European Conference on Computational Fluid Dynamics
ECCOMAS CFD 2006
P. Wesseling, E. Oñate and J. Périaux (Eds)
© TU Delft, The Netherlands, 2006

ERROR CONVERGENCE OF THE LEAST-SQUARES SPECTRAL ELEMENT FORMULATION OF THE LINEAR ADVECTION-REACTION EQUATION

W.R. van Dalen* **and M.I. Gerritsma**†

*Delft University of Technology, Department of Aerospace Engineering,
Kluyverweg 1, 2629 HS Delft, The Netherlands
e-mail: W.R.vanDalen@student.tudelft.nl
†Delft University of Technology, Department of Aerospace Engineering,
Kluyverweg 1, 2629 HS Delft, The Netherlands
e-mail: M.I.Gerritsma@lr.tudelft.nl

Key words: Least-Squares formulation, spectral methods, Galerkin formulation, advection-reaction equation

Abstract. *This paper discusses the use of the Least-Squares Spectral Element Method in solving the linear, 1-dimensional advection-reaction equation. Well-posedness of the Least-Squares formulation will be derived. The formulation and its results will be compared to the standard Galerkin Spectral Element Method.*

1 INTRODUCTION

The Least-Squares formulation provides an interesting alternative to the conventional Galerkin formulation in combination with spectral elements. The advantage of Least-Squares (LS) is that the resulting system for well-posed problems is symmetric, positive definite (SPD), which makes the resulting algebraic systems to be solved amenable to well established direct and iterative solvers.[1–3] Furthermore, the Least-Squares formulation has proven to be inherently stable, thus avoiding the necessity to introduce stabilization terms such as upwinding, artificial diffusion or limiting.[4]

The combination of the Least-Squares formulation with spectral elements was first proposed by Proot en Gerritsma[7] en Pontaza and Reddy.[8] The method has been applied to a variety of flow problem and structural analysis.[5,6,9] However, the mechanism by which the Least-Squares formulation acts on the differential operator still requires attention.

It has been shown that the Galerkin and Least-Squares behave significantly different for linear advection equations; Galerkin is sup-optimal depending on the fact whether the polynomial degree employed is even or odd, whereas the Least-Squares formulation always yields optimal convergence behavior.[10]

A similar analysis is discussed in this paper for the linear advection-reaction equation. This equation plays an important role in hyperbolic systems and in the assessment of polynomial chaos[12] and is sufficiently simple to analyze in closed form.

The outline of this paper is as follows: in Section 2, the basic principles of the Least-Squares Spectral Element Method (LS-SEM) will be explained, as well as the inclusion of boundary conditions. In Section 3, the well-posedness of the linear advection-reaction equation will be demonstrated. In Section 4, a direct comparison between the behavior of the Galerkin formulation and the Least-Squares formulation will be presented. Furthermore, some characteristics of strongly and weakly imposed boundary conditions will be discussed.

2 THE LEAST-SQUARES SPECTRAL ELEMENT METHOD

As a starting point, the following linear boundary-value problem is considered:

$$\mathcal{L}u \;=\; f \quad \text{in } \Omega. \tag{1}$$

Here, $\mathcal{L}$ is a linear differential operator acting on the unknown variable u, defined over the domain Ω; the function f is a forcing function.

If the residual is denoted by R and the approximate solution by $\tilde{u}$, we have

$$R(\tilde{u}) \equiv \mathcal{L}\tilde{u} - f. \tag{2}$$

2.1 The Least-Squares formulation

The idea of Least-Squares (LS) is that the residual is minimized in a certain norm. The norm of the residual associated with function space Y can be written as

$$\|R(\tilde{u})\|_Y \,, \tag{3}$$

so that LS solves the following minimization problem:

$$\min_{\tilde{u}\in X} \|R(\tilde{u})\|_Y \,. \tag{4}$$

We only allow elements $\tilde{u}$ form a linear function space X, such that the associated Y-norm exists. It is advantageous to choose this function space X as large as possible. The function space should at least be large enough to contain the exact solution of the partial differential equation.

