Analysis of the order of accuracy for node-centered finite volume schemes

Sofia Eriksson and Jan Nordström

Original Publication:

Sofia Eriksson and Jan Nordström, Analysis of the order of accuracy for node-centered finite volume schemes, 2009, Applied Numerical Mathematics, (59), , 2659-2676.
http://dx.doi.org/10.1016/j.apnum.2009.06.001
Copyright: Elsevier Science B.V., Amsterdam.
http://www.elsevier.com/

Postprint available at: Linköping University Electronic Press
http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-68540

Analysis of the order of accuracy for node-centered finite volume schemes

Sofia Eriksson and Jan Nordström

March 10, 2009

Abstract

The order of accuracy of the node-centered finite volume methods is analyzed, and the analysis is based on an exact derivation of the numerical errors in one dimension. The accuracy for various types of grids are considered. Numerical simulations and analysis are performed for both a hyperbolic and a eliptic case, and the results agree. The impact of weakly imposed boundary conditions is analyzed and verified numerically. We show that the error contribution from the primal and dual grid can be treated separately.

1 Introduction

The order of accuracy of the widely used [11, 3, 8, 4, 16, 9] node-centered finite volume method (FVM) has been discussed recently. There are indications that the scheme yields at least first order accuracy for all types of grids [2, 14], while others have indications that for some meshes the order is less than one [12, 15, 6]. In this report the behaviour of the node-centered finite volume method on a one dimensional mesh is analyzed, using both hyperbolic and elliptic model problems. The scheme is formulated in terms of Summation-By-Parts operators and the boundary conditions are imposed weakly using penalty terms [1, 7, 10, 13].

The advantage with the one-dimensional analysis is that we can exactly compute the discretization error in terms of the truncation error. No assumptions or non-sharp estimates are needed. The drawback is of course that most applications are in multiple dimensions. However, without the direct link from truncation error to discretization error, no real understanding is possible. To be crystal clear, one has to be able to invert the operators.

The outline of this paper is as follows. First we make sure that our numerical procedure is stable for time dependent problems. Next, numerical simulations are performed for one hyperbolic and one elliptic ordinary differential equation (ODE), in order to find the rates of convergence. This is done using several types of grids. Thereafter the two problems are thoroughly analyzed, and the orders of accuracy obtained analytically is compared to the rates of convergence obtained experimentally.

2 Well-posedness and stability

Even though we here primarily focus on steady solutions we will first make sure that the corresponding time-dependent problem is well-posed and stable. This procedure guarantees that one can solve for the steady solution and that the matrix in the resulting linear equation system can always be inverted.

2.1 Well-posedness

The continuous problem that we consider is

$$\begin{aligned} &u_t + au_x = \varepsilon u_{xx} + F(x,t), \qquad 0 \leq x \leq 1 \\ &u(x,0) = f(x), \end{aligned} \tag{1}$$

where $\varepsilon \geq 0$, F is a forcing function and f is the initial condition. The problem in (1) is strongly well-posed if the solution is bounded in terms of all the data F, f, g_0, g_1, where g_0 and g_1 are the boundary data at the left and right boundary, respectively. We limit ourselves to show well-posedness, hence considering the data $F, g_0, g_1 = 0$, see [5].

To examine this we use the energy method. We multiply the differential equation in (1) by $2u$ and integrate over the spatial domain. We obtain

$$\frac{d}{dt}\|u\|^2 = au(0,t)^2 - au(1,t)^2 + 2\varepsilon\left(-u(0,t)u_x(0,t) + u(1,t)u_x(1,t) - \|u_x\|^2\right) \tag{2}$$

where $\|u\|^2 \equiv \int_0^1 u^2 dx$. If $\varepsilon = 0$ the differential equation will be hyperbolic and we will only need one boundary condition. If $a > 0$, $u(0,t) = g_0$ has to be given, if $a < 0$, $u(1,t) = g_1$ has to be given instead. Assuming $a > 0$ and $g_0 = 0$ we get

$$\frac{d}{dt}\|u\|^2 = -au(1,t)^2 \ . \tag{3}$$

Considering the parabolic case we have $\varepsilon \neq 0$ and hence we have to give data at both boundaries, i.e. $u(0,t) = g_0$ and $u(1,t) = g_1$. Inserting the zero boundary data yields

$$\frac{d}{dt}\|u\|^2 = -2\varepsilon\|u_x\|^2 \ . \tag{4}$$

Time integration of (3) and (4) gives that the problem in (1) is well-posed in the classical sense, assuming that the correct number of boundary conditions is used.

2.2 The discrete formulation

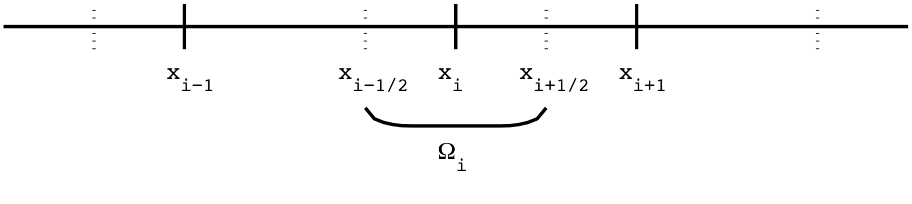

Figure 1: The control volume Ω_i on an unstructured mesh. The unknowns are located at positions with an integer index, i.e. x_{i-1}, x_i, x_{i+1}. The positions of the flux points are denoted with half indices, e.g. $x_{i-1/2}$ and $x_{i+1/2}$. The control volumes Ω_i are defined as $x_{i+1/2} - x_{i-1/2}$.

To discretize we integrate the differential equation in (1) over the control volumes, Ω_i. The volumes Ω_i corresponds to positions x_i and are defined as $\Omega_i = x_{i+1/2} - x_{i-1/2}$, where $x_{i+1/2}$ is the flux point between the positions x_i and x_{i+1}, see Figure 1. At the boundaries we have $\Omega_0 = (x_{1/2} - x_0)$ and $\Omega_N = (x_N - x_{N-1/2})$.

Next the term $\int_{\Omega_i} u_t dx$ is approximated by $\Omega_i u_t(x_i)$, and the flux on the control volume boundaries by centered approximations. The numerical scheme becomes

$$\Omega_i(v_i)_t = -a\frac{(v_{i+1} - v_{i-1})}{2} + \varepsilon\left(\frac{v_{i+1} - v_i}{x_{i+1} - x_i} - \frac{v_i - v_{i-1}}{x_i - x_{i-1}}\right) + \Omega_i F \tag{5}$$

where v_i denotes the discrete representation of $u(x_i)$. In matrix form, including a weak implementation of the boundary conditions, this is written as

$$\begin{aligned} v_t &= -aP^{-1}Qv + \varepsilon P^{-1}(-A + BS)v + F \\ &\quad + \tau_0 P^{-1}e_0(v_0 - g_0) + \tau_N P^{-1}e_N(v_N - g_N) \\ v(0) &= f \end{aligned} \tag{6}$$

where $v = [v_0\ v_1\ \ldots\ v_N]^T$ is a vector containing the discretized numerical solution. We have also introduced the vectors $e_0 = [1\ 0\ \ldots\ 0]^T$ and $e_N = [0\ \ldots\ 0\ 1]^T$. $P^{-1}Q$ represents d/dx and $P^{-1}(-A + BS)$ represents d^2/dx^2. The boundary conditions are included using a weak penalty treatment, see [1, 10, 13] for details. The penalty parameters are denoted $\tau_{0,N}$. P contains the control volumes Ω_i. The difference operator Q is nearly skew-symmetric as $Q + Q^T = B$, where $B = E_N - E_0$. E_0 and E_N are matrices containing only zeros, except for $E_0(0,0) = E_N(N,N) = 1$. P and Q are shown in (7).

$$P = \begin{bmatrix} \Omega_0 & & & & & \\ & \Omega_1 & & & & \\ & & \Omega_2 & & & \\ & & & \ddots & & \\ & & & & \Omega_{N-1} & \\ & & & & & \Omega_N \end{bmatrix}, \quad Q = \frac{1}{2}\begin{bmatrix} -1 & 1 & & & & \\ -1 & 0 & 1 & & & \\ & -1 & 0 & 1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & -1 & 0 & 1 \\ & & & & -1 & 1 \end{bmatrix}. \tag{7}$$

Here S is such that $(Sv)_0 = (v_1 - v_0)/\Delta x_1$ and $(Sv)_N = (v_N - v_{N-1})/\Delta x_N$. The interior of S is equal to the identity matrix. A and S are given in (8), where $\Delta x_i = x_i - x_{i-1}$.

$$A = \begin{bmatrix} \frac{1}{\Delta x_1} & \frac{-1}{\Delta x_1} & & & \\ \frac{-1}{\Delta x_1} & \frac{1}{\Delta x_1} + \frac{1}{\Delta x_2} & \frac{-1}{\Delta x_2} & & \\ & \frac{-1}{\Delta x_2} & \frac{1}{\Delta x_2} + \frac{1}{\Delta x_3} & \frac{-1}{\Delta x_3} & \\ & \ddots & \ddots & \ddots & \\ & & \frac{-1}{\Delta x_{N-1}} & \frac{1}{\Delta x_{N-1}} + \frac{1}{\Delta x_N} & \frac{-1}{\Delta x_N} \\ & & & \frac{-1}{\Delta x_N} & \frac{1}{\Delta x_N} \end{bmatrix}, \quad S = \begin{bmatrix} \frac{-1}{\Delta x_1} & \frac{1}{\Delta x_1} & & & \\ & 1 & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 & \\ & & & & \frac{-1}{\Delta x_N} & \frac{1}{\Delta x_N} \end{bmatrix}. \tag{8}$$

2.3 Stability in the discrete case

For simplicity we again let $F = g_0 = g_N = 0$. We multiply (6) by $v^T P$ from the left, add the transpose, introduce $A = S^T R S$ and use the relation $\frac{d}{dt}(\|v\|_P^2) = \frac{d}{dt}(v^T P v)$.

