Numerical Solution of Partial Differential Equations

Endre Süli

Mathematical Institute
University of Oxford

2024

Lecture 2

Elliptic boundary-value problems

A second-order linear PDE for a function $u = u(x, y)$:

$$a(x,y)\frac{\partial^2 u}{\partial x^2} + 2b(x,y)\frac{\partial^2 u}{\partial x \partial y} + c(x,y)\frac{\partial^2 u}{\partial y^2}$$
$$+ d(x,y)\frac{\partial u}{\partial x} + e(x,y)\frac{\partial u}{\partial y} = f(x,y) \qquad \text{is}$$

- ELLIPTIC if $b^2 - ac < 0$;
- PARABOLIC if $b^2 - ac = 0$; (and at least one of a or c is nonzero);
- HYPERBOLIC if $b^2 - ac > 0$.

Elliptic boundary-value problems

A second-order linear PDE for a function $u = u(x, y)$:

$$a(x,y)\frac{\partial^2 u}{\partial x^2} + 2b(x,y)\frac{\partial^2 u}{\partial x \partial y} + c(x,y)\frac{\partial^2 u}{\partial y^2} + d(x,y)\frac{\partial u}{\partial x} + e(x,y)\frac{\partial u}{\partial y} = f(x,y) \qquad \text{is}$$

- ELLIPTIC if $b^2 - ac < 0$;
- PARABOLIC if $b^2 - ac = 0$; (and at least one of a or c is nonzero);
- HYPERBOLIC if $b^2 - ac > 0$.

Ellipticity amounts to requiring that a and c are of the same sign, say $a > 0$ and $c > 0$ (or $a < 0$ and $c < 0$), and $ac - b^2 > 0$, which is equivalent (by Sylvester's criterion) to demanding that

$$A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$$

is a positive definite matrix, i.e. $\xi^{\mathrm{T}} A \xi > 0$ for all $\xi \in \mathbb{R}^2 \setminus \{0\}$.

Example (Elliptic equations)

(a) Laplace's equation: $\Delta u = 0$;

Example (Elliptic equations)

(a) Laplace's equation: $\Delta u = 0$;

(b) Poisson's equation $-\Delta u = f$;

Example (Elliptic equations)

(a) Laplace's equation: $\Delta u = 0$;

(b) Poisson's equation $-\Delta u = f$;

(c) More generally, let Ω be a bounded open set in $\mathbb{R}^n$, and consider the (linear) second-order partial differential equation

$$-\sum_{i,j=1}^{n} \frac{\partial}{\partial x_j}\left(a_{i,j}(x)\frac{\partial u}{\partial x_i}\right) + \sum_{i=1}^{n} b_i(x)\frac{\partial u}{\partial x_i} + c(x)u = f(x), \quad x \in \Omega,$$

where the coefficients $a_{i,j}$, b_i, c and f are such that

$$\begin{aligned} a_{i,j} &\in C^1(\overline{\Omega}), \qquad && i,j = 1,\dots,n; \\ b_i &\in C(\overline{\Omega}), && i = 1,\dots,n; \\ c &\in C(\overline{\Omega}), && f \in C(\overline{\Omega}), \quad \text{and} \end{aligned}$$

$$\sum_{i,j=1}^{n} a_{i,j}(x)\xi_i\xi_j \geq \tilde{c}\sum_{i=1}^{n}\xi_i^2 \qquad \forall\, \xi = (\xi_1,\dots,\xi_n) \in \mathbb{R}^n, \quad \forall\, x \in \overline{\Omega};$$

here $\tilde{c}$ is a positive constant independent of x and ξ.

An elliptic equation is usually supplemented with one of the following boundary conditions:

(a) $u = g$ on $\partial\Omega$ (*Dirichlet boundary condition*);

An elliptic equation is usually supplemented with one of the following boundary conditions:

(a) $u = g$ on $\partial\Omega$ (*Dirichlet boundary condition*);

(b) $\frac{\partial u}{\partial \nu} = g$ on $\partial\Omega$, where ν denotes the unit outward normal vector to $\partial\Omega$ (*Neumann boundary condition*);

An elliptic equation is usually supplemented with one of the following boundary conditions:

(a) $u = g$ on $\partial\Omega$ (*Dirichlet boundary condition*);

(b) $\frac{\partial u}{\partial \nu} = g$ on $\partial\Omega$, where ν denotes the unit outward normal vector to $\partial\Omega$ (*Neumann boundary condition*);

