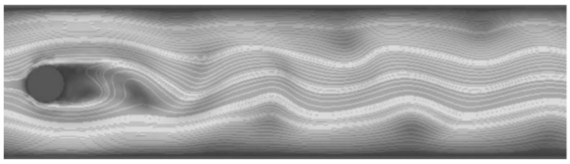
Advanced Solvers for Numerical Partial Differential Equations (PDEs)



Lecture 3

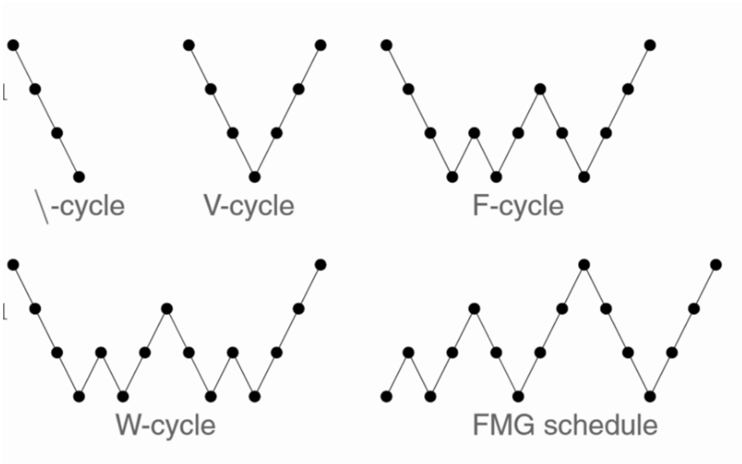
Nikolay Yavich
n.yavich@skoltech.ru

Part 1

Generalization of Multigrid

2

Multigrid Cycles



J. Gopalakrishnan

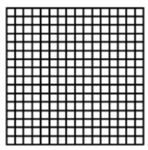
All these cycles satisfy the optimal $O(N)$ work estimate.

3

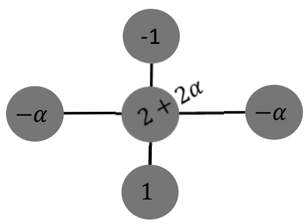
Model Anisotropic Discrete Problem

$0 < \alpha \ll 1$

- Anisotropic coefficients and equidistant grid
  $h_x = h_y$
  $-u_{xx} - \alpha\, u_{yy} = f$

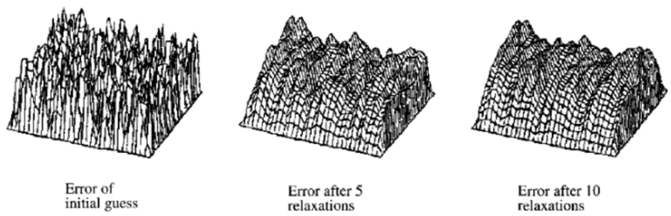


- Isotropic coefficients and stretched grid
  $h_y = \dfrac{h_x}{\sqrt{\alpha}}$
  $-u_{xx} - u_{yy} = f$



4

Anisotropy Impact

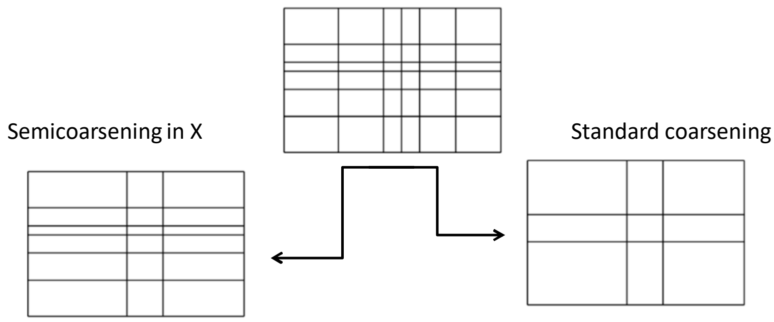


U. Trottenberg (2001)

The error does not get smoother in all directions.
We can not coarse all directions, only in those where the error got smoother.

