Project 2: NumDiff
NUMERICAL SOLUTION OF A 1D STATIONARY CONVECTION DIFFUSION PROBLEM WITH A FINITE ELEMENT METHOD

Hanna Bu Kvaløy, Henrik Olaussen, Sigrid Maria Nese, Kjersti Kvistad Dengerud

March 2023

1 Introduction

This report considers a 1D convection-diffusion-reaction model in the stationary regime. The model is given by a boundary value problem for a second order partial differential equation (Poisson-like equation):

$$-\partial_x(\alpha(x)\partial_x u) + \partial_x(b(x)u) + c(x) = f(x), \tag{1}$$

The equation includes the concentration of the substance $u(x)$, the diffusion coefficient $\alpha(x)$, the convective/fluid velocity $b(x)$, the decay rate of the substance $c(x)$, and the source term $f(x)$. The problem is independent of time and scaled to the domain $\Omega = (0, 1)$. In this report, we model the solution $u(x)$ on Ω with a finite element method (FEM), which is a special case of Galerkin method.

This report has two main parts. The first part focuses on formulating a weak form for problem (1), which is required in the FEM. In the second part, we use our results to construct and implement numerical schemes that approximate the solution $u(x)$ for various cases where the exact solution $u_{exact}(x)$ is known. At the end of the report, we consider a solution that is shifted towards the boundary $x = 0$ (what we get with negative values of b). For this scenario, we will examine whether a numerical solution can be improved by using an adapted grid.

2 A 1D stationary convection diffusion problem

As stated in the introduction, we have to reformulate problem (1) to a weak form in order to derive the FEM. That is:

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

Hence, setting up the weak form involves derivation of $F(v)$ and $a(u, v)$. In addition, we want to ensure that problem (V) has a unique solution $u \in H_0^1(0, 1)$. By the Lax-Milgram theorem, (V) has a unique solution if $F(v)$ is a linear continuous functional and $a(u, v)$ is of continuous, coercive bilinear form. In this part of the report, we derive $F(v)$ and $a(u, v)$, and then we show that by the Lax-Milgram theorem, (V) has a unique solution.

Firstly, we derive $F(v)$ and $a(u,v)$. The first step is to multiply (1) with a test solution $v \in H_0^1(0,1)$. We then integrate both sides from 0 to 1, and get the following:

$$\int_0^1 (-\partial_x(\alpha(x)\partial_x u)v + \partial_x(b(x)u)v + c(x)uv)dx = \int_0^1 fvdx$$

We define the integral on the right hand side as $F(v)$. Henceforward, we use the left hand side of the equation to derive $a(u,v)$. To find an expression, we use integration by parts on the first and second term:

$$\int_0^1 (-\partial_x(\alpha(x)\partial_x u)v+\partial_x(b(x)u)v+c(x)uv)dx = [-\alpha u_x v]_0^1 + \int_0^1 \alpha u_x v_x dx + [buv]_0^1 - \int_0^1 buv_x dx + \int_0^1 cuvdx$$

Since $v \in H_0^1(1,0)$, we have that $[-\alpha u_x v]_0^1 = [buv]_0^1 = 0$. The resulting integral yields the bilinear form $a(u,v)$:

$$a(u,v) = \int_0^1 (\alpha u_x v_x - buv_x + cuv)dx$$

We now move on to showing that (V) has a unique soltion. The first step is to show that $a(u,v)$ is a bilinear, continuous and coersive form on $H^1 \text{x} H^1$. We start by showing bilinearity. In order to show bilinearity, we have to show linearity in both arguments. We start by considering the first argument, where r_1, r_2 are real numbers:

$$\begin{aligned}
a(r_1u_1 + r_2u_2, v) &= \int_0^1 (\alpha\partial_x(r_1u_1 + r_2u_2)v_x - b(r_1u_1 + r_2u_2)v_x + c(r_1u_1 + r_2u_2)v)dx \\
&= \int_0^1 (\alpha r_1(u_1)_x v_x + \alpha r_2(u_2)_x v_x - bv_x r_1 u_1 - bv_x r_2 u_2 v_x + cvr_1u_1 + cvr_2u_2)dx \\
&= r_1 \int_0^1 (\alpha(u_1)_x v_x - bv_x u_1 + cvu_1)dx + r_2 \int_0^1 (\alpha(u_2)_x v_x - bu_2 v_x + cvu_2)dx \\
&= r_1 a(u_1, v) + r_2 a(u_2, v)
\end{aligned}$$

