FEM for 1d stationary convection diffusion problem

Håvard Fossdal, Henning Hegstad and Eivind Jensløkken

March 2024

Abstract

Abstact goes here

Introduction

We want to study a model for the transport, mixing, and decay of a chemical substance in a fluid moving in a finite tube. The concentration u of the substance is modeled as a one-dimensional convection-diffusion problem. The problem is assumed to be stationary and the domain is scaled to the unit interval. By conservation of mass, Fick's law for diffusion and the transport theorem, the resulting model can be expressed as

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

where $\alpha(x) > 0$ is the diffusion coefficient, $b(x)$ the convective/fluid velocity, $c(x) \geq 0$ the decay rate of the substance and $f(x)$ is a source term. We consider only Dirichlet boundary values $u(0) = 0 = u(1)$.

In order to solve the problem numerically, we rewrite it in a variational form, and then employ a Galerkin approximation to the solution space. The resulting system is solved as a linear system by using the $\mathbb{P}_1$ Lagrange finite element method.

Existence of solutions

To find the variational form of (1), multiply by an arbitrary test function $v \in V$. Consider V as the Sobolev space $H_0^1(0,1) := \{v \in L^2 : v_x \in L^2,\ v(0) = 0 = v(1)\}$, where $v_x := \partial_x v$, which is a Banach space with the norm $\|v\|_{H^1}^2 = \|v\|_{L^2} + \|v_x\|_{L^2}$.
Integrating both sides of (1) over Ω, and using integration by parts yields

$$\int_0^1 -\partial_x(\alpha(x)\partial_x u)v + \partial_x(b(x)u)v + c(x)uv \,\mathrm{d}x = \int_0^1 f(x)v \,\mathrm{d}x$$
$$\int_0^1 \alpha(x)u_x v_x - b(x)uv_x + c(x)uv \,\mathrm{d}x = \int_0^1 f(x)v \,\mathrm{d}x.$$

Define the left-hand side as the bilinear form $a(u,v)$, and the right-hand side as the linear functional $F(v)$. Thus, any classical solution u of (1) solves the variational problem:

$$\text{Find } u \in H_0^1 \text{ such that} \qquad a(u,v) = F(v) \quad \forall\, v \in H_0^1(0,1). \tag{2}$$

Furthermore, assume that $\alpha(x) = \cos(\frac{\pi}{3}x)$, $c = 5$ and that $\|b\|_{L^\infty} + \|f\|_{L^2} < \infty$. We want to use the Lax-Milgram theorem to show that there exist a unique solution of the variational problem given in (2). We will not prove this theorem, but state it nonetheless.

Theorem 1 (Lax-Milgram). *Suppose that F is a continuous linear functional, and that a is bilinear, bounded and coercive. Then the variational problem defined in (2) admits a unique solution.*

Thus, we need to show that the assumptions needed for Lax-Milgram hold.

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

Proof. We start by showing that $a(u, v)$ is bilinear:

$$\begin{aligned} a(c_1u_1 + c_2u_2, v) &= \int_0^1 \Big(\alpha(x)(c_1u_1 + c_2u_2)_x v_x - b(x)(c_1u_1 + c_2u_2)v_x + c(x)(c_1u_1 + c_2u_2)v\Big)\,\mathrm{d}x \\ &= c_1 \int_0^1 \Big((\alpha(x)u_1)_x v_x - b(x)u_1v_x + c(x)u_1v\Big)\,\mathrm{d}x + c_2 \int_0^1 \Big((\alpha(x)u_2)_x v_x - b(x)u_2v_x + c(x)u_2v\Big)\,\mathrm{d}x \\ &= c_1a(u_1, v) + c_2a(u_2, v), \end{aligned}$$

by the linearity of the integral and differential operators. Similar computation holds for $a(u, c_1v_1 + c_2v_2)$, and thus $a(u, v)$ is bilinear.

