# Numerical Solution of Partial Differential Equations

*Endre Süli*

Mathematical Institute
University of Oxford
2021

Lecture 3

# Finite difference approximation of a two-point b.v.p.

We illustrate the method of finite difference approximation on a simple two-point boundary-value problem for a second-order linear (ordinary) differential equation:

$$
\begin{aligned}
& -u'' + c(x)u = f(x), \quad x \in (0,1), \\
& u(0) = 0, \;\; u(1) = 0,
\end{aligned}
\tag{1}
$$

where $f$ and $c$ are real-valued functions, which are defined and continuous on the interval $[0,1]$ and $c(x) \geq 0$ for all $x \in [0,1]$.

# The first step

The first step in the construction of a finite difference scheme for this boundary-value problem is to define the mesh.

# The first step

The first step in the construction of a finite difference scheme for this boundary-value problem is to define the mesh.

Let $N$ be an integer, $N \geq 2$, and let $h = 1/N$ be the mesh-size; the mesh-points are $x_i = ih$, $i = 0, \ldots, N$.

# The first step

The first step in the construction of a finite difference scheme for this boundary-value problem is to define the mesh.

Let $N$ be an integer, $N \geq 2$, and let $h = 1/N$ be the mesh-size; the mesh-points are $x_i = ih$, $i = 0, \ldots, N$.

We define the set of interior mesh-points:

$$\Omega_h := \{x_i : i = 1, \ldots, N-1\}$$

the set of boundary mesh-points:

$$\Gamma_h := \{x_0, x_N\},$$

and the set of all mesh-points:

$$\overline{\Omega}_h := \Omega_h \cup \Gamma_h.$$

## The second step

Suppose that $u$ is sufficiently smooth (e.g. $u \in C^4([0,1])$).

## The second step

Suppose that $u$ is sufficiently smooth (e.g. $u \in C^4([0,1])$). Then, by Taylor series expansion,

$$\begin{aligned} u(x_{i\pm 1}) &= u(x_i \pm h) \\ &= u(x_i) \pm h u'(x_i) + \frac{h^2}{2} u''(x_i) \pm \frac{h^3}{6} u'''(x_i) + \mathcal{O}(h^4), \end{aligned}$$

so that

$$D_x^+ u(x_i) := \frac{u(x_{i+1}) - u(x_i)}{h} = u'(x_i) + \mathcal{O}(h),$$

$$D_x^- u(x_i) := \frac{u(x_i) - u(x_{i-1})}{h} = u'(x_i) + \mathcal{O}(h),$$

and

$$\begin{aligned} D_x^+ D_x^- u(x_i) &= D_x^- D_x^+ u(x_i) \\ &= \frac{u(x_{i+1}) - 2u(x_i) + u(x_{i-1})}{h^2} \\ &= u''(x_i) + \mathcal{O}(h^2). \end{aligned}$$

$D_x^+$ and $D_x^-$ are called the *forward* and *backward first divided difference* operator, respectively, and $D_x^+ D_x^-$ $(= D_x^- D_x^+)$ is called the (symmetric) *second divided difference* operator.

$D_x^+$ and $D_x^-$ are called the *forward* and *backward first divided difference* operator, respectively, and $D_x^+ D_x^-$ ($= D_x^- D_x^+$) is called the (symmetric) *second divided difference* operator.

Thus we replace the second derivative $u''$ in the differential equation by the second divided difference $D_x^+ D_x^- u(x_i)$; hence,

$$\begin{aligned} -D_x^+ D_x^- u(x_i) + c(x_i)u(x_i) &\approx f(x_i), \quad i = 1, \ldots, N-1, \\ u(x_0) = 0, \ \ u(x_N) &= 0. \end{aligned} \tag{2}$$

$D_x^+$ and $D_x^-$ are called the *forward* and *backward first divided difference* operator, respectively, and $D_x^+ D_x^-$ ($= D_x^- D_x^+$) is called the (symmetric) *second divided difference* operator.

Thus we replace the second derivative $u''$ in the differential equation by the second divided difference $D_x^+ D_x^- u(x_i)$; hence,

$$\begin{aligned} -D_x^+ D_x^- u(x_i) + c(x_i)u(x_i) &\approx f(x_i), \quad i = 1, \dots, N-1, \\ u(x_0) = 0, \ \ u(x_N) &= 0. \end{aligned} \tag{2}$$

Now (2) motivates us to seek the approximate solution $U$ as the solution of the system of difference equations:

$$\begin{aligned} -D_x^+ D_x^- U_i + c(x_i)U_i &= f(x_i), \quad i = 1, \dots, N-1, \\ U_0 = 0, \ \ U_N &= 0. \end{aligned} \tag{3}$$

This is a system of $N-1$ linear algebraic equations for the $N-1$ unknowns, $U_i$, $i = 1, \ldots, N-1$.