We then consider the second argument:

$$\begin{aligned}
a(u, r_1v_1 + r_2v_2) &= \int_0^1 (\alpha u_x \partial_x(r_1v_1 + r_2v_2) - bu_x\partial(r_1v_1 + r_2v_2) + cu(r_1v_1 + r_2v_2))dx \\
&= \int_0^1 (\alpha r_1 u_x(v_1)_x + \alpha r_2 u_x(v_2)_x - br_1u(v_1)_x - br_2u(v_2)_x + cur_1v_1 + cur_2v_2)dx \\
&= r_1 \int_0^1 (\alpha u_x(v_1)_x - bu(v_1)_x + cv_1u)dx + r_2 \int_0^1 (\alpha u_x(v_2)_x - bu(v_2)_x + cv_2u)dx \\
&= r_1 a(u, v_1) + r_2 a(u, v_2)
\end{aligned}$$

We have linearity in both arguments hence $a(u,v)$ is bilinear. Secondly, we show that $a(u,v)$ is continuous. $a(u,v)$ is continuous on $H^1 \times H^1$ if there exists a constant M such that $a(u,v) \leq M\|u\|_{H^1}\|v\|_{H^1}$ for all $u, v \in H^1$.

$$\begin{aligned}
a(u,v) &= \int_0^1 (\alpha(x)u_xv_x - b(x)uv_x + c(x)uv)\, dx \leq \int_0^1 (|\alpha(x)u_xv_x| + |b(x)uv_x| + |c(x)uv)|)\, dx \\
&= |<\alpha(x)u_x, v_x>_{L^2}| + |<b(x)u, v_x>_{L^2}| + |<c(x)u, v>_{L^2}| \\
&\leq \|\alpha\|_{L^\infty}\|u_x\|_{L^2}\|v_x\|_{L^2} + \|b\|_{L^\infty}\|u\|_{L^2}\|v_x\|_{L^2} + \|c\|_{L^\infty}\|u\|_{L^2}\|v\|_{L^2} \\
&\leq \|\alpha\|_{L^\infty}\|u\|_{H^1}\|v\|_{H^1} + \|b\|_{L^\infty}\|u\|_{H^1}\|v_{\|H^1} + \|c\|_{L^\infty}\|u\|_{H^1}\|v\|_{H^1} \\
&= (\|\alpha\|_{L^\infty} + \|b\|_{L^\infty} + \|c\|_{L^\infty})\|u\|_{H^1}\|v\|_{H^1} = M\|u\|_{H^1}\|v\|_{H^1}
\end{aligned}$$

This implies that $a(u,v)$ is continuous with $M = \|\alpha\|_{L^\infty} + \|b\|_{L^\infty} + \|c\|_{L^\infty} < \infty$. For the remaining part of the report let α, b, c be nonzero constants. Thirdly, we show that $a(u,v)$ is coersive. In order to do this, we first have to show that $a(u,v)$ satisfies the Gårdings inequality:

$$a(u,u) = \int_0^1 (\alpha u_x^2 - buu_x + cu^2)\,dx, \text{ and } uu_x \leq \frac{1}{2\epsilon}u^2 + \frac{\epsilon}{2}u_x^2\ \epsilon > 0$$

This gives $a(u,v) \geq \int_0^1 \alpha u_x^2 - |b|(\frac{1}{2\epsilon}u^2 + \frac{\epsilon}{2}u_x^2) + cu^2 dx = (\alpha - |b|\frac{\epsilon}{2}) \int_0^1 u_x^2 dx + (c - \frac{1}{2\epsilon}) \int_0^1 u^2 dx.$

Moreover, $a(u,v)$ is coercive on H_0^1 if there exists a constant $C > 0$ such that $a(u,u) :\geq C\|u\|_{H_0^1}^2$ for all $u \in H_0^1$.