That yields

$$\frac{d}{dt}(\|v\|_P^2) = a(v_0^2 - v_N^2) + 2\varepsilon\left(-v_0(Sv)_0 + v_N(Sv)_N) - 2\varepsilon(Sv)^T R(Sv)\right) \\ + 2\tau_0 v_0^2 + 2\tau_N v_N^2 \ . \tag{9}$$

First consider the hyperbolic case. In the continuous version we had $a > 0$ and should hence only give data at the left boundary. In the discrete case, having $\varepsilon = 0$ and $a > 0$, the corresponding action is to let $\tau_N = 0$. We get

$$\frac{d}{dt}(\|v\|_P^2) = v_0^2(a + 2\tau_0) - av_N^2 \ . \tag{10}$$

In the hyperbolic case the semi-discrete scheme in (6) will be stable if $\tau_0 \leq -a/2$. For the parabolic case we will use another technique, see [1]. Rewrite (9) as

$$\begin{aligned} \frac{d}{dt}(\|v\|_P^2) \quad &= -2\varepsilon(Sv)^T\tilde{R}(Sv) \\ &+ \begin{bmatrix} v_0 \\ (Sv)_0 \end{bmatrix}^T \begin{bmatrix} a+2\tau_0 & -\varepsilon \\ -\varepsilon & -2\varepsilon\gamma_0 \end{bmatrix} \begin{bmatrix} v_0 \\ (Sv)_0 \end{bmatrix} \\ &+ \begin{bmatrix} v_N \\ (Sv)_N \end{bmatrix}^T \begin{bmatrix} -a+2\tau_N & \varepsilon \\ \varepsilon & -2\varepsilon\gamma_N \end{bmatrix} \begin{bmatrix} v_N \\ (Sv)_N \end{bmatrix} , \end{aligned} \tag{11}$$

where $\tilde{R} = R - \gamma_0 E_0 + \gamma_N E_N$. To make (6) stable we need $\tilde{R} \geq 0$ and also that the two quadratic forms in (11) are negative semi-definite. This will be fulfilled if

$$\tau_0 \leq -a/2 - \varepsilon/(4\Delta x_1), \qquad \tau_N \leq a/2 - \varepsilon/(4\Delta x_N) \ . \tag{12}$$

These stability requirements are derived in Appendix A.

3 Numerical experiments

The partial differential equation (PDE) is well-posed and the numerical scheme is stable. The next question is: How accurate will the results be when doing simulations? We begin be looking and the hyperbolic and parabolic problems separately. At steady-state the time derivative is zero, and we mimic this by numerically implementing the ODE's

$$u_x = F_1(x), \qquad u(0) = g_0 \tag{13}$$

$$-u_{xx} = F_2(x), \qquad u(0) = g_0, \quad u(1) = g_1 \tag{14}$$

in the domain $0 \leq x \leq 1$. We consider the manufactured solution $u(x) = \sin(5\pi x/2) + x^2 + 1$. Consequently $F_1 = 5\pi\cos(5\pi x/2)/2 + 2x$ and $F_2 = (5\pi)^2\sin(5\pi x/2)/4 - 2$. (The solution u was chosen in order to not be 2π-periodic or having zero valued derivatives on the boundaries.)

A discrete version of (13) and (14) becomes

$$P^{-1}Qv = F_1 + \tau_0 P^{-1}e_0(v_0 - g_0) \tag{15}$$

$$-P^{-1}Mv = F_2 + \tau_0 P^{-1}e_0(v_0 - g_0) + \tau_N P^{-1}e_N(v_N - g_N) \tag{16}$$

where $M = -A + BS$. Equation (15) and (16) was implemented using the stability requirements on τ_0 and τ_N in (12), and solved using five differently designed grids.

The grids are specified in terms of the primal and dual mesh. The primal mesh consists of the integer-index solution points and the dual mesh of the flux points, see Figure 1. For the primal mesh we will require the condition stated below

Assumption 3.1. *Assume that all Δx_i have the same probability distribution (this restriction excludes stretched meshes). Hence*

$$E(\Delta x_i) = 1/N, \qquad E\left((\Delta x_i)^p\right) = m_p/N^p \tag{17}$$

where N is the number of grid points and the constant m_p is dependent on the probability distribution chosen. For a deterministic mesh such that $\Delta x_i \equiv 1/N$ we have $m_p \equiv 1$.

For the primal grid the following notation will be used. If all $\Delta x_i = h = 1/N$ then the primal mesh is said to be 'equidistant' (or uniform). If the Δx_i's are randomly perturbed, the primal mesh is called 'random'. For the control volumes we have a corresponding notation. If the flux points $x_{i+1/2} \equiv (x_i + x_{i+1})/2$ the dual mesh will be called 'centered'. If $x_{i+1/2}$ is perturbed by a constant we call the dual mesh 'shifted' and if the flux points are perturbed randomly we call it 'random'.

The discretization errors are defined as $e_i = u(x_i) - v_i$ and the truncation errors as $T_e = P^{-1}Qu - F_1$ and $T_e = -P^{-1}Mu - F_2$, respectively. The resulting convergence rates for e and T_e for the two cases are given in Table 1 and Table 2. In Figure 2 and Figure 3 the L_2-norms of e are plotted. The slopes in the figures correspond to the rates listed in the tables.

Primal mesh:	uniform	uniform	uniform	random	random
Dual mesh:	centered	shifted	random	centered	random
Rate of Conv. $L_2(e)$	2	1	0.5	1.5	0.5
Rate of Conv. $L_2(T_e)$	1.5	0.5	0	1	0
R.o.C. $L_2(T_e)$(inner points)	2	2	0	1	0

Table 1: Hyperbolic case. Order of accuracy for $e = u - v$ when solving $u_x = F_1$.

Primal mesh:	uniform	uniform	uniform	random	random
Dual mesh:	centered	shifted	random	centered	random
Rate of Conv. $L_2(e)$	2	2	1.5	2	1.5
Rate of Conv. $L_2(T_e)$	0.5	0.5	0	0.5	0
R.o.C. $L_2(T_e)$(inner points)	2	2	0	1	0

Table 2: Elliptic case. Order of accuracy for $e = u - v$ when solving $-u_{xx} = F_2$.

One can conclude that a centering of the dual mesh, i.e. to have $x_{i+1/2} \equiv (x_i + x_{i+1})/2$, always gives the best rate of convergence. Another observation is that the order of accuracy of the truncation error does not seem to have a clear relation to that of the discretization error, see Table 1 and 2 . This was also the observation in [2], [14].

Remark: Solving with randomly distributed grid points naturally gives rise to random errors. Therefore we have for those meshes (i.e. 'equidistant, random', 'random, centered' and 'random, random') made 500 simulation runs. In the figures the mean of the 500 L_2-norms of the discretization errors are plotted.

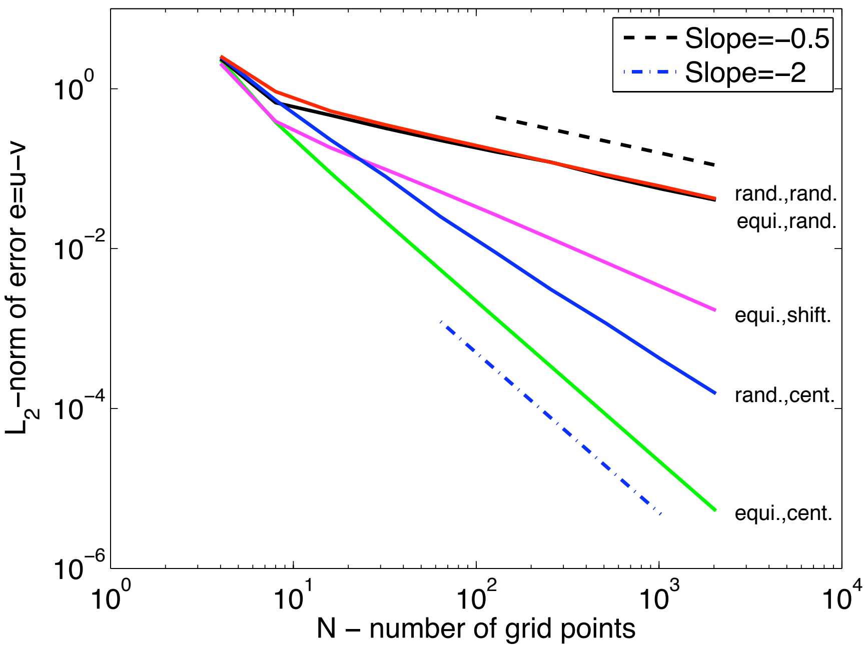

Figure 2: Hyperbolic case. Discretization error $e = u - v$ as a function of number of gridpoints N, when solving $u_x = F_1$. For the random grids a mean value after 500 test runs is shown.