It is common to demand that the residual is measured in the L^2-norm: the residual should be squared integrable. An even larger function space, such as H^{-1}, is not desirable: it is very complicated to perform calculations in the associated norms.

The residual norm of the partial differential equation given in (1) is defined by

$$\|R(\tilde{u})\|_0^2 = \int_\Omega (\mathcal{L}\tilde{u} - f)^2 \, d\Omega, \tag{5}$$

where $\|.\|_0$ denotes the norm that belongs to L^2.

The partial differential equation, (1), can now be cast into a minimization problem. If $I(\tilde{u})$ denotes the quadratic functional $\|R(\tilde{u})\|_0^2$, the problem can be written as

$$\min_{\tilde{u}\in X} I(\tilde{u}) \tag{6}$$

In order to solve this problem, it is necessary that the first variation of $\tilde{u}$ should vanish[1]:

$$\frac{d}{dt} I(\tilde{u}+vt) = \int_\Omega \mathcal{L}v \left(\mathcal{L}\tilde{u} - f\right) d\Omega = 0 \quad \forall v \in X \ . \tag{7}$$

2.2 Equivalence of the norms

Least-Squares is based on the requirement that if the residual tends to zero in the L^2-norm, the error in the X-norm will go to zero, as well. The error measured in the X-norm is given by

$$\|\tilde{u} - u\|_X^2 \, , \ u \in X, \tag{8}$$

where u denotes the exact solution of the partial differential equation.

If it can be shown that two positive constants C_1 and C_2 exist, such that

$$C_1 \|w\|_X \leq \|\mathcal{L}w\|_{L^2} \leq C_2 \|w\| \, , \ \forall w \in X, \tag{9}$$

then the two norm $\|\cdot\|_X$ and $\|\mathcal{L}\cdot\|_{L^2}$ are equivalent, which means that convergence in one norm implies convergence in the other norm.

Taking for $w = \tilde{u} - u$, the norm equivalence reads

$$C_1 \|\tilde{u} - u\|_X \leq \|\mathcal{L}\left(\tilde{u} - u\right)\|_{L^2} = \|\mathcal{L}\tilde{u} - f\|_{L^2} \leq C_2 \|\tilde{u} - u\| \, , \ \forall \tilde{u} \in X, \tag{10}$$

The left inequality – *coercivity* – implies that when the residual converges to zero in the L^2-norm, the approximate solution converges to the exact solution in the X-norm.

The right inequality – *boundedness, continuity* – ensures that when the approximate solution converges to the exact solution in the X-norm, the residual converges to zero in the L^2-norm.

Once norm equivalence is established between the residuals and the error, it makes sense to approximate the exact solution by minimizing the residuals

The numerical method based on the Least-Squares formulation restricts the function space to a subspace $X^h \subset X$. This leads to the problem of finding $\tilde{u}^h$ such that

$$\int_\Omega \mathcal{L}v^h \left(\mathcal{L}\tilde{u}^h - f\right) d\Omega = 0, \ \forall v^h \in X^h. \tag{11}$$

[1] For linear problems this requirement is sufficient, as well.

2.3 Boundary conditions

When approximating one-dimensional functions with polynomials, there are as many coefficients (degrees of freedom) as the polynomial degree plus one. Imposing a boundary condition strongly means that one degree of freedom is necessary to satisfy each boundary condition. Therefore, less degrees of freedom are available to approximate the differential equation in the interior of the domain. Therefore, the total approximation could be worse than when applying weak boundary conditions.

For weak boundary conditions, consider the following problem:

$$\begin{aligned} \mathcal{K}u &= g \quad \text{in } \Omega, \\ \mathcal{R}u &= h \quad \text{on } \Gamma. \end{aligned} \tag{12}$$

Here, $\mathcal{K}$ is a linear differential operator and $\mathcal{R}$ is a trace operator, both acting on the variable u. The forcing function g is defined on domain Ω and h is a function on boundary Γ. The problem can be written in a Least-Squares sense as a minimization problem:

$$\min_{\tilde{u}} \|\mathcal{K}\tilde{u} - g\|_0^2 + \lambda \min_{\tilde{u}} \|\mathcal{R}\tilde{u} - h\|_0^2 , \tag{13}$$

where λ is an arbitrarily chosen, strictly positive constant: the weight factor of the boundary condition. This formulation tries to find the minimum value of both the residual of the differential equation and the residual of the boundary conditions simultaneously.