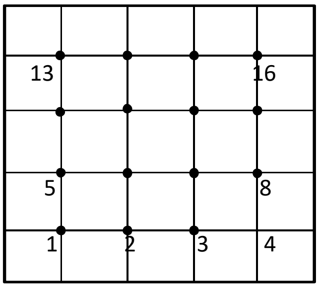
5

Semicoarsening



6

Line smoothers



Standard/point Jacobi smoothing

$u^{k+1} = u^k + D^{-1}(g - Au^k)$

D - diagonal

4	4	4	4
3	3	3	3
2	2	2	2
1	1	1	1

(a)

1	2	3	4
1	2	3	4
1	2	3	4
1	2	3	4

(b)

X-line and Y-line smoothing

D_x and D_y block tridiagonal

General Anisotropy

- Robustness for *mildly* anisotropic problem can gained with line smoothing and semicoarseing or their combinations
- Designing an optimal multigrid solver for heterogeneous *strongly* anisotropic problems is still an open question

Part 2
Basics of algebraic multigrid

The dream

$$-div\ (K\ \nabla u) = f$$
$$u = 0$$

$$A_h u_h = f_h$$

The earlier described multigrid (MG)
- Assume we know the underlying mesh X_h, and matrix A_h
- Generate coarse mesh points X_H
  analogous to taking every other point in regular mesh
- Retriangulate to get a new mesh (for unstructured grid)
- Use FE/FD on coarse mesh
  to project fine matrix to coarse one

Geometric MG

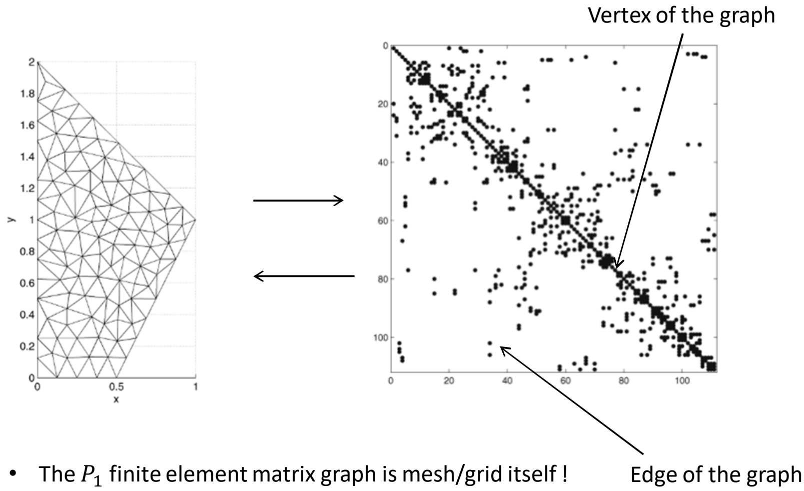
The dream multigrid
- Don't even have the underlying mesh, just matrix A_h
  Define X_H from A_h
  Define prolongation operator P from X_H to X_h

- With this, we can work out the rest:

$$R = P^T$$
$$A_H = R\ A_h\ P$$

Algebraic MG

Grid, Sparse Matrix, Graph



- The P_1 finite element matrix graph is mesh/grid itself !
- We don't actually need mesh,
  The matrix already stores connectivity information!

Coarsening Strategies

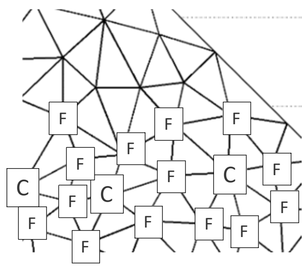
Splitting the mesh into coarse and fine nodes is achieved from the following considerations

$$X_h = C \cup F \qquad C \equiv X_H \qquad C \cap F = \emptyset$$

Simple greedy strategy based on magnitude of matrix entries
- C-point selected (point with largest "value")
- Neighbors of C-point become F-points
- Next C-point selected (after updating "values")
- F-points selected, etc.