$$a(u,u) \geq (\alpha - |b|\frac{\epsilon}{2}) \int_0^1 u_x^2 dx + (c - |b|\frac{1}{2\epsilon}) \int_0^1 u^2 dx = (\alpha - |b|\frac{\epsilon}{2})\|u_x\|_{L^2}^2 + (c - |b|\frac{1}{2\epsilon})\|u^2\|_{L^2}^2$$

Let $C \leq (\alpha - |b|\frac{\epsilon}{2})$ and $C \leq (c - \frac{1}{2\epsilon})$. Then $a(u,u) \geq C(\|u_x\|_{L^2}^2 + \|u\|_{L^2}^2) = C\|u\|_{H_0^1}^2$

As $C > 0$ it requires for $\alpha - |b|\frac{\epsilon}{2}$ and $c - \frac{1}{2\epsilon}$ to be greater than zero.

$$\alpha - |b|\frac{\epsilon}{2} > 0 \Rightarrow \frac{1}{\epsilon} > \frac{|b|}{2\alpha}, \text{ and } c - |b|\frac{1}{2\epsilon} > 0 \Rightarrow c > \frac{|b|^2}{4\alpha}$$

$a(u,v)$ is coercive if $c > \frac{|b|^2}{4\alpha}$, which makes it coercive for $c > \frac{|b|^2}{2\alpha}$.

Finally, we have to show that $F(v)$ is a linear continuous functional on H^1. First and foremost, $F(v)$ is an integral. Since integrals are linear, $F(v)$ is also linear. Moreover, we have that $F(v)$ is continuous if and only if $F(v)$ is bounded:

$$||F|| = \max_{0 \neq v \in H^1} \frac{|F(v)|}{||v||_{H^1}} \leq \max_{0 \neq v \in H^1} \frac{||f||_{L^2}||v||_{L^2}}{||v||_{H^1}}$$

Above, we use the definition of the dual norm combined with the Cauchy-Schwartz inequality. Since $||v||_{L^2} \leq ||v||_{H^1}$, we have: $\max_{0 \neq v \in H^1} \frac{||f||_{L^2}||v||_{L^2}}{||v||_{H^1}} \leq ||f||_{L^2} \cdot 1 < \infty$

Hence $F(v)$ is a linear continuous (bounded) functional on H^1.

To sum up, we have in this section stated the weak form (V) of problem (1), and then showed that (V) has a unique solution (by the Lax-Milligram Theorem).

3 Numerical solution using FEM

For the second part of the report, we solve problem (V) numerically with the FEM. Throughout this section, we implement and test numerical solvers for different cases of $u(x)$ and compare them to the known exact solutions $u_{exact}(x)$. To get the FEM, we need the discretization (V_h) of the weak form (V) and a basis $\{\phi_i\}$ for a finite dimentional vector space $V_h \subset H_0^1(0,1)$. That is:

$$\text{Find } u_h \in V_h \text{ such that } a(u_h, v_h) = F(v_h)\ \forall\ v_h \in V_h \tag{V_h}$$

where $V_h = X_h^1(0,1) \cap H_0^1(0,1)$ is the space of continuous functions with zero boundary values that are piecewise linear on the "triangulation" given by the grid.

The problem can be described as a linear system $AU = F$, where A is an $M \times M$ matrix (a stiffness matrix) with coefficients $A_{ij} = a(\phi_i, \phi_j)$. ϕ_i is the triangular basis function centered at x_i. U is the weak, numerical solution and F is the right hand side of the equation(RHS). Using the basis functions we get that the numerical solution is: $u_h = \sum_j U_j \phi_j$

3.1 Galerkin orthogonality, Cea's lemma and error bounds

Before solving (V_h) for different functions satisfying the given conditions we want to present some properties regarding the problems well-posedness, provide an error bound and show what rates of convergence one would expect when solving the problem.

Galerkin orthogonality is an important property of the P1 (linear) FEM. It ensures that the error is minimised in the sense of least squares, which provides a more accurate numerical solution. Furthermore it also ensures that the numerical solution is stable.