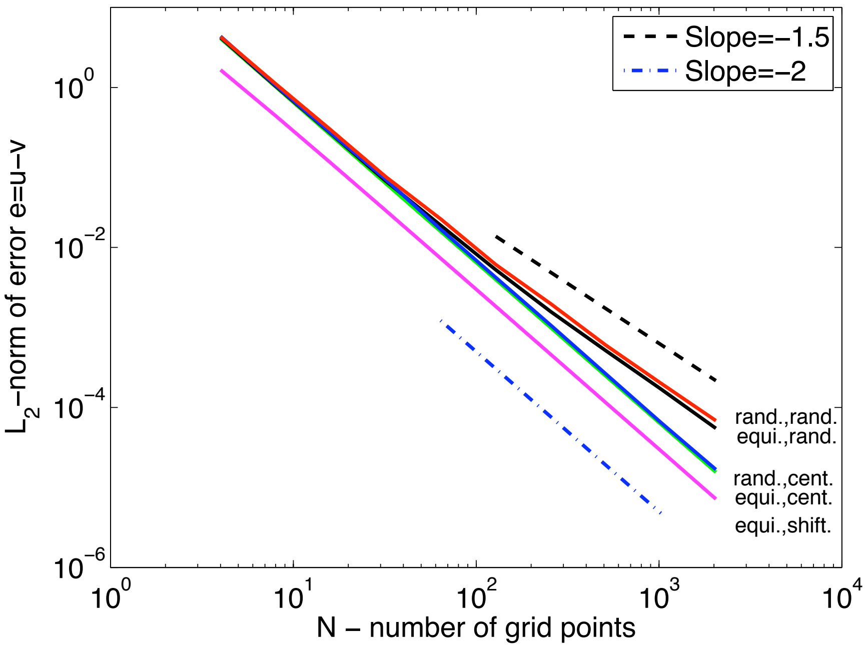

Figure 3: Parabolic case. Discretization error $e = u - v$ as a function of number of gridpoints N, when solving $-u_{xx} = F_2$. For the random grids a mean value after 500 test runs is shown.

4 Order of accuracy in the hyperbolic case

To find out how the discretization errors depend on the truncation errors, we analyze the model problems further. The discrete formulation of the hyperbolic problem in (13) is given in (15) where v_i is the numerical approximation of $u(x_i)$. Let $\tilde{Q} = Q - \tau_0 E_0$ and rewrite (15) as

$$v = \tilde{Q}^{-1}(PF - \tau_0 e_0 g) \ . \tag{18}$$

The truncation error $T_e = P^{-1}\tilde{Q}u - F + \tau_0 P^{-1} e_0 g$ is obtained by inserting the exact solution u into (15). That leads to

$$u = \tilde{Q}^{-1}(PF - \tau_0 e_0 g + PT_e) \ . \tag{19}$$

Combining equation (18) and (19), we obtain

$$e = u - v = \tilde{Q}^{-1}PT_e \ . \tag{20}$$

$\tilde{Q}$ is non-singular if $\tau_0 \neq 0$, and is inverted explicitly in Appendix B. The result is

$$\tilde{Q}^{-1} = \frac{1}{\tau_0}\begin{bmatrix} -1 & 1 & -1 & 1 & & 1 & -1 \\ -1 & 1 & -1 & 1 & \cdots & 1 & -1 \\ -1 & 1 & -1 & 1 & & 1 & -1 \\ -1 & 1 & -1 & 1 & & 1 & -1 \\ & \vdots & & & & \vdots & \\ -1 & 1 & -1 & 1 & \cdots & 1 & -1 \\ -1 & 1 & -1 & 1 & & 1 & -1 \end{bmatrix} + 2\begin{bmatrix} 0 & 0 & 0 & 0 & & 0 & 0 \\ 0 & 1 & -1 & 1 & \cdots & 1 & -1 \\ 0 & 1 & 0 & 0 & & 0 & 0 \\ 0 & 1 & 0 & 1 & & 1 & -1 \\ & & \vdots & & & \vdots & \\ 0 & 1 & 0 & 1 & \cdots & 1 & -1 \\ 0 & 1 & 0 & 1 & & 1 & 0 \end{bmatrix} . \tag{21}$$

Note that $\tilde{Q}^{-1}$ has one constant part and one part that varies from row to row.

4.1 Computation of the discretization error

The point-wise discretization error $e_i = (\tilde{Q}^{-1}PT_e)_i$ can now be split into two parts as

$$e_i = \frac{1}{\tau_0}e_\tau + 2\tilde{e}_i \tag{22}$$

where $e_\tau = -\sum_{j=0}^{N}(-1)^j(PT_e)_j$ is constant in all grid points. The error part $\tilde{e}_i$ is more complex and varies with the index i. It is given in Appendix C.

To compute e_τ we need PT_e. First define $x_{i-1/2} = (x_{i-1} + x_i)/2 - \xi_i$, where $|\xi_i| < \Delta x_i/2$. (I.e. ξ_i introduces an arbitrary perturbation around a centered dual mesh). The control volumes then become

$$\Omega_i = \frac{\Delta x_{i+1} + \Delta x_i}{2} - \xi_{i+1} + \xi_i \ , \qquad \Omega_0 = \frac{\Delta x_1}{2} - \xi_1 \ , \qquad \Omega_N = \frac{\Delta x_N}{2} + \xi_N \ . \tag{23}$$

We Taylor expand $PT_e = Qu - PF$ and denote $F^i \equiv F(x_i)$. Recalling (7) and inserting $u_x = F$ and $u_{xx} = F_x$, we can write

$$\begin{aligned}
(PT_e)_0 &= \xi_1 F^0 + \frac{\Delta x_1^2}{4} F_x^0 + \mathcal{O}(h^3) \\
(PT_e)_i &= (\xi_{i+1} - \xi_i)\, F^i + \left(\frac{\Delta x_{i+1}^2 - \Delta x_i^2}{4}\right) F_x^i + \mathcal{O}(h^3) \\
(PT_e)_N &= -\xi_N F^N - \frac{\Delta x_N^2}{4} F_x^N + \mathcal{O}(h^3)
\end{aligned} \tag{24}$$

which yields

$$e_\tau = \underbrace{\sum_{i=1}^{N} (-1)^i \xi_i (F^{i-1} + F^i)}_{e_\xi} + \underbrace{\frac{1}{4} \sum_{i=1}^{N} (-1)^i \Delta x_i^2 (F_x^{i-1} + F_x^i)}_{e_{\Delta x}} + \mathcal{O}(h^2)\ . \tag{25}$$

Note that e_ξ depends on the dual mesh and $e_{\Delta x}$ on the primal mesh.

Remark: Since the error can be divided into a dual mesh dependent part an a primal mesh dependent part, one can analyze the error contribution from the dual and primal mesh separately, rather than analyzing all combination of grids.

4.2 Dual mesh dependent error

First consider the dual mesh (computation of e_ξ). When the control volumes are centered then $\xi_i = 0$ and therefore $e_\xi = 0$. If the dual mesh is shifted ($\xi_i = Ch$) cancellation yields $e_\xi = Ch(F^N - F^0)$ which means that an $\mathcal{O}(h)$ contribution is added to the discretization error.

The random case requires more explanation. The expected value of e_ξ, $E(e_\xi)$, will be equal to zero if $E(\xi_i) = 0$. On the other hand, the "worst case scenario" when one considers $\xi_i \sim \mathcal{O}(h)$, leads to e_ξ being $\mathcal{O}(1)$. The deviation from the mean value, $\sigma(e_\xi)$, will give the correct answer from a statistical point of view.

Proposition 4.1. *Consider solving the problem* (13) *using* (15) *on a mesh where the control volumes are randomly distributed, so that $\xi_i = kr_i\Delta x_i$ and where Assumption 3.1 holds. Assume that the random variable r_i is uniformly distributed in* $[-1, 1]$ *and that $|k| < 1/2$ is a constant. Then we have*

$$E(e_\xi) = 0, \qquad \sigma(e_\xi) \sim \mathcal{O}(h^{0.5}) \tag{26}$$

where $E(e_\xi)$ is the expected value of e_ξ and $\sigma(e_\xi)$ the standard deviation. Hence the contribution from the dual mesh to the discretization error is on average $e_\xi \sim \mathcal{O}(h^{0.5})$.

Proof of proposition 4.1. The relation $\xi_i = kr_i\Delta x_i$ inserted into (25) gives

$$e_\xi = k \sum_{i=1}^{N} (-1)^i r_i \Delta x_i (F^{i-1} + F^i)\ . \tag{27}$$

The mean of e_ξ, $E(e_\xi)$, is zero since $E(r_i) = 0$. To compute the standard deviation $\sigma(e_\xi)$ we need

$$e_\xi^2 = k^2 \left(\sum_{i=1}^{N} r_i^2 \Delta x_i^2 (F^{i-1} + F^i)^2 + \sum_{i \neq j} (-1)^{i+j} r_i r_j \Delta x_i \Delta x_j (F^{i-1} + F^i)(F^{j-1} + F^j) \right) \quad (28)$$

and

$$E(r_i^2) = \int_{-\infty}^{\infty} p(r_i) r_i^2 dr = \int_{-1}^{1} \frac{1}{2} r_i^2 dr = \frac{1}{3} \ , \qquad E(r_i r_j) = E(r_i)E(r_j) = 0 \ . \quad (29)$$

Using (28), (29) and Assumption 3.1 we obtain

$$E(e_\xi^2) = \frac{k^2}{3} \frac{m_2}{N^2} \underbrace{\sum_{i=1}^{N} (F^{i-1} + F^i)^2}_{\varphi N} = \frac{k^2 m_2 \varphi}{3N} \quad (30)$$

where φ is an $\mathcal{O}(1)$ constant depending on the function F. (For example if $F = c_0$ then $\varphi = 4c_0^2$ and if $F = c_1 x$ then $\varphi = c_1^2(4/3 - \mathcal{O}(h^2))$.) The standard deviation $\sigma(e_\xi)$ is obtained by taking the square root of the variance V,

$$\begin{aligned} V(e_\xi) &= E(e_\xi^2) - E(e_\xi)^2 = \frac{k^2 m_2 \varphi}{3N} \\ \sigma(e_\xi) &= \sqrt{V(e_\xi)} = k \sqrt{\frac{m_2 \varphi}{3N}} \sim \mathcal{O}(h^{0.5}) \ . \end{aligned} \quad (31)$$

Hence the error contribution e_ξ is on average $\mathcal{O}(h^{0.5})$, when using random dual mesh. □

The error contributions introduced by the dual meshes are summarized in Table 3.

4.3 Primal mesh dependent error

Now consider the primal mesh (the component $e_{\Delta x}$ in (25)). The equidistant case, i.e. $\Delta x_i = h$, is straight-forward. Due to cancellation $e_{\Delta x} = h^2(F_x^N - F_x^0)/4$, and we conclude that an equidistant primal mesh contributes with an $\mathcal{O}(h^2)$ error.

