Project 2 TMA4212 - Finite element methods for a fluid model

Reidar Bråthen Kristoffersen, Viktor Sandve, Trond Skaret Johansen

September 30, 2024

Abstract

In this paper we present or findings from numerically solving a fluid model, using finite element methods. We present some theoretical analysis. The numerical experiments include error study, non-smooth solutions and a grid refinement. We observe that non-smooth solutions significantly reduces the convergence rate, while grid refinement can reduce errors in specific scenarios.

1 Introduction

Our goal is to study a convective and diffusive substance in one dimension. Combining conservation of mass, Fick's law for diffusion, and Reynold's transport theorem we obtain the following model:

$$-\partial_x(\alpha(x)\partial_x u) + \partial_x(b(x))u + c(x)u = f(x) \quad \text{in} \quad \Omega = (0,1). \tag{1}$$

Here u is the concentration, $\alpha(x) > 0$ is the diffusion coefficient, $b(x)$ the convective/fluid velocity, $c(x) \geq 0$ is the decay rate of the substance, and $f(x)$ is a source term. The problem is scaled to the domain (0, 1), and there is no time dependence since we look at the stationary time limit.

Finite element methods are a powerful tool for solving such equations, providing flexibility in step size and ease of implementation compared to finite difference methods. Several results are proved, which are used to conclude with a uniqueness result of the solution. We derive an error bound and convergence rate of the $\mathbb{P}_1$ finite element method scheme, given sufficiently smooth solutions. Finally we experiment with non-smooth solutions and tactical variations in step size.

2 Theory and methods

2.1 The weak formulation

Our first goal is to rewrite the equation to the weak form that any classical solution of (1) must satisfy:

$$a(u,v) = F(v) \quad \forall\, v \in V. \tag{2}$$

Where V is the space:

$$V = H_0^1(0,1) = \{f : f, f' \in L^2(0,1), f(0) = 0 = f(1)\}.$$

To obtain the weak form, multiply each side of (1) by some $v \in H_0^1(0,1)$ and integrate over $(0,1)$ using integration by parts to obtain:

$$\begin{aligned} -\int_0^1 (\alpha u_x)_x v \,\mathrm{d}x + \int_0^1 (bu)_x v \,\mathrm{d}x + \int_0^1 cuv dx &= \int_0^1 fv \,\mathrm{d}x \\ -[\alpha u_x v]_0^1 + \int_0^1 \alpha u_x v_x \,\mathrm{d}x + [buv]_0^1 - \int_0^1 buv_x \,\mathrm{d}x + \int_0^1 cuv \,\mathrm{d}x &= \int_0^1 fv \,\mathrm{d}x \\ \underbrace{\int_0^1 (\alpha u_x v_x - buv_x + cuv) \,\mathrm{d}x}_{=:a(u,v)} &= \underbrace{\int_0^1 fv \,\mathrm{d}x.}_{=:F(v)} \end{aligned}$$

The evaluations vanished since $v(0) = 0 = v(1)$ for $v \in H_0^1(0,1)$. Having obtained the weak form, we can prove some statements.

Proposition 1. *$a(u,v)$ is a bilinear and continuous form on $H^1 \times H^1$.*

Proof. First to show bilinearity of $a(u,v)$:

$$\begin{aligned} a(k_1u_1 + k_2u_2, v) &= \int_0^1 \alpha(x)(k_1u_1 + k_2u_2)_x v_x - b(x)(k_1u_1 + k_2u_2)v_x + c(k_1u_1 + k_2u_2)v \,\mathrm{d}x \\ &= k_1 \int_0^1 \alpha(x)(u_1)_x v_x - b(x)u_1 v_x + cu_1 v \,\mathrm{d}x + k_2 \int_0^1 \alpha(x)(u_2)_x v_x - b(x)u_2 v_x + cu_2 v \,\mathrm{d}x \\ &= k_1 a(u_1, v) + k_2 a(u_2, v). \end{aligned}$$