Galerkins orthogonality for the P1 FEM (V_h) is stated as follows:

"Let u and u_h be the solutions of the infinite and finite dimensional variational problems respectively. Then $a(u-u_h, v_h) = 0$." From (V) and (V_h) we have that $a(u, v) = F(v)$ and $a(u_h, v_h) = F(v_h)$ with $v \in H_0^1(0,1)$ and $v_h \in V_h$. As $V_h \subset H_0^1(0,1)$ we have that$a(u, v_h) = F(v_h)$. Galerkins orthogonality therefore holds for the problem (V_h):

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

Furthermore, we want to show that Cea's lemma also holds for this problem, and use this to find an H^1 error bound. Ceas lemma states that there exists constants M and C such that $\|u - u_h\|_{H^1} \leq \frac{M}{C}\|u - v_h\|_{H^1}$. As $a(u, v)$ is coercive there exists a constant C such that:

$$\frac{1}{C}a(u - u_h, u - u_h) \geq \|u - u_h\|_{H^1}^2$$

Moreover, $a(u - u_h, u - u_h) = a(u - u_h, u - v_h) + a(u - u_h, v_h - u_h)$ because of the bilinearity of $a(\cdot, \cdot)$. Due to the galerkin orthogonality, we have $a(u - u_h, v_h - u_h) = 0$ as $v_h - u_h \in V_h$. Inserting this we get:

$$\frac{1}{C}a(u - u_h, u - v_h) \geq \|u - u_h\|_{H^1}^2$$

As $a(u, v)$ is continuous, there exists a constant M such that $a(u-u_h, u-v_h) \leq M\|u-u_h\|_{H^1}\|u-v_h\|_{H^1}$. Inserting this we get:

$$\|u - u_h\|_{H^1}^2 \leq \frac{1}{C}a(u - u_h, u - v_h) \leq \frac{M}{C}\|u - u_h\|_{H^1}\|u - v_h\|_{H^1}$$

$$\Rightarrow \|u - u_h\|_{H^1} \leq \frac{M}{C}\|u - v_h\|_{H^1}$$

Which shows that Cea's lemma holds for this problem. Additionally, we want to derive an error bound in H^1 and present the theoretical rates of convergence of our problem. From Cea's lemma we have that $\|u - u_h\|_{H^1} \leq \frac{M}{C} inf_{v_h \in V_h}\|u - v_h\|_{H^1}$. We then define $E_h(u) := inf_{v_h \in V_h}||u - v_h||_V$, with $V = H_0^1$, where $E_h(u)$ can be interpereted as "how well $u \in V$ can be approximated by $v_h \in V_h$". For convergence we need $u_h \to u$ as $h \to 0$, we need for $E_h(u) \to 0$.

The interpolation of $u(x)$ can be stated as follows $I_h u(x) = \sum_{i=0}^{n} u(x_i)\phi_i(x)$. We have that

$u \in H_0^1(0,1)$ implies that u is continuous (in 1D only), and $I_h u$ is well defined in X_h^1, which makes $I_h u \in V_h$. Choosing $v_h = I_h u$ we have that $E_h(u) \leq ||u - I_h u||_V$. We then define the interpolation error $\tilde{E_h}(u) := ||u - I_h u||_V$, with the bound $\tilde{E_h} u \leq 2h||u''||_{L^2(0,1)}$ given in class. An error bound in H^1 can now be derived:

$$E_h(u) = ||u - u_h||_{H^1} \leq \frac{M}{C} \inf_{I_h u \in V_h} ||u - I_h u||_{H^1} = \frac{M}{C} \tilde{E_h(u)} \leq \frac{2M}{C} ||u''||_{L^2(0,1)} h$$

Hence the rate of convergence in H^1 is 1. The interpolation error in L^2 is however of order 2. The error function converges faster then the derivative which makes the error in L^2 decrease faster then the error in H^1 as the norm in H^1 also is dependent on the derivative of the function.

3.2 Implementing and applying the method

Now, we want to solve for two known exact solutions; $u_1 = x(1-x)$ and $u_2 = \sin(3\pi x)$. To implement the $(M+1)\times(M+1)$ stiffness matrix A, we introduce the matrices $A^{k_i}, i = 1, 2, .., M$, where A^{k_0} contains the values for $a_{00}, a_{0,1}, a_{10}, a_{11}$, A^{k_1} the values of $a_{11}, a_{12}, a_{21}, a_{22}$ and so on. In other words, we assemble the A^{k_i} matrices, so that the A^{k_i} overlap with the next matrix $A^{k_{i+1}}$ along the diagonal of A. When solving the linear system numerically, we solve the equation for $\tilde{A}\tilde{U} = \tilde{F} + g$ where $\tilde{A}$ is a $(M-1) \times (M-1)$ submatrix of A ($\tilde{A}$ contains the inner points of A, i.e without the boundary points) and $\tilde{U}\tilde{F}, g$ are $(M-1)$ dimentional vectors. The vector g contains the boundary conditions.