This is a system of $N-1$ linear algebraic equations for the $N-1$ unknowns, $U_i$, $i = 1, \dots, N-1$. Using matrix notation,

$$AU = F,$$

where $A$ is the $(N-1) \times (N-1)$ matrix

$$A = \begin{bmatrix} \frac{2}{h^2} + c(x_1) & -\frac{1}{h^2} & & & \mathbf{0} \\ -\frac{1}{h^2} & \frac{2}{h^2} + c(x_2) & -\frac{1}{h^2} & & \\ & \ddots & \ddots & \ddots & \\ & & -\frac{1}{h^2} & \frac{2}{h^2} + c(x_{N-2}) & -\frac{1}{h^2} \\ \mathbf{0} & & & -\frac{1}{h^2} & \frac{2}{h^2} + c(x_{N-1}) \end{bmatrix}$$

$$U = (U_1, U_2, \dots, U_{N-2}, U_{N-1})^{\mathrm{T}}$$

and

$$F = (f(x_1), f(x_2), \dots, f(x_{N-2}), f(x_{N-1}))^{\mathrm{T}}.$$

# Existence and uniqueness of a solution

We begin the analysis of the finite difference scheme (3) by showing that it has a unique solution. It suffices to show that the matrix $A$ is non-singular (i.e. $\det A \neq 0$), and therefore invertible.

# Existence and uniqueness of a solution

We begin the analysis of the finite difference scheme (3) by showing that it has a unique solution. It suffices to show that the matrix $A$ is non-singular (i.e. $\det A \neq 0$), and therefore invertible.

We shall develop a technique which we shall, in subsequent sections, extend to the finite difference approximation of PDEs.

# Existence and uniqueness of a solution

We begin the analysis of the finite difference scheme (3) by showing that it has a unique solution. It suffices to show that the matrix $A$ is non-singular (i.e. $\det A \neq 0$), and therefore invertible.

We shall develop a technique which we shall, in subsequent sections, extend to the finite difference approximation of PDEs.

For this purpose, we introduce, for two functions $V$ and $W$ defined at the interior mesh-points $x_i$, $i = 1, \ldots, N-1$, the inner product

$$(V, W)_h = \sum_{i=1}^{N-1} h V_i W_i,$$

which resembles the $L_2((0,1))$-inner product

$$(v, w) = \int_0^1 v(x) w(x) \, \mathrm{d}x.$$

The argument is based on mimicking, at the discrete level, the following procedure based on integration-by-parts, noting that the solution of the boundary-value problem (1) satisfies the homogeneous boundary conditions $u(0) = 0$ and $u(1) = 0$:

$$\begin{aligned} \int_0^1 \left(-u''(x) + c(x)u(x)\right) u(x) \,\mathrm{d}x &= \int_0^1 |u'(x)|^2 + c(x)|u(x)|^2 \,\mathrm{d}x \\ &\geq \int_0^1 |u'(x)|^2 \,\mathrm{d}x, \end{aligned} \tag{4}$$

because $c(x) \geq 0$ for all $x \in [0, 1]$.

The argument is based on mimicking, at the discrete level, the following procedure based on integration-by-parts, noting that the solution of the boundary-value problem (1) satisfies the homogeneous boundary conditions $u(0) = 0$ and $u(1) = 0$:

$$\begin{aligned}\int_0^1 (-u''(x) + c(x)u(x))\, u(x)\,\mathrm{d}x &= \int_0^1 |u'(x)|^2 + c(x)|u(x)|^2\,\mathrm{d}x \\ &\geq \int_0^1 |u'(x)|^2\,\mathrm{d}x,\end{aligned} \tag{4}$$

because $c(x) \geq 0$ for all $x \in [0, 1]$. Thus if, for example, $f \equiv 0$ on $[0, 1]$, then $-u'' + c(x)u \equiv 0$ on $[0, 1]$, and therefore by (4) also $u' \equiv 0$ on $[0, 1]$.

The argument is based on mimicking, at the discrete level, the following procedure based on integration-by-parts, noting that the solution of the boundary-value problem (1) satisfies the homogeneous boundary conditions $u(0) = 0$ and $u(1) = 0$:

$$\begin{aligned} \int_0^1 (-u''(x) + c(x)u(x))\, u(x)\, \mathrm{d}x &= \int_0^1 |u'(x)|^2 + c(x)|u(x)|^2\, \mathrm{d}x \\ &\geq \int_0^1 |u'(x)|^2\, \mathrm{d}x, \end{aligned} \tag{4}$$

because $c(x) \geq 0$ for all $x \in [0, 1]$. Thus if, for example, $f \equiv 0$ on $[0, 1]$, then $-u'' + c(x)u \equiv 0$ on $[0, 1]$, and therefore by (4) also $u' \equiv 0$ on $[0, 1]$. Consequently, $u$ is a constant function on $[0, 1]$, but because $u(0) = 0$ and $u(1) = 0$, necessarily $u \equiv 0$ on $[0, 1]$.

