NTNU

Project

Least Squares Finite Element Method

Author:
Magnus Aarskaug Rud

Supervisor:
Anne Kvernø

Research Group Name
IME

May 2015

NTNU

Abstract

Faculty Name
IME

Project

Least Squares Finite Element Method

by Magnus Aarskaug Rud

The Thesis Abstract is written here (and usually kept to just this page). The page is kept centered vertically so can expand into the blank space above the title too...

Contents

Notation

CONVENTION	we let subscript h denote the discretized variables
u	Solution of the pde
$\mathbf{w} = [w_1\ w_2]$	Negative gradient of u
$\mathbf{u} = [\mathbf{w}\ u]^T$	Solution to the first order transfomation
R_g	Lifting function
$\tilde{\mathbf{u}} = \mathbf{u} - R_g$	Solution to the first order transfomation minus the lifting function
$\mathbf{b}$	Two dimensional vector field
f	loading function
$\mathbf{f} = (0, 0, f)$	loading function for the first order transformation
$\mathcal{L}, \mathcal{B}$	Linear operators
$(\cdot, \cdot)$	Inner product
$a(\cdot, \cdot)$	Bilinear form obtained from standard galerkin approach
$Q(\cdot, \cdot)$	Bilinear form obtained from the least squares approach
$\mathring{a}(\cdot, \cdot)$	Bilinear form obtained from the combined GLS-method
$F(\cdot)$	Linear form obtained from the least squares approach
$\tilde{F}(\cdot)$	Linear form obtained from the least squares approach including BC's
$\mathring{F}(\cdot)$	Linear form obtained from the combined GLS-method
$\mathcal{F}(\cdot, \cdot)$	non-linear functional
$\mathcal{J}$	Jacobian of $\mathcal{F}$
Linear Algebra	A good idea to denote matrices with an underscore ?
A	The stiffness matrix
G	The gradient matrix
F	The vector obtained from the linear functional F
B_1	The diagonal matrix with the first component of b evaluated in each node
B_2	The diagonal matrix with the second component of b evaluated in each node

F	The diagonal matrix with the loading function evaluated in each node
W	The diagonal matrix with the GLL-weights
L	The matrix with the derivative of the lagrange functions evaluated in each node
Φ	$W \otimes WL$
Ψ	$WL \otimes W$

Chapter 1

Theory

1.1 Informal introduction to least-squares

The least-squares finite element method is a numerical method with similarities to mixed Galerkin. However it has a fundamentally different approach regarding the definition of the bilinear functional. Let us look at a system of first order differential equations on the form

$$Au = f \text{ in } \Omega \tag{1.1}$$

$$u = g \text{ on } \partial\Omega. \tag{1.2}$$

Where A is a partial differential operator defined as

$$A = \sum_{i=1}^{n} A_i \frac{\partial}{\partial x_i} + A_0. \tag{1.3}$$

n being the number of dimensions of the domain Ω. If u happens to be a vector function of say k dimensions then A_i will be a matrix with k columns and k or more rows. Let us initially assume without loss of generality that $g = 0$. Further we require $f \in L_2(\Omega)$ and choose $V = \{v \in L_2(\Omega) | v = 0 \text{ on } \partial\Omega\}$. A residual is defined

$$R(v) = Av - f, \tag{1.4}$$

and a functional

$$J(v) = \frac{1}{2}||R(v)||_0^2. \tag{1.5}$$

The solution u and it's gradient needs to be in L^2 for the functional to make sense, hence u is restricted to the space $H_0^1(\Omega)$. The homogeneous boundary condition is now baked into the definition of the search space. By minimizing J we obtain

$$\lim_{t\to 0}\frac{d}{dt}J(u+tv) = \int_\Omega (Av)^T(Au-f)d\Omega = 0 \ , \forall v \in V. \tag{1.6}$$

We can now write a variational formulation of the least-squares method: Find $u \in V$ such that

$$Q(u,v) = F(v) \quad , \quad \forall v \in V, \tag{1.7}$$

where

$$Q(u,v) = (Au, Av), \tag{1.8}$$

$$F(v) = (f, Av). \tag{1.9}$$

Notice that the bilinear form Q is symmetric, this is an important advantage least-squares has over regular Galerkin methods. The bilinear form that surged from a first-order problem by the least-squares leads us to a variational formulation similar to the one obtained from a second order problem by regular FEM. Generally the bilinear form from least-squares will correspond to a bilinear form of a problem of twice the order obtained using FEM. Bold statement? In order to avoid problems of large complexity a higher order PDE should therefore be transformed to a system of first order PDE's (similar to a mixed Galerkin approach) before defining the least squares functional. [1]

1.2 Formal formulation of least-squares

Let us look at a general boundary value problem where $f \in Y(\Omega)$, $g \in B(\partial\Omega)$, $\mathcal{B}\colon X(\partial\Omega) \to B(\partial\Omega)$ and $\mathcal{L}\colon X(\Omega) \to Y(\Omega)$. Find $u \in X(\Omega)$ such that

$$\mathcal{L}u = f \quad \text{in } \Omega \tag{1.10}$$

$$\mathcal{B}u = g \quad \text{on } \partial\Omega. \tag{1.11}$$

Whenever this BVP has a unique solution, a least-squares functional can be defined as

$$J(u; f, g) = ||\mathcal{L}u - f||_Y^2 + ||\mathcal{B}u - g||_B^2 \tag{1.12}$$

and the corresponding minimization problem is then given as

$$\min_{u \in X} J(u; f, g) \tag{1.13}$$

For any well-posed problem $\exists \alpha, \beta > 0$ such that

$$\alpha||u||_X^2 \leq J(u; 0, 0) = (\mathcal{L}u, \mathcal{L}u)_Y + (\mathcal{B}u, \mathcal{B}u)_B \leq \beta||u||_X^2. \tag{1.14}$$

The fact that our functional is norm-equivalent is of crucial importance to a successful LS-method. It is therefore important that the spaces X, Y and B is chosen such that the LS-functional defines a norm is equivalent to $||\cdot||_X$. Minimizing this functional is equivalent to solving the Euler-Lagrange equations formulated as

$$\text{find } u \in X \text{ such that } Q(u, v) = F(v) \;\; \forall v \in X \tag{1.15}$$

Where $Q(u, v)$ and $F(v)$ are defined as

$$\begin{aligned} Q(u, v) &= (\mathcal{L}u, \mathcal{L}v)_Y + (\mathcal{B}u, \mathcal{B}v)_B, \\ F(v) &= (f, \mathcal{L}v)_Y + (g, \mathcal{L}v)_B. \end{aligned} \tag{1.16}$$

Notice that $Q(u, v)$ defines an inner product and $Q(u, u)^{1/2} = J(u; 0, 0)^{1/2}$ defines the corresponding norm, which we will name the *energy norm*,

$$|||u||| := Q(u,u)^{1/2} \tag{1.17}$$

In order to solve the BVP numerically we define discrete function spaces $X^h \subset X$, $Y^h \subset Y$ and $B^h \subset B$ and the corresponding variational formulation is then written as

$$\text{find } u^h \in X^h \text{ such that } Q(u^h, v^h) = F(v^h) \;\; \forall v^h \in X^h. \tag{1.18}$$

[2]

