Numerical methods for the forward Diffusion Approximation

Martin Rodriguez-Vega

December 29, 2025

1 Preface

In this document, we will develop numerical schemes for solving the Diffusion Approximation (DA)

$$-\nabla \cdot D(\mathbf{r})\nabla\phi(\mathbf{r}) + \mu_a(\mathbf{r})\phi(\mathbf{r}) = s(\mathbf{r}) \tag{1}$$

where $\mathbf{r}$ is a point in some domain Ω, ϕ [W/m^2] is the fluence or energy flux, μ_a [m^{-1}] is the absorption coefficient, D [m] is the diffusion coefficient, and s is a source term.

As this is intended as an introductory primer, we will initially restrict out analysis to the Simplified DA, which assumes the diffusion coefficient is homogeneous in space, so (1) becomes

$$-\nabla^2\phi + k(\mathbf{r})^2\phi = \frac{s(\mathbf{r})}{D} \tag{2}$$

where $k^2 = \mu_a/D$ (the squared is a convention commonly used in the literature and in the study of differential equations). Furthermore, we will consider the two dimensional case and a slab geometry *i.e.* $\Omega = [0, L_x] \times [0, L_y]$, using a uniform rectangular discretization. However, hints will be included on how to generalize.

It is expected that the reader is proficient with vector calculus and linear algebra.

2 Finite Element Method

The Finite Element Method (FEM) consists on approximating the solution using piece-wise polynomial functions on a discretized mesh. An element is a region of the domain defined by a collection of vertices as shown in Fig. 1. For each vertex i there is a basis function $\psi_i(\mathbf{r})$ which is (usually) polynomial inside each element and continuous across boundaries.
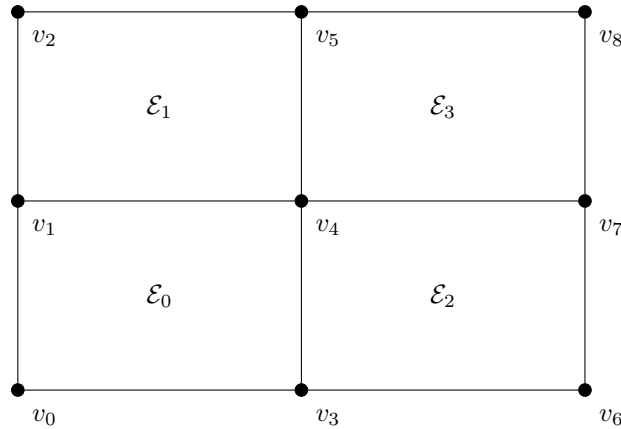

Figure 1: Simple mesh consisting of 9 vertices and 4 elements.

For instance, the basis function associated with vertex 4 in the figure is

$$\psi_4(x,y) = \begin{cases} (1-x)(1-y) & \text{if} \quad (x,y) \in [-1,0]^2 \\ (1-x)(1+y) & \text{if} \quad (x,y) \in [-1,0] \times [0,1] \\ (1+x)(1-y) & \text{if} \quad (x,y) \in [0,1] \times [-1,0] \\ (1+x)(1+y) & \text{if} \quad (x,y) \in [0,1]^2 \end{cases} \tag{3}$$

assuming the grid covers $[-1,1]^2$. This are the quadrilateral degree 1 basis function, and is a biliear function.

We want to approximate

$$\phi(\mathbf{r}) = \sum_i \phi_i \psi_i(\mathbf{r}) \tag{4}$$

where ϕ_i are weights that have to be determined. Note that by the form of basis functions that we chose, $\psi_i(\mathbf{r})$ is 1 at the position of v_i and 0 for any other vertex, so ϕ_i is simply the value of the function at v_i.

In general, the elements used to define the basis functions are not squares of side length 1. To use the same definition, we introduce the normalized coordinates $\mathbf{p}$ such that $\psi_i(\mathbf{p})$ is defined as before, and there is an (invertible) mapping $\xi : \mathbb{R}^2 \to \mathbb{R}^2$ such that $\mathbf{r} = \xi(\mathbf{p})$ Thus

$$\psi_i(\mathbf{r}) = \psi_i(\xi(\mathbf{p})) \tag{5}$$

For general quadrilaterals with points vertices located at $\mathbf{r}_0, \mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$, we have

$$\xi(\mathbf{p}) = \sum_{i=0}^3 \psi_i(\mathbf{p})\mathbf{r}_i \tag{6}$$

inside each element. When the grid is regular consisting of parallelograms, ξ is a linear function: $\xi(\mathbf{p}) = \Xi\mathbf{p}$ where Ξ is a 2×2 matrix. In the simple case of rectangular grids

$$\Xi = \begin{bmatrix} \Delta x & 0 \\ 0 & \Delta y \end{bmatrix} \tag{7}$$

Now, to analyze (2), we introduce the concept of weak formulation. The idea behind this, is that numerical methods are better behaved when dealing with integrals than derivatives, even more so when they are of higher order. The weak form of a differential equation is obtained by multiplying it by a "well-behaved" (in the sense that it is bounded, continuous and differentiable in a given domain) test function $w(\mathbf{r})$ and then we integrate over Ω:

$$-\int_\Omega w(\mathbf{r})\nabla^2\phi(\mathbf{r})\mathrm{d}^2\mathbf{r} + \int_\Omega k(\mathbf{r})^2 w(\mathbf{r})\phi(\mathbf{r})\mathrm{d}^2\mathbf{r} = \int_\Omega w(\mathbf{r})\frac{s(\mathbf{r})}{D}\mathrm{d}^2\mathbf{r} \tag{8}$$

and using Green's first identity

$$\int_\Omega \nabla w \cdot \nabla\phi\,\mathrm{d}^2\mathbf{r} - \int_{\partial\Omega} w\frac{\partial\phi}{\partial\mathbf{n}}\mathrm{d}\mathbf{r} + \int_\Omega k^2 w\phi\,\mathrm{d}^2\mathbf{r} = \int_\Omega w\frac{s}{D}\mathrm{d}^2\mathbf{r} \tag{9}$$

The interest of the weak formulation is that we were able to "move" one of the derivatives of the Laplacian to the test function. The second term encodes boundary conditions. For instance, if we use homogeneous Dirichlet conditions, *i.e.* $\phi(\mathbf{r}) = 0$ for $\mathbf{r} \in \partial\Omega$, we can force the test function to satisfy them as well, so

$$\int_\Omega \nabla w \cdot \nabla\phi\,\mathrm{d}^2\mathbf{r} + \int_\Omega k^2 w\phi\,\mathrm{d}^2\mathbf{r} = \int_\Omega \frac{s}{D}\mathrm{d}^2\mathbf{r} \tag{10}$$

Expanding the approximation of $\phi(\mathbf{r})$ using the basis functions:

$$\sum_i \phi_i \int_\Omega \nabla w \cdot \nabla\psi_i\,\mathrm{d}^2\mathbf{r} + \sum_i \phi_i \int_\Omega k^2 w\psi_i\,\mathrm{d}^2\mathbf{r} = \int_\Omega w\frac{s}{D}\mathrm{d}^2\mathbf{r} \tag{11}$$

This has to satisfied for any test function w, so we choose for simplicity one of the basis functions ψ_i, so

$$\sum_i \phi_i \int_\Omega \nabla\psi_j(\mathbf{r}) \cdot \nabla\psi_i(\mathbf{r})\mathrm{d}^2\mathbf{r} + \sum_i \phi_i \int_\Omega k(\mathbf{r})^2\psi_j(\mathbf{r})\psi_i(\mathbf{r})\mathrm{d}^2\mathbf{r} = \int_\Omega \psi_j(\mathbf{r})\frac{s}{D}\mathrm{d}^2\mathbf{r} \tag{12}$$

To apply the change of coordinates, we denote the Jacobian matrix as $J_\xi(\mathbf{p}) = \partial \xi / \partial \mathbf{p}$. Furthermore, we use the numerator layout for matrix calculus, so gradients are row vectors. With this, the change of coordinates for the gradient is

$$\nabla \psi(\mathbf{r}) = \nabla_\mathbf{p} \psi(\mathbf{p}) J_{\xi^{-1}}(\mathbf{r}) = \nabla_\mathbf{p} \psi(\mathbf{p}) J_\xi^{-1}(\mathbf{p}) \tag{13}$$

where the last term uses the Inverse Function Theorem.

$$\sum_i \phi_i \int_\Omega \nabla_\mathbf{p} \psi_j(\mathbf{p}) J_\xi^{-1}(\mathbf{p}) J_\xi^{-\top}(\mathbf{p}) \nabla_\mathbf{p}^\top \psi_i(\mathbf{p}) \left| \det J_\xi(\mathbf{p}) \right| \mathrm{d}^2\mathbf{p}$$

$$+ \sum_i \phi_i \int_\Omega k^2(\mathbf{p}) \psi_j(\mathbf{p}) \psi_i(\mathbf{p}) \left| \det J_\xi(\mathbf{p}) \right| \mathrm{d}^2\mathbf{p} \tag{14}$$

$$= \int_\Omega \psi_j(\mathbf{p}) \frac{s(\mathbf{p})}{D} \left| \det J_\xi(\mathbf{p}) \right| \mathrm{d}^2\mathbf{p}$$