We display one realization of the stiffness matrix A in figure 1, with $\alpha = 1, b = 1, c = 1, M = 11$ and the corresponding Xk-vector. The plots in figure 2 and 3 show the exact solutions and the numerical approximations of the two functions u_1 and u_2 and their respective convergence rates in both H^1 and L^2.

The implemented FEM seems to give good approximations to the exact solutions, and the convergens rates are 1.03 in H^1, and 2.00 in L^2 for both cases of $u_{exact}(x)$. This supports the teoretical results discussed in 3.1.

We now want to look at some cases where different properties of the potentials require some adjustments to the method.

3.3 Non-smooth potentials

To solve non-smooth potentials with the FEM, the solver needs some adjustments as there exists non-differentiable points. Two non-smooth potential exact solutions are given:

$$w_1(x) = \begin{cases} 2x, & \text{if } x \in (0, \frac{1}{2}) \\ 2(1-x), & \text{if } x \in (\frac{1}{2}, 1) \end{cases} \text{, and } w_2(x) = x - |x|^{\frac{2}{3}}$$

$w_1(x)$ is not differentiable in $x = \frac{1}{2}$, and $w_2(x)$ is not differentiable in $x = 0$, as the functions have breaking points in these x-values, and therefore a discontinuity in the first derivative. Both functions are in $H^1(0,1)$, but not in $H^2(0,1)$. From the definition of the Sobolev spaces "$H^k(\Omega) = f : f, \nabla f, ..., \nabla^k f \in L^2(\Omega$", one have that for $w_1(x)$ and $w_2(x)$ to be in $H^1(0,1)$, and not in $H^2(0,1)$, the square integral of $w_1(x)$, $w_1'(x)$, $w_2(x)$ and $w_2'(x)$ must exist and be finite on the interval $[0,1]$, while this should not hold for $w_1''(x)$ and $w_2''(x)$. $\implies w_1(x)$ and $w_2(x)$ are both in $L^2(0,1)$, as they are squared integrable on $[0,1]$. The first derivatives are:

$$w_1'(x) = \begin{cases} 2, & \text{if } x \in (0, \frac{1}{2}) \\ -2, & \text{if } x \in (\frac{1}{2}, 1) \end{cases} \text{, and } w_2'(x) = 1 - \frac{2}{3}x^{\frac{-1}{3}}, \text{if } x > 0$$

$w_1'(x)$ is discontinuous at $x = \frac{1}{2}$, however it is still squared integrable on $[0, 1]$. $w_2'(x)$ is not defined for $x = 0$, however it is still squared integrable as x approches 0, so $w_2'(x) \in L_2(0, 1)$. Hence both $w_1(x)$ and $w_2(x)$ are in $H^1(0, 1)$. Now we look at the second derivatives:

$$w_1''(x) = -4\delta(x - \frac{1}{2}) \text{ , and } w_2''(x) = \begin{cases} -\frac{2}{9}x^{-\frac{4}{3}}, & \text{if } x < 0 \\ \frac{2}{9}x^{-\frac{4}{3}}, & \text{if } x > 0 \end{cases}$$

$\delta(x)$ being the Dirac delta function, where $\delta(x - a) = \infty$, if $x = a$, 0 elsewhere. $w_1(x)'$ can be written as $-4 * H(x - 1/2) + 2$ where H(x) is the heaviside step function. The derivative of the heaviside step function is the Dirac delta function, which is not an element of L^2, which means $w_1(x)''$ is not in L^2.

Now, consider $w_2''(x)$. We observe the integral of $w_2''(x)^2$ on the domain:

$$\lim_{a \to 0^+} \int_a^1 (w_2''(x))^2 dx = \lim_{a \to 0^+} \int_a^1 (\frac{2}{9})^2 x^{-\frac{8}{3}} dx = \lim_{a \to 0^+} [-\frac{4}{135}x^{-\frac{5}{3}}]_a^1 = \infty$$

$\implies w_2''$ is not in L^2 either. $\implies$ neither $w_1''(x)$ nor $w_2''(x)$ are in $H^2(0, 1)$, as $w_1''(x)$ and $w_2''(x)$ are not in L^2.