1.3 Error analysis

Let u be the analytical solution of a problem of the type (1.11), u^h is our numerical solution to (1.18) and $u^h_\perp$ is the orthogonal projection of u in X_h.

$$\begin{aligned} Q(u-u^h, u-u^h) &= Q(u-u^h, u-u^h_\perp) + Q(u-u^h, u^h_\perp - u^h) \\ &= Q(u-u^h, u-u^h_\perp) \\ &\leq \beta ||u-u^h||_{X_h} \, ||u-u^h_\perp||_{X_h}. \end{aligned} \tag{1.19}$$

The first equality is due to adding and subtracting $u^h_\perp$, because both u^h and $u^h_\perp$ solves the variational formulation we can cancel the last term, and by using the norm-equivalency from (1.14) and Schwartz inequality we get the last expression. Now by applying the first inequality of (1.14) we end up with

$$||u-u^h||_{X_h} \leq \frac{\beta}{\alpha} ||u-u^h_\perp||_{X_h} = \min_{w^h \in X_h} \frac{\beta}{\alpha} ||u-w^h||_{X_h}. \tag{1.20}$$

Hence we can show that the Least squares method provides a convergence result of similar order as the Finite element method. Add the convergence result of spectral methods

1.4 Least squares applied on some common PDE's

Before we start elaborating to much on each test case, I will provide a quick overview of the problems investigated in this project. All of them are BVP with either Neumann

or Dirichlet boundary conditions, the results can be found in chapter 4. The problems attempted to be solved in this project are simply expansions of the Poisson problem,

$$-\Delta u = f \text{ in } \Omega. \tag{1.21}$$

Then by adding a transport term you obtain the diffusion transport equation

$$-\mu\Delta u + \mathbf{b} \cdot \nabla u = f \text{ in } \Omega \tag{1.22}$$

This can be further expanded by adding a reaction term

$$-\mu\Delta u + \mathbf{b} \cdot \nabla u + \sigma u = f \text{ in } \Omega \tag{1.23}$$

And as a final complication the vector field $\mathbf{b}$ can be made dependent on u and hence give us the nonlinear diffusion transport reaction equation

$$-\mu\Delta u + \mathbf{b}(u) \cdot \nabla u + \sigma u = f \text{ in } \Omega \tag{1.24}$$

By defining $\mathbf{w} = -\nabla u$, and $\mathbf{u} = \mathbf{w} \oplus u$, all the equations listed above are transformed into a first order system of PDE's

$$\mathcal{L}\mathbf{u} = \mathbf{f}. \tag{1.25}$$

We will now investigate the partial differential operator $\mathcal{L}$ for each of the mentioned problems.

1.4.1 Poisson problem

The Poisson problem is defined as

$$-\Delta u = f \text{ in } \Omega \tag{1.26}$$

$$u = g \text{ on } \partial\Omega \tag{1.27}$$

Let us first consider the homogeneous case. The straight forward least-squares approach is to define $\mathbf{w} = -\nabla u$ and solve the system of equations

$$\mathbf{w} + \nabla u = 0 \text{ in } \Omega \tag{1.28}$$

$$\nabla \cdot \mathbf{w} = f \text{ in } \Omega \tag{1.29}$$

$$u = 0 \text{ on } \partial\Omega. \tag{1.30}$$

which can be written in the same form as (1.11) with $\mathbf{u} = \mathbf{w} \oplus u$, $\mathbf{f} = (0, 0, f)$, $g = 0$, $\mathcal{B} = (0, 0, 1)^T$ and $\mathcal{L}$ given as

$$\mathcal{L} = \begin{bmatrix} 1 & 0 & \frac{\partial}{\partial x} \\ 0 & 1 & \frac{\partial}{\partial y} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & 0 \end{bmatrix} \tag{1.31}$$

We define the search space $X = H^1(\Omega; \text{div}) \times H_0^1(\Omega)$ and the solution space $Y \times B = [L^2(\Omega)]^3 \times L^2(\partial\Omega)$. Is it ok to write Y x B like this ?? The functional defined in (1.12) will for our problem be

$$J(u; f) = ||\mathcal{L}u - f||_Y^2. \tag{1.32}$$

The term corresponding to the boundary conditions is imposed directly in the definition of our search space and is therefore not included in our functional as implied by equation (1.12). We will later show that this results in a better numerical solution. The corresponding variational formulation defined in (1.16) can now be stated. Find $\mathbf{u} \in X$ such that

$$Q(\mathbf{u}, \mathbf{v}) = F(\mathbf{v}) \ \ \forall \ \ \mathbf{v} \in X. \tag{1.33}$$

We require that $\mathbf{f} \in Y$. get a reference on this !!! Notice that the spaces X and Y chosen as described above fulfill the condition (1.14).

1.4.2 Diffusion transport reaction problem

The diffusion transport reaction problem to be analyzed is given as

$$-\mu\Delta u + \mathbf{b}\cdot\nabla u + \sigma u = f \text{ in } \Omega \tag{1.34}$$

$$u = g \text{ on } \partial\Omega \tag{1.35}$$

where μ is the diffusion constant, $\mathbf{b} = [b_1, b_2]$ is a vector field and σ is some reaction constant. By following the same approach as for the Poisson problem we end up with $\mathcal{L}$ on the form

$$\mathcal{L} = \begin{bmatrix} 1 & 0 & \frac{\partial}{\partial x} \\ 0 & 1 & \frac{\partial}{\partial y} \\ \mu\frac{\partial}{\partial x} - b_1 & \mu\frac{\partial}{\partial y} - b_2 & \sigma \end{bmatrix} \tag{1.36}$$

which leads to a similar but slightly different linear system than the one created by the Poisson problem.

1.4.3 Nonlinear problem

Now let the vector field $\mathbf{b}$ from the diffusion transport reaction problem be a function of u.

$$-\mu\Delta u + \mathbf{b}(u)\cdot\nabla u + \sigma u = f \text{ in } \Omega \tag{1.37}$$

$$u = g \text{ on } \partial\Omega \tag{1.38}$$

This gives us a nonlinear variational formulation that does not allow us to solve our system of equations directly. In order to find a solution Newtons method was used. A notable difference between standard Galerkin and least-squares is that the linear functional $F(\cdot)$ will also be non-linear in the LS case.

