r,s)(\frac{\partial \hat v_N}{\partial s}(r,s)(4-\frac{4}{a}s)drds - (((4-\frac{4}{a}s)\widehat{\frac{\partial u}{\partial y}}, \frac{\partial \hat v_N}{\partial s}))_M$$

$$B_{3,1} = (((\frac{\omega_2}{J})\frac{\partial \hat u}{\partial r}, \frac{\partial \hat v_N}{\partial s}))_M - (((\frac{\omega_2}{J})\frac{\partial \hat w_N}{\partial r}, \frac{\partial \hat v_N}{\partial s}))_M$$

$$B_{3,2} = (((\frac{\omega_3}{J})\frac{\partial \hat u}{\partial s}, \frac{\partial \hat v_N}{\partial s}))_M - (((\frac{\omega_3}{J})\frac{\partial \hat w_N}{\partial s}, \frac{\partial \hat v_N}{\partial s}))_M.$$

In the decomposition (58), the terms $A_-$ are small from the accuracy of the quadrature formula; the terms $B_-$ are small since $\hat w_N$ is a polynomial that is close to $\hat u$. More precisely, let us denote by $\hat C$ the supremum of $|1/J|$, $|(1-\omega_1)/J|$, $|\omega_2/J|$, $|\omega_3/J|$, on $\hat\Omega$. It is an easy matter to derive that, for any index $\alpha$, where $\alpha = 1,(2,1),(2,2),(3,1),(3,2)$ we have

$$|B_\alpha| \le \hat C \|I_M \frac{\partial}{\partial t}(\hat u - \hat w_N)\|_{L^2(\hat\Omega)} \|\hat v_N\|_{H^1(\hat\Omega)},$$

where $t$ in the derivation stands for $s$ or $r$. Using (33), with $\mu = 1$, we derive

$$\begin{aligned} |B_\alpha| &\le \hat C[\|\hat u - \hat w_N\|_{H^1(\hat\Omega)} + N^{-1}\|\frac{\partial}{\partial t}(\hat u - \hat w_N)\|_{H^1(\hat\Omega)}]\|\hat v_N\|_{H^1(\hat\Omega)}, \\ &\le C[\|\hat u - \hat w_N\|_{H^1(\hat\Omega)} + N^{-1}\|\hat u - \hat w_N\|_{H^2(\hat\Omega)}]\|\hat v_N\|_{H^1(\hat\Omega)}. \end{aligned} \tag{59}$$

Let us denote now by $\hat c$ the supremum of 1, $|(4-4s/a)|$, $|4r/a|$, on $\hat\Omega$. We derive as in section 2.2 that

$$|A_i| \le \hat c(2M-N-1)^{1-\sigma}[\|\widehat{\frac{\partial u}{\partial x}}\|_{H^{\sigma-1}(\hat\Omega)} + \|\widehat{\frac{\partial u}{\partial y}}\|_{H^{\sigma-1}(\hat\Omega)}]\|\hat v_N\|_{H^1(\hat\Omega)} \tag{60}$$

and the Lemma follows from (58), (59) and (60).

In order to analyse the error in the approximation of problem (52) with (56), we first note that (16) (with obvious changes of notations) also holds in this case. This leads to the following:

**Theorem 3.2.** *Let us suppose that the solution $u$ of problem (52) belongs to $H^\sigma(\Omega)$ and that $f$ belongs to $H^\rho(\Omega)$ and let us suppose in addition that the real numbers $\sigma - 1, \rho$ are larger than 1. Then the following inequality holds*

$$\begin{aligned} \|\hat u - \hat u_N\|_{H^1(\hat\Omega)} \le C[N^{1-\sigma}\|u\|_{H^\sigma(\Omega)} &+ C(2M-N-1)^{1-\sigma}\|u\|_{H^\sigma(\Omega)} \\ &+ C(2M-N)^{-\rho}\|f\|_{H^\rho(\Omega)} \end{aligned} \tag{61}$$

**Proof**: This is an easy consequence of (48) and of Lemma 3.1, if we choose $\hat w_N$ as being the best fit of $\hat u$ in $H^\sigma(\hat\Omega)$ by elements of $\mathbb{P}_N(\hat\Omega)$.