Similar steps can be done to show $a(u, k_1v_1 + k_2v_2) = k_1a(u,v_1) + k_2a(u,v_2)$. I.e. $a(u,v)$ is bilinear. Next we show continuity:

$$\begin{aligned} a(u,v) &= \int_0^1 \alpha(x)u_x v_x - b(x)uv_x + cuv \,\mathrm{d}x \\ &\leq \|\alpha(x)\|_{L^\infty} \|u_x\|_{L^2} \|v_x\|_{L^2} + \|b(x)\|_{L^\infty} \|u\|_{L^2} \|v_x\|_{L^2} + c\|u\|_{L^2} \|v\|_{L^2} \\ &\leq (\|\alpha(x)\|_{L^\infty} + \|b(x)\|_{L^\infty} + c)\|u\|_{H^1} \|v\|_{H^1} = M\|u\|_{H^1} \|v\|_{H^1}. \end{aligned}$$

In the first inequality we use Cauchy Schwartz' inequality, and in the last we use the fact that $\|u\|_{L^2} \leq \|u\|_{H^1}$ and $\|u_x\|_{L^2} \leq \|u\|_{H^1}$, which follows directly from the definition of $\|\cdot\|_{H^1}$.

This means that $a(u,v)$ is a bilinear and continuous form on $H^1 \times H^1$. □

Proposition 2. *$F(v)$ is a linear and continuous(bounded) functional on H^1.*

Proof. We first write:

$$F(u+v) = \int_0^1 f(x)(u(x) + v(x)) \,\mathrm{d}x = \int_0^1 f(x)u(x) \,\mathrm{d}x + \int_0^1 f(x)v(x) \,\mathrm{d}x = F(u) + F(v).$$

Meaning F is linear. To show that F is bounded we compute the operator norm of F:

$$\|F\| = \max_{0 \neq v \in H^1} \frac{|F(v)|}{\|v\|_{H^1}} = \max_{0 \neq v \in H^1} \frac{|\int_0^1 fv \,\mathrm{d}x|}{\|v\|_{H^1}} \leq \max_{0 \neq v \in H^1} \frac{\|f\|_{L^2} \|v\|_{L^2}}{\|v\|_{H^1}} \leq \|f\|_{L^2}.$$

Where we in the last inequality use $\|v\|_{H^1} = \|v\|_{L^2} + \|v_x\|_{L^2} \geq \|v\|_{L^2}$. Since F is a bounded functional, it is also continuous, indeed:

$$|F(v) - F(v_0)| = |F(v - v_0)| \leq \|f\|_{L^2}\|v - v_0\|_{H^1}.$$

□

Lemma 1. (Gårding inequality) *a satisfies the following Gårding inequality:*

$$a(u,u) \geq (\alpha_0 - \frac{\varepsilon}{2}\|b\|_{L^\infty})\|u_x\|_{L^1}^2 + (c_o - \frac{1}{2\varepsilon}\|b\|_{L^\infty})\|u\|_{L^1}^2, \quad \forall \varepsilon > 0.$$

Proof. Let $\varepsilon > 0$ and put $\alpha_0 = \min_{x\in[0,1]} \alpha(x)$ and $c_0 = \min_{x\in[0,1]} c(x)$. Then

$$\begin{aligned} a(u,u) &= \int_0^1 \alpha(x)u_x^2 - b(x)uu_x + c(x)u^2 \,\mathrm{d}x \geq \int_0^1 \alpha_0 u_x^2 - \|b\|_{L^\infty}(\frac{1}{2\varepsilon}u^2 + \frac{\varepsilon}{2}u_x^2) + c_0 u^2 \,\mathrm{d}x \\ &= (\alpha_0 - \frac{\varepsilon}{2}\|b\|_{L^\infty}) \int_0^1 u_x^2 \,\mathrm{d}x + (c_o - \frac{1}{2\varepsilon}\|b\|_{L^\infty}) \int_0^1 u^2 \,\mathrm{d}x. \end{aligned}$$

In the inequality we use a variant of Young's inequality which for any $\varepsilon > 0$ reads $ab \leq \frac{1}{2\varepsilon}a^2 + \frac{\varepsilon}{2}b^2$. □

Proposition 3. *a(u,v) is coercive when* $\|b\|_{L^\infty} \leq \sqrt{2\alpha_0 c_0}$.