1.5 Boundary conditions

1.5.1 non-homogeneous Dirichlet boundary conditions

If $g \neq 0$ then we can simply define a lifting function $R_g \in X$ such that $R_g(\partial\Omega) = g(\partial\Omega)$. By defining $\tilde{\mathbf{u}} = \mathbf{u} - R_g$ we can replace $\mathbf{u}$ in the variation formulation and get

$$Q(\tilde{\mathbf{u}} + R_g, \mathbf{v}) = F(\mathbf{v}) \tag{1.39}$$

$$Q(\tilde{\mathbf{u}}, \mathbf{v}) + Q(R_g, \mathbf{v}) = F(\mathbf{v}) \tag{1.40}$$

$$Q(\tilde{\mathbf{u}}, \mathbf{v}) = F(\mathbf{v}) - Q(R_g, \mathbf{v}) \tag{1.41}$$

$$Q(\tilde{\mathbf{u}}, \mathbf{v}) = \tilde{F}(\mathbf{v}) \tag{1.42}$$

1.5.2 non-homogeneous Neumann boundary conditions

Because of the geometry of our problem and the fact that we define the flux as an extra variable we can transform the Neumann conditions to a Dirichlet condition on the flux.

$$\frac{\partial u}{\partial \mathbf{n}} = h \text{ on } \partial\Omega \tag{1.43}$$

$$\nabla u \cdot \mathbf{n} = h \tag{1.44}$$

$$\mathbf{w} \cdot \mathbf{n} = -h. \tag{1.45}$$

Let us define $\hat{x}$ and $\hat{y}$ as the unit vectors in each direction. Notice that for the west ($x = 0$) and east ($x = 1$) edges the normal vector $\mathbf{n} = \pm\hat{x}$, and at the north ($y = 1$) and south($y = 0$) edges $\mathbf{n} = \pm\hat{y}$. This way we can write the Neumann conditions as a Dirichlet condition on the first and second component of $\mathbf{w} = [w_1 \ w_2]$.

$$w_1 = \pm h \text{ for } y = 0 \text{ and } y = 1 \tag{1.46}$$

$$w_2 = \pm h \text{ for } x = 0 \text{ and } x = 1 \tag{1.47}$$

1.5.3 Boundary as an additional functional

The boundary conditions can also be implemented as the functional described in equation (1.12). For a BVP with Dirichlet BC's this will correspon to adding $||u - g||_0^2$ to our

functional J. By minimizing this term we end up with the following contribution to the variational form

$$(u, v)_{\partial\Omega} = (g, v)_{\partial\Omega} \tag{1.48}$$

Similarly for a BVP with Neumann boundary conditions on Γ_N and Dirichlet conditions on Γ_D the contribution to the functional will be

$$(u, v)_{\Gamma_D} + (\mathbf{n} \cdot \mathbf{w}, v)_{\Gamma_N} = (g_D, v)_{\Gamma_D} + (g_N, v)_{\Gamma_N} \tag{1.49}$$

Chapter 2

Stability properties of least-squares

2.1 Instability with the Galerkin formulation

With regular Galerkin approach for the diffusion transport equation you end up with the variational formulation

$$a(u,v) = (f,v) \quad \forall v \in V. \tag{2.1}$$

where V is some closed subspace of H^1, and the bilinear functional is given as

$$a(u,v) = \mu \int_\Omega \nabla u \cdot \nabla v \, d\Omega + \int_\Omega v(b \cdot \nabla u) d\Omega \tag{2.2}$$

We can define a new norm from this functional in the following manner

$$\begin{aligned} a(u,u) &= \mu \int_\Omega \nabla u \cdot \nabla u \, d\Omega + \int_\Omega b \cdot (u \nabla u) d\Omega \\ &= \mu ||\nabla u||_0^2 + \frac{1}{2} \int_\Omega b \cdot \nabla u^2 d\Omega \\ &= \mu ||\nabla u||_0^2 - \frac{1}{2} \int_\Omega u^2 (\nabla \cdot b) d\Omega \end{aligned} \tag{2.3}$$

Where we have used Greens theorem and the assumption that $u = 0$ on the boundary in the last equality. Let us assume that the divergence of our vector field can be bounded within some interval, say $\gamma_0 \leq -\frac{1}{2}\nabla \cdot b \leq \gamma_1$, we can now make a lower and upper bound

for the norm surging from the bilinear form

$$\mu||\nabla u||_0^2 + \gamma_0||u||_0^2 \leq a(u,u) \leq \mu||\nabla u||_0^2 + \gamma_1||u||_0^2 \tag{2.4}$$

It is clear that for negative γ_0 and sufficiently small μ the bilinear form is no longer coercive and thus our convergence requirements are no longer valid. Standard Galerkin method is therefore not a suited way to solve this type of problem.

2.2 Adding least squares in order to obtain stability

Now remember from equation (1.14) that the least squares formulation guarantees us a coercive bilinear form, given that the BVP has a solution and that the search- and solution space are chosen correctly. For this particular problem we can find α, β such that $\alpha||u||_1^2 \leq Q(\mathbf{u},\mathbf{u}) \leq \beta||u||_1^2$. Now, let us define $\mathring{a}(\cdot,\cdot)$ and $\mathring{f}(\cdot)$ as the as the linear combinations of the linear and bilinear form surging from standard Galerkin and least-squares method.

$$\begin{aligned} \mathring{a}(u,v) &= a(u,v) + \delta Q(\mathbf{u},\mathbf{v}) \\ \mathring{f}(v) &= (f,v) + \delta F(\mathbf{v}) \end{aligned} \tag{2.5}$$

Let us study the coerciveness of the bilinear form $\mathring{a}(\cdot,\cdot)$

$$\begin{aligned} \mathring{a}(u,u) &\geq \mu||\nabla u||_0^2 + \gamma_0||u||_0^2 + \delta\alpha||u||_1^2 \\ &\geq \mu||\nabla u||_0^2 + \gamma_0||u||_0^2 + \delta\alpha||u||_0^2 \\ &\geq \mu||\nabla u||_0^2 + \mu||u||_0^2 \\ &= \mu||u||_1^2 \end{aligned} \tag{2.6}$$

Don't like the way Q has to be defined with **u** In the third inequality we make the assumption that $\gamma_0 + \delta\alpha \geq \mu$ in other words δ, (the amount of smoothing from LS) has to be chosen such that $\delta \geq (\mu - \gamma_0)/\alpha$.

Using (1.16) with homogeneous boundary conditions we obtain the identities $Q(\mathbf{u},\mathbf{u}) = (\mathcal{L}\mathbf{u}, \mathcal{L}\mathbf{u})$ and $F(\mathbf{u}) = (\mathbf{f}, \mathcal{L}\mathbf{u})$ we can derive the following stability result for our discrete solution of the variational formulation,

$$\mathring{a}(u_h, u_h)_h \leq C||f||^2 \tag{2.7}$$

The proof for a similar method can be found in [3] Ch.12. Let us start by assuming that the Galerkin formulation provides a bilinear form that can be stated as earlier, $a(u_h, u_h) = \mu||\nabla u_h||_0^2 + \gamma||u_h||_0^2$, with $\gamma > 0$. Could be done with negative gamma as well

Proof:

$$\begin{aligned}
\mathring{a}(u_h, u_h) &= \mathring{f}(u_h) \\
&= (f, u_h) + \delta(\mathbf{f}, \mathcal{L}\mathbf{u_h}) \\
&= (\frac{1}{\sqrt{\gamma}} f, \sqrt{\gamma} u_h) + \delta(\mathbf{f}, \mathcal{L}\mathbf{u_h}) \\
&\leq ||\frac{1}{\sqrt{\gamma}} f||\ ||\sqrt{\gamma} u_h|| + \delta||\mathbf{f}||\ ||\mathcal{L}\mathbf{u_h}|| \\
&\leq \frac{1}{\gamma}||f||^2 + \frac{1}{4}\gamma||u_h||^2 + \delta||\mathbf{f}||^2 + \frac{\delta}{4}||\mathcal{L}\mathbf{u_h}||^2 \\
&\leq \frac{1}{\gamma}||f||^2 + \delta||\mathbf{f}||^2 + \frac{1}{4}\gamma||u_h||^2 + \frac{1}{4}\mu||\nabla u_h||^2 + \frac{\delta}{4}||\mathcal{L}\mathbf{u_h}||^2 \\
&= \frac{1}{\gamma}||f||^2 + \delta||f||^2 + \frac{1}{4}\mathring{a}(u_h, u_h).
\end{aligned} \tag{2.8}$$

Which Allows us to determine the constant $C = \frac{4}{3}(\frac{1}{\theta} + \delta)$

2.3 Estimation of the coercivity constant α

In order to determine the amount of smoothing acquired it is necessary to know the coercivity constant α. In this section I will determine this constant for the Poisson problem on $\Omega = (0, 1)^2$.