The random case requires statistical analysis. For the simplest case $F_x = c_1$ we get $e_{\Delta x} = c_1(-\Delta x_1^2 + \Delta x_2^2 - \ldots + \Delta x_N^2)/2$. At best, all Δx_i's have the same size and $e_{\Delta x} = 0$, and at worst they vary such that $e_{\Delta x} \sim \mathcal{O}(h)$. The statistical result (for a general F) is given in Proposition 4.2.

Proposition 4.2. *Consider solving the problem* (13) *using* (15) *on a mesh with randomly distributed primal mesh, such that Assumption 3.1 holds. Then we have*

$$E(e_{\Delta x}) \sim \mathcal{O}(h^2), \qquad \sigma(e_{\Delta x}) \sim \mathcal{O}(h^{1.5}), \quad (32)$$

i.e. the primal mesh contribution to the discretization error is on average $e_{\Delta x} \sim \mathcal{O}(h^{1.5})$.

Proof of proposition 4.2. Consider $e_{\Delta x}$ in (25). Assumption 3.1 leads to

$$E(\Delta x_i^2) = \frac{m_2}{N^2}, \qquad E(\Delta x_i^4) = \frac{m_4}{N^4} \tag{33}$$

which yields $E(e_{\Delta x}) = m_2(F_x^N - F_x^0)/(4N^2)$ due to cancellation. In order to compute the standard deviation we need

$$e_{\Delta x}^2 = \frac{1}{16}\sum_{i=1}^{N} \Delta x_i^4 (F_x^{i-1} + F_x^i)^2 + \frac{2}{16}\underbrace{\sum_{i \neq j} (-1)^{i+j} \Delta x_i^2 \Delta x_j^2 (F_x^{i-1} + F_x^i)(F_x^{j-1} + F_x^j)}_{\zeta} \tag{34}$$

where ζ consists of $N(N-2)/4$ terms with positive sign, and $N^2/4$ negative terms. We use the relations in (33) to obtain

$$\begin{aligned} E(e_{\Delta x}^2) &= \frac{1}{16}\frac{m_4}{N^4}\underbrace{\sum_{i=1}^{N} (F_x^{i-1} + F_x^i)^2}_{\varphi_1 N} - \frac{2}{16}\frac{m_2^2}{N^4}\underbrace{\sum_{i \neq j} (F_x^{i-1} + F_x^i)(F_x^{j-1} + F_x^j)}_{\varphi_2 N/2} \\ &= \frac{m_4\varphi_1 - m_2^2\varphi_2}{16N^3} \ . \end{aligned} \tag{35}$$

The constants m_2 and m_4 depends on how Δx_i is distributed. The constants φ_1 and φ_2 depends on F_x, and if $F_x = c_1$ then $\varphi_1 = \varphi_2 = 4c_1^2$. The constants $m_2, m_4, \varphi_1, \varphi_2$ will automatically be such that (35) is positive, since we are computing the mean value of a square. Hence we can put $K = \sqrt{m_4\varphi_1 - m_2^2\varphi_2}$. Then

$$\sigma(e_{\Delta x}) = \sqrt{E(e_{\Delta x}^2) - E(e_{\Delta x})^2} = \frac{K}{4\sqrt{N^3}} + \mathcal{O}(h^{2.5}) \ . \tag{36}$$

Thus a random primal mesh gives an $\mathcal{O}(h^{1.5})$ error contribution since $\sigma(e_{\Delta x}) \sim 1/N^{1.5}$. □

4.4 Result of the analysis of the hyperbolic case

In (22) we have split the discretization error e into one constant part e_τ and one varying part $\tilde{e}$. We divide $\tilde{e}_i$ into a dual mesh dependent and a primal mesh dependent part just as we did with e_τ

$$\tilde{e}_i = \tilde{e}_\xi + \tilde{e}_{\Delta x} + \mathcal{O}(h^2) \tag{37}$$

and perform similar analysis as in the previous sections. This is done in Appendix C and the result is listed in table 3 and 4. From the tables it is clear that e_τ and $\tilde{e}_i$ have the same order of accuracy.

Remark: The contribution from e_τ to the discretization error e is equal in every grid point. $\tilde{e}$ on the other hand varies from grid point to grid point. In particular, $\tilde{e}_0 = 0$. If $\tau_0 \to \infty$ then $e_0 \to 0$, i.e. with increased penalty the numerical solution v_0 tends to the exact solution $u(0)$. However, in the rest of the domain one can not (in the general case) say whether increased penalty leads to small errors.

Dual mesh		**Error part**		**Collective**
Notation	Form of ξ_i	e_ξ	$\tilde{e}_\xi$	Order of acc.
Centered	0	0	0	∞
Shifted	Ch	$Ch(F^N - F^0)$	$\frac{1-(-1)^i}{2}ChF^N$	1
Random	$k\Delta x_i r_i$	$\sigma(e_\xi) \sim kh^{0.5}$	$\sigma(\tilde{e}_\xi) \sim kh^{0.5}$	0.5

Table 3: Error contribution from the dual mesh when solving a hyperbolic problem.

Primal mesh		**Error part**		**Collective**
Notation	Form of Δx_i	$e_{\Delta x}$	$\tilde{e}_{\Delta x}$	Order of acc.
Equidist.	h	$h^2(F_x^N - F_x^0)/4$	$\frac{1-(-1)^i}{2}h^2F_x^N/4$	2
Random	$E(\Delta x_i) = h$	$\sigma(e_{\Delta x}) \sim h^{1.5}$	$\sigma(\tilde{e}_{\Delta x}) \sim h^{1.5}$	1.5

Table 4: Error contribution from the primal mesh when solving a hyperbolic problem.

As we can see from table 5 the results from the analysis agree perfectly with the rates of convergence obtained experimentally. This is due to the fact that the inverse of the discrete operator, i.e. the direct link between the truncation error and the discretization error, can be computed exactly.

Primal mesh:	uniform	uniform	uniform	random	random
Dual mesh:	centered	shifted	random	centered	random
R.o.C. $L_2(e)$	2	1	0.5	1.5	0.5
Analysis	$\min(\infty, 2)$	$\min(1, 2)$	$\min(0.5, 2)$	$\min(\infty, 1.5)$	$\min(0.5, 1.5)$

Table 5: Hyperbolic case. Order of accuracy for $e = u - v$ when solving $u_x = F_1$.

Note that $\Delta x_i \sim 1/N$ is assumed. For a linearly stretched mesh such that $\Delta x_i = k\Delta x_{i-1}$ this is not the case, since $\Delta x_i = k^{i-1}(k-1)/(k^N - 1)$ is not proportional to $1/N$.

5 Order of accuracy in the elliptic case

Next, we consider the case with second derivatives. Equation (14) in discretized form is given in (16). Let $\tilde{M} = M + \tau_0 E_0 + \tau_N E_N$ and rewrite (16) as

$$v = -\tilde{M}^{-1}\left(PF - \tau_0 e_0 g_0 - \tau_N e_N g_N\right). \tag{38}$$

Analogous to the hyperbolic case we get

$$e = -\tilde{M}^{-1}PT_e\ . \tag{39}$$

In order to express the discretization error e in terms of the truncation error T_e we need the inverse of $\tilde{M}$. $\tilde{M}$ is inverted exactly in Appendix D, and the result is

$$\left(\tilde{M}^{-1}\right)_{i0} = \frac{1}{\tau_0}\sum_{k=i+1}^{N}\Delta x_k, \qquad \left(\tilde{M}^{-1}\right)_{iN} = \frac{1}{\tau_N}\sum_{k=1}^{i}\Delta x_k \tag{40}$$

$$\text{For } i \geq j: \qquad \left(\tilde{M}^{-1}\right)_{ij} = -\sum_{k=1}^{j}\Delta x_k \sum_{k=i+1}^{N}\Delta x_k \tag{41}$$

$$\text{For } i \leq j: \qquad \left(\tilde{M}^{-1}\right)_{ij} = -\sum_{k=1}^{i}\Delta x_k \sum_{k=j+1}^{N}\Delta x_k \tag{42}$$

The row index i goes from 0 to N, and so does the column index j. On the diagonal (41) and (42) coincide.

5.1 Computation of the discretization error

We Taylor expand $PT_e = -Mu - PF$ and insert $-u_{xx} = F$, $-u_{xxx} = F_x$. Denote $F^i \equiv F(x_i)$ and use $\Omega_0 = \frac{\Delta x_1}{2} - \xi_1$ and $\Omega_N = \frac{\Delta x_N}{2} + \xi_N$ from (23). That leads to

$$(PT_e)_0 = \left(\xi_1 - \frac{\Delta x_1}{2}\right) F^0 \tag{43}$$

$$(PT_e)_i = (\xi_{i+1} - \xi_i)\, F^i + \left(\frac{\Delta x_{i+1}^2 - \Delta x_i^2}{6}\right) F_x^i + \mathcal{O}(h^3) \tag{44}$$

$$(PT_e)_N = -\left(\xi_N + \frac{\Delta x_N}{2}\right) F^N \ . \tag{45}$$

The error $e_i = -(\tilde{M}^{-1}PT_e)_i$ is computed and divided into one dual mesh dependent part and three primal mesh dependent parts.

$$e_i = e_\xi + e_{\Delta_x} + \mathcal{O}(h^2) \tag{46}$$

where

$$e_\xi = \sum_{j=1}^{N}\xi_j\left((\tilde{M}^{-1})_{ij}F^j - (\tilde{M}^{-1})_{ij-1}F^{j-1}\right) \tag{47}$$

$$e_{\Delta_x} = \underbrace{(\tilde{M}^{-1})_{i0}\frac{\Delta x_1}{2}F^0}_{e_{i0}} + \underbrace{(\tilde{M}^{-1})_{iN}\frac{\Delta x_N}{2}F^N}_{e_{iN}} + \underbrace{\sum_{j=1}^{N-1}(\tilde{M}^{-1})_{ij}\left(\frac{\Delta x_j^2 - \Delta x_{j+1}^2}{6}\right)F_x^j}_{\tilde{e}_{\Delta x}} \tag{48}$$

Here the error contribution e_ξ is dependent on the dual grid, whereas e_{i0}, e_{iN} and $\tilde{e}_{\Delta x}$ depends on the primal grid.