3.2 — General deformations.

Let us suppose in this subsection that the domain $\Omega$ is the image of $\hat\Omega$ by some mapping $F : (r,s) \mapsto (x = F_1(r,s), y = F_2(r,s))$. We no longer assume that $F$ is bilinear, nevertheless we assume it has some regularity, more precisely, that it is in $H^\mu(\hat\Omega)$ for some $\mu > 1$, and that its Jacobian $J$ is larger than some $p_1 > 0$. The problem (52) is transferred onto $\hat\Omega$ as before and reads :

*Find $\hat u \in H^1_0(\hat\Omega)$ such that (54) is satisfied,*

where the geometric factors $\omega_i$, $i = 1,2,3$ are defined as follows

$$\begin{aligned} \omega_1 &= (\frac{\partial F_1}{\partial s})^2 + (\frac{\partial F_2}{\partial s})^2 \\ \omega_2 &= -[\frac{\partial F_1}{\partial s}\frac{\partial F_1}{\partial r} + \frac{\partial F_2}{\partial s}\frac{\partial F_2}{\partial r}] \\ \omega_3 &= (\frac{\partial F_1}{\partial r})^2 + (\frac{\partial F_2}{\partial r})^2 \end{aligned} \tag{62}$$

The discrete problem is now:

*Find $\hat u_N \in \mathbb{P}_N(\hat\Omega) \cap H^1_0(\hat\Omega)$ such that (56) is satisfied.*

Using here the same technique as in the proof of Lemma 3.1, and theorem 2.4, we obtain

**Lemma 3.3.** *Let us suppose that the solution $u$ to problem (52) belongs to $H^\sigma(\Omega)$, for some real number $\sigma > 2$, that the deformation mapping $F$ belongs to $H^\mu(\hat{\Omega})$ for some real number $\mu > 1$. Then the following inequality holds*

$$\sup_{\hat{v}_N \in \mathbb{P}_N(\hat{\Omega})} \frac{a(\hat{u},\hat{v}_N) - a_M(\hat{w}_N,\hat{v}_N)}{\|\hat{v}_N\|_{H^1(\hat{\Omega})}} \leq C[\|\hat{u}-\hat{w}_N\|_{H^1(\hat{\Omega})} + N^{-1}\|\hat{u}-\hat{w}_N\|_{H^2(\hat{\Omega})} + (2M-N-1)^{-min(\mu,\sigma-1)}[\|F\|_{H^\mu(\hat{\Omega})}\|\widehat{\frac{\partial u}{\partial x}}\|^2_{H^{min(\mu,\sigma-1)}(\hat{\Omega})} + \|\widehat{\frac{\partial u}{\partial y}}\|^2_{H^{min(\mu,\sigma-1)}(\hat{\Omega})}]] \tag{64}$$

**Remark 3.2:** It is important, at this stage, to note that the regularity of $\hat{u}$ and the regularity of $\widehat{\frac{\partial u}{\partial t}}$, where $t$ is either $s$ or $r$ is bounded by the regularity of both $u$ and $F$, which means that

$$\inf_{\hat{w}_N \in \mathbb{P}_N(\hat{\Omega})} \|\hat{u}-\hat{w}_N\|_{H^1(\hat{\Omega})} \leq CN^{-min(\mu-1,\sigma-1)}\|F\|_{H^\mu(\hat{\Omega})}\|u\|_{H^{min(\mu,\sigma)}(\hat{\Omega})} \tag{65}$$

is the best result we can expect. Note also (but this is natural after the analysis of section 3.1) that it is only the regularity of $F$ that is important and not at all the regularity of its inverse (as one could previously fear from the factor $1/J$).