We start by proving the Poincaré inequality on our domain Ω, a similar proof can be found in [3]. Let $\mathbf{g} = \frac{1}{\sqrt{2}}[x, y]$ such that $k = \nabla \cdot \mathbf{g} = \frac{2}{\sqrt{2}}$ and let $u \in H_0^1$ then we can outline the following

$$
\begin{aligned}
||u||_0^2 &= k^{-1} \int_\Omega k|u(\mathbf{x})|^2 d\Omega \\
&= -k^{-1} \int_\Omega \mathbf{g} \cdot \nabla(|u(\mathbf{x})|^2) d\Omega \\
&= -2k^{-1} \int_\Omega \mathbf{g} \cdot [u(\mathbf{x}) \nabla(u(\mathbf{x}))] d\Omega \\
&\leq 2k^{-1} ||\mathbf{g}||_\infty ||u||_0 \, |u|_1 \\
&\leq 2k^{-1} \frac{1}{\sqrt{2}} ||u||_0 \, |u|_1
\end{aligned}
\tag{2.9}
$$

Dividing both sides with the L^2-norm leaves us with

$$
|u|_1 \geq ||u||_0 \tag{2.10}
$$

Further we can show the following result by using the definition of $||\cdot||_1$

$$
\begin{aligned}
||u||_1 &= \sqrt{||u||_0^2 + |u|_1^2} \\
&\leq \sqrt{2|u|_1^2} \\
&= \sqrt{2}|u|_1
\end{aligned}
\tag{2.11}
$$

This result can be used to show equivalency of the norm surging from the bilinear functional $Q(\cdot, \cdot)$ and the $||\cdot||_1$-norm . By using the inequality obtained in the previous section we can make the following argument,

$$
\begin{aligned}
|||u||| &= Q(\mathbf{u}, \mathbf{u})^{1/2} = (\mathcal{L}\mathbf{u}, \mathcal{L}\mathbf{u})_0^{1/2} \\
&= \sqrt{||\nabla u + \mathbf{w}||_0^2 + ||\nabla \cdot \mathbf{w}||_0^2} \\
&\geq ||\nabla \cdot \mathbf{w}||_0 = |\mathbf{w}|_1 \\
&\geq \frac{1}{\sqrt{2}} ||\mathbf{w}||_1 \\
&\geq \frac{1}{\sqrt{2}} ||\mathbf{w}||_0 = \frac{1}{\sqrt{2}} |u|_1 \\
&\geq \frac{1}{2} ||u||_1
\end{aligned}
\tag{2.12}
$$

Hence $\alpha = \frac{1}{2}$. We have then proven coercivity and found the coercivity constant of the energy norm obtained from the least-squares formulation for the Poisson problem. should find this constant for the difftrans problem

Chapter 3

Implementation

All the implementation has been done in two dimensions on the unit square $(0,1)^2$. The goal of the implementation has been to investigate the virtues of the least-squares method. It is implemented with finite element and spectral basis functions and is compared to the results from standard Galerkin formulation. Finally the combined GLS-method described in chapter 2 is also tested against the separate methods.

3.1 The bilinear form obtained from least-squares

For the general problem (1.11) the functional Q will take the form

$$Q(u,v) = \int_\Omega (\mathcal{L}v)^T(\mathcal{L}u)d\Omega. \tag{3.1}$$

Implementing Q requires two sets of basis functions $\{N_i\}$ that describes the search and solution space. In this project assignment the search and solution space will be described by the same set of basis functions which will depend on the method applied. u is discretized as

$$u_h = \sum_{I=0}^{K} a_I N_I. \tag{3.2}$$

Where K denotes the number of discretization points, which is the same as the number of basis functions. Is this statement OK? Since equation (1.15) requires equality for all test functions in the search space we simply solve it for each basis function. We are

therefore left with a system of K equations. Equation (3.1) can then be written for each test function as

$$\begin{aligned} Q(u_h, N_I) &= \int_\Omega (\mathcal{L}N_I)^T (\mathcal{L}u_h) d\Omega \\ &= \int_\Omega (\mathcal{L}N_I)^T (\mathcal{L}\sum_{J=1}^{K} a_J N_J) d\Omega \\ &= \sum_{J=1}^{K} \int_\Omega (\mathcal{L}N_I)^T (\mathcal{L}a_J N_J) d\Omega \\ &= \sum_{J=1}^{K} \int_\Omega (\mathcal{L}N_I)^T (\mathcal{L}a_J N_J) d\Omega \\ &= \sum_{J=1}^{K} \int_\Omega (\mathcal{L}N_I)^T (\mathcal{L}N_J) d\Omega \ \cdot a_J. \end{aligned} \tag{3.3}$$

The total system of equation for all test functions can then be written as a matrix equation

$$K^{LS}\mathbf{u} = F^{LS}. \tag{3.4}$$

Where $K^{LS}_{I,J} = \int_\Omega (\mathcal{L}N_I)^T (\mathcal{L}N_J) d\Omega$. Should somehow show the importance of this formula

Written explicitly for the **Poisson and diffusion transport equation** it will be a 3-by-3 matrix on the form

$$K^{LS}_{I,J} = \int_\Omega \begin{bmatrix} N_I N_J + \partial_x N_I \partial_x N_J & \partial_x N_I \partial_y N_J & N_I N_{J,x} \\ \partial_y N_I \partial_x N_J & N_I N_J + \partial_y N_I \partial_y N_J & N_I N_{J,y} \\ N_{I,x} N_J & N_{I,y} N_J & N_{I,x} N_{J,x} + N_{I,y} N_{J,y} \end{bmatrix} d\Omega \text{ where}$$

$\partial_x = \partial/\partial x$ for the Poisson problem and $\partial_x = \mu\partial/\partial x - b_1$ for the diffusion transport problem. Similarly F^{LS}_I will be given as

$$F^{LS}_I = \int_\Omega (\mathcal{L}N_I)^T \vec{f} d\Omega = \int_\Omega \begin{bmatrix} \partial_x N_I \\ \partial_x N_I \\ 0 \end{bmatrix} f d\Omega \tag{3.5}$$

Notice that by doing the splitting of variables we obtain a system of equations three times as big as if we were to solve the equation directly.

Note that the gradient of our solution must be restricted to the space $H^1(\text{div}, \Omega) = \{\mathbf{w} \in L^2(\Omega) | \nabla \cdot \mathbf{w} \in L^2(\Omega)\}$. In order to simplify the implementation I have in this project chosen my basis functions to a subspace such that $\mathbf{w} \in [H^1(\Omega)]^2$. In other words, the basis functions for the solution, and both components of the gradients are the same.