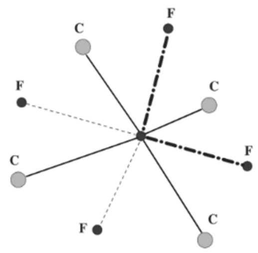
More advanced strategies to form a group of fine grid nodes
- Maximal Independent Sets
- Graph Partitioning
- Magnitude of matrix entries

And then associate one coarse grid node to each group



Operator-induced Prolongation

$$A_h = \begin{pmatrix} \cdots & & \\ & a_{ii} & a_{ij} \\ & & \cdots \end{pmatrix}$$



Y.Saad, 2003

<u>Claim</u> The error gets smoother in the directions where $\left|{}^{a_{ij}}/_{a_{ii}}\right|$ is large.

We thus can distinguish
- $\left|{}^{a_{ij}}/_{a_{ii}}\right| \geq \sigma$ *strong* connections.
- $\left|{}^{a_{ij}}/_{a_{ii}}\right| < \sigma$ *weak* connections.

We will have
- C-C connections
- Strong C-F connections
- Weak C-F connections

$$(Pu)_i = \begin{cases} u_i & i \in C \\ \sum_{k \in C_i} w_{ik} u_k & i \in F \end{cases}$$

with w_{ik} depending on a_{ii} and a_{ij}

13

Summary on AMG

- Much harder to implement than conventional geometric multigrid
- Yet much more universal
- Might be less robust (payment for universality)
  e.g. in case of anisotropy or strong heterogeneity
- Performance depends on a bunch of parameters, σ and some other
- Available from public-domain and proprietary packages ,
  e.g. `PyAMG` or `SAMG`

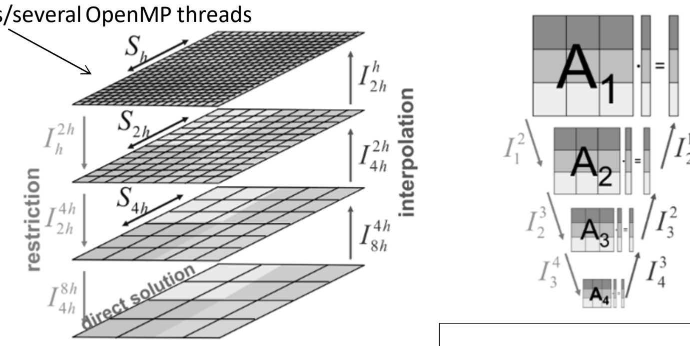
14

Part 3

Parallel Performance of Multigrid

Hybrid OpenMP/MPI Parallelization

1 MPI process/several OpenMP threads



K.Stueben

Partitioning of 1st grid induces partitioning on all levels

Parallelization Possibilities and Problems

Scalability of multigrid components

- Residual calculation – fairly parallel
- Smoothing – good smoothers are serial (eg, GS, ILU)…
- Restriction – fully parallel
- Prolongation

More importantly, at coarser levels computations will be highly communication-intensive

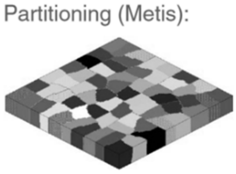
17

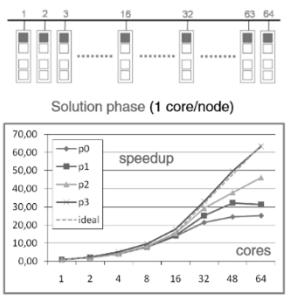
Example

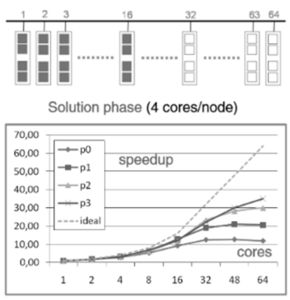
Test case:
3D Poisson problem,
27-point FE stencil