The argument is based on mimicking, at the discrete level, the following procedure based on integration-by-parts, noting that the solution of the boundary-value problem (1) satisfies the homogeneous boundary conditions $u(0) = 0$ and $u(1) = 0$:

$$\begin{aligned}\int_0^1 (-u''(x) + c(x)u(x))\, u(x)\, \mathrm{d}x &= \int_0^1 |u'(x)|^2 + c(x)|u(x)|^2\, \mathrm{d}x \\ &\geq \int_0^1 |u'(x)|^2\, \mathrm{d}x,\end{aligned} \tag{4}$$

because $c(x) \geq 0$ for all $x \in [0,1]$. Thus if, for example, $f \equiv 0$ on $[0,1]$, then $-u'' + c(x)u \equiv 0$ on $[0,1]$, and therefore by (4) also $u' \equiv 0$ on $[0,1]$. Consequently, $u$ is a constant function on $[0,1]$, but because $u(0) = 0$ and $u(1) = 0$, necessarily $u \equiv 0$ on $[0,1]$. Hence, the only solution to the homogeneous boundary-value problem is the function $u(x) \equiv 0$, $x \in [0,1]$.

For the finite difference approximation of the boundary-value problem, if we can show by an analogous argument that the homogeneous system of linear algebraic equations corresponding to $f(x_i) = 0$, $i = 1, \ldots, N-1$, has the trivial solution $U_i = 0$, $i = 0, \ldots, N$, as its unique solution, then the desired invertibility of the matrix $A$ will directly follow.

For the finite difference approximation of the boundary-value problem, if we can show by an analogous argument that the homogeneous system of linear algebraic equations corresponding to $f(x_i) = 0$, $i = 1, \dots, N-1$, has the trivial solution $U_i = 0$, $i = 0, \dots, N$, as its unique solution, then the desired invertibility of the matrix $A$ will directly follow.

Our key tool is a summation-by-parts identity, which is the discrete counterpart of the integration-by-parts identity

$$(-u'', u) = (u', u') = \|u'\|^2_{L_2((0,1))} = \int_0^1 |u'(x)|^2 \, \mathrm{d}x$$

satisfied by the function $u$, obeying the homogeneous boundary conditions $u(0) = 0$, $u(1) = 0$, used in (4) above.

# Summation by parts identity

## Lemma

*Suppose that $V$ is a function defined at the mesh-points $x_i$, $i = 0, \ldots, N$, and let $V_0 = V_N = 0$; then,*

$$(-D_x^+ D_x^- V, V)_h = \sum_{i=1}^{N} h \left| D_x^- V_i \right|^2 . \tag{5}$$

## Proof.

By the definitions of $(\cdot,\cdot)_h$ and $D_x^+ D_x^- V_i$ we have that

$$
\begin{aligned}
(-D_x^+ D_x^- V, V)_h &= -\sum_{i=1}^{N-1} h\,(D_x^+ D_x^- V_i)V_i \\
&= -\sum_{i=1}^{N-1} \frac{V_{i+1}-V_i}{h} V_i + \sum_{i=1}^{N-1} \frac{V_i - V_{i-1}}{h} V_i \\
&= -\sum_{i=2}^{N} \frac{V_i - V_{i-1}}{h} V_{i-1} + \sum_{i=1}^{N-1} \frac{V_i - V_{i-1}}{h} V_i \\
&= -\sum_{i=1}^{N} \frac{V_i - V_{i-1}}{h} V_{i-1} + \sum_{i=1}^{N} \frac{V_i - V_{i-1}}{h} V_i \\
&= \sum_{i=1}^{N} \frac{V_i - V_{i-1}}{h} (V_i - V_{i-1}) = \sum_{i=1}^{N} h \left| D_x^- V_i \right|^2 .
\end{aligned}
$$

In the transition to the 3rd line we shifted the index in the first sum; in the transition to the 4th line used that $V_0 = V_N = 0$. $\square$

Returning to the finite difference scheme (3), let $V$ be as in the above lemma and note that as, by hypothesis, $c(x) \geq 0$ for all $x \in [0, 1]$, we have

$$\begin{aligned}(AV, V)_h &= (-D_x^+ D_x^- V + cV, V)_h \\ &= (-D_x^+ D_x^- V, V)_h + (cV, V)_h \\ &\geq \sum_{i=1}^{N} h \left| D_x^- V_i \right|^2 .\end{aligned} \tag{6}$$

Returning to the finite difference scheme (3), let $V$ be as in the above lemma and note that as, by hypothesis, $c(x) \geq 0$ for all $x \in [0,1]$, we have

$$\begin{aligned}(AV, V)_h &= (-D_x^+ D_x^- V + cV, V)_h \\ &= (-D_x^+ D_x^- V, V)_h + (cV, V)_h \\ &\geq \sum_{i=1}^{N} h \left| D_x^- V_i \right|^2 .\end{aligned} \tag{6}$$

Thus, if $AV = 0$ for some $V$, then $D_x^- V_i = 0$, $i = 1, \ldots, N$.

