Scientific Computing: A New Way of Looking at Mathematics

Amiya Kumar Pani

Industrial Mathematics Group
Department of Mathematics
IIT Bombay, Powai, Mumbai-400076 (India)
Email: akp@math.iitb.ac.in

Roughly Questions to be answered (from the abstract)?

- **Why there is a need to relook at Mathematics ?**
- **In what ways questions asked in the new frame are different from traditional Mathematics ?**
- **What are its major objectives ?**
- **Does it answer the major concerns of the user community like ‘ How do we beleive the numbers crunched by the machine ?’**
- **Does it preserve the aesthetic beauty that is abstraction of the traditional Mathematics ?**

Organization of the Talk

- **On Scientific Computing**
  - Why there is a need to relook at Mathematics
  - Disaster caused by bad numerics
  - Role of Scientific Computing in Industry
  - Questions posed in Computational PDEs

Organization of the Talk—-

- **On Computational Issues**
  - Some questions which influence other related areas
  - Two important Concepts :Stability and consistency & their role
  - Convergence: what does it mean to user community
  - Some mathematical questions and possible attempt to provide answer

- Let us start with the following system of linear algebraic equations:

$$A\mathbf{X} = \mathbf{b}, \tag{1}$$

where **b** is a given vector and A is a **Hilbert matrix** given by

$$A_{ij} = \frac{1}{i + j - 1}.$$

In other words

$$A_{n\times n} = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \cdots & \frac{1}{n} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \cdots & \frac{1}{n+1} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \cdots & \frac{1}{n+2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{1}{n} & \frac{1}{n+1} & \frac{1}{n+2} & \cdots & \frac{1}{2n-1} \end{bmatrix}$$

- Elementary algebra shows

$$\det A \neq 0,$$

and hence, we obtain:

$$\mathbf{X} := A^{-1}\mathbf{b}. \tag{2}$$

- Using a finite machine, a machine with finite precession: Solve $AX = b$ and $A\tilde{X} = \tilde{b}$, where

$$\mathbf{b}' = (1, 0, 0, \cdots, 0)_{1\times 50}$$

and a slight perturbation of $\mathbf{b}$:

$$\tilde{\mathbf{b}}' = (1 + 10^{-3}, 0, 0, \cdots, 0)_{1\times 50}.$$

Note:

$$\|\mathbf{b} - \tilde{\mathbf{b}}\|_{\max} := \max_{1\leq i\leq 50} |b_i - \tilde{b}_i| \leq 10^{-3},$$

but let us look at the errors.

- For computing solution of

$$A\mathbf{X} = \mathbf{b}, \quad \text{and} \quad A\tilde{\mathbf{X}} = \tilde{\mathbf{b}},$$

using a finite machine, we observe that

$$\|\mathbf{X} - \tilde{\mathbf{X}}\|_{\max} = \max_{1 \le i \le 50} |X_i - \tilde{X}_i| \le 2.9329 \times 10^6.$$

- A small change in **b** gives rise to drastic change in the solution. Therefore, if we use low end machine and then high end machine, there will be be a drastic change in the solution and it is not acceptable as an approximate solution.

On Computational Mathematics—

- To probe a bit: it is the effect of round off error.

> Therefore, a new way of looking at Mathematics has emerged in the last sixty to seventy years and now a new branch called 'Computational Mathematics' has come into existence.

- When a given problem is approximated by using a finite machine (machine with a finite precession), we commit round off errors or truncation errors at every stage of computation. It is, therefore, essential to have a control on these errors.
- In this talk, we concentrate on 'Computational PDEs'.

Some disasters attributable to bad numerics

Here are some real life examples: The Patriot Missile failure in Dharan, Saudi Arabia, on February 25, 1991 which resulted in 28 deaths, is ultimately attributable to poor handling of rounding errors.

Some disasters attributable to bad numerics...