Proof. Let $\varepsilon = \sqrt{\alpha_0/c_0}$. Using Gårding's inequality we get

$$\begin{aligned} a(u,u) &\geq (\alpha_0 - \frac{\sqrt{\alpha_0/c_0}}{2}\sqrt{2\alpha_0 c_0})\|u_x\|_{L^2}^2 + (c_0 - \frac{1}{2\sqrt{2\alpha_0/c_0}}\sqrt{2\alpha_0 c_0})\|u\|_{L^2}^2 \\ &= \alpha_0 \left(1 - \frac{1}{\sqrt{2}}\right)\|u_x\|_{L^2}^2 + c_0 \left(1 - \frac{1}{\sqrt{2}}\right)\|u\|_{L^2}^2 \geq k\|u\|_{H^1}^2. \end{aligned}$$

Where $k = \frac{2-\sqrt{2}}{2}\min\{\alpha_0, c_0\}$ and we again used that $\|u\|_{H^1}^2 = \|u\|_{L^2}^2 + \|u_x\|_{L^2}^2$. □

We introduce the following theorem without proof, which alows us to conclude when our weak problem has a unique solution.

Theorem 1. (Lax-Milgram) *Let V be a Hilbert space. Suppose that F is a continuous linear functional, and that a is a continuous, coercive bilinear form. Then the variational problem, find u such that $\forall v \in V$, $a(u,v) = F(v)$, admits a unique solution.*

Corollary 1. *When a(u,v) is coercive, Equation* (2) *has a unique solution.*

As we have seen, this is the case when $\|b\|_{L^\infty} \leq \sqrt{2\alpha_0 c_0}$. It is also possible to show that $c > 0$ is sufficient for coercivity.

2.2 $\mathbb{P}_1$ finite element method scheme

The goal of the project is to approximate solutions of 2 in a smaller space $V_h \subset V = H_0^1(0,1)$. Before defining the space, we partition our domain $[0,1]$ with nodes $0 = x_0 < x_1 < \cdots < x_M = 1$. These define the segments $K_i = (x_{i-1}, x_i)$. With this, we define:

$$V_h = \{f \in C^1(0,1) : f|_{K_i} \in \mathbb{P}_1, i = 1, ..., M\} \cap V.$$

This is the space of piecewise polynomials that are 0 on the endpoints. As our approximated solution we take $u_h \in V_h$ satisfying:

$$a(u_h, v_h) = F(v_h) \quad \forall\, v_h \in V_h. \tag{3}$$

To solve for u_h numerically we introduce basis functions φ_i for this space. These are constructed to be linear the intervals (x_{i-1}, x_i) and $(x_i, xi + 1)$. They take the value 1 on x_i and 0 on the other nodes x_j. Examples of these functions are shown in Figure 1. Using these basis functions, any $u \in V_h$ can be written as $u = \sum_{i=1}^{M-1} U_i \varphi_i(x)$ with $U_i = u(x_i)$. Hence, we have reduced the problem to finding $U = [U_1, ..., U_{M-1}]^T$ satisfying $AU = F$, where:

$$A_{i,j} = a(\varphi_j, \varphi_i) = \begin{cases} -\frac{\alpha}{h_i} + \frac{b}{2} + \frac{ch_i}{6}, & j = i-1 \\ \alpha\left(\frac{1}{h_i} + \frac{1}{h_i+1}\right) + c\left(\frac{h_i}{3} + \frac{h_i+1}{3}\right), & j = i \\ -\frac{\alpha}{h_{i+1}} - \frac{b}{2} + \frac{ch_{i+1}}{6}, & j = i+1 \\ 0, & \text{else} \end{cases}$$

$$F_j = F(\varphi_j) = \int_0^1 f(x)\varphi_j(x)\,\mathrm{d}x.$$

F is computed with numerical integration using the trapezoidal rule. As the trapezoidal rule has second order convergence, it will not affect the convergence of the method. A printout of an example matrix A can be found as Figure 5 in the appendix.

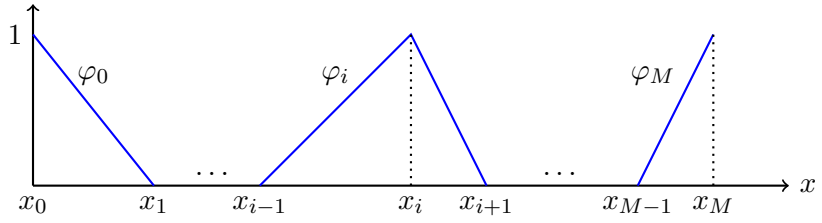


Figure 1: The basis functions φ_0, φ_i and φ_M.

With the implementation set up, we move on to finding an error bound for this method.

Definition 1. A method is said to be Galerkin orthogonal if $u \in V$ and $u_h \in V_h$ solve (2) and (3) respectively implies that $a(u - u_h, v_h) = 0$.