3.2 Least squares with finite element basis functions

The same triangular grid has been chosen for the implementation of both regular Galerkin FEM and the least-squares FEM. By choosing 1st order hat functions each element consists of three non-zero basis functions. For the GFEM this leads to calculating a 3×3 matrix for each element whereas for LSFEM a 9×9 matrix has to be calculated, although it has a relatively simple structure it still leads to additional computational costs and is a lot more tedious to implement.

What can I say here... not worth to show any derivations

3.3 Least squares with spectral basis functions

The spectral implementation is done using Gauss Lobatto nodes and quadrature rule. The basis functions are then chosen as the lagrange functions based on the GL nodes. Before we continue let us define the two matrices W and L that will help us construct the final block matrices.

$$W = \begin{cases} w_i & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \tag{3.6}$$

Where w_i is the ith GLL-weight, with this notation W is the $n \times n$ diagonal matrix with the GLL-weights along the diagonal, further L is the matrix containing the derivatives of the lagrange polynomials,

$$L_{i,j} = l'_j(x_i). \tag{3.7}$$

Now, with the helping matrices defined let us look at the linear system surging from a least squares formulation with spectral basis functions.

Notice that the discrete solution $\mathbf{u}_h$ of the system $Q(\mathbf{u}_h, \mathbf{v}_h) = F(\mathbf{v}_h)$ consist of the discretization of both u and $\mathbf{w} = -\nabla u$. $\mathbf{u}_h$ can be structured block wise or node wise, by choosing a block wise representation the final system of equations surging from the Poisson problem can be written as

$$\begin{bmatrix} K_{1,1} & K_{1,2} & K_{1,3} \\ K_{2,1} & K_{2,2} & K_{2,3} \\ K_{3,1} & K_{3,2} & K_{3,3} \end{bmatrix} \begin{bmatrix} u^h \\ w_1^h \\ w_2^h \end{bmatrix} = \begin{bmatrix} F_1 \\ F_2 \\ 0 \end{bmatrix}.$$

Where each block $K_{m,n}$ corresponds to calculating element m, n in the matrix (3.1) for all combinations of I, J. In order to keep track of the indexing we use the expressions $I = i + jN$ and $J = k + lN$ where the helping indices denotes the position in x, y-coordinates. Let us take a closer look at element $K_{1,3}$ in order to achieve a more compact notation.

$$\begin{aligned} (K_{1,3})_{I,J} &= \int_\Omega N_I N_{J,x} d\Omega \\ &= \int_\Omega l_i(x) l_j(y) l_k'(x) l_l(y) d\Omega \\ &= \sum_\alpha \sum_\beta w_\alpha w_\beta l_i(x_\alpha) l_j(y_\beta) l_k'(x_\alpha) l_l(y_\beta). \end{aligned} \tag{3.8}$$

The sum is obtained by using Gauss Lobatto quadrature rule. Now notice that the lagrange polynomials $l_j(y_\beta)$ are non-zero only when $\beta = j$, this implies that for the non-zero terms in the sum we have $\beta = j = l$, and $\alpha = i$. With these considerations the double sum above simplifies to

$$(K_{1,3})_{I,J} = w_i w_j l_k'(x_i) \tag{3.9}$$

Remember that $I = i + jN$ and $J = k + lN$, with N being the number of nodes in each spacial direction. This means that $K_{1,3}$ will consist of blocks where i and k goes from 1 to N while j and k are constant within each block. Since we require that $j = k$ we can immediately conclude that $K_{1,3}$ is nonzero only in the blocks along the diagonal. Notice that the factor $w_i l_k'(x_i)$ is the same for each block. And it can be written in matrix form as WL. Since each block WL is multiplied with w_j the whole matrix can simply be written as the Kronecker tensor product $(W \otimes WL)$. Similar reasoning can be made with all the other block matrices $K_{m,n}$ and for the Poisson problem we end up with a matrix on the form

$$K^{LS} = \begin{bmatrix} W \otimes (L^T W L + W) & WL \otimes L^T W & W \otimes WL \\ L^T W \otimes WL & (L^T W L + W) \otimes W & WL \otimes W \\ W \otimes L^T W & L^T W \otimes W & L^T W L \otimes W + W \otimes L^T W L \end{bmatrix}$$

Note that without the reformulation of the PDE as a first order system and with regular Galerkin formulation the stiffness matrix will simply be $K_{3,3} = W \otimes L^T W L + L^T W L \otimes W$.

3.4 LS spectral method for the diffusion transport reaction equation

The matrix corresponding to the discretized variational formulation can be divided into three parts due to its dependency on μ,$\mathbf{b}$ and σ. When using galerkin formulation one can easily add extra terms, and the bilinear form expands as a superposition of the bilinear form from each term. In the least-squares setting this is not the case, each term added creates a bit more chaos, I have however tried to illustrate a way one can divide up the total matrix resulting from the diffusion transport equation

$$-\mu \Delta u + \mathbf{b} \cdot \nabla u + \sigma u = f \text{ in } \Omega \tag{3.10}$$

3.4.1 Laplacian part

The part of the matrix which is closest related to the laplacian operator can be generated as

$$A^{LS} = \begin{bmatrix} W \otimes (\mu L^T W L + W) & \mu WL \otimes L^T W & W \otimes WL \\ \mu L^T W \otimes WL & (\mu L^T W L + W) \otimes W & WL \otimes W \\ W \otimes L^T W & L^T W \otimes W & L^T W L \otimes W + W \otimes L^T W L \end{bmatrix}$$

3.4.2 Gradient part

The gradient part will probvide us with a matrix that is dependent of $\mathbf{b}$. Let B_1 and B_2 be diagonal $n^2 \times n^2$ matrices with the values of the first and second component of $\mathbf{b}$

evaluated in each spacial node along the diagonal.

$$\begin{aligned} G_{1,1} &= -\mu B_1(W \otimes WL) - \mu(W \otimes L^T W)B_1 + B_1(W \otimes W)B_1 \\ G_{1,2} &= -\mu B_2(W \otimes WL) - \mu(L^T W \otimes W)B_1 + B_1(W \otimes W)B_2 \\ G_{2,1} &= -\mu B_1(WL \otimes W) - \mu(W \otimes L^T W)B_2 + B_2(W \otimes W)B_1 \\ G_{2,2} &= -\mu B_2(WL \otimes W) - \mu(L^T W \otimes W)B_2 + B_2(W \otimes W)B_2. \end{aligned} \tag{3.11}$$

Notice that $G_{2,1} = G_{1,2}^T$, the least-squares formulation always provide a symmetric system of equations. The full contribution from the gradient term given as a matrix is then given as

$$G^{LS} = \begin{bmatrix} G_{11} & G_{12} & 0 \\ G_{21} & G_{22} & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

3.4.3 Reaction part

The reactional term will consist of the following three $n^2 \times n^2$ sub matrices

$$\begin{aligned} R_{1,3} &= \sigma(\mu W \otimes L^T W - B_1 W \otimes W), \\ R_{2,3} &= \sigma(\mu L^T W \otimes W - B_2 W \otimes W), \\ R_{3,3} &= \sigma^2(W \otimes W). \end{aligned} \tag{3.12}$$

Now, because of the symmetry guaranteed by the least-squares formulation $R_{3,1} = R_{1,3}^T$ and $R_{3,2} = R_{2,3}^T$. Hence the total attribution from the extra reactional term is

$$R^{LS} = \begin{bmatrix} 0 & 0 & R_{1,3} \\ 0 & 0 & R_{2,3} \\ R_{3,1} & R_{3,1} & R_{3,3} \end{bmatrix}.$$

3.4.4 The loading function

Adding the gradient and reactional term in our equation also affects the loading function. In a compact notation the discretized loading vector from the variational formulation can

be written as $F^{LS} = \begin{bmatrix} \mu(W \otimes L^T W)F_m - (W \otimes W)B_1 F_m \\ \mu(L^T W \otimes W)F_m - (W \otimes W)B_2 F_m \\ \sigma(W \otimes W)F_m \end{bmatrix}$, where F_m is the vector with the loading function evaluated in each spacial node.