The explosion of the Ariane 5 rocket just after lift-off on its maiden voyage off French Guiana on June 4, 1996, was ultimately the consequence of a simple overflow. costing: $7 billion (construction cost) + $500 million (cargo cost + rocket)

Some disasters attributable to bad numerics...

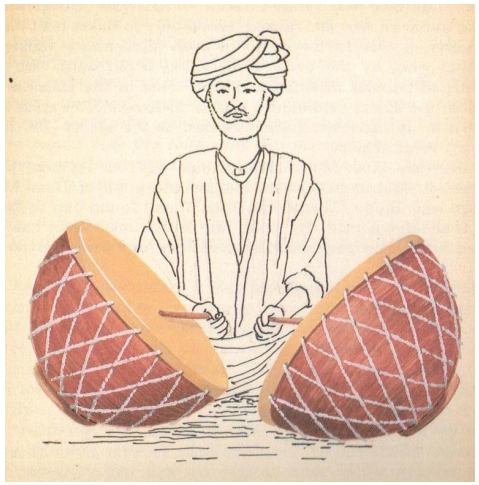
The sinking of the Sleipner A offshore platform in Gandsfjorden near Stavanger, Norway, on August 23, 1991, resulted in a loss of nearly one billion dollars. It was found to be the result of inaccurate finite element analysis. cost: $700 million

and so on ————-

Toy Problem

Vibration of Drum: (Modelling)

Ω : thin membrane of unit mass which is wrapped to the bream of a shallow wooden structure.



Problem

Given an external force f (the force exerted by the drummer through the wooden stick), find displacement $u = u(x, y)$, for $(x, y) \in \Omega$ satisfying

$$\begin{aligned} \Delta u &= -f \text{ in } \Omega \\ u &= 0 \text{ on } \partial\Omega \end{aligned} \tag{3}$$

where, $\Delta = \dfrac{\partial^2}{\partial x^2} + \dfrac{\partial^2}{\partial y^2}$, $\partial\Omega =$ boundary of Ω

Why Numerical Methods

- When Ω is square or disc and f is nice, we can expect an analytical solution otherwise not possible.
- For a simple linear elliptic problem like one discussed above, we have difficulties in calculating an exact solution if the domain is not a square or disc. However, problems comming from industry and or research & development organisations are not necessarily linear or not as simple as our toy problem.
- Even when analytical solutions are available, the final expression may contain certain complicated integral or expression which may be difficult to calculate analytically.