Proposition 4. *$a(u, v)$ is Galerkin orthogonal.*

Proof. Using bilinearity, and the fact that u and u_h solve (2) and (3) respectively, we obtain:

$$a(u - u_h, v_h) = a(u, v_h) - a(u_h, v_h) = F(v_h) - F(v_h) = 0.$$

□

Lemma 2. (Cea's Lemma) *Suppose $a(u, v)$ is bilinear, continuous and coercive and that $u \in V$ and $u_h \in V_h$ solve* (2) *and* (3) *respectively. We then have:*

$$\|u - u_h\|_V \leq \frac{M}{k} \inf_{v_h \in V_h} \|u - v_h\|_V.$$

Proof. Let $v_h \in V_h$, and note that using coercitivity we may obtain:

$$\begin{aligned} k\|u-u_h\|_V^2 &\leq a(u-u_h, u-u_h) \\ &= a(u-u_h, u-v_h+v_h-u_h) \\ &= a(u-u_h, u-v_h) + a(u-u_h, v_h-u_h). \end{aligned}$$

Where we use bilinearity in the second equality. By Galerkin othogonality the second summand is zero. Thus using continuity we have:

$$k\|u-u_h\|_V^2 \leq a(u-u_h, u-v_h) \leq M\|u-u_h\|_V\|u-v_h\|_V.$$

Since this holds $\forall v_h \in V_h$ we may rearange and take the infimum to obtain the result. □

In the lectures [1, Lecture 22] we have seen that:

Lemma 3. *For an interpolating polynomial $v_h \in V_h$ of $v \in V$ the error can be bounded as:*

$$\|v_h - v\|_V \leq \sqrt{2}h\|v_{xx}\|_{L^2}$$

if we impose $h := \max_i h_i < 1$.

Using that this cannot be less than the infimum over all $v_h \in V_h$, we obtain our error bound as:

$$\|u-u_h\|_V \leq \sqrt{2}\frac{M}{k}\|u_{xx}\|_{L^2}h.$$

For our numerical experiments, we have constant α, b, c, so inserting M, k gives the bound:

$$\|u-u_h\|_V \leq 2(\sqrt{2}+1)\frac{a+b+c}{\min\{a,c\}}\|u_{xx}\|_{L^2}h. \tag{4}$$

So the method is expected to converge linearly in H^1-norm, which we have seen is a norm taking into account both v and its gradient.

2.3 Non-smooth solutions

If the solution to our problem is non-smooth, problems quickly arise, as either the first or second derivative could be undefined on the domain. We define the functions

$$w_1(x) = \begin{cases} \frac{x}{\frac{\sqrt{2}}{2}} & , x \in [0, \frac{\sqrt{2}}{2}] \\ \frac{1-x}{1-\frac{\sqrt{2}}{2}} & , x \in (\frac{\sqrt{2}}{2}, 1] \end{cases} \quad \text{and} \quad w_2(x) = x - x^{\frac{3}{4}}.$$

One can see that the first derivative of w_1 is undefined in $\frac{\sqrt{2}}{2}$, which also means the second derivative will not be defined here. Similarly for w_2 the second summand leads to an undefined point at 0 for both the first and second derivative. However one can compute the weak derivative of both:

$$w_1'(x) = \begin{cases} \sqrt{2} & , x \in [0, \frac{\sqrt{2}}{2}] \\ -\frac{1}{1-\frac{\sqrt{2}}{2}} & , x \in (\frac{\sqrt{2}}{2}, 1] \end{cases} \quad \text{and} \quad w_2'(x) = 1 - \frac{3}{4}x^{-\frac{1}{4}}.$$

As both these are square integrable on our domain, we can conclude that w_1 and $w_2 \in H^1$. However the weak derivatives of w_1' does not exist and for w_2' we have $w_2''(x) = \frac{3}{16}x^{-\frac{5}{4}} \notin L^2$. Thus w_1 and w_2 are in H^1, but not H^2. To see how this affects the convergence of the method, we construct the right hand side of (1) by computing the left hand side with $u = w_i$, using integration by parts to compute the $u_{xx}v$ term. We can now solve the model with the finite element method as described above.