¿From Lemma 3.3 and (65), we derive in a (now) standard manner the

**Theorem 3.4.** *Let us suppose that the solution $u$ belongs to problem (52), belongs to $H^\sigma(\Omega)$, for some real number $\sigma > 2$, that the deformation mapping $F$ belongs to $H^\mu(\hat{\Omega})$ for some real number $\mu > 1$ and that the forcing function $f$ belongs to $H^\rho(\Omega)$ for some real number $\rho > 1$. Then the following inequality holds*

$$\|\hat{u}-\hat{u}_N\|_{H^1(\hat{\Omega})} \leq C\{N^{-min(\mu-1,\sigma-1)}\|F\|_{H^\mu(\hat{\Omega})}\|u\|_{H^{min(\mu,\sigma)}(\Omega)} + (2M-N-1)^{-min(\mu,\sigma-1)}\|F\|_{H^\mu(\hat{\Omega})}\|u\|_{H^{min(\mu,\sigma-1)}(\Omega)} + (2M-N)^{-min(\mu,\rho)}\|F\|_{H^\mu(\hat{\Omega})}\|f\|_{H^{min(\mu,\rho)}(\Omega)}\} \tag{66}$$

**Remark 3.3:** In the light of the previous theorem, it appears that there is no need to overintegrate in the case where the deformation is responsible for non-constant coefficients in the formulation of the discrete problem, even if the geometry is very distorted. This is readily seen by comparing the results obtained in the theorem by choosing $M >> N$ and $M = N$ (the gain of just one degree is not, as explained in the remark 2.4 sufficient to start the overintegration machinery). We can explain it a posteriori now as overintegration could allow to improve part of the consistency error term (57) in the case where the geometry mapping $F$ is less regular then the solution $u$ itself ( contrary to the case in section 2, this can happen, especially if domain decomposition is used). However, as pointed out in the previous remark, loss of regularity of $F$ induces also a loss of regularity of $\hat{u}$ which is the function that is of importance since it is this one that is approximated by polynomials.

**Remark 3.4:** We have not considered here the case where the geometry mapping $F$ itself is approximated by isoparametric polynomials. This effect can easily be analysed by following the same lines as previously and including the standard arguments adapted to the finite element discretization [18]

APPENDIX (in collaboration with A. T. PATERA.)

In this appendix we consider the computational cost of using overintegration. We show that the matrix-vector products can still be evaluated efficiently using tensor product forms, and that for a fixed discretization (polynomial degree $N$), the increase in operation count is at least a factor of $d$ in $\mathbb{R}^d$. Overintegration is attractive to use only if the convergence rate (error as a function of polynomial degree) improves substantially. In this case we can achieve a fixed accuracy either with a low polynomial degree and overintegration, or a higher polynomial degree with consistent quadrature.

With no loss of generality we consider here the evaluation of the left hand side of (46),

$$a_M(u_N, v_N) = ((p\nabla u_N \nabla v_N))_M \tag{67}$$

Here $p$ represents a non-constant coefficient which either comes from the original strong formulation (as in (42)), or represents a geometric factor in the case of deformed geometry, or both. The evaluation of (67) represents the computationally most expensive part in an iterative solution of (46).

In order to implement (46) we require a basis for our high-order polynomial space $X_N$. The choice of basis does not effect the error estimates, however it greatly effects the conditioning and sparsity of the resulting set of algebraic equations, and is critical for the efficiency of parallel iterative solution procedures. We choose an interpolant basis to represent $w_N \in X_N$,

$$w_N(r,s) = \sum_{p=0}^{N}\sum_{q=0}^{N} w_{pq} h_p(r) h_q(s), \tag{68}$$

where $(x,y) \in \Omega \rightarrow (r,s) \in ]-1,1[^2$. Here the $h_p(z)$ are the one-dimensional $N^{th}$-order Lagrangian interpolants through the Gauss-Lobatto Legendre points $\xi_p$ ($h_p \in \mathbb{P}_N(]-1,1[), h_p(\xi_q) = \delta_{pq}$), and $w_{pq}$ is the value of $w_N$ at the local node $(\xi_p, \xi_q)$. In addition to (68) we require $w_N \in X_N$ to honor the homogeneous Dirichlet boundary conditions.