Way out: Resort to Numerical Approximation

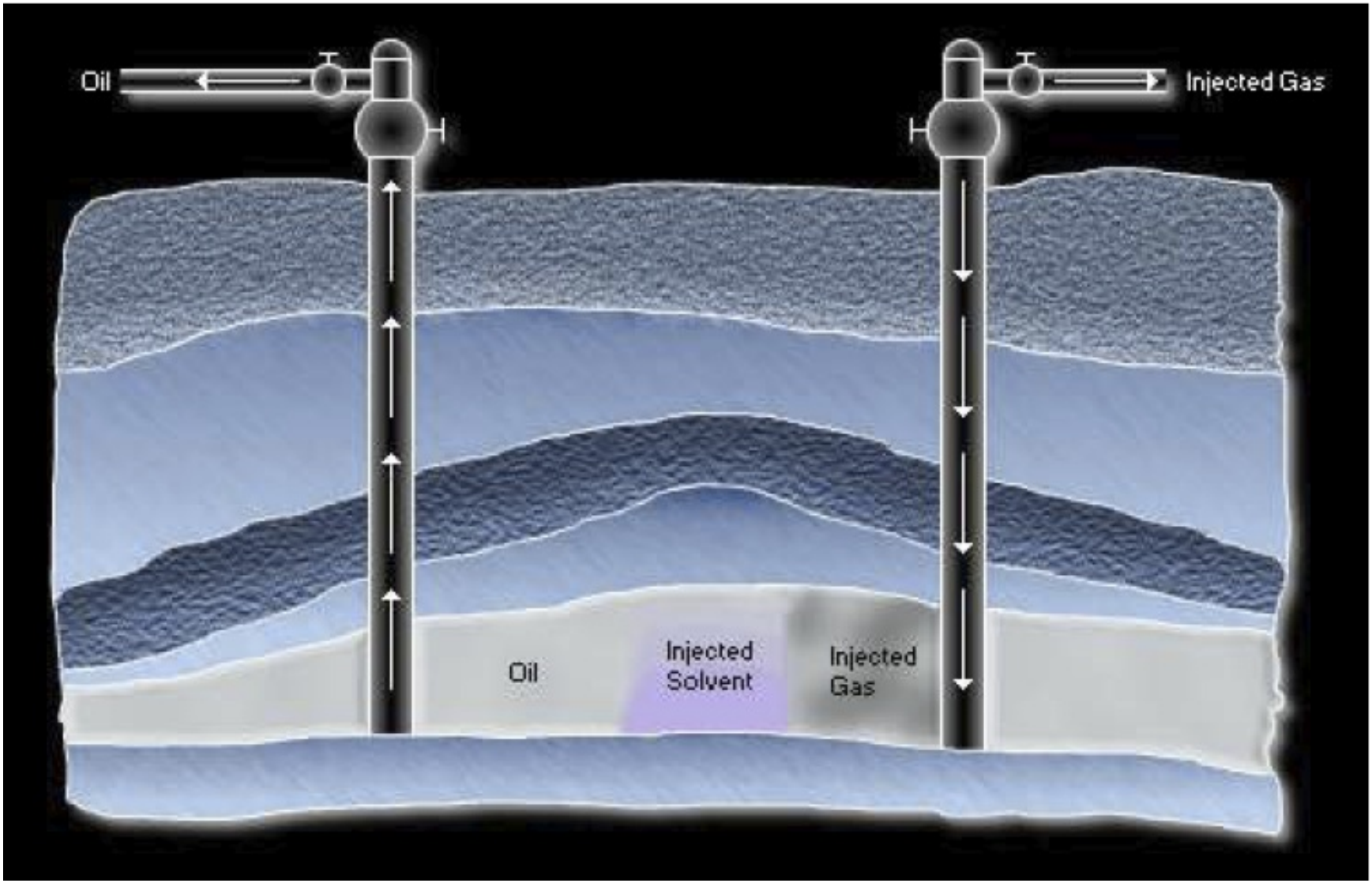
Buzz Word in Industry: Modelling and Simulation

- **Why Modelling and Simulations?**
  - To check Economically Viability.
  - Optimization to reduce operational costs.
  - To replace ultimately Pilot Plant by a Simulator (long term objective)

Heart of Modelling and Simulations: Mathematics

Basis of Simulation: Scientific Computing and Numerical Analysis

Enhanced Oil Recovery: One Example



Numerical Analysis

Broad Focus:

- To provide **mathematical justifications in terms of stability and convergence** to widely used numerical schemes for computational PDEs.
  It provides confidence to the user sothat he/she can repose faith on the numbers being crunched
- To design and develop **reliable and efficient algorithms** for numerical solutions to PDEs. By reliability, we mean that for a given tolerance and measurement, the computed solution stays near to the exact unknown solution within the prescribed tolerance with respect to the given measurement. By efficient, we understand that this can be achieved with minimal computational effort.
  It provides an answer to the question posed by the user such as 'Is it possible to find a computational solution which is up to some decimal point correct to the exact unknown solution with minimal effort ?'