2.4 Graded grids

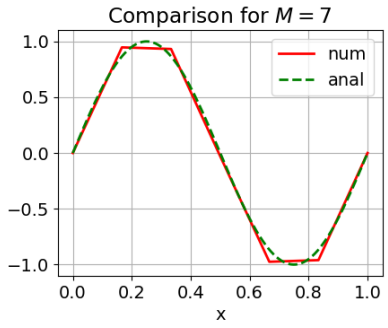
For $x \in (0,1)$ we define the functions $f_1(x) = x^{-\frac{2}{5}}$ and $f_2(x) = x^{-\frac{7}{5}}$. We note that $f_1 \in L^2(0,1)$, $f_2 \notin L^2(0,1)$, and $(f_2)_x \in L^2(0,1)$. Also, $f_2 \in H^{-1}(0,1)$. We will see that solution with such right hand sides have sharp gradients near $x = 0$. Uniform grids therefore might have poor convergence here, so we experiment with refining the grid near this point. Our strategy is to take $r \in (0,1)$ and define our $M+1$ nodes by $x_0 = 0$, $x_i = r^{M-i}$. We experiment with different r, and immediately remark that smaller r correspond to more nodes near $x = 0$. This grid will be referred to as the *graded grid*.

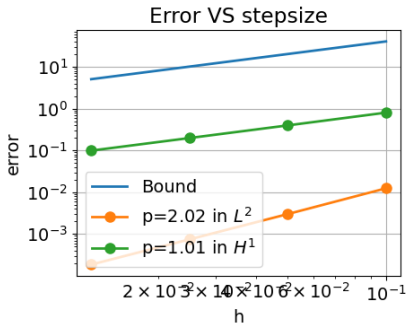
For error analysis, we use a very fine grid to get a reference solution. We also include the grid points of the poorer solutions to compare in these points.

3 Results and discussion

3.1 Convergence experiment

To test the convergence of the method, we use the manufactured solution $u(x) = \sin(2\pi x)$ with appropriate right hand side f. The results of a convergence study are shown in Figure 2. From the plot we see that the results are consistent with the error bound in Equation (4). The rate in L^2 is better than H^1, which is expected as the function is approximated by a piecewice linear function while the derivative contributing to the H^1-norm is approximated by a piecewise constant function.





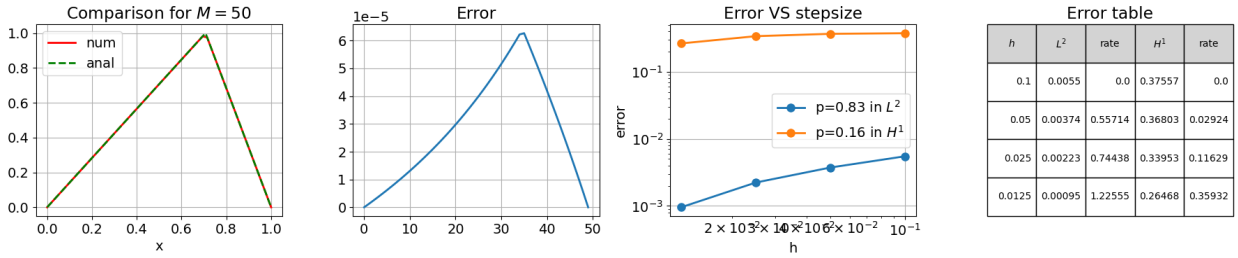
Error table

h	L^2	rate	H^1	rate
0.1	0.01238	0.0	0.8007	0.0
0.05	0.003	2.04746	0.39632	1.0146
0.025	0.00074	2.01194	0.19761	1.00402
0.0125	0.00019	2.00295	0.09876	1.00064

Figure 2: Convergence for the manufactured solution $u(x) = \sin(2\pi x)$ using parameters $a = b = c = 1$.

3.2 Reduced order for non-smooth solutions

As discussed in Section 2.3, non-smooth solutions can affect the convergence. To test this we use manufactured solutions w_1 and w_2. The results from a convergence study of w_1 are given in Figure 3. The one for w_2 is similar, and can be found in the appendix.