The bases (68) are now inserted into (67), and the test functions are systematically chosen to be unity at one global node and zero at all the other Gauss-Lobatto Legendre points. We then arrive at the following discrete statement,

$$a_M(u_N, v_N) = \sum_{\alpha=0}^{M}\sum_{\beta=0}^{M}\sum_{m=0}^{N}\sum_{n=0}^{N} \rho_\alpha \rho_\beta p_{\alpha\beta} D_{\alpha i} I_{\beta j} D_{\alpha m} I_{\beta n} u_{mn} + \ldots\ldots \tag{69}$$

where we for convenience only have included the derivatives with respect to the first (local) spatial direction. Here $\rho_p$ are the one-dimensional quadrature weights, and $D$ and $I$ are the one-dimensional derivative and interpolation operators defined as:

$$D_{pq} = \frac{dh_q}{d\xi}(\xi_p), \tag{70}$$

$$I_{pq} = h_q(\xi_p). \tag{71}$$

Using tensor-product sum-factorization techniques [5] we can now evaluate (69) efficiently as follows:

$$(D_{\alpha i}(I_{\beta j}(\rho_\alpha \rho_\beta (p_{\alpha\beta}(D_{\alpha m}(I_{\beta n} u_{mn})))))),$$

where the expression in the innermost partenthesis is evaluated first. An evaluation in this order results in the following operation count,

$$MN^2 + M^2N + M^2 + M^2 + M^2N + MN^2 = 2(M^2N + MN^2 + M^2).$$

In the case where $M = N$ (no overintegration), the interpolation operator becomes $I_{pq} = \delta_{pq}$ and (69) reduces to the following expression,

$$a_N(u_N, v_N) = \sum_{\alpha=0}^{M}\sum_{m=0}^{N} \rho_\alpha \rho_j p_{\alpha j} D_{\alpha i} D_{\alpha m} u_{mj} + \ldots.$$

which can be efficiently evaluated in $2(N^3 + N^2)$ operations.

We are now in a position to compare the computational cost with and without overintegration. In the case $M = N+1$ the operation count is approximately twice the cost of using consistent quadrature ($M = N$) (in three dimensions the cost would increase with a factor of three). If $M = 2N$, say, the cost to evaluate (67) is four times the cost with $M = N$. The factors which determine whether overintegration is economical are the the difference in convergence rate with and without overintegration, and the specified accuracy of the discrete solution.

## FIGURES

Figure 2.1. A plot of various numerical results relevant to the approximation of the one dimensional problem (1), in order to analyze the effect of overintegration. We have considered here the case where $q = 0$ and different values of $p$ and $f$ are simulated and the effect of overintegration is adressed. The relation between $p$ and $f$ are always chosen in such a way that the exact solution is $u = sin(3\pi x)e^{-x}$ (of course, we do not use the knowledge of the exact solution in the simulations !!). The discretisation is based on $K = 4$ equal spectral elements for different values of the polynomial degree, N. First of all, since this is the aim of the numerical scheme, we have computed the best fit of $u$ in $H_0^1(0,1)$ by piecewise polynomials of degree $\leq N$ (plot ▲); the plot proves a convergence better then exponential that is consistant with the fact that the solution is entire. Next we consider the case $p = 1$ with no overintegration (plot ○) and we can check that the approximation is optimal and very close to the best fit.

We then choose $p = \frac{1}{x+a}$ for various values of $a$. The case where $a = .25$, treated with no overintegration, is represented on the plot ⊕ and we can remark that the approximation is still very good. The convergence rate does not however follow the best fit of $u$, but the best fit of $p$ (consistently with the corollary 2.5) that is a straight line in this scale since $p$ has a pole outside (0,1). Let us now use a more singular $p$, corresponding to $a = .05$. The plot ⊙ represents the results when no overintegration in used. It is very close to three other plots representing the results when overintegration ($M = 3N/2$) is used to compute only the right-hand side (the contribution of $f$) (□) or only the left-hand side (corresponding to $p$) (plot △), while the third one represents the best fit of $p$ (plot •) in the $L^2(0,1)$-norm by the discrete functions. In order to recover optimal results, overintegration must be used on both the left and the right-hand side of the equation (plot ◇).