Returning to the finite difference scheme (3), let $V$ be as in the above lemma and note that as, by hypothesis, $c(x) \geq 0$ for all $x \in [0, 1]$, we have

$$\begin{aligned}(AV, V)_h &= (-D_x^+ D_x^- V + cV, V)_h \\ &= (-D_x^+ D_x^- V, V)_h + (cV, V)_h \\ &\geq \sum_{i=1}^{N} h \left| D_x^- V_i \right|^2 .\end{aligned} \tag{6}$$

Thus, if $AV = 0$ for some $V$, then $D_x^- V_i = 0$, $i = 1, \ldots, N$. Because $V_0 = V_N = 0$, this implies that $V_i = 0$, $i = 0, \ldots, N$.

Returning to the finite difference scheme (3), let $V$ be as in the above lemma and note that as, by hypothesis, $c(x) \geq 0$ for all $x \in [0, 1]$, we have

$$\begin{aligned}
(AV, V)_h &= (-D_x^+ D_x^- V + cV, V)_h \\
&= (-D_x^+ D_x^- V, V)_h + (cV, V)_h \\
&\geq \sum_{i=1}^{N} h \left| D_x^- V_i \right|^2 .
\end{aligned} \tag{6}$$

Thus, if $AV = 0$ for some $V$, then $D_x^- V_i = 0$, $i = 1, \ldots, N$. Because $V_0 = V_N = 0$, this implies that $V_i = 0$, $i = 0, \ldots, N$. Hence $AV = 0$ if and only if $V = 0$.

Returning to the finite difference scheme (3), let $V$ be as in the above lemma and note that as, by hypothesis, $c(x) \geq 0$ for all $x \in [0,1]$, we have

$$\begin{aligned}(AV, V)_h &= (-D_x^+ D_x^- V + cV, V)_h \\ &= (-D_x^+ D_x^- V, V)_h + (cV, V)_h \\ &\geq \sum_{i=1}^{N} h \left| D_x^- V_i \right|^2 .\end{aligned} \tag{6}$$

Thus, if $AV = 0$ for some $V$, then $D_x^- V_i = 0$, $i = 1, \ldots, N$. Because $V_0 = V_N = 0$, this implies that $V_i = 0$, $i = 0, \ldots, N$. Hence $AV = 0$ if and only if $V = 0$.

It therefore follows that $A$ is a non-singular matrix, and thereby (3) has a unique solution, $U = A^{-1}F$.

We record this result in the next theorem.

**Theorem**

*Suppose that $c$ and $f$ are continuous real-valued functions defined on the interval $[0, 1]$, and $c(x) \geq 0$ for all $x \in [0, 1]$; then, the finite difference scheme (3) possesses a unique solution $U$.*

We record this result in the next theorem.

**Theorem**

*Suppose that $c$ and $f$ are continuous real-valued functions defined on the interval $[0,1]$, and $c(x) \geq 0$ for all $x \in [0,1]$; then, the finite difference scheme (3) possesses a unique solution $U$.*

We note in passing that, thanks the Lax–Milgram theorem (cf. the Lecture Notes), the boundary-value problem (1) has a unique (weak) solution under the hypotheses on $c$ and $f$ assumed in the above theorem.

# Stability, consistency, and convergence

Next, we investigate the approximation properties of the finite difference scheme (3).

# Stability, consistency, and convergence

Next, we investigate the approximation properties of the finite difference scheme (3). A key ingredient in our analysis is that the scheme (3) is stable (or discretely well-posed) in the sense that "small" perturbations in the data result in "small" perturbations in the corresponding finite difference solution.

# Stability, consistency, and convergence

Next, we investigate the approximation properties of the finite difference scheme (3). A key ingredient in our analysis is that the scheme (3) is stable (or discretely well-posed) in the sense that "small" perturbations in the data result in "small" perturbations in the corresponding finite difference solution.

To prove this, we define the *discrete $L_2$-norm*

$$\|U\|_h := (U, U)_h^{1/2} = \left( \sum_{i=1}^{N-1} h|U_i|^2 \right)^{1/2},$$

and the *discrete Sobolev norm*

$$\|U\|_{1,h} := (\|U\|_h^2 + \|D_x^- U]|_h^2)^{1/2},$$

where

$$\|V]|_h^2 := \sum_{i=1}^{N} h|V_i|^2 .$$

Using this notation, the inequality (6) can be rewritten as follows:

$$(AV, V)_h \geq \|D_x^- V]|_h^2. \tag{7}$$

Using this notation, the inequality (6) can be rewritten as follows:

$$(AV, V)_h \geq \|D_x^- V]|_h^2. \tag{7}$$

In fact, by employing a discrete version of the Poincaré–Friedrichs inequality, stated in the next lemma, we shall be able to prove that

$$(AV, V)_h \geq c_0 \|V\|_{1,h}^2,$$

where $c_0$ is a positive constant, independent of $h$.

Using this notation, the inequality (6) can be rewritten as follows:

$$(AV, V)_h \geq \|D_x^- V]|_h^2. \qquad (7)$$

In fact, by employing a discrete version of the Poincaré–Friedrichs inequality, stated in the next lemma, we shall be able to prove that

$$(AV, V)_h \geq c_0 \|V\|_{1,h}^2,$$

where $c_0$ is a positive constant, independent of $h$.