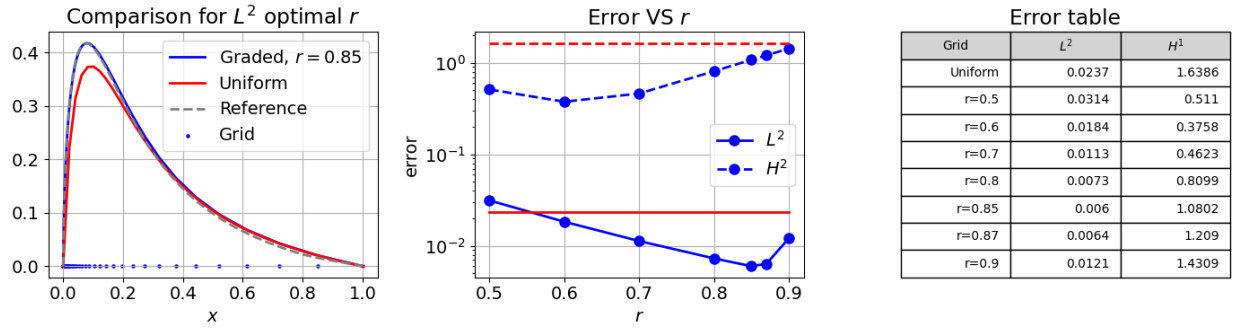
h	L^2	rate	H^1	rate
0.1	0.0055	0.0	0.37557	0.0
0.05	0.00374	0.55714	0.36803	0.02924
0.025	0.00223	0.74438	0.33953	0.11629
0.0125	0.00095	1.22555	0.26468	0.35932

Figure 3: Convergence study for w_1 using $a = b = c = 1$.

As expected the error is largest near the point where the derivative is undefined. The convergence rate in both norms are also significantly worse than in the previous example. This is also to be expected as the $||u_{xx}||_{L^2}$ term in the error bound, in Equation (4), is unbounded for these non-smooth functions. This also means that we cannot theoretically guarantee convergence, even though it does seem to be the case here. Drawing further conclusions as to whether this holds in general would require further study.

3.3 Graded grids

Our final experiment was with the use of graded grids. The results from this experiment using righ-hand-side f_2 are shown in Figure 4. The plots all show that refining the grid in this case gives a more accurate solution in both norms. We note that the optimal r for H^1 is much lower, which is expected as the derivative is small for large x and very large near $x = 0$, so there is more to be gained for the derivative compared to the solution.



Grid	L^2	H^1
Uniform	0.0237	1.6386
r=0.5	0.0314	0.511
r=0.6	0.0184	0.3758
r=0.7	0.0113	0.4623
r=0.8	0.0073	0.8099
r=0.85	0.006	1.0802
r=0.87	0.0064	1.209
r=0.9	0.0121	1.4309

Figure 4: Experiment with different r-values for the right-hand-side f_2 using $a = c = 1$ and $b = -10$.

4 Conclusion

In this paper we analyzed the use of finite element methods to model 1-dimensional diffusion. We have found both analytically and experimentally experimentally a linear convergence in the H^1 norm. Experiments show quadratic convergence in the L^2 norm. This convergence seems to be severely lessened in the case of non-smooth solutions. However using one of the main advantages of the finite element methods, namely the ease of implementing a varying step length, one can achieve better results. We analysed situations where the solution had rapid change around $x = 0$ and observed smaller errors for specific congestions of nodes near origo.

References

[1] Jakobsen, E. R. *Various lecture notes for the course TMA4212*. NTNU, 2024.

5 Appendix

5.1 The matrix A

```
[[12. -5.  0.  0.  0.]
 [-7. 12. -5.  0.  0.]
 [ 0. -7. 12. -5.  0.]
 [ 0.  0. -7. 12. -5.]
 [ 0.  0.  0. -7. 12.]]
```

Figure 5: Printout of the matrix A in the case $a = 1, b = 2, c = 3$ for a grid with 7 nodes. Rounded to no decimals.

5.2 Additional results

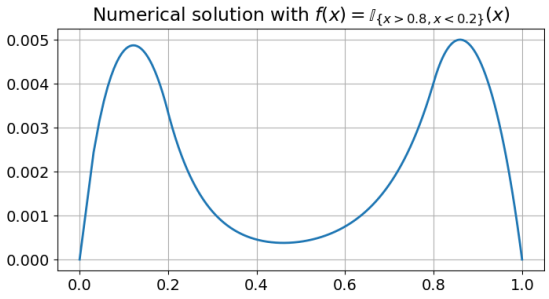
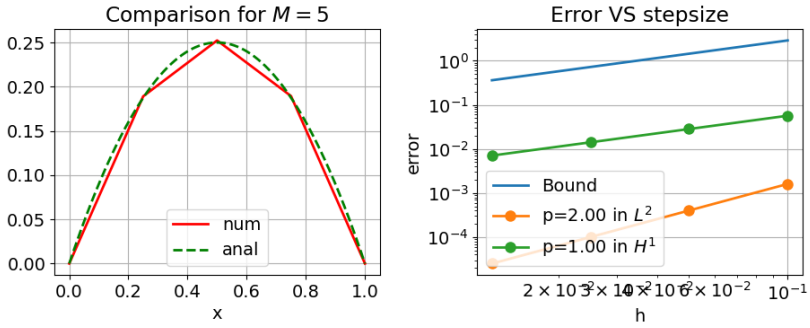