$a(u, v)$ is continuous on $H^1 \times H^1$ if there exists an $M > 0$ such that

$$|a(u, v)| \leq M\|u\|_{H^1}\|v\|_{H^1}, \quad \forall\, u, v \in H^1.$$

$$\begin{aligned} |a(u, v)| &= \Big|\int_0^1 \big(\alpha(x)u_xv_x - b(x)uv_x + c(x)uv\big)\,\mathrm{d}x\Big| \leq \int_0^1 |u_x||v_x| + \|b\|_{L^\infty}|u||v_x| + 5|u||v|\,\mathrm{d}x \\ &\overset{\text{C.S.}}{\leq} \|u_x\|_{L^2}\|v_x\|_{L^2} + \|b\|_{L^\infty}\|u\|_{L^2}\|v_x\|_{L^2} + 5\|u\|_{L^2}\|v\|_{L^2} \leq M\|u\|_{H^1}\|v\|_{H^1}, \end{aligned}$$

where $M := \max\{5, \|b\|_{L^\infty}\}$ and C.S denotes the Cauchy-Schwarz inequality. □

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

Proof. $F(v)$ is a linear functional on H^1: $F : H^1 \to \mathbb{R}$ as it is a linear mapping from a normed space to the real numbers. For linearity, let $v_1, v_2 \in H^1$, $c_1, c_2 \in \mathbb{R}$:

$$F(c_1v_1 + c_2v_2) = \int_0^1 f(c_1v_1 + c_2v_2)\,\mathrm{d}x = c_1 \int_0^1 fv_1\,\mathrm{d}x + c_2 \int_0^1 fv_2\,\mathrm{d}x = c_1F(v_1) + c_2F(v_2).$$

Recall that a linear operator is continuous if and only if it is bounded. Thus, it is sufficient to show that $F(v)$ is bounded:

$$|F(v)| \overset{\text{C.S}}{\leq} \|f\|_{L^2}\|v\|_{L^2} = C\|v\|_{L^2}, \quad \forall\, v \in H^1.$$

Additionally, the constant C can be bounded from below by the dual norm of F

$$C \geq \|F\| = \sup_{0 \neq v \in H^1} \frac{|F(v)|}{\|v\|_{H^1}}$$

□

Furthermore, we formulate a preparatory result in order to show that $a(u, v)$ is coercive.

Proposition 3. *a(u,v) satisfies the Gårding inequality*

$$a(u, u) \geq (\alpha_0 - \frac{\varepsilon}{2}\|b\|_{L^\infty}) \int_0^1 u_x^2\,\mathrm{d}x + (c_0 - \frac{1}{2\varepsilon}\|b\|_{L^\infty}) \int_0^1 u^2\,\mathrm{d}x, \quad \forall\, \varepsilon > 0,$$

where $\alpha_0 = \min\limits_{x\in[0,1]} \alpha(x) = \min\limits_{x\in[0,1]} \cos(\frac{\pi}{3}x) = \frac{1}{2}$ *and* $c_0 = \min\limits_{x\in[0,1]} c(x) = 5.$

Proof. Recall Young's inequality (Y.I.): $ab \leq \frac{1}{2\varepsilon}a^2 + \frac{\varepsilon}{2}b^2$ for any $\varepsilon > 0$, and thus $-uu_x \geq -(\frac{1}{2\varepsilon}u^2 + \frac{\varepsilon}{2}u_x^2)$. Then

$$a(u,u) = \int_0^1 \alpha u_x^2 - buu_x + cu^2 \,\mathrm{d}x \geq \alpha_0 \int_0^1 u_x^2 \,\mathrm{d}x - b\int_0^1 uu_x \,\mathrm{d}x + c_0 \int_0^1 u^2 \,\mathrm{d}x$$
$$\overset{\text{Y.I}}{\geq} \Big(\alpha_0 - \frac{\varepsilon}{2}\|b\|_{L^\infty}\Big)\int_0^1 u_x^2 \,\mathrm{d}x + \Big(c_0 - \frac{1}{2\varepsilon}\|b\|_{L^\infty}\Big)\int_0^1 u^2 \,\mathrm{d}x, \quad \forall\, \varepsilon > 0.$$

$\square$

Lemma 1. *a(u,v) is coercive when* $\|b\|_{L^\infty} < \sqrt{2\alpha_0 c_0} = \sqrt{5}$.