**Lemma (Discrete Poincaré–Friedrichs inequality)**

*Let $V$ be a function defined on the mesh $\{x_i, i = 0, \ldots, N\}$, and such that $V_0 = V_N = 0$; then, there exists a positive constant $c_\star$, independent of $V$ and $h$, such that*

$$\|V\|_h^2 \leq c_\star \|D_x^- V]|_h^2 \qquad (8)$$

*for all such $V$.*

PROOF. Thanks to the definition of $D_x^- V_i$ and by use of the Cauchy–Schwarz inequality,

$$|V_i|^2 = \left|\sum_{j=1}^{i} h\,(D_x^- V_j)\right|^2 \leq \left(\sum_{j=1}^{i} h\right) \sum_{j=1}^{i} h\left|D_x^- V_j\right|^2 = ih \sum_{j=1}^{i} h\left|D_x^- V_j\right|^2.$$

PROOF. Thanks to the definition of $D_x^- V_i$ and by use of the Cauchy–Schwarz inequality,

$$|V_i|^2 = \left|\sum_{j=1}^{i} h\,(D_x^- V_j)\right|^2 \leq \left(\sum_{j=1}^{i} h\right) \sum_{j=1}^{i} h\left|D_x^- V_j\right|^2 = ih \sum_{j=1}^{i} h\left|D_x^- V_j\right|^2.$$

Thus, because $\sum_{i=1}^{N-1} i = \frac{1}{2}(N-1)N$ and $Nh = 1$, we have that

$$\begin{aligned}
\|V\|_h^2 = \sum_{i=1}^{N-1} h|V_i|^2 &\leq \sum_{i=1}^{N-1} ih^2 \sum_{j=1}^{i} h\left|D_x^- V_j\right|^2 \\
&\leq \frac{1}{2}(N-1)Nh^2 \sum_{j=1}^{N} h\left|D_x^- V_j\right|^2 \\
&\leq \frac{1}{2}\|D_x^- V]|_h^2.
\end{aligned}$$

PROOF. Thanks to the definition of $D_x^- V_i$ and by use of the Cauchy–Schwarz inequality,

$$|V_i|^2 = \left| \sum_{j=1}^{i} h \left( D_x^- V_j \right) \right|^2 \le \left( \sum_{j=1}^{i} h \right) \sum_{j=1}^{i} h \left| D_x^- V_j \right|^2 = ih \sum_{j=1}^{i} h \left| D_x^- V_j \right|^2 .$$

Thus, because $\sum_{i=1}^{N-1} i = \frac{1}{2}(N-1)N$ and $Nh = 1$, we have that

$$\begin{aligned} \|V\|_h^2 = \sum_{i=1}^{N-1} h|V_i|^2 &\le \sum_{i=1}^{N-1} ih^2 \sum_{j=1}^{i} h \left| D_x^- V_j \right|^2 \\ &\le \frac{1}{2}(N-1)Nh^2 \sum_{j=1}^{N} h \left| D_x^- V_j \right|^2 \\ &\le \frac{1}{2} \| D_x^- V]|_h^2. \end{aligned}$$

We note that the constant $c_\star = 1/2$ in the inequality (8). □

Using the inequality (8) to bound the right-hand side of the inequality (7) from below we obtain

$$(AV, V)_h \geq \frac{1}{c_\star} \|V\|_h^2. \tag{9}$$

Using the inequality (8) to bound the right-hand side of the inequality (7) from below we obtain

$$(AV, V)_h \geq \frac{1}{c_\star}\|V\|_h^2. \tag{9}$$

Adding the inequality (7) to the inequality (9) we arrive at the inequality

$$(AV, V)_h \geq (1 + c_\star)^{-1}\left(\|V\|_h^2 + \|D_x^- V]|_h^2\right).$$

Using the inequality (8) to bound the right-hand side of the inequality (7) from below we obtain

$$(AV, V)_h \geq \frac{1}{c_\star} \|V\|_h^2. \tag{9}$$

Adding the inequality (7) to the inequality (9) we arrive at the inequality

$$(AV, V)_h \geq (1 + c_\star)^{-1} \left( \|V\|_h^2 + \|D_x^- V]|_h^2 \right).$$

Letting $c_0 = (1 + c_\star)^{-1}$ it follows that

$$(AV, V)_h \geq c_0 \|V\|_{1,h}^2. \tag{10}$$

Now the stability of the finite difference scheme (3) easily follows.

**Theorem**

*The scheme (3) is stable in the sense that*

$$\|U\|_{1,h} \leq \frac{1}{c_0}\|f\|_h. \tag{11}$$

Now the stability of the finite difference scheme (3) easily follows.

**Theorem**

*The scheme (3) is stable in the sense that*

$$\|U\|_{1,h} \leq \frac{1}{c_0}\|f\|_h. \tag{11}$$

PROOF. From (10) and (3) we have that

$$\begin{aligned} c_0\|U\|_{1,h}^2 &\leq (AU, U)_h = (f, U)_h \leq \left|(f, U)_h\right| \\ &\leq \|f\|_h\|U\|_h \leq \|f\|_h\|U\|_{1,h}, \end{aligned}$$

and hence (11). $\square$

Using this stability result it is easy to derive an estimate of the error between the exact solution $u$, and its finite difference approximation, $U$.