(c) $\frac{\partial u}{\partial \nu} + \sigma u = g$ on $\partial\Omega$, where $\sigma(x) \geq 0$ on $\partial\Omega$ (*Robin boundary cond.*);

An elliptic equation is usually supplemented with one of the following boundary conditions:

(a) $u = g$ on $\partial\Omega$ (*Dirichlet boundary condition*);

(b) $\frac{\partial u}{\partial \nu} = g$ on $\partial\Omega$, where ν denotes the unit outward normal vector to $\partial\Omega$ (*Neumann boundary condition*);

(c) $\frac{\partial u}{\partial \nu} + \sigma u = g$ on $\partial\Omega$, where $\sigma(x) \geq 0$ on $\partial\Omega$ (*Robin boundary cond.*);

(d) A more general version of (b) and (c) is

$$\sum_{i,j=1}^{n} a_{i,j} \frac{\partial u}{\partial x_i} \cos \alpha_j + \sigma(x) u = g \qquad \text{on } \partial\Omega,$$

where α_j is the angle between the unit outward normal vector ν to $\partial\Omega$ and the Ox_j axis (*oblique derivative boundary cond.*).

Classical solutions

Consider the homogeneous Dirichlet boundary-value problem:

$$-\sum_{i,j=1}^{n} \frac{\partial}{\partial x_j}\left(a_{i,j}(x)\frac{\partial u}{\partial x_i}\right) + \sum_{i=1}^{n} b_i(x)\frac{\partial u}{\partial x_i} + c(x)u = f(x) \quad \text{for } x \in \Omega, \tag{1}$$

$$u = 0 \quad \text{on } \partial\Omega, \tag{2}$$

where $a_{i,j}$, b_i, c and f are as stated earlier.

Classical solutions

Consider the homogeneous Dirichlet boundary-value problem:

$$-\sum_{i,j=1}^{n} \frac{\partial}{\partial x_j}\left(a_{i,j}(x)\frac{\partial u}{\partial x_i}\right) + \sum_{i=1}^{n} b_i(x)\frac{\partial u}{\partial x_i} + c(x)u = f(x) \quad \text{for } x \in \Omega, \tag{1}$$

$$u = 0 \qquad \text{on } \partial\Omega, \tag{2}$$

where $a_{i,j}$, b_i, c and f are as stated earlier.

A function $u \in C^2(\Omega) \cap C(\overline{\Omega})$ satisfying (1) and (2) is called a *classical solution* of this problem.

Classical solutions

Consider the homogeneous Dirichlet boundary-value problem:

$$-\sum_{i,j=1}^{n} \frac{\partial}{\partial x_j}\left(a_{i,j}(x)\frac{\partial u}{\partial x_i}\right) + \sum_{i=1}^{n} b_i(x)\frac{\partial u}{\partial x_i} + c(x)u = f(x) \quad \text{for } x \in \Omega, \tag{1}$$

$$u = 0 \qquad \text{on } \partial\Omega, \tag{2}$$

where $a_{i,j}$, b_i, c and f are as stated earlier.

A function $u \in C^2(\Omega) \cap C(\overline{\Omega})$ satisfying (1) and (2) is called a *classical solution* of this problem.

The theory of partial differential equations tells us that (1), (2) has a unique classical solution, provided that $a_{i,j}$, b_i, c, f and $\partial\Omega$ are sufficiently smooth.

Weak solutions

If these smoothness requirements on the coefficients are violated, the classical theory of partial differential equations is inappropriate.

Weak solutions

If these smoothness requirements on the coefficients are violated, the classical theory of partial differential equations is inappropriate.

Example

Take, for example, Poisson's equation on the cube $\Omega = (-1,1)^n$ in $\mathbb{R}^n$, subject to a zero Dirichlet boundary condition:

$$\left.\begin{array}{rcll} -\Delta u & = & \operatorname{sgn}\left(\frac{1}{2} - |x|\right), & x \in \Omega, \\ u & = & 0, & x \in \partial\Omega. \end{array}\right\} \qquad (*)$$

Weak solutions

If these smoothness requirements on the coefficients are violated, the classical theory of partial differential equations is inappropriate.