Proof. a is coercive if there exists a constant $\beta > 0$ such that $a(u,u) \geq \beta\|u\|_{H^1}^2$.
From Proposition 3, define

$$\gamma = \min\Big\{\alpha_0 - \frac{\varepsilon}{2}\|b\|_{L^\infty}, c_0 - \frac{1}{2\varepsilon}\|b\|_{L^\infty}\Big\} < \min\Big\{\frac{1}{2} - \frac{\varepsilon}{2}\sqrt{5}, 5 - \frac{1}{2\varepsilon}\sqrt{5}\Big\}.$$

Let $\varepsilon \in \Big(\frac{\sqrt{5}}{10}, \frac{1}{\sqrt{5}}\Big)$, i.e. between the roots of the two arguments within the min-function, and thus $\gamma > 0$. Then

$$a(u,u) \geq \gamma\int_0^1 u^2 \,\mathrm{d}x + \gamma\int_0^1 u_x^2 \,\mathrm{d}x \geq \gamma\|u\|_{H^1}.$$

$\square$

Proposition 1, 2 and Lemma 1 shows that the assumptions for the Lax-Milgram theorem holds. Thus, there exists a unique solution $u \in H_0^1(0,1)$ of (2).

Numerical scheme and error bound

Galerkin approximation

To solve (2), we use a Galerkin approximation on the solution space V, such that it becomes a finite dimensional space denoted by V_h. For the grid, consider a subdivision τ_h of Ω defined as $\tau_h = \{K_i\}_{i=0}^M$ such that $\bigcup_{K\in\tau_h} \bar{K} = \Omega$. For the method to be $\mathbb{P}_1$ Lagrange, let $V_h \subset X_h^1$, where X_h^1 is the space of piecewise continuous linear functions on the subdivision τ_h. To retain the property of the zero Dirichlet boundary conditions, let $V_h = X_h^1 \cap H_0^1$.
Thus, we obtain the Galerkin approximation of the variational problem

$$\text{Find } u_h \in H_0^1 \text{ such that} \qquad a(u_h, v_h) = F(v_h) \quad \forall\, v_h \in V_h. \tag{3}$$

Numerical scheme

As V_h is finite dimensional, there exists a basis, $\mathcal{B}$, for the space. The elements in the subdivision $K_i \in \tau_h,\ i = 1,\dots,M$, are defined as $K_i = [x_{i-1}, x_i]$ and
$0 = x_0 < x_1 < \cdots < x_{M+1} = 1$. We use the hat-basis functions, denoted by φ_i:

$$\varphi_i = \begin{cases} \frac{x - x_{i-1}}{h_i}, & x \in [x_{i-1}, x_i] \\ \frac{x_{i+1} - x}{h_{i+1}}, & x \in [x_i, x_{i+1}] \\ 0 & \text{otherwise.} \end{cases}$$

Furthermore, u_h and v_h can be written as a linear combination of the basis elements $\{\varphi_i\}_{i=0}^M$ by some constants U_j and V_i

$$u_h = \sum_{j=0}^{M} U_j \varphi_j, \qquad v_h = \sum_{i=0}^{M} V_i \varphi_i. \tag{4}$$

By inserting (4) into the Galerkin approximation (3), we obtain the linear system

$$A\vec{U} = \vec{F}, \tag{5}$$

where $A_{ij} = a(\varphi_j, \varphi_i)$, $\vec{U} = (U_0, U_1, \ldots, U_M)^T$ and $F_i = F(\varphi_i)$. The matrix A is called the stiffness matrix, and the vector F is called the load vector.
To summarize, the $\mathbb{P}_1$ FEM scheme can be written as

$$\begin{gathered} u_h = \sum_{j=0}^{M} U_j \varphi_j, \text{ where } \vec{U} = (U_0, \ldots, U_M)^T \text{ solves} \\ \begin{cases} U_0 = 0 = U_M \\ A\vec{U} = \vec{F}, \quad A_{ij} = a(\varphi_j, \varphi_i), \ F_i = F(\varphi_i). \end{cases} \end{gathered} \tag{6}$$

Lemma 2. *The $\mathbb{P}_1$ FEM scheme satisfies Galerkin orthogonality.*

Proof. Let $u \in V$ and $u_h \in V_h$ solve (2) and (3) respectively. Let $v_h \in V_h$ be arbitrary, then

$$a(u - u_h, v_h) \overset{\text{Proposition 1}}{=} a(u, v_h) - a(u_h, v_h) = F(v_h) - F(v_h) = 0,$$

which holds for all $v_h \in V_h$. □

Error bound

To formulate an error bound, we need to utilize Cea's lemma.