Using this stability result it is easy to derive an estimate of the error between the exact solution $u$, and its finite difference approximation, $U$. We define the *global error*, $e$, by

$$e_i := u(x_i) - U_i, \quad i = 0, \ldots, N.$$

Using this stability result it is easy to derive an estimate of the error between the exact solution $u$, and its finite difference approximation, $U$. We define the *global error*, $e$, by

$$e_i := u(x_i) - U_i, \quad i = 0, \dots, N.$$

Obviously $e_0 = 0$, $e_N = 0$, and

$$\begin{aligned} Ae_i &= Au(x_i) - AU_i = Au(x_i) - f(x_i) \\ &= -D_x^+ D_x^- u(x_i) + c(x_i)u(x_i) - f(x_i) \\ &= u''(x_i) - D_x^+ D_x^- u(x_i), \qquad i = 1, \dots, N-1. \end{aligned}$$

Using this stability result it is easy to derive an estimate of the error between the exact solution $u$, and its finite difference approximation, $U$. We define the *global error*, $e$, by

$$e_i := u(x_i) - U_i, \quad i = 0, \ldots, N.$$

Obviously $e_0 = 0$, $e_N = 0$, and

$$\begin{aligned} Ae_i &= Au(x_i) - AU_i = Au(x_i) - f(x_i) \\ &= -D_x^+ D_x^- u(x_i) + c(x_i)u(x_i) - f(x_i) \\ &= u''(x_i) - D_x^+ D_x^- u(x_i), \qquad i = 1, \ldots, N-1. \end{aligned}$$

Thus,

$$\begin{aligned} Ae_i = \varphi_i, \qquad & i = 1, \ldots, N-1, \\ & e_0 = 0, \quad e_N = 0, \end{aligned} \tag{12}$$

where $\varphi_i := u''(x_i) - D_x^+ D_x^- u(x_i)$ is the *consistency error* (sometimes also called the *truncation error*).

By applying ineq. (11) to the finite difference scheme (12):

$$\|u - U\|_{1,h} = \|e\|_{1,h} \leq \frac{1}{c_0}\|\varphi\|_h. \tag{13}$$

By applying ineq. (11) to the finite difference scheme (12):

$$\|u - U\|_{1,h} = \|e\|_{1,h} \le \frac{1}{c_0}\|\varphi\|_h. \tag{13}$$

It remains to bound $\|\varphi\|_h$. We showed that, if $u \in C^4([0,1])$, then

$$\varphi_i = u''(x_i) - D_x^+ D_x^- u(x_i) = \mathcal{O}(h^2),$$

By applying ineq. (11) to the finite difference scheme (12):

$$\|u - U\|_{1,h} = \|e\|_{1,h} \leq \frac{1}{c_0}\|\varphi\|_h. \tag{13}$$

It remains to bound $\|\varphi\|_h$. We showed that, if $u \in C^4([0,1])$, then

$$\varphi_i = u''(x_i) - D_x^+ D_x^- u(x_i) = \mathcal{O}(h^2),$$

i.e. there exists a positive constant $C$, independent of $h$, such that

$$|\varphi_i| \leq Ch^2.$$

By applying ineq. (11) to the finite difference scheme (12):

$$\|u - U\|_{1,h} = \|e\|_{1,h} \leq \frac{1}{c_0}\|\varphi\|_h. \tag{13}$$

It remains to bound $\|\varphi\|_h$. We showed that, if $u \in C^4([0,1])$, then

$$\varphi_i = u''(x_i) - D_x^+ D_x^- u(x_i) = \mathcal{O}(h^2),$$

i.e. there exists a positive constant $C$, independent of $h$, such that

$$|\varphi_i| \leq Ch^2.$$

Consequently,

$$\|\varphi\|_h = \left( \sum_{i=1}^{N-1} h|\varphi_i|^2 \right)^{1/2} \leq Ch^2. \tag{14}$$

By applying ineq. (11) to the finite difference scheme (12):

$$\|u - U\|_{1,h} = \|e\|_{1,h} \leq \frac{1}{c_0}\|\varphi\|_h. \tag{13}$$

It remains to bound $\|\varphi\|_h$. We showed that, if $u \in C^4([0,1])$, then

$$\varphi_i = u''(x_i) - D_x^+ D_x^- u(x_i) = \mathcal{O}(h^2),$$

i.e. there exists a positive constant $C$, independent of $h$, such that

$$|\varphi_i| \leq Ch^2.$$

Consequently,

$$\|\varphi\|_h = \left(\sum_{i=1}^{N-1} h|\varphi_i|^2\right)^{1/2} \leq Ch^2. \tag{14}$$