Example

Take, for example, Poisson's equation on the cube $\Omega = (-1, 1)^n$ in $\mathbb{R}^n$, subject to a zero Dirichlet boundary condition:

$$\left.\begin{array}{rcll} -\Delta u & = & \operatorname{sgn}\left(\frac{1}{2} - |x|\right), & x \in \Omega, \\ u & = & 0, & x \in \partial\Omega. \end{array}\right\} \qquad (*)$$

This problem has no classical solution, $u \in C^2(\Omega) \cap C(\overline{\Omega})$, for otherwise Δu would be a continuous function on Ω, which is not possible because $\operatorname{sgn}(1/2 - |x|)$ is not a continuous function on Ω.

Definition (Weak solution)

Let $a_{i,j} \in C(\overline{\Omega})$, $i, j = 1, \ldots, n$, $b_i \in C(\overline{\Omega})$, $i = 1, \ldots, n$, $c \in C(\overline{\Omega})$, and let $f \in L^2(\Omega)$. A function $u \in H_0^1(\Omega)$ satisfying

$$\sum_{i,j=1}^{n} \int_\Omega a_{i,j}(x) \frac{\partial u}{\partial x_i} \frac{\partial v}{\partial x_j} \, \mathrm{d}x + \sum_{i=1}^{n} \int_\Omega b_i(x) \frac{\partial u}{\partial x_i} v \, \mathrm{d}x + \int_\Omega c(x) u v \, \mathrm{d}x = \int_\Omega f(x) v(x) \, \mathrm{d}x \qquad \forall\, v \in H_0^1(\Omega)$$

is called a *weak solution* of (1), (2).

Example

Suppose that $\Omega = (a, b) \times (c, d) \subset \mathbb{R}^2$ and let $f \in L^2(\Omega)$. We wish to state the weak formulation of the elliptic boundary-value problem

$$\begin{aligned} -\Delta u + u &= f \qquad \text{in } \Omega, \\ u &= 0 \qquad \text{on } \partial\Omega. \end{aligned}$$

Solution. Note that $-\Delta u = -\operatorname{div}(\nabla u)$ and

$$\int_\Omega (-\Delta u)\, v \,\mathrm{d}x = -\int_\Omega \operatorname{div}(\nabla u)\, v \,\mathrm{d}x = \int_\Omega \nabla u \cdot \nabla v \,\mathrm{d}x$$

for all $v \in H_0^1(\Omega)$ by the divergence theorem.

Example

Suppose that $\Omega = (a, b) \times (c, d) \subset \mathbb{R}^2$ and let $f \in L^2(\Omega)$. We wish to state the weak formulation of the elliptic boundary-value problem

$$\begin{aligned} -\Delta u + u &= f \qquad \text{in } \Omega, \\ u &= 0 \qquad \text{on } \partial\Omega. \end{aligned}$$

Solution. Note that $-\Delta u = -\operatorname{div}(\nabla u)$ and

$$\int_\Omega (-\Delta u)\, v \,\mathrm{d}x = -\int_\Omega \operatorname{div}(\nabla u)\, v \,\mathrm{d}x = \int_\Omega \nabla u \cdot \nabla v \,\mathrm{d}x$$

for all $v \in H_0^1(\Omega)$ by the divergence theorem.

Hence, the weak formulation of the boundary-value problem is: find $u \in H_0^1(\Omega)$ such that

$$\int_\Omega \nabla u \cdot \nabla v + u\, v \,\mathrm{d}x = \int_\Omega f\, v \,\mathrm{d}x \qquad \forall v \in H_0^1(\Omega).$$

Introduction to the theory of finite difference schemes

Let Ω be a bounded open set in $\mathbb{R}^n$ and suppose that we wish to solve the boundary-value problem

$$\begin{aligned} \mathcal{L}u &= f \qquad && \text{in } \Omega, \\ \mathcal{B}u &= g \qquad && \text{on } \Gamma := \partial\Omega, \end{aligned} \tag{3}$$

where $\mathcal{L}$ is a linear partial differential operator, and $\mathcal{B}$ is a linear operator which specifies the boundary condition.

Introduction to the theory of finite difference schemes

Let Ω be a bounded open set in $\mathbb{R}^n$ and suppose that we wish to solve the boundary-value problem

$$\begin{aligned} \mathcal{L}u &= f \qquad \text{in } \Omega, \\ \mathcal{B}u &= g \qquad \text{on } \Gamma := \partial\Omega, \end{aligned} \tag{3}$$

where $\mathcal{L}$ is a linear partial differential operator, and $\mathcal{B}$ is a linear operator which specifies the boundary condition. For example,

$$\mathcal{L}u \equiv -\sum_{i,j=1}^{n} \frac{\partial}{\partial x_j}\left(a_{i,j}(x)\frac{\partial u}{\partial x_i}\right) + \sum_{i=1}^{n} b_i \frac{\partial u}{\partial x_i} + cu,$$

and

$$\mathcal{B}u \equiv u \quad \text{(Dirichlet boundary condition)},$$

or

$$\mathcal{B}u \equiv \frac{\partial u}{\partial \nu} \quad \text{(Neumann boundary condition)},$$

or some other boundary condition.