Mesh sizes:
p0: 884,736
p1: 1,560,896
p2: 3,176,523
p3: 6,331,625

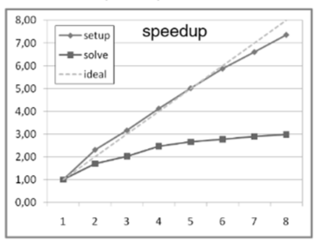
Partitioning (Metis):







Poblem with pure OpenMP:



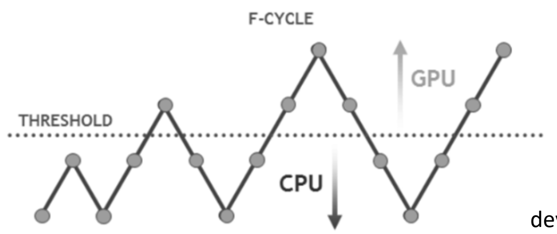
K.Stueben

Efficient cache utilization is very difficult

18

Hybrid GPU/CPU Parallelization – 1

- Top (fine) levels have lots of grid points and can run efficiently on throughput-oriented parallel architectures like the GPU.
- Lower (coarse) grids are better suited for latency-optimized processors like CPUs.



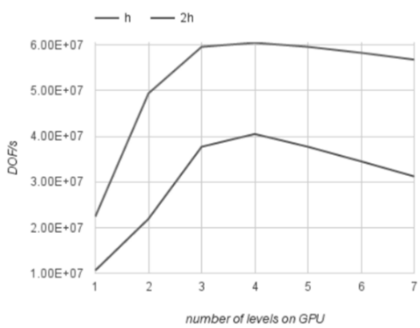
developer.nvidia.com

- What is the optimal threshold?

19

Hybrid GPU/CPU Parallelization – 2

Larger (h) and smaller (2h) problem sizes using NVIDIA Tesla K40 and dual-socket Intel Xeon E5-2690 v2.



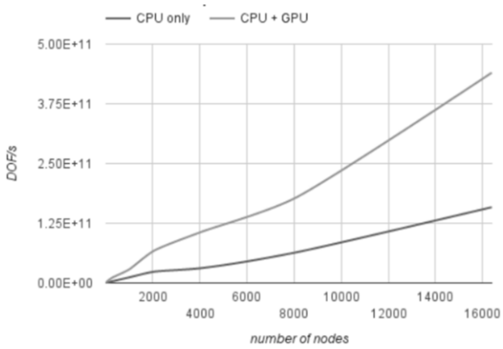
developer.nvidia.com

For this system configuration it's best to keep only the first 4 levels on the GPU and then switch over to the CPU for level 5 and higher.

20

Hybrid GPU/CPU Parallelization – 3

ORNL Titan supercomputer



Hybrid implementation scales linearly and provides a significant boost to overall system performance, up to **2.8x** speed-up compared to the CPU-only version.

Part 4

Solving Pure Neumann Problems and Singular Systems in General

1.13.1 Existence of a Solution

Consider the *linear system*

$$Ax = b. \qquad (1.73)$$

Here, x is termed the ***unknown*** and b the ***right-hand side.*** When solving the linear system (1.73), we distinguish three situations.

Case 1 The matrix A is nonsingular. There is a unique solution given by $x = A^{-1}b$.

Case 2 The matrix A is singular and $b \in \mathrm{Ran}(A)$. Since $b \in \mathrm{Ran}(A)$, there is an x_0 such that $Ax_0 = b$. Then $x_0 + v$ is also a solution for any v in $\mathrm{Null}(A)$. Since $\mathrm{Null}(A)$ is at least one-dimensional, there are infinitely many solutions.

Case 3 The matrix A is singular and $b \notin \mathrm{Ran}(A)$. There are no solutions.