Combining the inequalities (13) and (14), it follows that

$$\|u - U\|_{1,h} \leq \frac{C}{c_0} h^2. \tag{15}$$

In fact, a more careful treatment of the remainder term in the Taylor series expansion on p.4 reveals that

$$\varphi_i = u''(x_i) - D_x^+ D_x^- u(x_i) = -\frac{h^2}{12} u^{IV}(\xi_i), \quad \xi_i \in [x_{i-1}, x_{i+1}].$$

In fact, a more careful treatment of the remainder term in the Taylor series expansion on p.4 reveals that

$$\varphi_i = u''(x_i) - D_x^+ D_x^- u(x_i) = -\frac{h^2}{12} u^{IV}(\xi_i), \quad \xi_i \in [x_{i-1}, x_{i+1}].$$

Thus

$$|\varphi_i| \le h^2 \frac{1}{12} \max_{x \in [0,1]} \left| u^{IV}(x) \right|,$$

In fact, a more careful treatment of the remainder term in the Taylor series expansion on p.4 reveals that

$$\varphi_i = u''(x_i) - D_x^+ D_x^- u(x_i) = -\frac{h^2}{12} u^{IV}(\xi_i), \quad \xi_i \in [x_{i-1}, x_{i+1}].$$

Thus

$$|\varphi_i| \leq h^2 \frac{1}{12} \max_{x \in [0,1]} \left| u^{IV}(x) \right|,$$

and hence

$$C = \frac{1}{12} \max_{x \in [0,1]} \left| u^{IV}(x) \right|$$

in inequality (14).

In fact, a more careful treatment of the remainder term in the Taylor series expansion on p.4 reveals that

$$\varphi_i = u''(x_i) - D_x^+ D_x^- u(x_i) = -\frac{h^2}{12} u^{IV}(\xi_i), \quad \xi_i \in [x_{i-1}, x_{i+1}].$$

Thus

$$|\varphi_i| \leq h^2 \frac{1}{12} \max_{x \in [0,1]} \left| u^{IV}(x) \right|,$$

and hence

$$C = \frac{1}{12} \max_{x \in [0,1]} \left| u^{IV}(x) \right|$$

in inequality (14). Recalling that $c_0 = (1 + c_\star)^{-1}$ and $c_\star = 1/2$, we deduce that $c_0 = 2/3$. Substituting the values of the constants $C$ and $c_0$ into inequality (15) it follows that

$$\|u - U\|_{1,h} \leq \frac{1}{8} h^2 \|u^{IV}\|_{C([0,1])}.$$

Thus we have proved the following result.

**Theorem**

*Let $f \in C([0,1])$, $c \in C([0,1])$, with $c(x) \geq 0$ for all $x \in [0,1]$, and suppose that the corresponding (weak) solution of the boundary-value problem (1) belongs to $C^4([0,1])$; then*

$$\|u - U\|_{1,h} \leq \frac{1}{8} h^2 \|u^{IV}\|_{C([0,1])}. \tag{16}$$

# Some general observations

The analysis of the finite difference scheme (3) contains the key steps of a general error analysis for finite difference approximations of (elliptic) partial differential equations:

# Some general observations

The analysis of the finite difference scheme (3) contains the key steps of a general error analysis for finite difference approximations of (elliptic) partial differential equations:

Consider the finite difference scheme:

$$\begin{aligned} \mathcal{L}_h u &= f_h, \qquad && \text{in } \Omega_h, \\ \mathcal{B}_h u &= g_h, \qquad && \text{on } \Gamma_h. \end{aligned}$$

(1) The first step is to prove the stability of the scheme in an appropriate mesh-dependent norm. A typical stability result for a finite difference scheme is

$$|||U|||_{\Omega_h} \leq C_1 \left(\|f_h\|_{\Omega_h} + \|g_h\|_{\Gamma_h}\right), \tag{17}$$

where $|||\cdot|||_{\Omega_h}$, $\|\cdot\|_{\Omega_h}$ and $\|\cdot\|_{\Gamma_h}$ are mesh-dependent norms involving mesh-points of $\Omega_h$ (or $\overline{\Omega}_h$) and $\Gamma_h$, respectively, and $C_1$ is a positive constant, independent of $h$.

(2) The second step is to estimate the size of the *consistency error*,

$$\begin{aligned} \varphi_{\Omega_h} &:= \mathcal{L}_h u - f_h, && \text{in } \Omega_h, \\ \varphi_{\Gamma_h} &:= \mathcal{B}_h u - g_h, && \text{on } \Gamma_h. \end{aligned}$$

(in the case of the finite difference scheme (1) $\varphi_{\Gamma_h} = 0$, and therefore $\varphi_{\Gamma_h}$ never appeared explicitly in our error analysis).