The first step

Suppose that we have 'approximated' $\overline{\Omega} = \Omega \cup \Gamma$ by a finite set of points

$$\overline{\Omega}_h = \Omega_h \cup \Gamma_h,$$

where $\Omega_h \subset \Omega$ and $\Gamma_h \subset \Gamma$.

The first step

Suppose that we have 'approximated' $\overline{\Omega} = \Omega \cup \Gamma$ by a finite set of points

$$\overline{\Omega}_h = \Omega_h \cup \Gamma_h,$$

where $\Omega_h \subset \Omega$ and $\Gamma_h \subset \Gamma$.

- $\overline{\Omega}_h$ is called a mesh;
- Ω_h is the set of interior mesh-points; and
- Γ_h the set boundary mesh-points.

The first step

Suppose that we have 'approximated' $\overline{\Omega} = \Omega \cup \Gamma$ by a finite set of points

$$\overline{\Omega}_h = \Omega_h \cup \Gamma_h,$$

where $\Omega_h \subset \Omega$ and $\Gamma_h \subset \Gamma$.

- $\overline{\Omega}_h$ is called a mesh;
- Ω_h is the set of interior mesh-points; and
- Γ_h the set boundary mesh-points.

The parameter $h = (h_1, \ldots, h_n)$ measures the 'fineness' of the mesh (here h_i denotes the mesh-size in the coordinate direction Ox_i): the smaller $\max_{1 \leq i \leq n} h_i$ is, the finer the mesh.

The second step

Having constructed the mesh, we replace the derivatives in $\mathcal{L}$ by divided differences, and we approximate the boundary condition in a similar fashion. This yields the finite difference scheme

$$\begin{aligned} \mathcal{L}_h U(x) &= f_h(x), \qquad && x \in \Omega_h, \\ \mathcal{B}_h U(x) &= g_h(x), \qquad && x \in \Gamma_h, \end{aligned} \tag{4}$$

where f_h and g_h are suitable approximations of f and g.

Now (4) is a system of linear algebraic equations involving the values of U at the mesh-points, and can be solved by Gaussian elimination or an iterative method, provided that it has a unique solution.

Now (4) is a system of linear algebraic equations involving the values of U at the mesh-points, and can be solved by Gaussian elimination or an iterative method, provided that it has a unique solution.

The sequence

$$\{U(x) : x \in \overline{\Omega}_h\}$$

is an approximation to

$$\{u(x) : x \in \overline{\Omega}_h\},$$

the values of the exact solution at the mesh-points.

There are two fundamental problems to be considered:

There are two fundamental problems to be considered:

- the first, and most basic, is the problem of approximation, that is, whether (4) approximates the boundary-value problem (3) in some sense, and whether its solution $\{U(x) : x \in \overline{\Omega}_h\}$ approximates $\{u(x) : x \in \overline{\Omega}_h\}$, the values of the exact solution at the mesh-points.

There are two fundamental problems to be considered:

- the first, and most basic, is the problem of approximation, that is, whether (4) approximates the boundary-value problem (3) in some sense, and whether its solution $\{U(x) : x \in \overline{\Omega}_h\}$ approximates $\{u(x) : x \in \overline{\Omega}_h\}$, the values of the exact solution at the mesh-points.
- the second concerns the effective solution of the discrete problem (4) using techniques from Numerical Linear Algebra.

There are two fundamental problems to be considered:

- the first, and most basic, is the problem of approximation, that is, whether (4) approximates the boundary-value problem (3) in some sense, and whether its solution $\{U(x) : x \in \overline{\Omega}_h\}$ approximates $\{u(x) : x \in \overline{\Omega}_h\}$, the values of the exact solution at the mesh-points.
- the second concerns the effective solution of the discrete problem (4) using techniques from Numerical Linear Algebra.

Here we shall be primarily concerned with the first of these two problems — the question of approximation — although we shall also briefly consider the question of iterative solution of systems of linear algebraic equations by a simple iterative method.