Remark: Remember that we consider Dirichlet boundary conditions. For this particular boundary condition the stability requirement on the penalty parameters is that $\tau_0 \leq -1/(4\Delta x_1)$ and $\tau_N \leq -1/(4\Delta x_N)$ should hold, see equation (12). This means that $\tau_{0,N}$ should be proportional to N.

5.2 Dual mesh dependent error

The dual mesh dependent error contribution e_ξ is given in (47). For a centered dual mesh the contribution is $e_\xi = 0$ since $\xi_j = 0$. If the dual mesh is shifted ($\xi_i = Ch$) we get $e_\xi = Ch((\tilde{M}^{-1})_{iN}F^N - (\tilde{M}^{-1})_{i0}F^0)$ due to cancellation. We have $(\tilde{M}^{-1})_{i0} = (\sum_{k=i+1}^N \Delta x_k)/\tau_0 \sim \mathcal{O}(1)/\tau_0$ so consequently $(\tilde{M}^{-1})_{i0} \sim \mathcal{O}(h)$. The same is found for $(\tilde{M}^{-1})_{iN}$ and hence $e_\xi \sim \mathcal{O}(h^2)$ in the shifted grid case.

We continue by looking at the case with random dual mesh.

Proposition 5.1. *Consider solving the problem* (14) *using* (16) *with the dual mesh being random, i.e.* $\xi_i = kr_i\Delta x_i$, *where* $r_i \in [-1, 1]$ *is uniformly distributed and the constant* $|k| < 1/2$. *Assume that Assumption 3.1 holds. Then the error contribution from the dual mesh* e_ξ *has the following properties*

$$E(e_\xi) = 0, \qquad \sigma(e_\xi) \sim \mathcal{O}(h^{1.5}) \tag{49}$$

i.e. a random dual mesh error contributes statistically with an 1.5 order error.

Proof of proposition 5.1. From (46) we have

$$e_\xi = \sum_{j=1}^{N} \xi_j \left(G_{ij} - G_{ij-1}\right) \tag{50}$$

where we have defined $G_{ij} = (\tilde{M}^{-1})_{ij}F^j$. The expected value of e_ξ equals zero, i.e. $E(e_\xi) = 0$, since $E(\xi_j) = 0$ and G_{ij} being independent of ξ_j. To find the statistically most probable error we compute the standard deviation. We need

$$e_\xi^2 = \sum_{j=1}^{N} \xi_j^2 (G_{ij} - G_{ij-1})^2 + \sum_{j \neq n} \xi_j \xi_n \left(G_{ij} - G_{ij-1}\right)\left(G_{in} - G_{in-1}\right) \ . \tag{51}$$

Using the relations from (29) we get $E(\xi_j^2) = k^2 E(r_j^2) E(\Delta x_j^2) = k^2 m_2/(3N^2) \sim \mathcal{O}(h^2)$ and $E(\xi_j\xi_n) = 0$. From Appendix E we have $(G_{ij} - G_{ij-1})^2 \sim \mathcal{O}(h^2)$. We obtain

$$E(e_\xi^2) = \mathcal{O}(h^3) \tag{52}$$

$$\sigma(e_\xi) = \sqrt{E(e_\xi^2) - E(e_\xi)^2} = \mathcal{O}(h^{1.5}) \ . \tag{53}$$

Hence the dual mesh error is on average proportional to $\mathcal{O}(h^{1.5})$. □

5.3 Primal mesh dependent error

The primal mesh dependent error was split into three parts as $e_{\Delta_x} = e_{i0} + e_{iN} + \tilde{e}_{\Delta x}$. First consider the boundary errors e_{i0} and e_{iN} from (48). Recalling that $\tau_0 \sim 1/\Delta x_1$, we get

$$e_{i0} = \frac{1}{\tau_0}\left(\sum_{k=i+1}^{N} \Delta x_k\right)\frac{\Delta x_1}{2}F^0 \ \sim \ \Delta x_1^2 \tag{54}$$

and in the same way $e_{iN} \sim \Delta x_N^2$.

We now take a look at $\tilde{e}_{\Delta x}$. If the mesh is equidistant ($\Delta x_i = h$), it is easily seen that $\tilde{e}_{\Delta x} = 0$. For a general primal grid we can write

$$\tilde{e}_{\Delta x} = \frac{\Delta x_1^2}{6} G'_{i1} - \frac{\Delta x_N^2}{6} G'_{iN-1} + \underbrace{\sum_{j=2}^{N-1} \frac{\Delta x_j^2}{6} \left(G'_{ij} - G'_{ij-1} \right)}_{\zeta} \tag{55}$$

where G'_{ij} is defined as $G'_{ij} = (\tilde{M}^{-1})_{ij} F_x^j$. From Appendix E we have $\left(G'_{ij} - G'_{ij-1}\right) \sim \mathcal{O}(h)$, and hence $\zeta \sim \mathcal{O}(h^2)$. Furthermore G'_{i1} and G'_{iN-1} are proportional to $\mathcal{O}(h)$. Thus the error contribution from $\tilde{e}_{\Delta x}$ will never be worse than $\mathcal{O}(h^2)$, and we conclude that the same holds for e_{Δ_x}.

5.4 Result of the analysis for the elliptic case

The error contribution from the primal mesh was at worst $\mathcal{O}(h^2)$ which is the best possible result.

The main part of the error comes from the positioning of the flux points. Except if the control volumes are centered, since then there is no additional error introduced by the dual mesh. The dual mesh dependent errors are summarized in Table 6.

Dual mesh		**Error part**	**Order**
Notation	Form of ξ_i	e_ξ	of accuracy
Centered	0	0	∞
Shifted	Ch	$\sim Ch^2$	2
Random	$k\Delta x_i r_i$	$\sigma(e_\xi) \sim kh^{1.5}$	1.5

Table 6: Error contribution from the dual mesh when solving a hyperbolic problem.

In table 7 the numerical results shown earlier are listed together with the result from the analysis. Apparently they are in full agreement.

Primal mesh:	uniform	uniform	uniform	random	random
Dual mesh:	centered	shifted	random	centered	random
R.o.C. Analysis	2 $\min(\infty, 2)$	2 $\min(2,2)$	1.5 $\min(1.5, 2)$	2 $\min(\infty, 2)$	1.5 $\min(1.5, 2)$

Table 7: Elliptic case. Order of accuracy for $e = u - v$ when solving $-u_{xx} = F_2$.

6 Advection-diffusion equation

Having analyzed the hyperbolic and elliptic parts separately, it is interesting to see how they combine. One common application for FVM computations is to solve the Navier-Stokes equations where both the hyperbolic and elliptic operators are involved.

The corresponding one-dimensional model problem in the steady case is written

$$au_x = \varepsilon u_{xx} + F(x), \qquad u(0) = g_0, \qquad u(1) = g_1, \tag{56}$$

which we can solve by implementing the discrete formulation

$$aP^{-1}Qv = \varepsilon P^{-1}Mv + F + \tau_0 P^{-1}e_0(v_0 - g_0) + \tau_N P^{-1}e_N(v_N - g_N). \tag{57}$$

The truncation error T_e is obtained by putting the exact solution u into the numerical scheme. We write $PT_e = aQu - \varepsilon Mu - PF$ which using Taylor expansion becomes

$$\begin{aligned}
(PT_e)_0 &= \xi_1 F^0 + \varepsilon \frac{\Delta x_1}{2} u^0_{xx} + \mathcal{O}(h^2) \\
(PT_e)_i &= (\xi_{i+1} - \xi_i) F_i + \left(\Delta x^2_{i+1} - \Delta x^2_i\right)\left(\frac{a}{4}u^i_{xx} - \frac{\varepsilon}{6}u^i_{xxx}\right) + \mathcal{O}(h^3) \\
(PT_e)_N &= -\xi_N F_N + \varepsilon \frac{\Delta x_N}{2} u^N_{xx} + \mathcal{O}(h^2)
\end{aligned} \tag{58}$$

Hence $(T_e)_0$ and $(T_e)_N$ are always are of order $\mathcal{O}(1)$, whereas the inner points $(T_e)_i$ have either order $\mathcal{O}(1)$, $\mathcal{O}(h)$ or even $\mathcal{O}(h^2)$, depending on the function F and the mesh.

To be more precise, for a 'uniform, centered' and a 'uniform, shifted' mesh we have an inner order of $\mathcal{O}(h^2)$, and for a 'uniform, random' mesh the order is $\mathcal{O}(F)+\mathcal{O}(h^2)$. If the mesh is 'random, centered' the order will be $\mathcal{O}(h)$ and if it is 'random, random' the order will be $\mathcal{O}(F) + \mathcal{O}(h)$. We call the order of the truncation error p in the interior and q at the boundaries, and get

$$L_2(T_e) \sim \sqrt{h(h^q)^2 + \sum_{i=1}^{N-1} h(h^p)^2} \sim \mathcal{O}(h^{min(p,q+0.5)}) \; . \tag{59}$$

So for a general F we will have

Primal mesh:	uniform	uniform	uniform		random	random	
Dual mesh:	centered	shifted	random		centered	random	
Forcing func. F:	any	any	$F \neq 0$	$F = 0$	any	$F \neq 0$	$F = 0$
p	2	2	0	2	1	0	1
q+0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
R.o.C. $L_2(T_e)$	0.5	0.5	0	0.5	0.5	0	0.5

Table 8: Advection-Diffusion. Order of accuracy for $e = u - v$ when solving $au_x = \varepsilon u_{xx} + F$.