Y. Saad, 2003

Iterative Solution of Singular Systems

$Au = f$

Eigenvalues $0 \cdots 0 < \lambda_{s+1} \cdots \lambda_N$
Eigenvectors $w_1 \cdots w_s, w_{s+1} \cdots w_N$

- $\det(A) = 0$
- $f \perp Null(A)$
- $A = A^T \geq 0$

u^* one of the solutions $\qquad$ $u^{**} = \min \|u^*\|$ normal solution

Richardson method $\qquad$ $u^k = u^{k-1} + \tau(f - Au^{k-1})$

Error $\qquad$ $z^k = u^k - u^*$ $\qquad$ $z^k = z^{k-1} - \tau\, Az^{k-1}$

$z^k = \sum \xi_t^k\, w_t$ $\qquad \cdots \qquad$ $z^k \to z^\infty \in Null(A)$

- For any initial guess u^0, the method converges
- If $u^0 \in Null(A)$, we receive the normal solution ($z^\infty = 0$)
- The same result holds for all other iterative methods (CG, SD, …)

Pure Neumann Problem

$-\Delta U = F$ in Ω
$\frac{\partial U}{\partial \nu} = G$ on Γ

Compatibility condition

$$\int_{\Omega} F dV = -\int_{\Gamma} G dS$$

After FD or FE discretization,
$Au = f$

- $\det(A) = 0$
- $Null(A) = \{1, \cdots 1\} = e$
- $f \perp Null(A)$
- $\sigma(A) = \{0, \lambda_2 \cdots \lambda_N\}$
- $A = A^T \geq 0$

We can do Richardson iterations,
$u^k = u^{k-1} + \tau(f - Au^{k-1})$
$u^0 = 0$ or $(u^0, e) = 0$

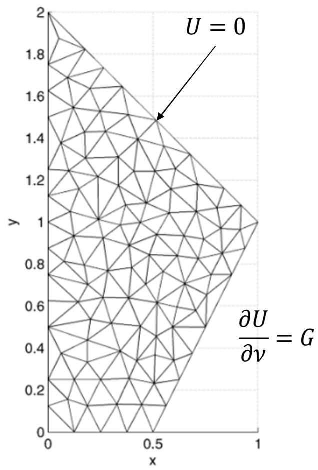
$$\tau_{\text{opt}} = \frac{2}{\lambda_2 + \lambda_N}$$

- The preconditioner has to be non-singular though

25

Make Matrix Non-singular

- It might be tempting to fix the solution at one boundary point
- This leads to a nonsingular problem
- We might expect a better convergence



26

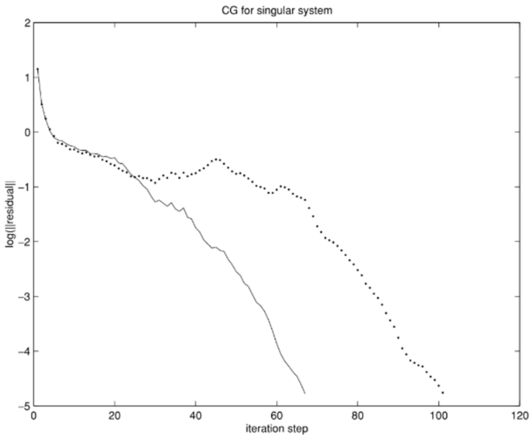


Figure 10.1. CG for a singular (—) and a near-singular (...) systems.

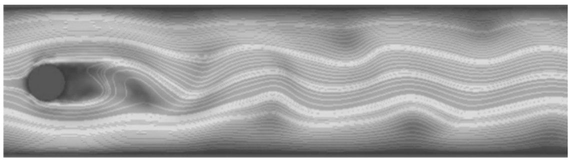
H. Vorst, 2003

Part 5
Stokes Problem

Stokes Problem

momentum equation

$$-\nu\Delta\boldsymbol{u} + \nabla p = \boldsymbol{f} \quad \text{in } \Omega$$
$$div\ \boldsymbol{u} = 0 \qquad \text{in } \Omega$$
$$u = 0 \qquad \text{on } \Gamma$$



continuity equation

$\boldsymbol{u}$ velocity
ν viscosity / “thickness”
p flow pressure, defined up to an additive constant
$\boldsymbol{f}$ external forces