2.4 The Spectral Element Method

In the spectral element method, the computational domain is divided in K non-overlapping sub-domains Ω^k, $k = 1, \ldots, K$ and in each sub-domain the solution is expanded in terms of orthogonal polynomials

$$u^k(x) \approx \sum_{i=0}^{N} \tilde{\alpha}_i^k P_i(x). \tag{14}$$

The set of basis functions P_i is called the *expansion basis.*

In this paper, Legendre polynomials are used to perform the calculations. These will be denoted by L_k, where k denotes the degree of the polynomial. The polynomials are a member of the well-known family of Jacobi polynomials and satisfy the following recursion relation:

$$L_{k+1}(x) = \frac{2k+1}{k+1} x L_k(x) - \frac{k}{k+1} L_{k-1}(x). \tag{15}$$

Furthermore, the polynomials are orthogonal with weighing function $w(x) \equiv 1$

$$\int_{-1}^{1} L_i L_k dx = \frac{2}{2k+1} \delta_{ik}. \tag{16}$$

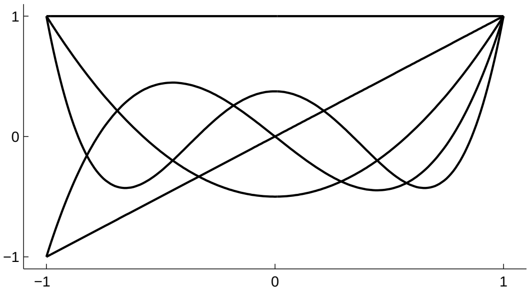


Figure 1: The curves that belong to the first five Legendre polynomials.

The first few Legendre polynomials are given by

$$\begin{array}{lcl} L_0 & = & 1, \\ L_1 & = & x, \\ L_2 & = & \frac{3}{2}x^2 - \frac{1}{2}, \\ L_3 & = & \frac{5}{2}x^3 - \frac{3}{2}x, \\ L_4 & = & \frac{35}{8}x^4 - \frac{30}{8}x^2 + \frac{3}{8}. \end{array} \tag{17}$$

The graphs of these polynomials are visualized in Figure 1.

3 WELL-POSEDNESS OF THE LINEAR ADVECTION-REACTION EQUATION

The linear advection-reaction equation in its 1-dimensional form can be written as

$$\frac{du}{dx} + cu = f, \quad x \in [-1, 1]. \tag{18}$$

Here, c is a constant and $f = f(x)$ is a forcing function. The solution is required to satisfy the following boundary condition:

$$u(-1) = u_0. \tag{19}$$

For this boundary value problem, the following norm equivalence will be derived

$$\tfrac{1}{2}e^{-2|c|} \left\| u \right\|^2_{H^1(-1,1)} \leq \left\| u' + cu \right\|^2_{H^0(-1,1)} \leq 2 \max(1, |c|) \left\| u \right\|^2_{H^1(-1,1)} . \tag{20}$$

This is valid for all $u \in H$, where H is the function space that contains the solution and the boundary condition:

$$H = \left\{ u \in H^1 \left(-1, 1\right) \; \mid u \left(-1\right) = u_0 \right\}. \tag{21}$$

Furthermore, let's assume that $u_0 = 0$, without any loss of generality. In this case, H becomes a linear function space.

Function space H^1 is defined as the set of functions, which satisfy

$$\left\| u \right\|^2_{H^1} \equiv \left\| u' \right\|^2_{H^0} + \left\| u \right\|^2_{H^0} < \infty. \tag{22}$$

For the following proofs, $\left\| \cdot \right\|_{H^p}$ is denoted as $\left\| \cdot \right\|_p$.