Since the double derivative does no longer exist, the RHS can not be calculated in the same way. The integral is no longer possible to calculate directly:

$$F(v) = \int_0^1 -\alpha u_{xx}v + bu_x v + cuv dx$$

To integrate the first joint, we use integration by parts:

$$\int_0^1 -\alpha u_{xx} v dx = -\alpha([u_x v]_0^1 - \int_0^1 u_x v_x dx) = 0 + \alpha \int_0^1 u_x v_x dx)$$

In our solver, we use this to find the RHS, assemblying F-vector by integrating:

$$F_i = \int_0^1 f(x)\phi_i(x) dx = \alpha \int_0^1 u_x v_x + bu_x v + cuv$$

The plots in figure 4 and 5 shows that the modified implementation seems to give good approximations to the exact solutions. The convergens rates are 0.56 in H^1 and 1.02 in L^2 for w_1 and 1.16 in H^1 and 1.83 in L^2 for w_2. Again we can see that the convergence rate in H^1 is lower then the convergence rate in L^2, which is as expected. For this case we can however see that the convergencerates are lower then in the previous case, especially for w_1. The results are worse as we now try to solve the problem for non smooth potensials. As w_1 and w_2 are not in H^2 the error bound derived earlier does not hold as $\|w''\|_{L^2}$ does not exist. We can therefore not theoretically guarantee a convergence rate of respectively 1 and 2. When the exact solution is non-smooth there can be jumps and discontinuities in the first and second derivatives. These can make the finite element solution wiggle and therefore give a lower convergence rate. Furthermore the intepolation and numerical integration, which is used in the FEM, can give higher errors with non-smooth functions. For these kinds of problems the FEM also might not be able to capture the behaviour of the solution.

3.4 Skewed function and not equidistant nodes

We now consider the function $f(x) = x^{-\frac{1}{4}}$, and our numerical scheme with $b = -15$. This way, we get a function that is shiftet towards 0(we get a steeper descent close to 0). We want to examine if our numerical solver give a smaller error on grids with more nodes and shorter neighbor distances near x = 0, compared to using a grid with equidistant nodes. First we check that $f(x)$ is in L^2.

$$||f||_{L^2}^2 = \int_0^1 |f(x)|^2 dx = \int_0^1 x^{-\frac{1}{2}} dx = [2x^{\frac{1}{2}}]_0^1 = 2$$

Hence $||f||_{L^2} = \sqrt{2} < \infty$, and we have that $f(x) \in L^2(0,1)$.

Now, we turn our attention to the numerical solutions. The resulting plot can be seen in Figure 6. In this case, an exact solution is not known. Hence, we use a solution on a refined grid as our refrence solution. Looking at the convergence plot(Figure 7), we see that the solution with more nodes close to 0 has lower error than the solution using equidistant nodes. This is as expected as we get more nodes to approximate the steepest part of the solution.

4 Conclusion

In summary, this report has examined a 1D convection-diffusion-reaction model in the stationary regime. The primary aim was solving the partial differential equation (1) using a finite element method. The first part of the project focused on deriving a weak form (V) of (1), and showed that it has a unique solution. In the second part, we implemented the FEM for various problems with known solutions $u_{exact}(x)$, so that the numerical results could be compared. For example, the numerical results for the convergence rates that we found for smooth test functions agree well with the expected theoretical values. When solving for non-smooth potentials, where we adjusted the solver, we observed that the convergence rates in H^1 was lower than for L^2, which was also the case for the smooth potentials. These results are consistent with theory. On the other hand, the convergence rates for the non-smooth potentials were worse than in for the smooth ones. One possible explanation is that the error bound that was established for the smooth potentials does not apply to the non-smooth potentials because these are not in H^2. Nonetheless, the numerical solutions were in agreement with the known exact solutions, and the method behaved as expected. In conclusion, it is reasonable to state that the FEM works well for approximating a solution to the partial differential equation (1).