(2) The second step is to estimate the size of the *consistency error*,

$$\varphi_{\Omega_h} := \mathcal{L}_h u - f_h, \qquad \text{in } \Omega_h,$$
$$\varphi_{\Gamma_h} := \mathcal{B}_h u - g_h, \qquad \text{on } \Gamma_h.$$

(in the case of the finite difference scheme (1) $\varphi_{\Gamma_h} = 0$, and therefore $\varphi_{\Gamma_h}$ never appeared explicitly in our error analysis). If

$$\|\varphi_{\Omega_h}\|_{\Omega_h} + \|\varphi_{\Gamma_h}\|_{\Gamma_h} \to 0 \quad \text{as} \quad h \to 0,$$

for a sufficiently smooth solution $u$ of the boundary-value problem, we say that the scheme is *consistent.*

(2) The second step is to estimate the size of the *consistency error*,

$$\varphi_{\Omega_h} := \mathcal{L}_h u - f_h, \qquad \text{in } \Omega_h,$$
$$\varphi_{\Gamma_h} := \mathcal{B}_h u - g_h, \qquad \text{on } \Gamma_h.$$

(in the case of the finite difference scheme (1) $\varphi_{\Gamma_h} = 0$, and therefore $\varphi_{\Gamma_h}$ never appeared explicitly in our error analysis). If

$$\|\varphi_{\Omega_h}\|_{\Omega_h} + \|\varphi_{\Gamma_h}\|_{\Gamma_h} \to 0 \quad \text{as} \quad h \to 0,$$

for a sufficiently smooth solution $u$ of the boundary-value problem, we say that the scheme is *consistent.* If $p$ is the largest positive integer such that

$$\|\varphi_{\Omega_h}\|_{\Omega_h} + \|\varphi_{\Gamma_h}\|_{\Gamma_h} \leq C_2 h^p \quad \text{as} \quad h \to 0,$$

(where $C_2$ is a positive constant independent of $h$) for all sufficiently smooth $u$, the scheme is said to have *order of accuracy* (or *order of consistency*) $p$.

The finite difference scheme is said to provide a *convergent* approximation to the solution $u$ of the boundary-value problem in the norm $|||\cdot|||_{\Omega_h}$, if

$$|||u - U|||_{\Omega_h} \to 0 \quad \text{as } h \to 0.$$

The finite difference scheme is said to provide a *convergent* approximation to the solution $u$ of the boundary-value problem in the norm $|||\cdot|||_{\Omega_h}$, if

$$|||u - U|||_{\Omega_h} \to 0 \quad \text{as } h \to 0.$$

If $q$ is the largest positive integer such that

$$|||u - U|||_{\Omega_h} \leq Ch^q \quad \text{as } h \to 0$$

(where $C$ is a positive constant independent of the mesh-size $h$), then the scheme is said to have *order of convergence* $q$.

We deduce the following fundamental theorem.

### Theorem

*Suppose that the finite difference scheme is stable (i.e. the inequality (17) holds for all $f_h$ and $g_h$ and the corresponding numerical solution $U$) and that the scheme is a consistent approximation of the boundary-value problem; then the finite difference scheme is a convergent approximation of the boundary-value problem, and the order of convergence $q$ is not smaller then the order of accuracy (order of consistency) $p$.*

PROOF. We define the *global error* $e := u - U$. Then,

$$\mathcal{L}_h e = \mathcal{L}_h(u - U) = \mathcal{L}_h u - \mathcal{L}_h U = \mathcal{L}_h u - f_h.$$

Thus

$$\mathcal{L}_h e = \varphi_{\Omega_h},$$

and similarly,

$$\mathcal{B}_h e = \varphi_{\Gamma_h}.$$

Proof. We define the *global error* $e := u - U$. Then,

$$\mathcal{L}_h e = \mathcal{L}_h(u - U) = \mathcal{L}_h u - \mathcal{L}_h U = \mathcal{L}_h u - f_h.$$

Thus

$$\mathcal{L}_h e = \varphi_{\Omega_h},$$

and similarly,

$$\mathcal{B}_h e = \varphi_{\Gamma_h}.$$

By stability of the scheme it then follows that

$$|||u - U|||_{\Omega_h} = |||e|||_{\Omega_h} \leq C_1(\|\varphi_{\Omega_h}\|_{\Omega_h} + \|\varphi_{\Gamma_h}\|_{\Gamma_h}),$$

and hence the stated result with $q \geq p$, thanks to the assumed consistency of order $p$ of the scheme. $\square$

PROOF. We define the *global error* $e := u - U$. Then,

$$\mathcal{L}_h e = \mathcal{L}_h(u - U) = \mathcal{L}_h u - \mathcal{L}_h U = \mathcal{L}_h u - f_h.$$

Thus

$$\mathcal{L}_h e = \varphi_{\Omega_h},$$

and similarly,

$$\mathcal{B}_h e = \varphi_{\Gamma_h}.$$

By stability of the scheme it then follows that

$$|||u - U|||_{\Omega_h} = |||e|||_{\Omega_h} \leq C_1(\|\varphi_{\Omega_h}\|_{\Omega_h} + \|\varphi_{\Gamma_h}\|_{\Gamma_h}),$$

and hence the stated result with $q \geq p$, thanks to the assumed consistency of order $p$ of the scheme. $\square$

In other words,

$$\textit{stability} + \textit{consistency} \Rightarrow \textit{convergence.}$$

This abstract result is at the heart of the convergence analysis of finite difference approximations of PDEs.