As we have chosen the basis functions to be compact, only few of the pairs (i, j) will have non-zero terms in the summation. Geometrically, this can be seen as the vertices i, j which share at least one element. Furthermore, inside each element, the functions are polynomial, and we assume as simplification that k^2 is piece-wise constant. Let $\mathcal{E}_i^j$ be the set of elements that have both vertices i and j. Then,

$$\sum_i \phi_i \sum_{\mathcal{E} \in \mathcal{E}_i^j} \int_\mathcal{E} \nabla_\mathbf{p} \psi_j(\mathbf{p}) J_\xi^{-1}(\mathbf{p}) J_\xi^{-\top}(\mathbf{p}) \nabla_\mathbf{p}^\top \psi_i(\mathbf{p}) \left| \det J_\xi(\mathbf{p}) \right| \mathrm{d}^2\mathbf{p}$$

$$+ \sum_i \phi_i \sum_{\mathcal{E} \in \mathcal{E}_i^j} k_\mathcal{E}^2 \int_\mathcal{E} \psi_j(\mathbf{p}) \psi_i(\mathbf{p}) \left| \det J_\xi(\mathbf{p}) \right| \mathrm{d}^2\mathbf{p} \tag{15}$$

$$= \int_\Omega \psi_j(\mathbf{p}) \frac{s(\mathbf{p})}{D} \left| \det J_\xi(\mathbf{p}) \right| \mathrm{d}^2\mathbf{p}$$

We introduce back the simplifying assumption that the physical grid consists of regular squares with side lengths Δx, so $J_\epsilon = \Delta x \mathbb{I}$. Thus

$$\sum_i \phi_i \sum_{\mathcal{E} \in \mathcal{E}_i^j} \int_\mathcal{E} \nabla_\mathbf{p} \psi_j \nabla_\mathbf{p}^\top \psi_i \mathrm{d}^2\mathbf{p} + \Delta x^2 \sum_i \phi_i \sum_{\mathcal{E} \in \mathcal{E}_i^j} k_\mathcal{E}^2 \int_\mathcal{E} \psi_j \psi_i \mathrm{d}^2\mathbf{p}$$

$$= \Delta x^2 \int_\Omega \psi_j \frac{s}{D} \mathrm{d}^2\mathbf{p} \tag{16}$$

To evaluate the integrals, we can use the symmetry of mesh. First, note that the basis functions are the product between two identical functions of a single variable. Moreover, for each element we can shift and rotate the coordinates as needed, so that

$$\psi(x, y) = \nu(x)\nu(y) \quad , \quad \nu(x) = 1 - x \quad , \quad x, y \in [0, 1] \tag{17}$$

and for any other vertex of the same element, the basis function is obtained by using $\nu(1 - x) = x$ for one or both terms as needed. Therefore, the only integrals required are

$$\int_0^1 \nu(x)^2 \mathrm{d}x = \frac{1}{3} \qquad \int_0^1 \nu(x)\nu(1 - x) \mathrm{d}x = \frac{1}{6}$$

$$\int_0^1 \nu'(x)^2 \mathrm{d}x = 1 \qquad \int_0^1 \nu'(x)\nu'(1 - x) \mathrm{d}x = -1 \tag{18}$$

Thus,

$$\int_\mathcal{E} \psi_i \psi_j \mathrm{d}^2\mathbf{p} = \int_0^1 \nu_i(x)\nu_j(x)\mathrm{d}x \int_0^1 \nu_i(y)\nu_j(y)\mathrm{d}y \tag{19}$$

$$\int_\mathcal{E} \nabla \psi_i \cdot \nabla \psi_j \mathrm{d}^2\mathbf{p} = \int_0^1 \nu_i'(x)\nu_j'(x)\mathrm{d}x \int_0^1 \nu_i(y)\nu_j(y)\mathrm{d}y + \int_0^1 \nu_i(x)\nu_j(x)\mathrm{d}x \int_0^1 \nu_i'(y)\nu_j'(y)\mathrm{d}y \tag{20}$$

There are three types of non-zero interactions between vertices i, j:

1. $i = j$ which yields 4 elements and $\nu_i(x) = \nu_j(x)$ (same for y), so

$$\int_{\mathcal{E}} \psi_i \psi_j \mathrm{d}^2 \mathbf{p} = \frac{1}{9} \tag{21}$$

$$\int_{\mathcal{E}} \nabla \psi_i \cdot \nabla \psi_j \mathrm{d}^2 \mathbf{p} = \frac{2}{3} \tag{22}$$

2. i and j are orthogonally adjacent yielding 2 elements, $\nu_i(x) = \nu_j(1-x)$ and $\nu_i(y) = \nu_j(y)$ (or vice-versa)

$$\int_{\mathcal{E}} \psi_i \psi_j \mathrm{d}^2 \mathbf{p} = \frac{1}{18} \tag{23}$$

$$\int_{\mathcal{E}} \nabla \psi_i \cdot \nabla \psi_j \mathrm{d}^2 \mathbf{p} = -\frac{1}{6} \tag{24}$$

3. i and j are diagonally adjacent which yields 1 element and $\nu_i(x) = \nu_j(1-x)$ (same for y), so

$$\int_{\mathcal{E}} \psi_i \psi_j \mathrm{d}^2 \mathbf{p} = \frac{1}{36} \tag{25}$$

$$\int_{\mathcal{E}} \nabla \psi_i \cdot \nabla \psi_j \mathrm{d}^2 \mathbf{p} = -\frac{1}{3} \tag{26}$$

For each vertex, there is one vertex of type 1 (itself), four of type 2, and four of type 3.