Proof of coercivity. In order to prove coercivity, note that every $u \in H^1(-1,1)$ with $u(-1) = 0$ can be written as $u = e^{-cx}v$, where $v \in H^1(-1,1)$ with $v(-1) = 0$. With this change of variables, we have

$$\|u' + cu\|_0^2 \;\; = \;\; \left\|e^{-cx}v'\right\|_0^2 . \tag{23}$$

Since $x \in [-1,1]$ we have $e^{-cx} \geq e^{-|c|}$. Therefore, we have the inequality

$$\|u' + cu\|_0^2 \;\; \geq \;\; e^{-2|c|}\,\|v'\|_0^2 . \tag{24}$$

Using the Poincaré inequality yields for the domain under consideration:

$$\|v'\|_0^2 \;\; \geq \;\; \tfrac{1}{2}\,\|v\|_1^2 . \tag{25}$$

Finally, using that $e^{cx} \geq e^{-|c|}$, it can be seen that

$$\|u' + cu\|_0^2 \geq \;\; \tfrac{1}{2}e^{-2|c|}\,\|u\|_1^2 . \tag{26}$$

Thus, $\forall u \in \{v \in H^1(-1,1)\,|\,v(-1) = 0\,\}$, it is valid that

$$\tfrac{1}{2}e^{-2|c|}\,\|u\|_{H^1(-1,1)}^2 \leq \|u' + cu\|_{L^2(-1,1)}^2 \tag{27}$$

Proof of continuity. Continuity of the differential operator follows from the triangle inequality:

$$\|u' + cu\|_{L^2(-1,1)}^2 \leq 2\,\|u'\|_{L^2(-1,1)}^2 + 2\,|c|\;\|u\|_{L^2(-1,1)}^2 \leq C_2\,\|u\|_{H^1(-1,1)} \;\; . \tag{28}$$

Here, constant $C_2 = 2\max(1,|c|)$.

4 RESULTS

Now that we have established a priori estimate that shows the norm equivalence between the L^2-norm of the residual and the H^1-norm of the error, it makes sense to apply the Least-Squares formulation to approximate the solution of the linear advection-reaction equation: $u' + cu = f$. In this section, the results will be presented for the Least-Squares and conventional Galerkin method in combination with Spectral Elements, both applied to this equation. For Least-Squares, two cases will be discussed: application of a weakly enforced boundary condition and the use of a strong boundary condition.

Let the approximate solution of either approximation scheme, Least-Squares or Galerkin, be given by

$$\tilde{u}(x) = \sum_{i=0}^{P} \tilde{\alpha}_i P_i(x) \;, \tag{29}$$

where the unknown coefficients $\tilde{\alpha}$ are computed by Least-Squares or the Galerkin formulation. Let the exact solution be given by

$$u(x) = \sum_{i=0}^{\infty} \alpha_i P_i(x) \ . \tag{30}$$

Then the error in the L^2-norm is given by

$$\|\tilde{u} - u\|_{L^2} = \sum_{i=0}^{P} \frac{2\left(\alpha_i - \tilde{\alpha}_i\right)^2}{2i+1} + \sum_{i=P+1}^{\infty} \frac{2\alpha_i^2}{2i+1} \ . \tag{31}$$

So the best approximation in the L^2-norm is obtained when $\tilde{\alpha}_i = \alpha_i$, or equivalently, when the error in the L^2-norm equals the truncation error