- Steady incompressible flow
- implicit schemes for Navie-Stokes equations of fluid mechanics
- slow motion of fluids with very high viscosity (low Reynolds number)

29

Weak and Variational Formulations

$V = H_0^1(\Omega)^2$
$Q = L_0^2(\Omega)$

$$\nu \int_\Omega \nabla\boldsymbol{u} \cdot \nabla\boldsymbol{v}\, dx - \int_\Omega p\, div\, \boldsymbol{v}\, dx = \int_\Omega \boldsymbol{f} \cdot \boldsymbol{v}\, dx \quad \forall v \in V$$

$$\int_\Omega q\, div\, \boldsymbol{u}\, dx = 0 \qquad \forall q \in Q$$

Equivalently:

$$\min_{\boldsymbol{v}\in V} \frac{1}{2}\int_\Omega |\nabla\boldsymbol{v}|^2\, dx - \int_\Omega \boldsymbol{f} \cdot \boldsymbol{v}\, dx \qquad \text{such that } div\, \boldsymbol{v} = 0$$

Or by KKT theorem

$$L(\boldsymbol{v}, q) = \frac{1}{2}\int_\Omega |\nabla\boldsymbol{v}|^2\, dx - \int_\Omega q\, div\, \boldsymbol{v}\, dx - \int_\Omega \boldsymbol{f} \cdot \boldsymbol{v}\, dx$$

Claim $(\boldsymbol{u}, p)$ is a saddle-point of the lagrangian $L(\boldsymbol{v}, q)$, i.e.

$$L(\boldsymbol{u}, p) = \min_{\boldsymbol{v}\in V} \max_{q\in Q} L(\boldsymbol{v}, q)$$

In this context, p is a Lagrange multiplier corresponding to the constraint $div\, \boldsymbol{u} = 0$

30

FEM Discretization

$V_h \subset V$
$Q_h \subset Q$

$$\nu \int_\Omega \nabla \boldsymbol{u}_h \cdot \nabla \boldsymbol{v}_h \, dx - \int_\Omega p_h \, div\, \boldsymbol{v}_h \, dx = \int_\Omega \boldsymbol{f} \cdot \boldsymbol{v}_h \, dx \quad \forall \boldsymbol{v}_h \in V_h$$

$$\int_\Omega q_h \, div\, \boldsymbol{u}_h \, dx = 0 \qquad \forall q_h \in Q_h$$

Interestingly, V_h and Q_h cannot be picked independently.

Intuitively, if (asymptotically) Q_h is "too large" compared to V_h , then:
- we have too many constraints on the velocity
- the velocity does not have enough degrees of freedom.

In this case the discrete solution may not converge / be unstable

Compatibility Condition

Compatibility (or inf-sup or Ladyzhenskaya-Babuška-Brezzi) condition:
For any $q_h \in Q_h$ there exists $\boldsymbol{v}_h \in V_h \; \boldsymbol{v}_h \neq 0$ such that

$$\int_\Omega q_h \, div\, \boldsymbol{v}_h \, dx \geq \beta ||\boldsymbol{v}_h|| \, ||q_h||$$

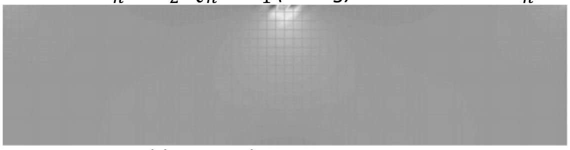
In other words, V_h is sufficiently rich compared with Q_h

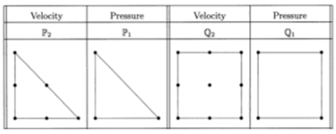
$||\boldsymbol{u}_h|| \leq C_1 ||f||$ and $||p_h|| \leq C_2 ||f||$

Under these assumptions we can prove existence and uniqueness of discrete solutions,