Elliptic PDE

Consider the following Poisson equation

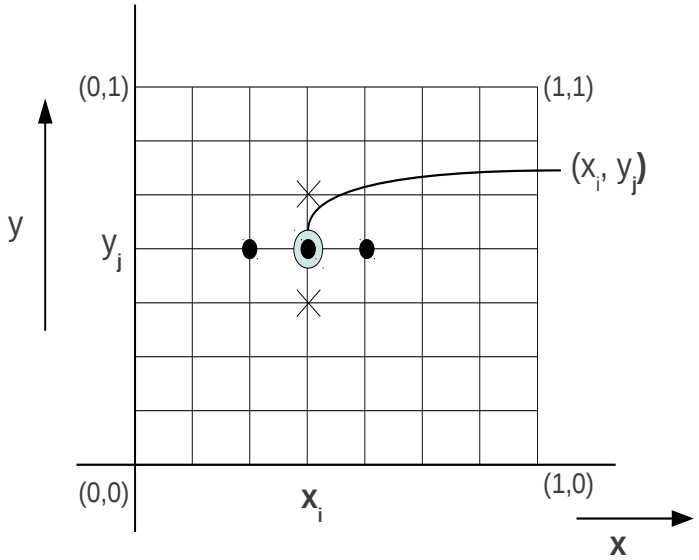
$$\begin{aligned} Lu := -\Delta u &= f \quad \text{in } \Omega = (0,1)\times(0,1) \\ u &= 0 \quad \text{on } \partial\Omega. \end{aligned} \tag{4}$$

where

- $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$,
- Ω is a square domain in $\mathbb{R}^2$.

Discretization of the domain Ω:

- $h_1 = \frac{1}{M_x+1}$ and $h_2 = \frac{1}{M_y+1}$ be the discretizing parameters in x and y directions,
- $\bar{\Omega}_h = \{(x_i, y_j) : x_i = ih_1, i = 0, ..., M_x + 1, y_j = jh_2, 0 \leq j \leq M_y + 1\}$ be a mesh on $\bar{\Omega}$,
  $\Omega_h = \{(x_i, y_j) : 1 \leq i \leq M_x, 0 \leq j \leq M_y\}$ be the set of interior nodal/mesh/grid points, and $\partial\Omega_h = \bar{\Omega}_h \setminus \Omega$ be the set of boundary points.



Discretization of the equation:

Replace the derivatives by difference quotients at the interior points (x_i, y_j) as

$$\left.\frac{\partial^2 u}{\partial x^2}\right|_{(x_i,y_j)} = \frac{u(x_{i-1}, y_j) - 2u(x_i, y_j) + u(x_{i+1}, y_j)}{h_1^2} + 0(h_1^2)$$

$$\left.\frac{\partial^2 u}{\partial y^2}\right|_{(x_i,y_j)} = \frac{u(x_i, y_{j-1}) - 2u(x_i, y_j) + u(x_i, y_{j+1})}{h_2^2} + 0(h_2^2).$$

Thus the equation (4) at (x_i, y_j) become:

$$\left.\begin{aligned} -\left[\frac{u(x_{i-1},y_j)-2u(x_i,y_j)+u(x_{i+1},y_j)}{h_1^2} + \frac{u(x_i,y_{j-1})-2u(x_i,y_j)+u(x_i,y_{j+1})}{h_2^2}\right] &\approx f_{ij} \\ 1 \le i \le M_x, 1 \le j \le M_y & \\ u(x_i, y_j) = 0 \quad (x_i, y_j) \in \partial\Omega_h. & \end{aligned}\right\} \quad (5)$$

where $f_{ij} = f(x_i, y_j)$.

Choose U_{ij} as an approximation of $u(x_i, y_j)$ such that

$$\left.\begin{aligned} L_h U_{ij} :\equiv & \\ & -\left[\frac{U_{i-1,j}-2U_{ij}+U_{i+1,j}}{h_1^2} + \frac{U_{i,j-1}-2U_{ij}+U_{i,j+1}}{h_2^2}\right] = f_{ij} \\ & 1 \le i \le M_x, 1 \le j \le M_y \\ & U_{ij} = 0 \quad \text{on} \quad \partial\Omega_h. \end{aligned}\right\} \quad (6)$$

Writing a vector U in lexicographic ordering, i.e.,

$$U = [U_{11}, U_{21}, ..., U_{M_x,1}, U_{12}, ..., U_{M_x,2}, ..., U_{M_x,M_y}]^T$$