$$\sum_{i=P+1}^{\infty} \frac{2\alpha_i^2}{2i+1}. \tag{32}$$

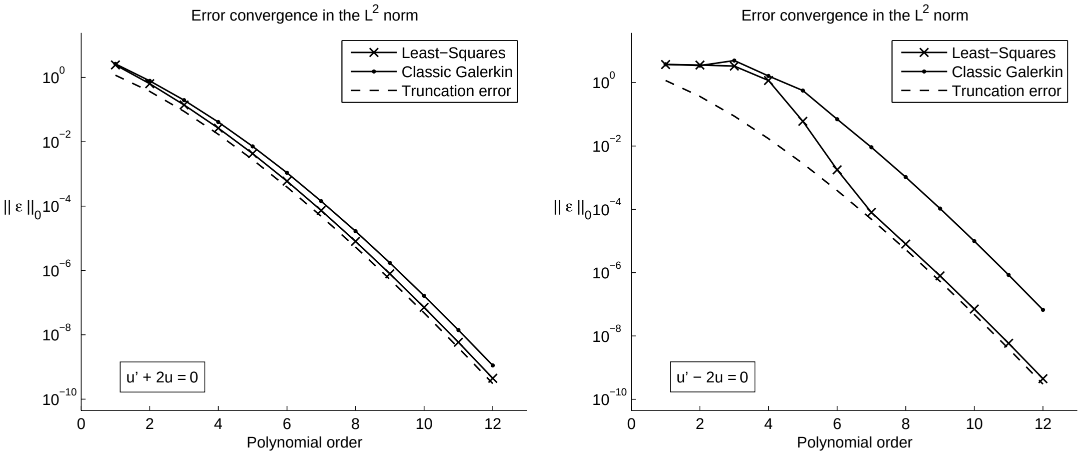
We will therefore compare the L^2-error of the solution obtained with Least-Squares and Galerkin, with the error of the best possible solution with polynomials of of degree P given by (32).

4.1 Comparison of Least-Squares and Galerkin

In Figure 2(a) an example is shown of the error convergence in the L^2-norm for $u'+2u = 0$, when using the strong boundary condition. Please note the semi-logarithmical scale; the error converges fast as a function of the polynomial degree. Although both Least-Squares and Galerkin converge exponentially fast to the exact solution, Least-Squares is more accurate and closer to the truncation error.

The next figure, Figure 2(b), globally shows the same characteristics as displayed in Figure 2(a). The error for Least-Squares is much smaller than Galerkin, although both methods display exponential convergence. However, in this figure, another effect is present. For $c < 0$, the convergence of both methods is quite poor for low order approximations. This phenomenon can be seen for all negative values of c. The more negative c is, the longer it takes before exponential convergence sets in. Clearly, in Figure 2(b), it can be seen that LS regains almost ideal convergence, where the error almost equals the truncation error at a polynomial of degree 7 and higher. There is hardly any difference between the LS solution for c equal to 2 and for c equal to -2. Galerkin, on the other hand, does converge exponentially fast to the exact solution, but does not seem to recover from the approximation error in the lower order coefficients. The convergence of this method never returns to the convergence that can be seen in Figure 2(a). This results in a less accurate approximation for Galerkin, even for high orders.

Another convergence graph is Figure 3. Here, c is set equal to zero, so that the solution to $u' = f$ is approximated. The phenomenon described by Gerritsma[10] is clearly visible:



(a) (b)

Figure 2: An example of the convergence of the error in the L^2-norm for both LS and Galerkin. Both methods are used to approximate the solution to $u' + 2u = 0$ (a) and $u' - 2u = 0$ (b) with a strong boundary condition.

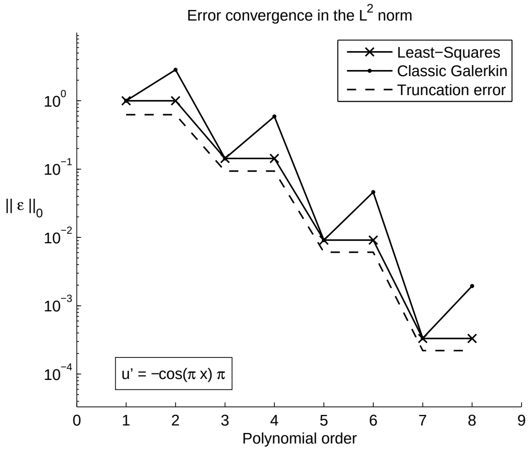
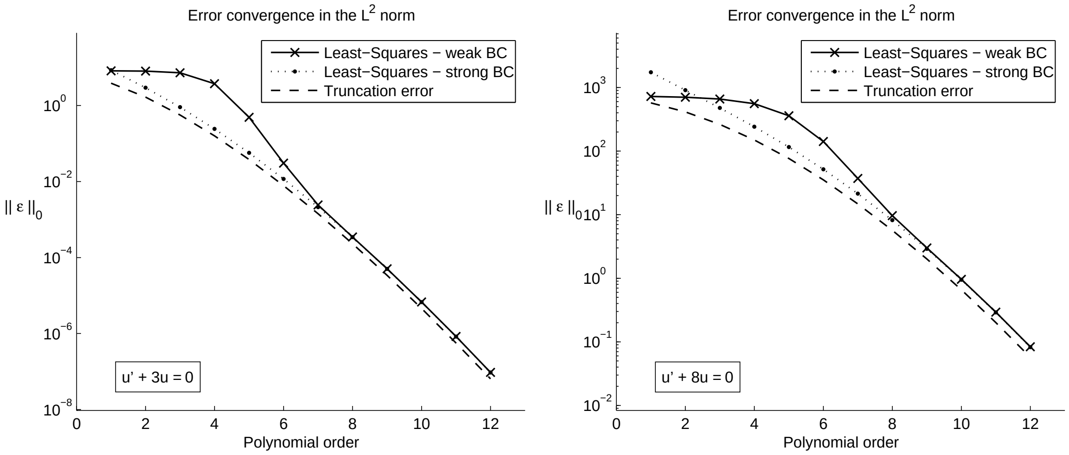