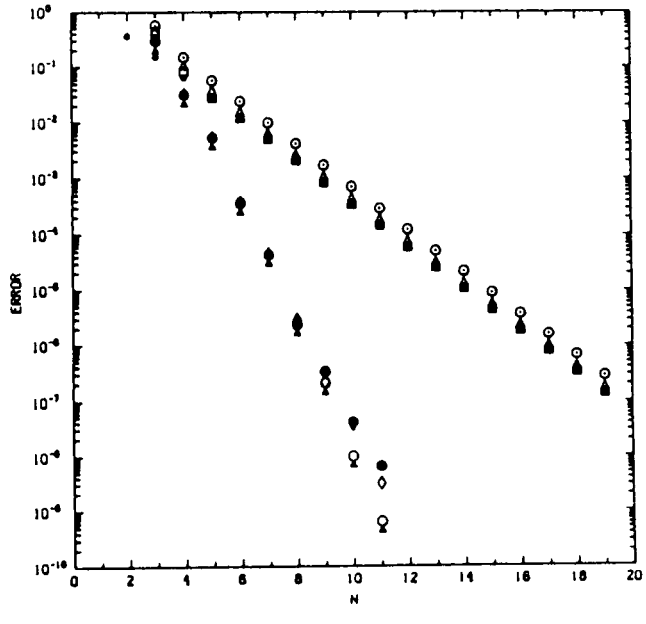
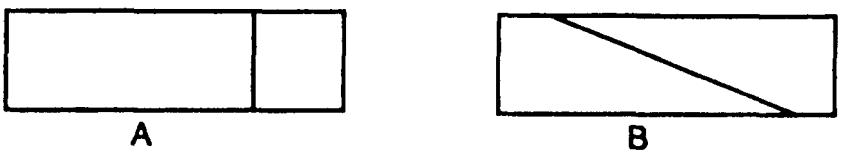



Figure 2.1

Figure 2.2. This Figure represents the equivalent of the previous case in a two dimensional domain. The equation is here (42) with $\mathbf{A} = p\mathrm{Id}$ and different values of $p$ and $f$, computed so that the solution is always $u = sin(\pi x)y(1 - e^{y-1})$. The domain of computation $\Omega$ is $(0,1)^2$ and decomposed in four squares. The plot • represents the case $p = 1$ with $M = N$, $M = N + 1$, $M = 3N$ (there is no difference between the two last experiments and the first case is a bit less precise but cannot really be distinguished from the two last cases). The plot ∘ represents the experiment with $p = e^{x+y}$ that is a very smooth function and a treatment with no overintegration gives the same accuracy as the previous plot (that certainly also corresponds to the best fit). Now comes the treatment of the case $p = 1 + cos(2\pi x + 3\pi y)$ which is a smooth function but is worse than $u$. The plot ▲ corresponds to no overintegration and we clealy see that some degree of overintegration is necessary to recover the optimal accuracy, as shown in the case where $M = N + 1$ (plot □) and $M = 3N$ (plot △).

Figure 3.1. On this plot we prove that the singularity of the function $\frac{1}{\text{Jacobian}}$ is not relevant for the need of overintegration. We consider the problem is (52) for two different domain decompositions A and B. They are




The three plots represent the discretisation error in the semi-norm with no overintegration. The plot ∘ corresponds to the solution $u = sin(2y)$ for the two cases A and B. They both coincide since the deformation is in the $x$–direction and the solution only depends on $y$. Next, we repeat the experiments with $u = sin(\frac{x}{3})$ and we arrive to the (a priori) surprising result: plot △ for A, and □ for B. We refer to section 3.1 for the explanation. The fact that B is even better in this case is due to the fact that the discretisation on the two domains is more equilibrated and that the left domain in case A is too long.