(6) may be written as

$$AU = b \quad (7)$$

where A is a block tridiagonal matrix:

$$A = \begin{bmatrix} B & -\frac{1}{h_2^2}I & 0 & 0 & \cdots & 0 & 0 \\ -\frac{1}{h_2^2}I & B & -\frac{1}{h_2^2}I & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & . & -\frac{1}{h_2^2}I \\ 0 & 0 & 0 & 0 & \cdots & -\frac{1}{h_2^2}I & B \end{bmatrix}_{M_y \times M_y},$$

where $I = I_{M_x \times M_x}$ identity matrix and

$$B = \begin{bmatrix} 2(\frac{1}{h_2^1} + \frac{1}{h_2^2}) & -\frac{1}{h_2^1} & 0 & 0 & \cdots & 0 & 0 \\ -\frac{1}{h_2^1} & 2(\frac{1}{h_2^1} + \frac{1}{h_2^2}) & -\frac{1}{h_1^2} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & . & -\frac{1}{h_1^2} \\ 0 & 0 & 0 & 0 & \cdots & -\frac{1}{h_1^2} & 2(\frac{1}{h_2^1} + \frac{1}{h_2^2}) \end{bmatrix}$$

Questions to be asked:

1. Is this system (6) or equivalently (7) solvable ?
2. If so, can we compute the solution efficiently ?
3. Does the discrete solution U converge to u ? If so in what sense ?
4. If it converges, then How fast it convergences ?

On Computational Complexity: $M_x = M_y = 10^2$. So size of the matrix A is 10^4. Gaussian Elimination needs 10^{12} operations If each operation takes $5_{\mu s} = 5 \times 10^{-6} sec$. Then time taken: $1.7 \times 10^6 \simeq 20$ days. Not acceptable. But this is hightly sparse matrix with eack row having atmost 5 nonzero elements,but again bandwidth is large. So iterative methods will come into rescue.

Existence and Uniqueness

. Since matrix A is diagonally dominant[a] with at least one row is strictly diagonally dominant and it is irreducible in the sense that A can not be decomposed as

$$A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix},$$

then A is invertible.

[a] the matrix $A = [a_{ij}]_{1 \le i,j \le N}$ is diagonally dominant if $|a_{ii}| \ge \sum_{i \ne j = 1} |a_{ij}|$

This piece of result may not be that much useful, unless we ensure a small change in the data does not give a drastic change in the solution.

Lax-Ritchmyer Equivalence Theorem (On Convergence)

Theorem

A consistent finite difference scheme for a linear well-posed differential equation is convergent if and only if it is stable.

- Consistency in terms of truncation error τ_h.
  - **Local Truncation Error** : The amount by which the exact solution does not satisfy the difference scheme at (x_i, y_j).
  - $\tau_{ij,h} \equiv L_h u_{i,j} - f_{i,j} = O(h^2)$.
- Concept of stability
  - Connected with the propagation of round off or chopping off errors during the course of computation
  - small change in the data does not give rise to a drastic change in the solution.

Lax-Ritchmyer Equivalence Theorem (On Convergence)

Theorem

A consistent finite difference scheme for a linear well-posed differential equation is convergent if and only if it is stable.

- Consistency in terms of truncation error τ_h.
  - **Local Truncation Error** : The amount by which the exact solution does not satisfy the difference scheme at (x_i, y_j).
  - $\tau_{ij,h} \equiv L_h u_{i,j} - f_{i,j} = O(h^2)$.
- Concept of stability
  - Connected with the propagation of round off or chopping off errors during the course of computation
  - small change in the data does not give rise to a drastic change in the solution.