Lemma 3. *Let $u \in V$ and $u_h \in V_h$ solve* (2) *and* (3) *respectively. Then*

$$\|u - u_h\|_{H^1} \leq \frac{M}{\gamma} \inf_{v_h \in V_h} \|u - v_h\|_{H^1},$$

where M and γ are the constants from the continuity and coercivity of the bilinear form.

Proof. $a(u, v)$ is bilinear, continuous and coercive by Proposition 1 and Lemma 1, and thus Cea's lemma hold. □

To find a bound for the right-hand side in Lemma 3, we use linear interpolation. Denote by $I_h(u)$ the piecewise linear interpolation of the solution of (2), u, in the nodes x_i. Note that $I_h(u) \in V_h$ if $u \in H_0^1$. The interpolation error, for functions $u \in H^2$, is given as

$$\|u - I_h(u)\|_{H^1} \leq Ch\|u_{xx}\|_{L^2}, \tag{7}$$

where $h = \max_i h_i$ and $C = \sqrt{1 + h^2}$. Thus, if $h \leq 1$ then $C \leq \sqrt{2}$, which is the case as $\Omega = (0, 1)$. Define the error of the scheme as $e = u - u_h$.

Theorem 2 (Error bound)**.** *The scheme is linearly convergent in the H^1 norm for functions in H^2, that is*

$$\|e\|_{H^1} \leq \frac{\sqrt{2}M}{\gamma} \|u_{xx}\|_{L^2}\, h.$$

Proof.

$$\|e\|_{H^1} = \|u - u_h\|_{H^1} \overset{\text{Lemma 3}}{\leq} \frac{M}{\gamma} \inf_{v_h \in V_h} \|u - v_h\|_{H^1} \leq \frac{M}{\gamma} \|u - I_h(u)\|_{H^1} \overset{(7)}{\leq} \frac{\sqrt{2}M}{\gamma} \|u_{xx}\|_{L^2}\, h.$$

□

Numerical problems and results

We have taken inspiration from the notebook '1dFEM_2024.ipynb' by prof. Espen R. Jakobsen in our own implementation.

To solve the linear system (5) numerically, we have to build the stiffness matrix and the load vector. The assembly process is done element wise, and by analytic computations the added i'th element is

$$\begin{bmatrix} \frac{\alpha}{h_i} + \frac{c}{3}h_i & -\frac{\alpha}{h_i} + \frac{b}{2} + \frac{c}{6}h_i \\ -\frac{\alpha}{h_i} - \frac{b}{2} + \frac{c}{6}h_i & \frac{\alpha}{h_i} + \frac{c}{3}h_i \end{bmatrix}.$$

In order to satisfy the properties of the space H_0^1 (Dirichlet boundary conditions), we change the first and last row of A and F such that the linear system is correct.
For testing the convergence of the method, we impose an exact solution and modify the load vector such that it conforms with it. The computation of F is done by the numerical quadrature Simpson's rule from the Scipy library.

The convergence rates are calculated in both L^2 and H^1 norms. The H^1 error is given by

$$\|e\|_{H^1}^2 = \|e\|_{L^2}^2 + \|e_x\|_{L^2}^2 = \left(\int_0^1 e^2 \, \mathrm{d}x\right)^2 + \left(\int_0^1 {e_x}^2 \, \mathrm{d}x\right)^2.$$

The integrals are solved numerically by the same Simpson's rule as for the load vector. Thus, to find the error rates, the grid is simply refined (equidistant grids are used).

Convergence for smooth solutions

To verify the convergence result from Theorem 2 numerically, we impose the manufactured solution $u = x(1-x) \in H^2$. The numerical solution u_h and the exact solution $u(x_i)$ is plotted in Figure 1.
Table 1 shows the result of the convergence test in both L^2 and H^1 norm. The H^1 rate seems to correspond quite well to the theoretical linear convergence. Additionally, the L^2 quadratic rate is a well known theoretical result.