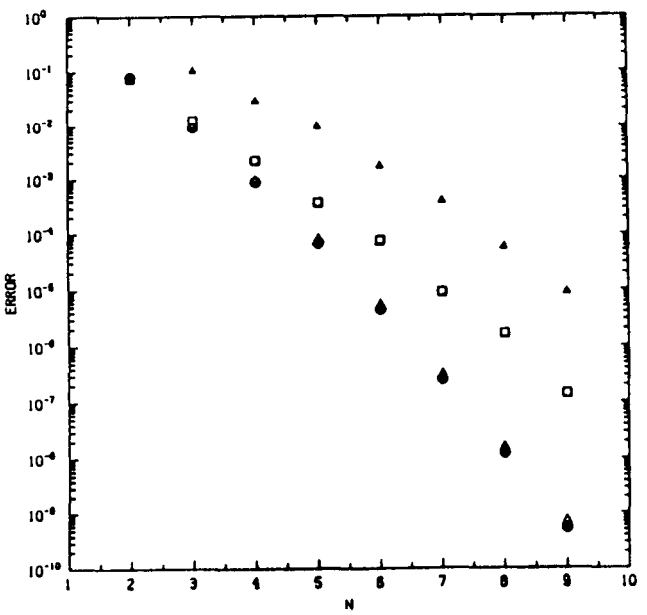



Figure 2.2

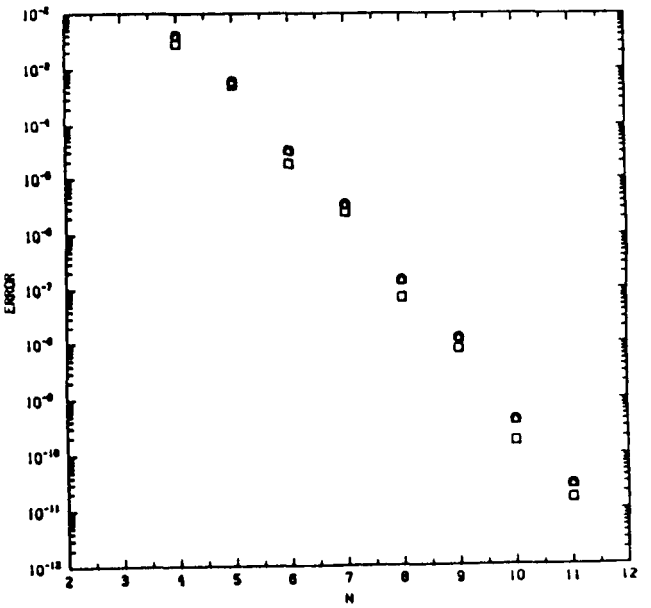
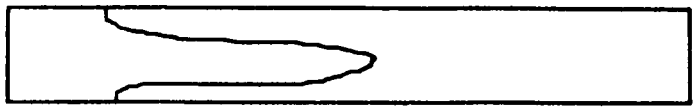


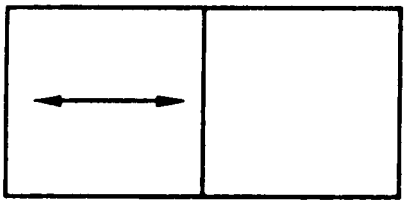

Figure 3.1

Figure 3.2. This plot is to prove that the overintegration can even have a negative effect. Again we consider equation (52) for $p = 1$, with a decomposition of $\Omega = [0, 7] \times [0, 1]$ into two subdomains:



The interface is described by the equation $x = \frac{7}{2} - \frac{5}{2}|y|^3$. This equation is not regular and its interpolant is more rough when the degree is higher. Since this interpolant is the only geometric factor for the discrete problem we can understand the plots $\Delta$ and $\blacktriangle$ that correspond to $M = N$ and $M = 2N$, respectively, for the approximation of the solution $u = sin(\frac{x}{3})$. As before, if the deformation and the solution are in different directions (this is a very particular case !!!), for example $u = sin2y$, the behaviour of the approximation is particular also and this is the only case where overintegration can help (in the case of deformed geometries) as is illustrated in the plots $\circ$ and $\bullet$ corresponding to $M = N$ and $M = 2N$ respectively.

Figure 3.3. A plot of the discretisation error in the semi-norm when solving (52) with $p = 0$. The domain $\Omega = [0, 2] \times [0, 1]$ is decomposed into two subdomains:



and this time the geometry is tortured in the following way. The subdomain on the left is mapped onto the reference square by the mapping $x = r + 0.5(|r| - 1)r^2$, while the mapping is affine in the other subdomain. The plot $\Delta$ correspond to a solution $u = sin(\frac{x}{3})$ that only depends on $x$ and is treated both with $M = N$ and $M = 2N$. The plot $\circ$ is corresponds again to a particular case as it is related to $u = sin(2y)$ in both cases $M = N$ and $M = 2N$.