Figure 3: An example of the convergence of the error in the L^2-norm for both LS and Galerkin. Both methods are used to approximate the solution to $u' = -\pi \cos(\pi x)$ with a strong boundary condition.

for odd orders, the approximate solution of Galerkin and LS is exactly the same. However, for even orders, the convergence of Galerkin is much worse. LS still converges well for these orders.

4.2 Comparison of weak and strong boundary conditions

Apart from a negative value of c, a very weak boundary condition ($\lambda \ll 1$, where λ is as defined in (13)), has an influence on the convergence of the lowest order approximations. In Figure 4(a), this phenomenon can be seen. Here, the LS solution is shown for a value



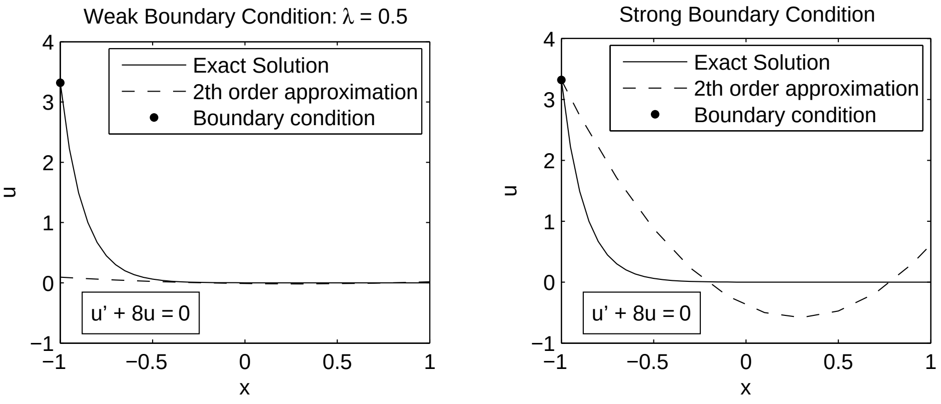
(a) (b)

Figure 4: An example of the convergence of the error in the L^2-norm for LS-SEM. The boundary condition is both weakly and strongly imposed. The results are the approximations of the solution to $u' + 3u = 0$ with $\lambda = 0.1$ (a) and $u' + 8u = 0$ with $\lambda = 0.5$ (b).

of λ of 0.1; c is equal to 3. The result is comparable to the outcomes for negative values of c, as shown before. Initially there is hardly any convergence. Then, for a particular polynomial degree, convergence suddenly sets in and the method converges almost ideal. This behavior can only be seen when λ is much smaller that c (given that $c > 0$).

A strong boundary condition is the same as a weakly imposed boundary condition with $\lambda \to \infty$. Therefore, such a boundary condition does never result in this kind of behavior, as long as $c \geq 0$. However, it is possible that a weakly imposed boundary condition results in a smaller error at low orders. This depends on the form of the solution. For instance, look at Figure 4(b). The weak boundary is imposed with a value of $\lambda = 0.5$, and c is set to 8. The error of the approximation with the weak boundary condition shows very limited convergence in the lowest orders. However, this is still better than the approximation with the boundary condition strongly imposed. At a polynomial degree equal to 3, the strong boundary condition gives better results. Around a polynomial degree of 9, the weak boundary condition approximation and the strong one give comparable results.