Non-smooth solutions

It is also interesting to test the convergence of our method for solutions that are not H^2. We want to consider the two non-smooth manufactured solutions

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

Theorem 3. *w_1 and w_2 belong to H^1, but not to H^2.*

Proof. Consider the first function, w_1. Its derivative is given by

$$w_1'(x) = \begin{cases} \frac{2}{\sqrt{2}}, & x \in [0, \frac{\sqrt{2}}{2}] \\ \frac{-1}{1-\frac{\sqrt{2}}{2}}, & x \in (\frac{\sqrt{2}}{2}, 1], \end{cases}$$

which is not differentiable at $x = \frac{\sqrt{2}}{2}$ as the left and right limits of the derivative do not coincide at this point. It is piecewise constant and well defined for $x \in (0, 1)$, and thus its integral also converges. Hence, $w_1 \in H^1$.
$w_1'(x)$ can be written as a step function, which in turn is not differentiable (excluding distributional derivatives). Hence, $w_1 \notin H^2$.

The derivatives of w_2 are

$$w_2'(x) = 1 - \frac{3}{4}x^{-\frac{1}{4}}, \qquad w_2''(x) = \frac{3}{16}x^{-\frac{5}{4}}.$$

The integral of $w_2'(x)$ clearly converges, whereas the integral of the second derivative has to be evaluated at $x^{-\frac{1}{4}}\big|_{x=0}$, which is not defined. Thus, $w_2 \in H^1$. □

Table 2 shows the numerical error rates in L^2 and H^1, which are notably lower than that of the smooth solution, in both norms.

Table 2 and 3 shows the numerical error rates in L^2 and H^1 for w_1 and w_2 respectively. It is notably lower than that of the smooth solution for both cases. This was expected as the two functions are not smooth enough for Theorem 2 to hold, but we do not know what the rate should be (theoretically) for these solutions.

Sharp gradients

We want to consider the two functions for $x \in (0, 1)$

$$f_1(x) = x^{-\frac{2}{5}} \quad \text{and} \qquad f_2(x) = x^{-\frac{7}{5}}.$$

Note that $f_1 \in L^2$ and $f_2 \notin L^2$ as the exponents of the anti derivatives are positive and negative, respectively. Moreover, one can see that $f_2(x) = -\frac{5}{2}(f_1)_x$, which means that $f_2 \in H^{-1}$, which is the dual space of H^1. As $f_2 \in H^{-1}$, there has to be a gradient term in the functional $F(v)$ As $f_2 \in H^{-1}$, the functional $F(v)$ has to be defined by using integration by parts

$$F(v) = \int_0^1 f_2 v \,\mathrm{d}x = -\frac{5}{2}\int_0^1 (f_1)_x v \,\mathrm{d}x \overset{v \in H_0^1}{=} \frac{5}{2}\int_0^1 f_1 v_x \,\mathrm{d}x.$$

Typically, these solutions of (1) contain sharp gradients near $x = 0$ (as the right hand side is singular here). In order to handle this, we introduce a graded grid

$$x_0 = 0 \quad \text{and} \quad x_i = r^{M-i},\ i = 1, \dots, M, \qquad \text{for some } r \in (0, 1).$$

This grid ensures that more nodes are placed near $x = 0$. Moreover, to (numerically) handle the fact that $x = 0$ is a singular point, we set $x_0 = 10^{-10}$. We assume that this method provides better results than a uniform grid, and thus we attempt to verify this numerically with L^2 and H^1 errors. The number of nodes used in the grid are called the degrees of freedom, and to compare different grids, we fix this number.

In order to find the errors, we have to compare to an exact solution. The exact solution is approximated by using a very refined grid. For our experiments we have used a refined grid with 5000 degrees of freedom, and comparatively, 50 degrees of freedom for the numerical tests. We used the coefficients $\alpha = 1$, $b = -1000$ and $c = 1$. The high negative value of b ensures that the gradient indeed is sharp near $x = 0$. Additionally, we have tested some values $r \in (0.5, 1)$.
The results for the two functions f_1 and f_2 are given in Tables 4 and 5 respectively. We can see that $r = 0.8$ is a good value for f_1 and $r = 0.7$ is a good value for f_2. Additionally, all selected r values outperform the uniform grid, which supports our claim that the graded grid is better.