We conclude that for the two meshes with random dual mesh, the order of the truncation error will improve if the source function F is zero. We examine this prediction by looking at the equation

$$au_x = \varepsilon u_{xx}, \qquad u(0) = 1, \quad u(1) = 0 \tag{60}$$

which has the exact solution

$$u(x) = \left(1 - e^{a(x-1)/\varepsilon}\right) / \left(1 - e^{-a/\varepsilon}\right). \tag{61}$$

Implementing and running numerical experiments using the five grids results in

Primal mesh:	uniform	uniform	uniform	random	random
Dual mesh:	centered	shifted	random	centered	random
Rate of Conv. $L_2(e)$	2	2	2	2	2
Rate of Conv. $L_2(T_e)$	0.5	0.5	0.5	0.5	0.5
R.o.C. $L_2(T_e)$(inner points)	2	2	2	1	1

Table 9: Advection-Diffusion. Order of accuracy for $e = u - v$ when solving $au_x = \varepsilon u_{xx}$.

Note that all five grids gives a second order error. Having the forcing term equal to zero is of course a special case, but nevertheless an often used and important one. Keeping the forcing term non-zero gives exactly the same result as the elliptic case, see Table 2.

Primal mesh:	uniform	uniform	uniform	random	random
Dual mesh:	centered	shifted	random	centered	random
Rate of Conv. $L_2(e)$	2	2	1.5	2	1.5
Rate of Conv. $L_2(T_e)$	0.5	0.5	0	0.5	0
R.o.C. $L_2(T_e)$(inner points)	2	2	0	1	0

Table 10: Advection-Diffusion. Order of accuracy for $e = u - v$ when solving $au_x = \varepsilon u_{xx} + F$.

Remark: We see that for the grids with random dual mesh the rate of convergence is improved by 0.5 if having the forcing function $F = 0$. This can be understood by looking at the truncation error in (58), where F_i is multiplied by $\xi_{i+1} - \xi_i$, i.e. the dual mesh perturbation.

7 Summary

The discrete versions of a hyperbolic and a elliptic differential equation with Dirichlet boundary conditions have been analyzed, and the numerical simulations agree perfectly with the analysis.

The corner stone in this analysis is that the discrete versions of the differential operators have been proven possible to invert. The fact that the errors introduced by random meshes are treated with probability theory is also of importance.

The grid design is clearly very important for the rate of convergence, especially for the hyperbolic case (the rate of convergence is 2 for a uniform mesh and 0.5 for a random mesh). For the elliptic problem the convergence rates found were 2 for a uniform mesh and at worst 1.5 for random meshes.

The most significant feature of a ”good” grid is that the control volumes are centered, i.e. that the flux points are positioned right between the solution points. Note that this centering is very easy to create in one dimension, but it is not clear how to achieve the same thing in multiple dimensions.

We also solved the advection-diffusion problem with and without a forcing term. When the forcing term was included the results were exactly the same as for the elliptic case. Without any forcing term the solution was second order accurate for all grids. If this is applicable in multi dimensions this can be of great interest when solving for example the Navier-Stokes equations.

A Penalty parameters for stability

To guarantee that the numerical scheme in (6) is stable we have to make sure that $\tilde{R} \geq 0$ and that the two quadratic forms in (11) are negative semi-definite. The matrix $\tilde{R} = R - \gamma_0 E_0 + \gamma_N E_N$, where $R = S^{-T} A S^{-1}$, has the following structure

$$\tilde{R} = \begin{bmatrix} \Delta x_1 - \gamma_0 & 0 & & & & & \\ 0 & \frac{1}{\Delta x_2} & \frac{-1}{\Delta x_2} & & & & \\ 0 & \frac{-1}{\Delta x_2} & \frac{1}{\Delta x_2} + \frac{1}{\Delta x_3} & \frac{-1}{\Delta x_3} & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & \frac{-1}{\Delta x_{N-2}} & \frac{1}{\Delta x_{N-2}} + \frac{1}{\Delta x_{N-1}} & \frac{-1}{\Delta x_{N-1}} & 0 \\ & & & & \frac{-1}{\Delta x_{N-1}} & \frac{1}{\Delta x_{N-1}} & 0 \\ & & & & & 0 & \Delta x_N - \gamma_N \end{bmatrix} . \tag{62}$$

If a matrix is symmetric, diagonally dominant and real, with non-negative diagonal entries, then the matrix is positive definite. For $\tilde{R}$ this will be fulfilled if $\tilde{R}_{ii} \geq \sum_{j \neq i} |\tilde{R}_{ij}|$ for every row i, which means that if (63) holds, then $\tilde{R} \geq 0$.

$$\Delta x_1 - \gamma_0 \geq 0 \qquad\qquad \Delta x_N - \gamma_N \geq 0 \tag{63}$$

Further, the two quadratic forms in equation (11) are negative semi-definite if

$$\begin{aligned} a + 2\tau_0 &\leq 0 & \qquad -a + 2\tau_N &\leq 0 \\ -2\varepsilon\gamma_0 &\leq 0 & -2\varepsilon\gamma_N &\leq 0 \\ \varepsilon^2 &\leq (a + 2\tau_0)(-2\varepsilon\gamma_0) & \varepsilon^2 &\leq (-a + 2\tau_N)(-2\varepsilon\gamma_N) \ . \end{aligned} \tag{64}$$

When implementing we have to choose the penalty parameters τ_0, τ_N according to equation (64). We can treat both boundaries simultaneously with the following equations:

$$-ca + 2\tau \leq 0 \tag{65}$$

$$-2\varepsilon\gamma \leq 0 \tag{66}$$

$$\varepsilon^2 \leq (-ca + 2\tau)(-2\varepsilon\gamma) \tag{67}$$

where $c = -1$ at the left boundary and $c = 1$ at the right boundary. We identify two special cases. First, $\varepsilon = 0$, the hyperbolic case: (65), (66) and (67) reduces to

$$-ca + 2\tau \leq 0 \ . \tag{68}$$

In the continuous counterpart of this case we have to bound the solution at the boundary where $ca \leq 0$. Therefore, if $ca > 0$ we put $\tau = 0$. If $ca < 0$ we instead put $\tau \leq ca/2$. Both scenarios are covered by

$$\tau = k\frac{ca - |a|}{4}, \quad k \geq 1 \ . \tag{69}$$

Secondly, $\varepsilon \neq 0$, the parabolic case: If we rewrite expression (67) using (66) we will get

$$-ca + 2\tau \leq \frac{\varepsilon}{-2\gamma} \leq 0, \tag{70}$$

hence expression (65) is unnecessary. Inserting $\gamma \leq \Delta x$ into equation (70) yields

$$\tau \leq -\frac{\varepsilon}{4\Delta x} + \frac{ca}{2} \tag{71}$$

If $ca > 0$ the penalty parameter τ may equal zero (even thought two boundary conditions are requiered). This can be solved by putting $\tau \leq -\varepsilon/(4\Delta x) - |ca|/2$.

Equation (69) in the hyperbolic case, or equation (71) in the parabolic case, gives the stability conditions on (6).

B Derivation of $\tilde{Q}^{-1}$

We have the difference operator Q including the penalty parameter τ_0 as

$$\tilde{Q} = Q - \tau_0 E_0 = \frac{1}{2}\begin{bmatrix} -1-2\tau_0 & 1 & & & & 0 \\ -1 & 0 & 1 & & & \\ & \ddots & & \ddots & \ddots & \\ & & & -1 & 0 & 1 \\ & 0 & & & -1 & 1 \end{bmatrix}. \tag{72}$$

The inverse $\tilde{Q}^{-1}$ is derived using Gaussian elimination. Start by multiplying by 2.

$$\left(\begin{array}{cccccc|cccccc} -1-2\tau_0 & 1 & & & & & 2 & & & & & \\ -1 & 0 & 1 & & & & & 2 & & & & \\ & -1 & 0 & 1 & & & & & 2 & & & \\ & & \ddots & \ddots & \ddots & & & & & \ddots & & \\ & & & -1 & 0 & 1 & & & & & 2 & \\ & & & & -1 & 1 & & & & & & 2 \end{array}\right) \sim \tag{73}$$

Subtract rows, starting at last row. Assume N even, where $N+1$ is the number of rows.

$$\left(\begin{array}{cccccc|cccccc} -2\tau_0 & & & & & & 2 & -2 & 2 & \cdots & -2 & 2 \\ -1 & 1 & & & & & & 2 & -2 & \cdots & 2 & -2 \\ & -1 & 1 & & & & & & 2 & \cdots & -2 & 2 \\ & & \ddots & \ddots & & & & & & \ddots & \vdots & \vdots \\ & & & -1 & 1 & & & & & & 2 & -2 \\ & & & & -1 & 1 & & & & & & 2 \end{array}\right) \sim \tag{74}$$

Define the help parameter $a = 1/\tau_0$ and write

$$\left(\begin{array}{cccccc|cccccc} 1 & & & & & & -a & a & -a & \cdots & a & -a \\ & 1 & & & & & -a & 2+a & -2-a & \cdots & 2+a & -2-a \\ & & 1 & & & & -a & 2+a & -a & \cdots & a & -a \\ & & & \ddots & & & \vdots & \vdots & \vdots & & \vdots & \vdots \\ & & & & 1 & & -a & 2+a & -a & \cdots & 2+a & -2-a \\ & & & & & 1 & -a & 2+a & -a & \cdots & 2+a & -a \end{array}\right), \tag{75}$$

which, if going back to τ_0, means that

$$\tilde{Q}^{-1} = \frac{1}{\tau_0}\begin{bmatrix} -1 & 1 & -1 & & 1 & -1 \\ -1 & 1 & -1 & \cdots & 1 & -1 \\ -1 & 1 & -1 & & 1 & -1 \\ & \vdots & & & \vdots & \\ -1 & 1 & -1 & \cdots & 1 & -1 \\ -1 & 1 & -1 & & 1 & -1 \end{bmatrix} + 2\begin{bmatrix} 0 & 0 & 0 & & 0 & 0 \\ 0 & 1 & -1 & \cdots & 1 & -1 \\ 0 & 1 & 0 & & 0 & 0 \\ & \vdots & & & \vdots & \\ 0 & 1 & 0 & \cdots & 1 & -1 \\ 0 & 1 & 0 & & 1 & 0 \end{bmatrix}. \quad (76)$$

Again, expression (76) assumes that N is even, where $\tilde{Q}^{-1}$ is $(N+1) \times (N+1)$.