3.4.5 Total matrix

We can now define the total matrix for the diffusion transport reaction problem K^{LS} as

$$K^{LS} = A^{LS} + G^{LS} + R^{LS} \tag{3.13}$$

Notice that most of the terms in the final matrix are cross terms, and does not surge solely from one of the terms in the equation. The way that the matrices have been divided into parts here are of now other than practical reasons and can be done in a different matter.

3.5 non-linear diffusion transport problem

The stepwise algorithm to solve the nonlinear equation is described in chapter 2. However there are several computational steps that needs to be taken care of, both with regular Galerkin and least squares. In both cases we obtain two matrices which we will name A and G and with superscript LS if they refer to the least squares formulation. In both cases only G will depend on the numerical solution u_h. An important difference however is that in the LS setting the F vector will depend on u_h while in the straight forward Galerkin setting it will not. For regular Galerkin spectral approach we obtain

$$A\tilde{u} + AR_g + G(\tilde{u} + R_g)(\tilde{u} + R_g) - F = 0 \tag{3.14}$$

Notice that for each iteration the matrix $G(\tilde{u} + R_g)$ needs to be evaluated, the homogeneous boundary conditions on $\tilde{u}$ needs to be imposed and the Jacobian needs to be

calculated. The Jacobian $\mathcal{F}$ will for this setting be given as

$$\mathcal{J}_{i,j} = A_{i,j} + G(\tilde{u} + R_g)_{i,j} + (\tilde{u} + R_g)_i \frac{\partial}{\partial \tilde{u}_j}(G(\tilde{u} + R_g))_i. \tag{3.15}$$

With the LS formulation we obtain

$$A^{LS}\tilde{u} + A^{LS}R_g + G^{LS}(\tilde{u} + R_g)(\tilde{u} + R_g) - F^{LS}(\tilde{u} + R_g) = 0 \tag{3.16}$$

$$A^{LS}\tilde{u} + A^{LS}R_g + G^{LS}(\tilde{u} + R_g)\tilde{u} - F^{LS}(\tilde{u} + R_g) = 0 \tag{3.17}$$

It is clear from this equation that the term surging from the loading function also needs to be handled when calculating the Jacobian. The lifting function R_g does only have nonzero values in the third "block", hence it belongs to the kernel of the G-matrix.

$$\mathcal{J}_{i,j} = A_{i,j} + G(\tilde{u} + R_g)_{i,j} + [\frac{\partial}{\partial \tilde{u}_j}(G(\tilde{u} + R_g))_{i,:}]\tilde{u} \; - \frac{\partial}{\partial \tilde{u}_j}F(\tilde{u} + R_g)_i. \tag{3.18}$$

Let us first consider the F-vector. The terms are given in equation (3.4.4) and it is clear that only the last terms in each block depends on u, notice that it does not depend on the components of the gradient $[w_1 \; w_2]$. Hence the contribution to the total Jacobi matrix will only be in block $(1, 3)$ and $(2, 3)$. Further since B_1 and B_2 are both diagonal matrices where $B_{i,i}(u) = B_{i,i}(u_i)$ the Jacobian can be calculated efficiently by creating the matrices dB_1 and dB_2 which has the partial derivative of B_1, B_2 wrt. u evaluated in each node.

Chapter 4

Results

4.1 The main differences

Using least squares will always give you a SPD system of equations which can be advantageous. However for second order equations this system is three times as big as if we were to solve it using more standard methods. Comparing the correctness of the solution as done in figure 4.1 shows that the convergence rate is the same as for standard methods, but the the value of the residual is slightly higher for the least squares method. This can be explained by the functional that is minimized. Notice that in the least squares methods you minimize the **square** of the residual, while with galerkin you minimize the residual itself. since the correctness of both methods are restricted by the smoothness of the solution and the number of discrete points LS-methods will minimize the residual squared down to a given precision and hence the residual itself to a slightly higher value.

4.1.1 Poisson problem

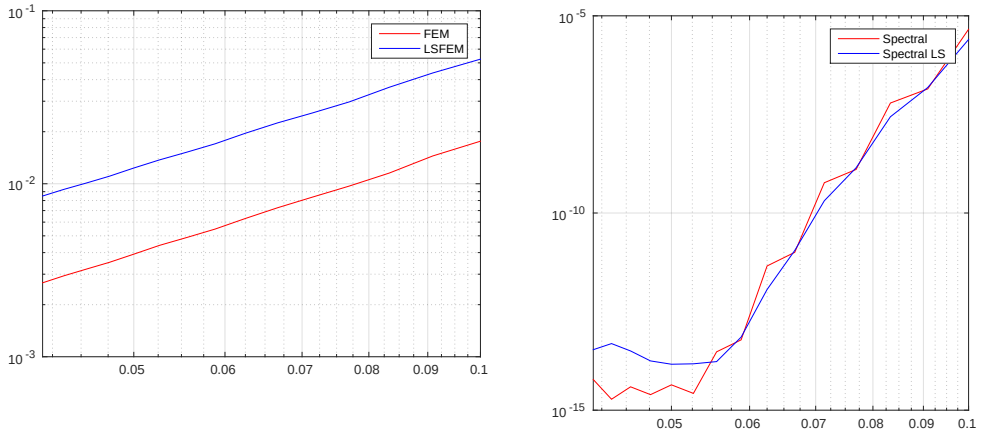

FIGURE 4.1: Convergence of Galerkin and corresponding least-squares formulation on the Poisson problem.

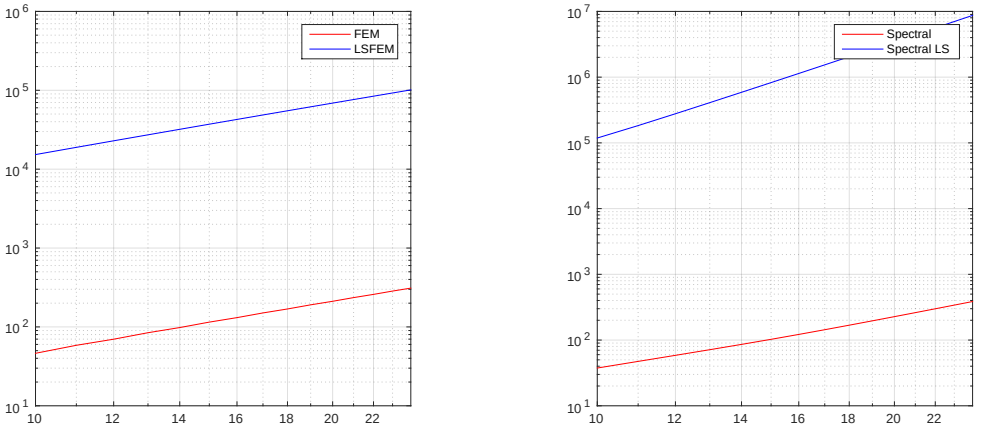

FIGURE 4.2: Condition number of Galerkin and corresponding least-squares formulation on the Poisson problem.