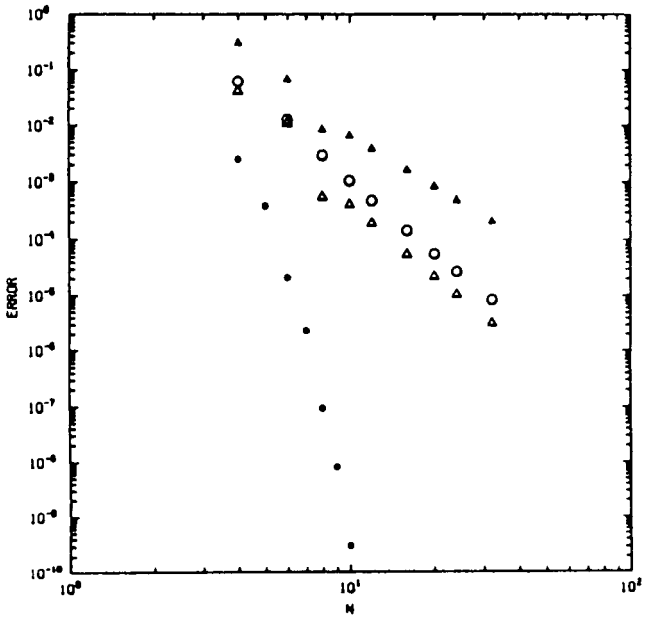



Figure 3.2

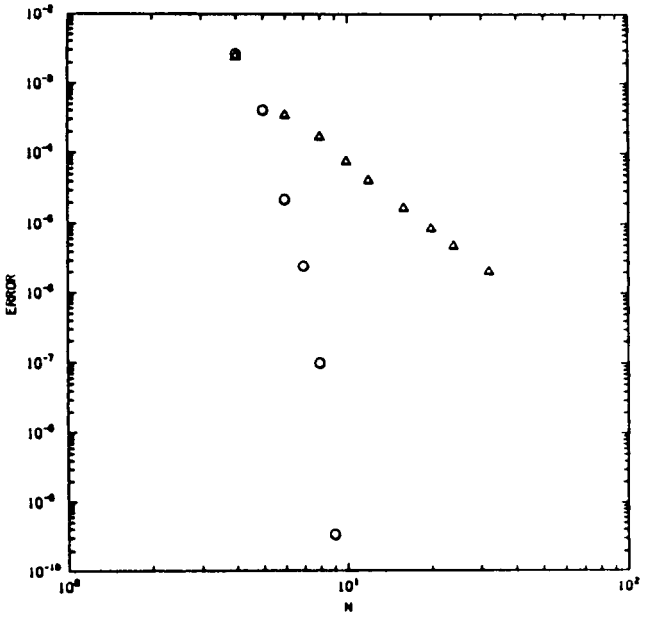



Figure 3.3

## REFERENCES


[1] D. GOTTLIEB, The Stability of Pseudospectral Chebyshev Methods, Math. Comput. 36 (1981) 107–118.

[2] B. MERCIER, Stabilité des Méthodes Spectrales polynômiales application à l'équation d'advection, R.A.I.R.O. Anal Num. 16 (1982) 67–100.

[3] Y. MADAY and A. QUARTERONI, Approximation of Burgers equations by Pseudospectral Methods, R.A.I.R.O., Num. Anal. 16 (1982) 375–404.

[4] C. CANUTO and A. QUARTERONI, Approximation Results for Orthogonal Polynomials in Sobolev Spaces, Math. Comput. 38 (1982) 67–86.

[5] S. A. ORSZAG, Spectral Methods for Problems in Complex Geometries, J. Comp. Phys. 37 (1980) 70–92.

[6] Y. MORCHOISNE, Résolution des Équations de Navier Stokes par une Méthode Spectrale de Sous Domaines. in: Proc. 3rd Int. Conf. Num. Meth. Sci. Eng. (GAMNI, PARIS 1983)

[7] B. MÉTIVET, Résolution Spectrales des Équations de Navier-Stokes par une Méthode de Sous-Domaines Courbes, Thèse de Doctorat, Université Pierre et Marie CURIE, (1987).

[8] C. CANUTO and D. FUNARO, The Schwarz Algorithm for Spectral Methods, SIAM J. of Numer. Anal. 25(1)(1988) 24–40.

[9] A. T. PATERA, A Spectral Element Method for Fluid Dynamics: Laminar Flow in a Channel Expension, J. Comput. Phys. 54 (1984) 468–488.