C Derivation of the hyperbolic error $\tilde{e}$

In (22) the hyperbolic error is divided into one constant part e_τ and one part $\tilde{e}$ that varies over the grid points i. The varying error $\tilde{e}_i$ can be written

$$\tilde{e}_i^{even} = \sum_{j=0}^{i} \frac{1-(-1)^j}{2}(PT_e)_j = \sum_{k=1}^{i/2}(PT_e)_{2k-1} \qquad \text{for even } i \qquad (77)$$

$$\tilde{e}_i^{odd} = \sum_{k=1}^{(i+1)/2}(PT_e)_{2k-1} - \sum_{j=i+1}^{N}(-1)^j(PT_e)_j \qquad \text{for odd } i \qquad (78)$$

Just as e_τ the error $\tilde{e}$ can be divided into one dual mesh dependent and one primal mesh dependent part. Recalling (24) yields

$$\tilde{e}_i^{even} = \underbrace{\sum_{k=1}^{i/2}\left(\xi_{2k}-\xi_{2k-1}\right)F^{2k-1}}_{\tilde{e}_\xi^{even}} + \underbrace{\sum_{k=1}^{i/2}\left(\frac{\Delta x_{2k}^2 - \Delta x_{2k-1}^2}{4}\right)F_x^{2k-1}}_{\tilde{e}_{\Delta x}^{even}} + \mathcal{O}(h^2) \qquad (79)$$

and

$$\tilde{e}_i^{odd} = \underbrace{\sum_{k=1}^{(i+1)/2}\left(\xi_{2k}-\xi_{2k-1}\right)F^{2k-1} - \sum_{j=i+1}^{N}(-1)^j(\xi_{j+1}-\xi_j)F^j}_{\tilde{e}_\xi^{odd}} \qquad (80)$$

$$+ \underbrace{\sum_{k=1}^{(i+1)/2}\left(\frac{\Delta x_{2k}^2 - \Delta x_{2k-1}^2}{4}\right)F_x^{2k-1} - \sum_{j=i+1}^{N}(-1)^j\left(\frac{\Delta x_{j+1}^2 - \Delta x_j^2}{4}\right)F_x^j}_{\tilde{e}_{\Delta x}^{odd}} + \mathcal{O}(h^2)\ .$$

First consider the dual mesh (computation of $\tilde{e}_\xi$). For $\xi_i = 0$ we have $\tilde{e}_\xi^{even} = \tilde{e}_\xi^{odd} = 0$. For $\xi_i = Ch$ cancellation yields that $\tilde{e}_\xi^{even} = 0$ and $\tilde{e}_\xi^{odd} = ChF^N$. Hence an $\mathcal{O}(h)$ error is added to the discretization error in every odd grid point i, in the shifted grid case.

For the random dual mesh we assume the control volumes to be distributed such that $\xi_i = kr_i\Delta x_i$, where $r_i \in (-1,1)$ and that $|k| < 1/2$. The relation in (79) is rewritten as

$$\tilde{e}_{\xi}^{even} = k\sum_{j=1}^{i}(-1)^j r_j \Delta x_j F^{\varphi(j)} \tag{81}$$

where $\varphi(j) = j - (1+(-1)^j)/2$. The expected value $E(\tilde{e}_{\xi}^{even}) = 0$ since $E(r_j) = 0$. We compute the deviation from the mean value, $\sigma(\tilde{e}_{\xi}^{even})$ by

$$(\tilde{e}_{\xi}^{even})^2 = k^2\left(\sum_{j=1}^{i} r_j^2 \Delta x_j^2 \left(F^{\varphi(j)}\right)^2 + \sum_{n\neq j}(-1)^{n+j} r_n r_j \Delta x_n \Delta x_j \left(F^{\varphi(n)}\right)\left(F^{\varphi(j)}\right)\right) . \tag{82}$$

From (29) we have $E(r_j^2) = 1/3$ and $E(r_n r_j) = 0$. Using Assumption 3.1 this yields

$$E((\tilde{e}_{\xi}^{even})^2) = \frac{k^2}{3}\sum_{j=1}^{i} E(\Delta x_j^2)\left(F^{\varphi(j)}\right)^2 = \frac{k^2 m_2 i\varphi}{3N^2} \sim k^2\mathcal{O}(h) \tag{83}$$

where φ is a constant depending on F. The standard deviation $\sigma(\tilde{e}_{\xi}^{even})$ is obtained by taking the square root of the variance. Hence we have

$$E(\tilde{e}_{\xi}^{even}) = 0 \qquad\qquad \sigma(\tilde{e}_{\xi}^{even}) \sim k\mathcal{O}(h^{0.5}) . \tag{84}$$

Combining this derivation with the methods in Proposition 4.1 will give us that the same result holds for $\tilde{e}_{\xi}^{odd}$.

Next consider the primal mesh (computation of $\tilde{e}_{\Delta x}$). For $\Delta x_i = h$ we will have $\tilde{e}_{\Delta x}^{even} = 0$ and $\tilde{e}_{\Delta x}^{odd} = h^2 F_x^N/4$ due to cancellation, and we conclude that an equidistant primal mesh contributes with an $\mathcal{O}(h^2)$ error in every odd point i. For the random case we consider $\tilde{e}_{\Delta x}^{even}$ in (79) and $\tilde{e}_{\Delta x}^{odd}$ in (80). Using Assumption 3.1 we obtain $E(\tilde{e}_{\Delta x}^{even}) = 0$ whereas $E(\tilde{e}_{\Delta x}^{odd}) = F_x^N m_2/(4N^2)$. In order to compute the standard deviations we take the squares of $\tilde{e}_{\Delta x}^{even}$ and get

$$(\tilde{e}_{\Delta x}^{even})^2 = \frac{1}{16}\sum_{j=1}^{i}\Delta x_j^4\left(F_x^{\varphi(j)}\right)^2 + \underbrace{\frac{2}{16}\sum_{n\neq j}(-1)^{n+j}\Delta x_n^2 \Delta x_j^2\left(F_x^{\varphi(n)}\right)\left(F_x^{\varphi(j)}\right)}_{\zeta} \tag{85}$$

where $\varphi(j) = j - (1+(-1)^j)/2$. Here ζ consists of $i(i-2)/4$ terms with positive sign and $i^2/4$ negative terms. Using (33) we achieve

$$E((\tilde{e}_{\Delta x}^{even})^2) = \frac{1}{16}\frac{m_4}{N^4}\sum_{j=1}^{i}\left(F_x^{\varphi(j)}\right)^2 - \frac{2}{16}\frac{m_2^2}{N^4}\sum_{n\neq j}(-1)^{n+j}\left(F_x^{\varphi(n)}\right)\left(F_x^{\varphi(j)}\right) \tag{86}$$

which we recognize from Proposition 4.2 to be proportinal to $i/N^4 \sim \mathcal{O}(h^3)$. Using the same methods we will obtain $E((\tilde{e}_{\Delta x}^{odd})^2) \sim \mathcal{O}(h^3)$. Then

$$\sigma(\tilde{e}_{\Delta x}^{even}) = \sqrt{\mathcal{O}(h^3)} = \mathcal{O}(h^{1.5}) \tag{87}$$

$$\sigma(\tilde{e}_{\Delta x}^{odd}) = \sqrt{\mathcal{O}(h^3) - \mathcal{O}(h^4)} = \mathcal{O}(h^{1.5}) . \tag{88}$$

Thus it holds that

$$E(\tilde{e}_{\Delta x}) \sim \mathcal{O}(h^2) \qquad\qquad \sigma(\tilde{e}_{\Delta x}) \sim \mathcal{O}(h^{1.5}) \tag{89}$$

i.e. the primal mesh contribution to the discretization error is on average $\tilde{e}_{\Delta x} \sim \mathcal{O}(h^{1.5})$.