The reason why the weakly imposed boundary condition performs better at the lower orders is as follows: the exact solution to the problem is Ce^{-8x}. On the computational domain, this function has a maximum at $x = -1$. At that location, it decays fast towards zero. The function is almost constant on the rest of the domain. An example of such a solution and the approximation with weak boundary condition can be found in Figure 5(a). The equivalent graph with strong boundary conditions can be found in Figure 5(b). It is clear that due to the large gradient at the leftmost part of the solution, the strong boundary condition yields a much worse approximation on the rest of the domain. The approximation with weakly imposed boundary conditions does not suffer from this effect.



(a) (b)

Figure 5: The influence of a weakly or strongly imposed boundary condition on a 2^{nd} order LS-SEM approximation.

Therefore, in this case, a weak boundary condition yields better results.

5 CONCLUSIONS

It can be concluded that the 1-dimensional, linear advection-reaction equation given by (18) and (19) is well-posed. Furthermore, if the residual converges to zero in the H^0-norm, the error will converge in the H^1-norm. Therefore, the Least-Squares formulation is proven to be able to get the desired convergence behavior of the approximate solution.

The LS-SEM approximation is generally more accurate than the conventional Galerkin SEM approximation. For both methods, the model equation is hard to solve for negative values of c. However, at higher polynomial degrees, LS returns to a convergence close to that of the truncation error, which is ideal. Galerkin performs much worse in this respect.

In combination with a weakly imposed boundary condition, LS performs worse at low polynomial degrees than with a strong boundary condition. However, at high polynomial degrees, the convergence of both methods is comparable. This might indicate that the boundary weighting factor should be a function of the polynomial degree P.

REFERENCES

[1] B.N. Jiang, The Least-Squares Finite Element Method, Ttheory and Applications in Computational Fluid Dynamics and Electromagnetics, Springer Verlag, 1998.

[2] B.N. Jiang, On The Least-Squares Finite Element Method, *Comput. Methods Appl. Mech. Engrg.*, **152** (1998). 239-257.

[3] B.N. Jiang and L.A. Povinelli, Least-Squares Finite Element Method for Fluid Dynamics, *Comput. Methods Appl. Mech. Engrg.*, **81** (1990). 13-37.

[4] B. de Maerschalck and M.I. Gerritsma, Space-Time Least-Squares Spectral Elements for Convection-Dominated Flows, *AIAA Journal*, **44**, no. 3 (2006). 558-565.

[5] M.M.J. Proot and M.I. Gerritsma, A Least-Squares Spectral Element formulation for the Stokes Problem, *J. Sci. Comput.*, **17)** (2002). 285-296.

[6] M.M.J. Proot and M.I. Gerritsma, Least-Squares Spectral Elements applied to the Stokes Problem, *J. Comp. Phys.*, **181)** (2002). 454-477.

[7] M.M.J. Proot, The Least-Squares Spectral Elements Method,Ph.D. thesis, Delft University of Technology, Department of Aerospace Engineering, Delft, The Netherlands, 2003

[8] J.P.Pontaza and J.N. Reddy, Space-time coupled spectral/*hp* least-squares finite element formulation for the incompressible Navier-Stokes equation, *J. Comput. Phys.*, **190)**, no. 2 (2003). 418-459.

[9] J.P.Pontaza and J.N. Reddy, Spectral/*hp* least-squares finite element formulation for the Navier-Stokes equation, *J. Comput. Phys.*, **197)**, no. 2 (2004). 523-549.

[10] M.I. Gerritsma, Missing Convergence Rates, *in preparation.*

[11] M.I. Gerritsma and Bart De Maerschalck, The Least-Squares Spectral Element Method, *Proceedings CFD–Higher Order Discretization Methods*, (2005).

[12] P.E.J. Vos, Application of Least-Squares Spectral Elements Method to Polynomial Choas *in preparation.*