4.1.2 Diffusion transport equation convergence and condition number

One of the disadvantages of least-squares methods is that the resulting linear system is badly condition number. Now, let $\mathbf{b}$ be the vector field $[x, y]$, and let us see how the condition number scales with the parameters μ and β in the equation

$$-\mu\Delta u + \beta\mathbf{b} \cdot \nabla u = f \tag{4.1}$$

FIGURE 4.3: plot of the condition number when the paramaters μ and β are varied, $N = 25, \mathbf{b} = -[x, y]$.

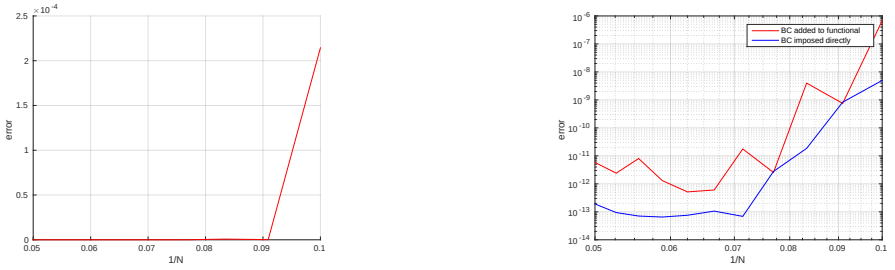

FIGURE 4.4: convergence plot $\mathbf{b} = -[x, y]$.

4.1.3 Diffusion transport equation surface plots

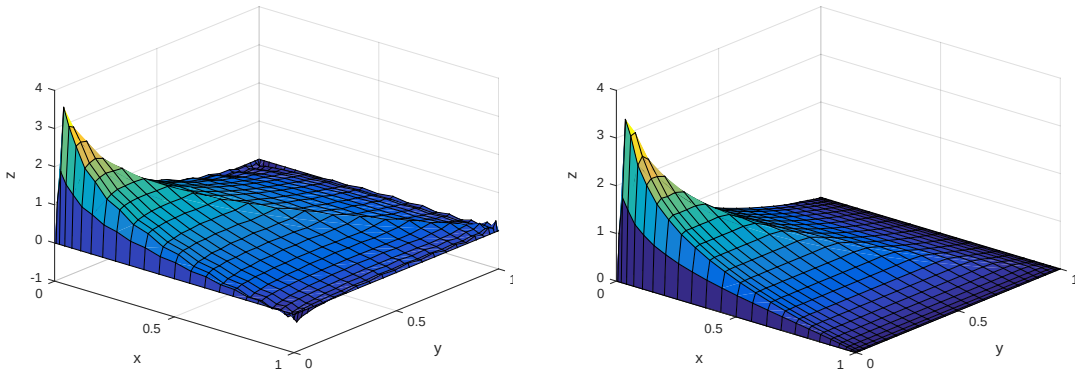

FIGURE 4.5: Surf plot of the numerical solution of the diffusion transport problem solved by galerkin spectral methods to the left and least-squares to the right, $\mu = 10^{-4}, N = 25, \mathbf{b} = -[x, y]$.

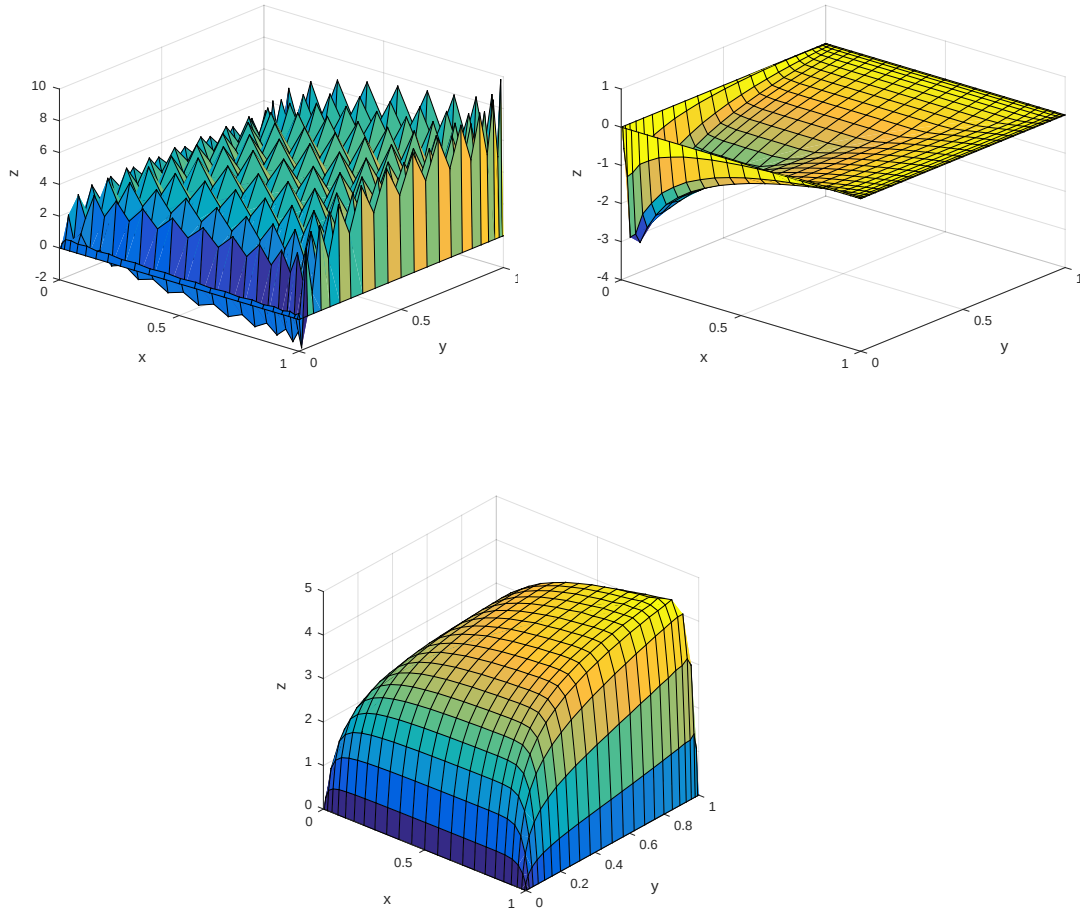

FIGURE 4.6: Surf plot of the numerical solution of the diffusion transport problem solved by galerkin to the left ,least-squares to the right and the combined GLS method below with $\delta = 0.05$, all the plots use the same parameters $\mu = 10^{-4}, N = 25, \mathbf{b} = [x, y]$.