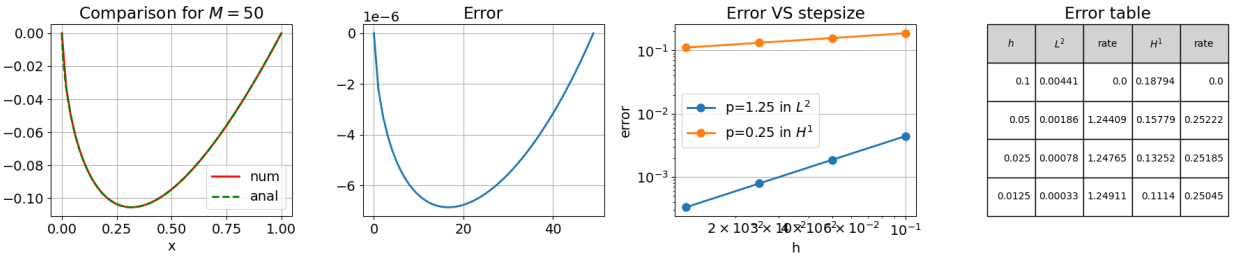
Figure 6: The solution with a step-function as right-hans side. This shows how f acts as a *source term*. Parameters $a = 1, b = -3, c = 100$.



Error table

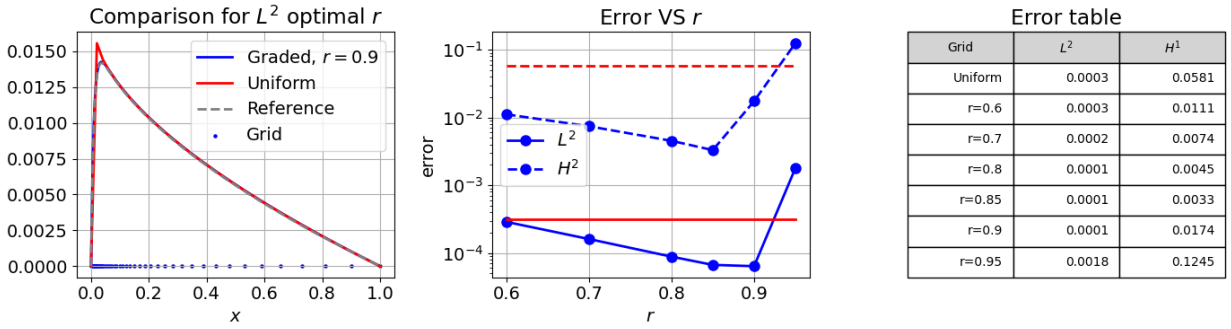
h	L^2	rate	H^1	rate
0.1	0.00161	0.0	0.05679	0.0
0.05	0.0004	2.00133	0.02834	1.00277
0.025	0.0001	2.00033	0.01416	1.00133
0.0125	3e-05	2.00008	0.00708	1.00065

Figure 7: Another convergence study using manufactured $u(x) = x(1 - x)$ and parameters $a = b = c = 1$.



h	L^2	rate	H^1	rate
0.1	0.00441	0.0	0.18794	0.0
0.05	0.00186	1.24409	0.15779	0.25222
0.025	0.00078	1.24765	0.13252	0.25185
0.0125	0.00033	1.24911	0.1114	0.25045

Figure 8: Convergence study for w_2 using $a = b = c = 1$.



Grid	L^2	H^1
Uniform	0.0003	0.0581
r=0.6	0.0003	0.0111
r=0.7	0.0002	0.0074
r=0.8	0.0001	0.0045
r=0.85	0.0001	0.0033
r=0.9	0.0001	0.0174
r=0.95	0.0018	0.1245

Figure 9: Experiment with different r-values for the right-hand-side f_1 using $a = c = 1$ and $b = -100$.