D Derivation of $\tilde{M}^{-1}$

Define $c_i = 1/\Delta x_i$. Then we can write $\tilde{M}$ as

$$\tilde{M} = \begin{bmatrix} \tau_0 & & & & & \\ c_1 & -c_1-c_2 & c_2 & & & \\ & \ddots & \ddots & \ddots & & \\ & & & c_{N-1} & -c_{N-1}-c_N & c_N \\ & & & & & \tau_N \end{bmatrix}. \tag{90}$$

First, consider only the inner 1 to $N-1$ rows (ignore row 0 and N for a while)

$$\left(\begin{array}{cccccc|cccc} c_1 & -c_1-c_2 & c_2 & & & & 0 & 1 & & \\ & \ddots & \ddots & \ddots & & & & & \ddots & \\ & & & c_{N-1} & -c_{N-1}-c_N & c_N & & & 1 & 0 \end{array}\right) \sim$$

and start Gaussian elimination.

$$\left(\begin{array}{ccccccc|ccccccc} c_1 & -c_1-c_2 & c_2 & & & & & 0 & 1 & & & & & \\ c_1 & -c_1 & -c_3 & c_3 & & & & 0 & 1 & 1 & & & & \\ c_1 & -c_1 & 0 & -c_4 & c_4 & & & 0 & 1 & 1 & 1 & & & \\ \vdots & \vdots & & & \ddots & \ddots & & & \vdots & & & \ddots & & \\ c_1 & -c_1 & & & & -c_N & c_N & 0 & 1 & & \dots & & 1 & 0 \end{array}\right) \sim$$

Divide by c_{i+1}

$$\left(\begin{array}{ccccccc|ccccccc} \frac{c_1}{c_2} & -\frac{c_1}{c_2}-1 & 1 & & & & & 0 & \Delta x_2 & & & & & \\ \frac{c_1}{c_3} & -\frac{c_1}{c_3} & -1 & 1 & & & & 0 & \Delta x_3 & \Delta x_3 & & & & \\ \frac{c_1}{c_4} & -\frac{c_1}{c_4} & 0 & -1 & 1 & & & 0 & \Delta x_4 & \Delta x_4 & \Delta x_4 & & & \\ \vdots & \vdots & & & \ddots & \ddots & & & \vdots & & & \ddots & & \\ \frac{c_1}{c_N} & -\frac{c_1}{c_N} & & & & -1 & 1 & 0 & \Delta x_N & & \dots & & \Delta x_N & 0 \end{array}\right) \sim$$

and add the rows together (first row is the sum of all rows, last row is the sum of itself).

$$\left(\begin{array}{ccccccc|ccccccc} c_1 \sum_2^N & -c_1 \sum_1^N & & & & & 1 & 0 & \sum_2^N & \sum_3^N & \sum_4^N & \dots & \sum_N^N & 0 \\ c_1 \sum_3^N & -c_1 \sum_3^N & -1 & & & & 1 & 0 & \sum_3^N & \sum_3^N & \sum_4^N & & & \\ c_1 \sum_4^N & -c_1 \sum_4^N & 0 & -1 & & & 1 & 0 & \sum_4^N & \sum_4^N & \sum_4^N & & & \\ \vdots & \vdots & & & \ddots & & \vdots & & \vdots & & & & & \\ c_1 \sum_N^N & -c_1 \sum_N^N & & & & -1 & 1 & 0 & \sum_N^N & \sum_N^N & \dots & & \sum_N^N & 0 \end{array}\right) \sim$$

Here we have written $\sum_i^N$ instead of $\sum_{k=i}^N \Delta x_k$ in order to save space. We use the fact that $\sum_{k=1}^N \Delta x_k = 1$ and that $1 - \sum_{k=i+1}^N \Delta x_k = \sum_{k=1}^i \Delta x_k$. Multiply the first row by $\sum_{k=i+1}^N \Delta x_k$ and subtract it from row i. Thereafter multiply first row by Δx_1. We get

$$\left(\begin{array}{ccccccc|cccccc} \sum_2^N & -1 & & & & & \sum_1^1 & 0 & \sum_2^N \sum_1^1 & \sum_3^N \sum_1^1 & & \sum_N^N \sum_1^1 & 0 \\ \sum_3^N & & -1 & & & & \sum_1^2 & 0 & \sum_3^N \sum_1^1 & \sum_3^N \sum_1^2 & & \sum_N^N \sum_1^2 & 0 \\ \sum_4^N & & & -1 & & & \sum_1^3 & 0 & \sum_4^N \sum_1^1 & \sum_4^N \sum_1^2 & & \sum_N^N \sum_1^3 & 0 \\ \vdots & & & & \ddots & & \vdots & & \vdots & & & \vdots & \vdots \\ \sum_N^N & & & & & -1 & \sum_1^{N-1} & 0 & \sum_N^N \sum_1^1 & \sum_N^N \sum_1^2 & \dots & \sum_N^N \sum_1^{N-1} & 0 \end{array}\right). \tag{91}$$

Now we need the first and last row from the whole, "original", matrix (90). Combining them with row 1 to $N-1$ in (91), we get back to the full system. Note that the mid-rows differ below and above the diagonal.

$$\left(\begin{array}{ccc|ccc} \tau_0 & & & 1 & & \\ \sum_{i+1}^{N} & -I & \sum_{1}^{i} & 0 & \sum_{i+1}^{N}\sum_{1}^{j} \setminus \sum_{j+1}^{N}\sum_{1}^{i} & 0 \\ & & \tau_N & & & 1 \end{array}\right) . \tag{92}$$

Multiply first and last row in (92) by $\sum_{i+1}^{N}/\tau_0$ and $\sum_{1}^{i}/\tau_N$, respectively. Subtract them from all the mid-rows. We obtain $\tilde{M}^{-1}$ as

$$\text{Left column } (j=0): \qquad \left(\tilde{M}^{-1}\right)_{i0} = \frac{1}{\tau_0}\sum_{k=i+1}^{N} \Delta x_k \tag{93}$$

$$\text{Right column } (j=N): \qquad \left(\tilde{M}^{-1}\right)_{iN} = \frac{1}{\tau_N}\sum_{k=1}^{i} \Delta x_k \tag{94}$$

$$\text{Below diagonal } (i \geq j): \qquad \left(\tilde{M}^{-1}\right)_{ij} = -\sum_{k=1}^{j} \Delta x_k \sum_{k=i+1}^{N} \Delta x_k \tag{95}$$

$$\text{Above diagonal } (i \leq j): \qquad \left(\tilde{M}^{-1}\right)_{ij} = -\sum_{k=1}^{i} \Delta x_k \sum_{k=j+1}^{N} \Delta x_k \tag{96}$$

E The order of G_{ij}

We have defined $G_{ij} = (\tilde{M}^{-1})_{ij}F^j$. Assume $i \leq j$.

$$(\tilde{M}^{-1})_{ij} - (\tilde{M}^{-1})_{ij-1} = \Delta x_j \sum_{k=1}^{i} \Delta x_k \;\sim\; \Delta x_j \tag{97}$$

$$(\tilde{M}^{-1})_{iN} - (\tilde{M}^{-1})_{iN-1} = \left(\frac{1}{\tau_N} + \Delta x_N\right) \sum_{k=1}^{i} \Delta x_k \;\sim\; \Delta x_N \tag{98}$$

where we have used that $\tau_N \sim 1/\Delta x_N$. In the same way $(\tilde{M}^{-1})_{i1} - (\tilde{M}^{-1})_{i0} \sim \Delta x_1$. That is, $(\tilde{M}^{-1})_{ij} - (\tilde{M}^{-1})_{ij-1} \sim \Delta x_j$ for all j. The same result is also found for $i \geq j$. Together with $(F^j - F^{j-1}) = \Delta x_j F_x^j + \mathcal{O}(h^2)$ this yields

$$G_{ij} - G_{ij-1} = (\tilde{M}^{-1})_{ij}F^j - \left((\tilde{M}^{-1})_{ij} + \mathcal{O}(\Delta x_j)\right)\left(F^j + \mathcal{O}(\Delta x_j)\right) \;\sim\; \Delta x_j \; . \tag{99}$$

Hence $G_{ij} - G_{ij-1} \sim h$.

References

[1] M. H. Carpenter, J. Nordstrom, and D. Gottlieb. A stable and conservative interface treatment of arbitrary spatial accuracy. *Journal of Computational Physics*, (148):341–365, 1999.

[2] B. Diskin and J. L. Thomas. Accuracy analysis for mixed-element finite-volume discretization schemes. NIA Report 2007-08, August 2007.

[3] P. Eliasson. EDGE, a Navier-Stokes solver for unstructured grids. In *Finite Volumes for Complex Applications III*, pages 527–534, 2002.

[4] T. Gerhold, O. Friedrich, and J. Evans. Calculation of complex three-dimensional configurations employing the dlr-tau-code. *in: AIAA paper 97-0167*, 1997.

[5] B. Gustafsson, H. O. Kreiss, and J. Oliger. *Time Dependent Problems and Difference Methods*. Wiley-Interscience, 1995.

[6] A. Haselbacher, J. J. McGuirk, and G. J. Page. Finite volume discretization aspects for viscous flows on mixed unstructured grids. *AIAA Journal*, 37(2):177–184, 1999.

[7] K. Mattsson and J. Nordström. Summation by parts operators for finite difference approximations of second derivatives. *Journal of Computational Physics*, (199):503–540, 2004.

[8] D.J. Mavriplis and D.W. Levy. Transonic drag prediction using an unstructured multigrid solver. In *Technical report, Institute for Computer Applications in Science and Engineering*, 2002.

[9] E.J. Nielsen and W.K. Anderson. Recent improvements in aerodynamic design optimization on unstructured meshes. In *AIAA-2001-0596*, 2001.

[10] J. Nordström, K. Forsberg, C. Adamsson, and P. Eliasson. Finite volume methods, unstructured meshes and strict stability for hyperbolic problems. *Applied Numerical Mathematics*, 45(4):453–473, 2003.

[11] D. Schwamborn, T. Gerhold, and R. Heinrich. The DLR TAU-code: Recent applications in research and industry. In *European Conference on Computational Fluid Dynamics, ECCOMAS CFD*, 2006.

[12] M. Svärd, J. Gong, and J. Nordström. An accuracy evaluation of unstructured node-centered finite-volume methods. *Applied Numerical Mathematics*, 58:1142–1158, 2008.

[13] M. Svärd and J. Nordström. Stability of finite volume approximations for the laplacian operator on quadrilateral and triangular grids. *Applied Numerical Mathematics*, 51(1):101–125, 2004.

[14] J. L. Thomas, B. Diskin, and C. L. Rumsey. Towards verification of unstructured-grid solvers. *AIAA Journal*, 46(12):3070–3079, 2008.

[15] L. Tysell and J. Nordström. Accuracy evaluation of the unstructured finite volume method in aerodynamic computations. In *10th ISGG Conference on Numerical Grid Generation*, page 13, Curran Associates, Red Hook, NY 2007.

[16] J.M. Weiss, J.P. Maruszewski, and W.A. Smith. Implicit solution of preconditioned Navier-Stokes equations using algebraic multigrid. *AIAA Journal*, 37(1):29–36, 1999.