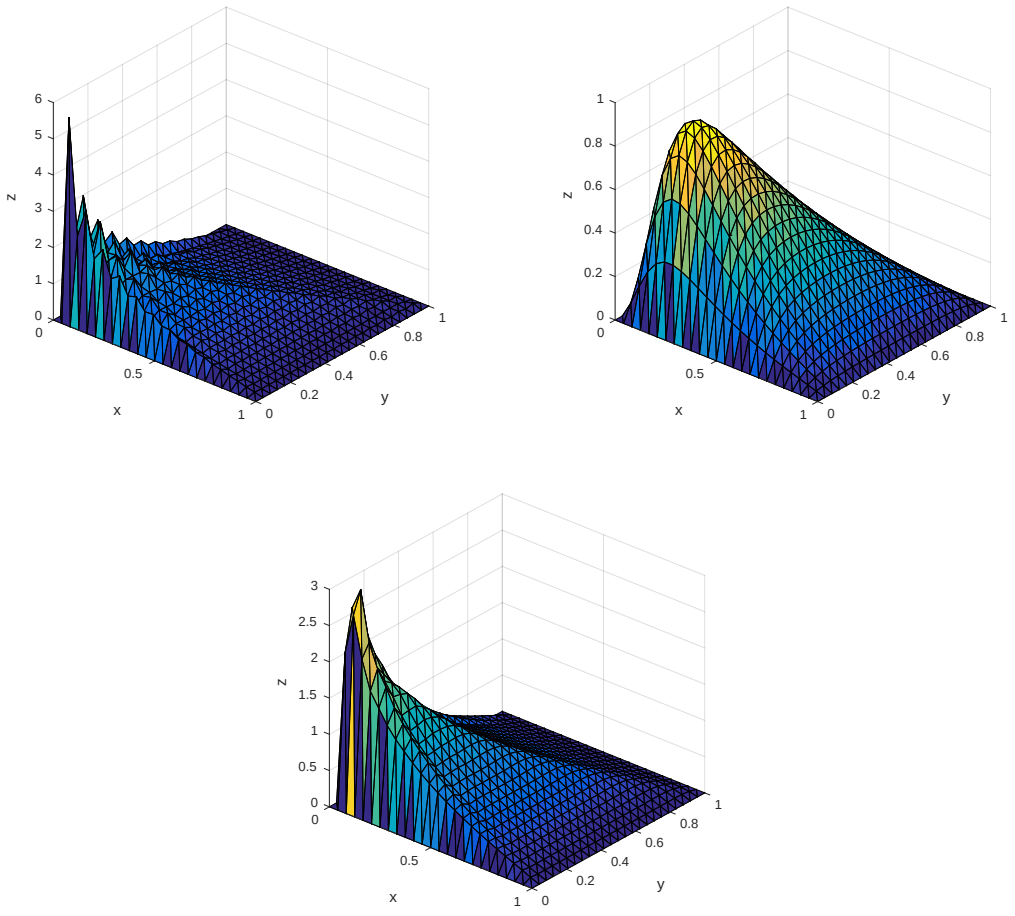

FIGURE 4.7: Surf plot of the numerical solution of the diffusion transport problem solved by galerkin FEM to the left, least-squares to the right and the combined GLS method below with $\delta = 0.05$, all the plots use the same parameters, $\mu = 10^{-4}, N = 25, \mathbf{b} = -[x, y]$.

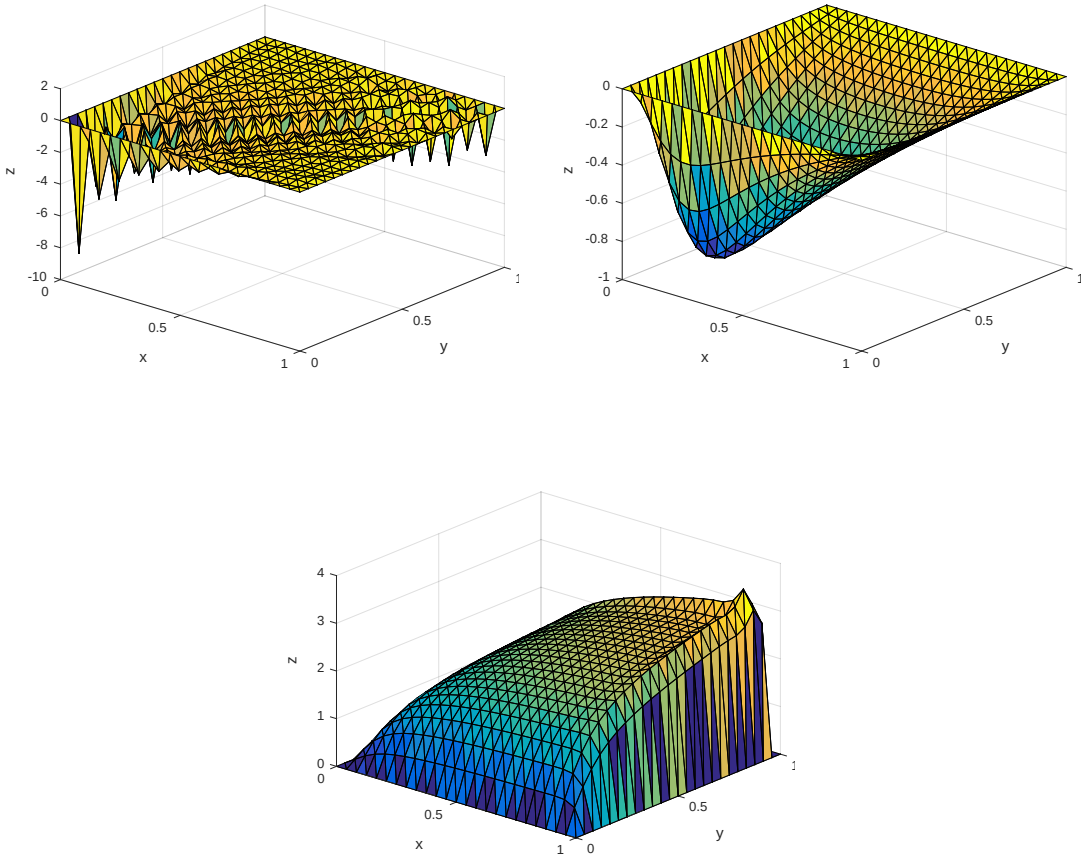

FIGURE 4.8: Surf plot of the numerical solution of the diffusion transport problem solved by galerkin FEM to the left ,least-squares to the right and the combined GLS method below with $\delta = 0.05$, all the plots use the same parameters $\mu = 10^{-4}, N = 25, \mathbf{b} = [x, y]$.

Appendix A

Appendix Title Here

Write your Appendix content here.

Bibliography

[1] Bo-nan Jiang. *The Least-Squares Finite Element Method.* Springer Berlin Heidelberg, 1998. ISBN http://id.crossref.org/isbn/978-3-662-03740-9. doi: 10.1007/978-3-662-03740-9. URL http://dx.doi.org/10.1007/978-3-662-03740-9.

[2] Max D. Gunzburger Pavel B. Bochev. *Least-Squares Finite Element Methods.* Springer, 2009.

[3] Alfio Quarteroni. *Numerical Models for Differential Problems, 2. edition.* Springer, 2014.