Conclusion

TODO

Spesifisere hvilke konstanter vi har brukt og sånn, slik at det er tydelig (α, b og c)

Appendix

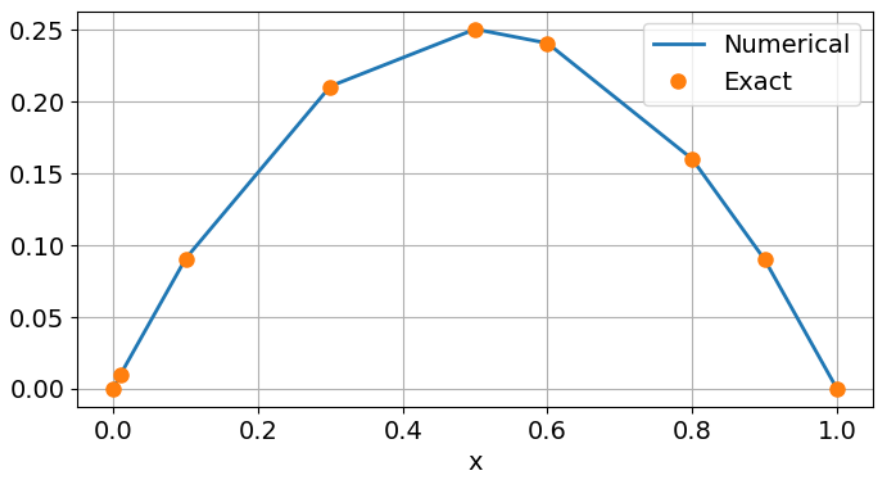

Figure 1: This figure shows our method tested by the test function $u = x(1 - x)$. It also creates the corresponding F by numerical integration.

h	L^2 Error	H^1 Error	L^2 Rate	H^1 Rate
0.1	1.7e-3	0.05773	N/A	N/A
0.05	4.3e-4	0.02886	2.00063	1.0008
0.025	1.1e-4	0.01442	2.00016	1.0007
0.0125	2.7e-5	0.00721	2.00004	1.00119

Table 1: This

h	L^2 Error	H^1 Error	L^2 Rate	H^1 Rate
0.1	0.11592	1.34554	N/A	N/A
0.05	0.06025	0.9327	0.94402	0.53104
0.025	0.02997	0.61503	1.00749	0.60205
0.0125	0.00965	0.4048	1.63503	0.60473

Table 2: Convergence w_1

h	L^2 Error	H^1 Error	L^2 Rate	H^1 Rate
0.1	1.65475	9.15071	N/A	N/A
0.05	0.8893	6.75142	0.89588	0.4495
0.025	0.46076	4.8777	0.94866	0.47499
0.0125	0.23447	3.48634	0.97461	0.48764