Lax-Ritchmyer Equivalence: Revisit in abstract form

$$Lu = f, \quad u = u(x),\ x \in \Omega$$

$\downarrow$ Discretize on to grid Ω_h. u_h interpolation of u on Ω_h

$$L_h U_h = f_h, \quad L_h : X_h \to Y_h$$

$\downarrow$ Definition

$e_h = u_h - U_h$ **Global Error (Convergence)**

$\downarrow$ Definition

$\tau_h = L_h u_h - f_h$ **Local Truncation Error**

Lax-Ritchmyer Equivalence: —

$\tau_h = L_h u_h - f_h$ **Local Truncation Error**

$\downarrow$ Taylor's Theorem

$\|\tau_h\| = O(h^p)$ **Consistency of order p**

$\downarrow$ Definition

$\|v_h\| \leq C\|L_h v_h\|, \quad \forall v_h \in X_h$ **Stability**

$\downarrow$ Lax Equivalence Theorem

$\downarrow$ Consistency + Stability

$\Updownarrow$

$\downarrow$ **Convergence**

Lax-Ritchmyer Equivalence: —

$\downarrow$

$\downarrow$

$\downarrow$

Lax Equivalence Theorem

Consistency + Stability

$\Updownarrow$

Convergence

$$\begin{aligned}\|e_h\| &\leq C\|L_h e_h\| \\ &= C\|L_h u_h - L_h U_h\| \\ &= C\|L_h u_h - f_h\| \\ &= C\|\tau_h\|\end{aligned}$$

Stability

Theorem

The FDM (6) is stable in the following sense:

$$\|U\|_{\infty,h} = \max_{(i,j)\in\bar{\Omega}_n} |U_{i,j}| \leq \frac{1}{8} \max_{(i,j)\in\Omega_n} |L_h U_{i,j}| = \frac{1}{8}\|L_h U\|_{\infty,h}$$

Theorem: (Convergence)

Let $u \in C^4(\bar{\Omega})$ be a solution of (4) and Let U be a solution of (6) or (7). Then

$$\max_{i,j \in \bar{\Omega}_h} |U_{i,j} - u(x_i, y_j)| \leq C(h_1^2 + h_2^2) \|\partial^4 u\|_{\infty,\bar{\Omega}}$$

where

$$\|\partial^4 v\|_{\infty,\bar{\Omega}} = \max_{(x,y) \in \bar{\Omega}} \{|D^\alpha v(x,y)| \ : \ |\alpha| \leq 4\}$$

Proof: Note that

$$\begin{aligned} L_h(U_{ij} - u(x_i, y_j) &= f_{ij} - L_h u(x_i, y_j) \\ &= Lu(x_i, y_j) - L_h u(x_i, y_j) \\ &= -\tau_{ij,\pi}(u) \end{aligned}$$

Proof—: With $E_{ij} = U_{ij} - u(x_i, y_j)$, we find that

$$L_h E_{ij} = -\tau_{ij,\pi}(u)$$

Using stability result, we arrive at

$$\max|E_{ij}| \leq \frac{1}{8} max|\tau_{ij,\pi}(u)|$$

With the help of Taylor series expansion, we observe that

$$|\tau_{ij,\pi}(u)| \leq C(h_1^2 + h_2^2)\|\partial^4 u\|_{\infty,\bar{\Omega}}$$

and this completes the rest of the proof.

Some related Mathematical Questions

- For $O(h^2)$ order convergence, we need $u \in C^4(\overline{\Omega})$. But, for square domain, PDE theory provides that the solution $u \in C^2(\Omega)$. So it is difficult to discuss order of convergence in case of square domain unless we deal with periodic BCs. However, extensive computational experiments suggest $O(h^2)$- order of convergence for problem in a square domain. Therefore, more pertinent question is

  > Is it possible to achieve $O(h^2)$-order of convergence using FDM for the Toy Problem in a square ?

  The answer is in affirmative.

Some related Mathematical Questions

- In 1968, Samarskii and Makarov proved some result for one dimesional situation which could not extended to higher dimensions. For Problem in higher dimensions, see[1].
  Other Researchers who have worked on this problem: E. Suli, Lazarov et al.

[1] A. K. Pani, S. K. Chung and R. S. Anderssen, On convergence of finite difference schemes for generalized solutions of parabolic and hyperbolic partial differential equations, CMA Report CMA-MRI-91

THANKS