[10] Y. MADAY and A. T. PATERA, Spectral Element Methods for the Incompressible Navier Stokes equations, in: A. K. NOOR, J. T. ODEN eds., State of the Art Survey in Computational Mechanics (1989) 71–143.

[11] E. M. RØNQUIST, Optimal Spectral Element Methods for the Unsteady Three-Dimensional Incompressible Navier-Stokes Equations, PhD thesis, MIT (1988).

[12] C. BERNARDI, Y. MADAY and A. T. PATERA, A New Nonconforming Approach to Domain Decomposition: The mortar Element Method. in: H. Brezis and J. L. LIONS, eds., Nonlinear Partial Differential Equations and their Applications (1989).

[13] Y. MADAY, C. MAVRIPLIS and A. T. PATERA, Non Conforming Mortar Element Methods: Application to Spectral Discretizations, in: T. CHAN, ed., Proceedings of the 7th International Conference on Domain Decomposition Techniques for P.D.E., (SIAM Philadelphia 1988).

[14] C. MAVRIPLIS, Nonconforming Discretizations and a Posteriori Error Estimators for Adaptative Spectral Element Techniques, PhD thesis, MIT (1989).

[15] D. FUNARO, A Multidomain Spectral Approximation of Elliptic Equations, Numer Meth. in P.D.E. 2 (1986) 187–202.

[16] Y. MADAY, Résultats d'Approximations des Opérateurs d'Interpolation aux Points de Type Gauss, submitted to Notes aux Comptes rendus de l'Académie des Sciences de Paris.

[17] C. BERNARDI and Y. MADAY, Non Conforming Spectral Element Methods, Analysis of some Projection Operators, to appear in Journal of Applied Numerical Mathematics (1989).

[18] P. G. CIARLET and P. A. RAVIART, The Combined Effect of Curved Boundaries and Numerical Integration in Isoparametric Finite Element Methods, in: A. K. AZIZ, ed., The Mathematical Foundation of the Finite Element Method with Applications to Partial Differential Equations (Academic Press, New York, 1970) 409–474.

NASA
National Aeronautics and Space Administration

# Report Documentation Page

| 1. Report No. | 2. Government Accession No. | 3. Recipient's Catalog No. |
|---|---|---|
| NASA CR-181902<br>ICASE Report No. 89-56 | | |

**4. Title and Subtitle**

OPTIMAL ERROR ANALYSIS OF SPECTRAL METHODS WITH EMPHASIS ON NON-CONSTANT COEFFICIENTS AND DEFORMED GEOMETRIES

**5. Report Date**

September 1989

**6. Performing Organization Code**

**7. Author(s)**

Yvon Maday
Einar M. Rønquist

**8. Performing Organization Report No.**

89-56

**10. Work Unit No.**

505-90-21-01

**9. Performing Organization Name and Address**

Institute for Computer Applications in Science and Engineering
Mail Stop 132C, NASA Langley Research Center
Hampton, VA 23665-5225

**11. Contract or Grant No.**

NAS1-18605

**13. Type of Report and Period Covered**

Contractor Report

**12. Sponsoring Agency Name and Address**

National Aeronautics and Space Administration
Langley Research Center
Hampton, VA 23665-5225

**14. Sponsoring Agency Code**

**15. Supplementary Notes**

Proceedings of ICOSAMOM Meeting 1989

Langley Technical Monitor:
Richard W. Barnwell

Final Report

**16. Abstract**

In this paper we present the numerical analysis of spectral methods when non-constant coefficients appear in the equation, either due to the original statement of the equations or to take into account the deformed geometry. A particular attention is devoted to the optimality of the discretization even for low values of the discretization parameter. The effect of some "overintegration" is also addressed, in order to possibly improve the accuracy of the discretization.

| 17. Key Words (Suggested by Author(s)) | 18. Distribution Statement |
|---|---|
| spectral methods, spectral element methods, nonconstat coefficients, quadrature formula | 64 - Numerical Analysis<br><br>Unclassified - Unlimited |

| 19. Security Classif. (of this report) | 20. Security Classif. (of this page) | 21. No. of pages | 22. Price |
|---|---|---|---|
| Unclassified | Unclassified | 23 | A03 |

**NASA FORM 1626** OCT 86