Table 3: Convergence w_2

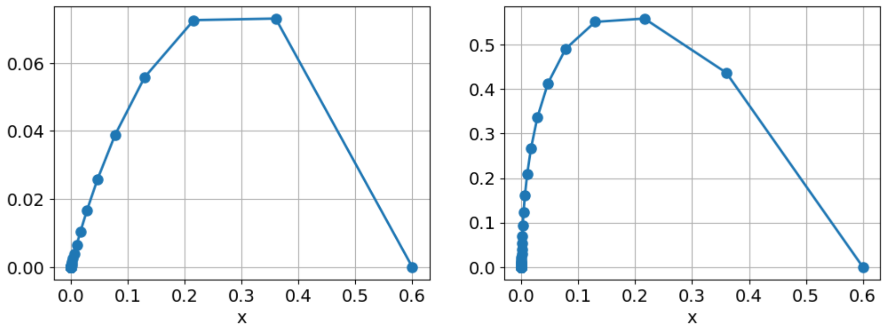

Figure 2: w_1 left, w_2 right

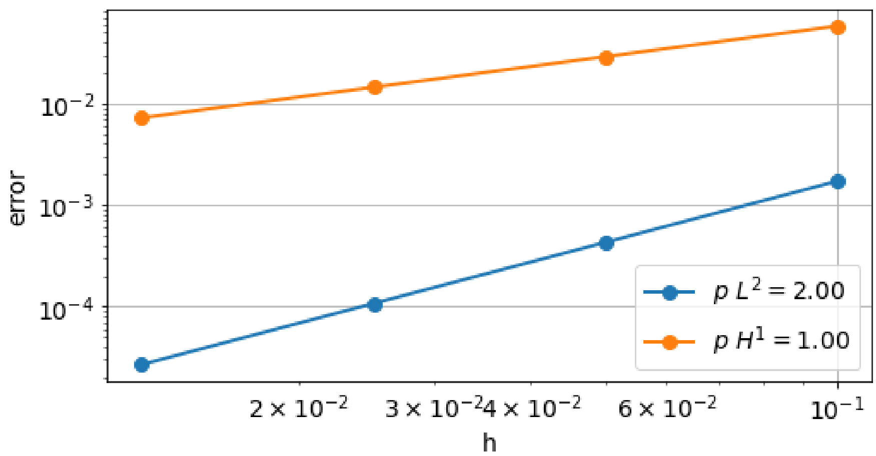

Figure 3: Caption

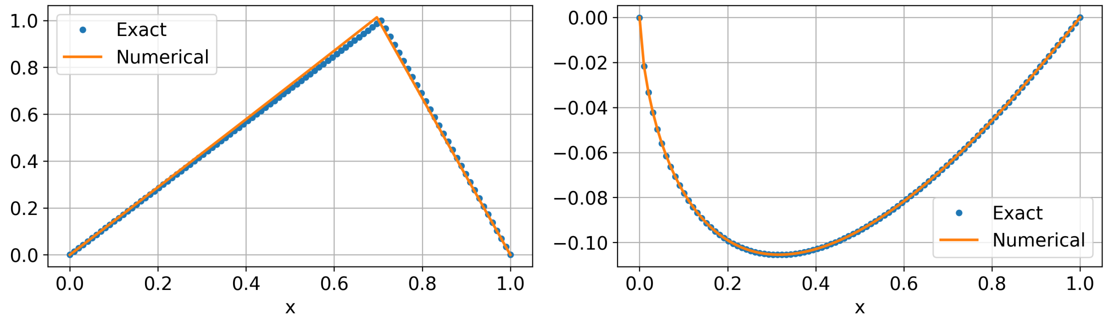

Figure 4: Comparison

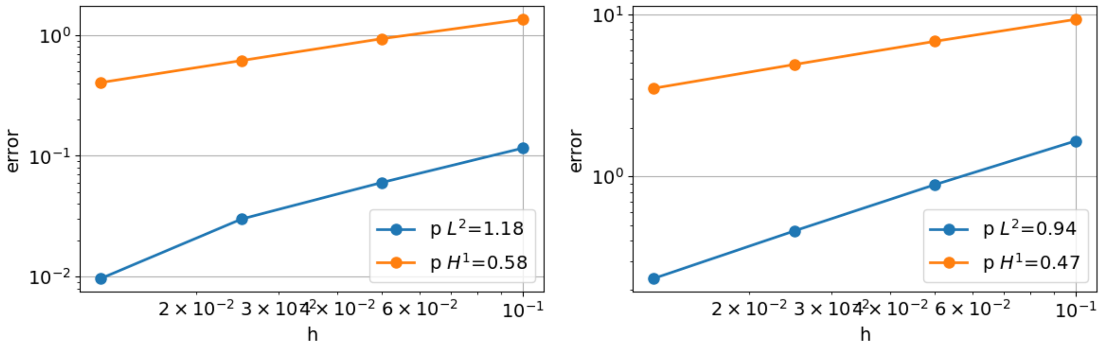

Figure 5: Left w_1, right w_2

r	L^2 error	H^1 error
0.5	1.4e-4	0.01552
0.6	9.2e-5	0.01116
0.7	4.5e-5	0.00748
0.8	1.7e-5	0.0044
0.9	8.0e-5	0.0047
Uniform	3.1e-4	0.06279

Table 4: Error of f_1

r	L^2 Error	H^1 error
0.5	1.7e-3	1.20447
0.6	1.5e-3	1.18714
0.7	1.5e-3	1.17829
0.8	1.8e-3	1.43885
0.9	8.6e-3	2.74769
Uniform	9.6e-3	3.16077

Table 5: Error of f_2