Examples of Pairs of V_h and Q_h

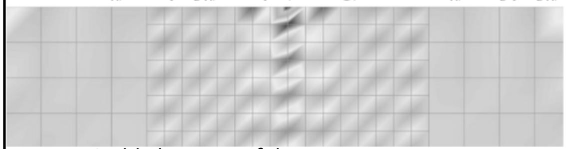
- Taylor-Hood elements (stable)
  $V_h = P_2\ Q_h = P_1$ *(triang)* $V_h = Q_2\ Q_h = Q_1$ *(quadr)*



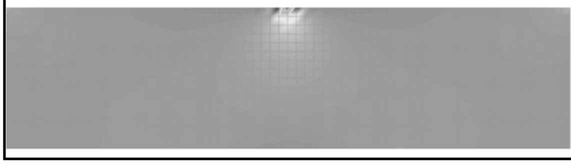


- Unstable example
  $V_h = P_1\ Q_h = P_1$ *(triang)* $V_h = Q_1\ Q_h = Q_1$ *(quadr)*

Admits generalization
$k \geq 1$



- Stable but wasteful
  $V_h = P_3\ Q_h = P_1$ *(triang)* $V_h = Q_3\ Q_h = Q_1$ *(quadr)*



W.Bangerth, TAMU

33

System Properties

$\boldsymbol{u}_h = \sum \bar{u}_i\, \boldsymbol{\eta}_i(x, y)$
$p_h = \sum \bar{p}_i\, \varphi_i(x, y)$

$$\bar{f}_i = \int_\Omega \boldsymbol{f} \cdot \boldsymbol{\eta}_i\, dx$$

$$\begin{pmatrix} M & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} \bar{u} \\ \bar{p} \end{pmatrix} = \begin{pmatrix} \bar{f} \\ 0 \end{pmatrix} \qquad \mathcal{A} = \begin{pmatrix} M & B^T \\ B & 0 \end{pmatrix}$$

- $\mathcal{A}$ indefinite symmetric
  very typical for constrained optimization
  real spectrum with both positive and negative eigenvalues

- M symmetric positive definite, simply two discrete laplacians

- B rectangular, discrete *div*
- B^T discrete *grad*
- LLB condition implies $Ker\ B^T = 0$, making $\mathcal{A}$ nonsinguar.

34

Solution Methods – 1

Pressure-matrix method

$$\begin{pmatrix} M & B^T \\ B & 0 \end{pmatrix}\begin{pmatrix} \bar{u} \\ \bar{p} \end{pmatrix} = \begin{pmatrix} \bar{f} \\ 0 \end{pmatrix}$$

$S = BM^{-1}B^T$

Schur complement system
symmetric positive-definite but dense
Cannot be stored explicitly!

Eliminate $\bar{u}$

$$S\,\bar{p} = M^{-1}\bar{f}$$

$$BM^{-1}B^T \approx div\,\Delta^{-1}grad \approx I$$

(i) compute first $\bar{q}_k = B^T\bar{p}_k$;
(ii) then solve $M\bar{w}_k = \bar{q}_k$
(iii) finally compute $\bar{r}_k = B\bar{w}_k$

- Schur complement matrix S is well-condition number.
- We can apply the conjugate gradient method,
  yet we'll have to solve $M\bar{w}_k = \bar{q}_k$ at every iteration with very high precision, not very universal

35

Solution Methods – 2

Preconditioned Uzawa method

$$\begin{pmatrix} M & B^T \\ B & 0 \end{pmatrix}\begin{pmatrix} \bar{u} \\ \bar{p} \end{pmatrix} = \begin{pmatrix} \bar{f} \\ 0 \end{pmatrix}$$

Pick $\bar{p}_0$, $\bar{u}_0$ and parameter ρ, $0 < \rho < 2\nu$

Integrate until convergence

P preconditioner to S

$$M\bar{u}_{\mathrm{k+1}} = \bar{f} - B^T\bar{p}_k$$

$$P(\bar{p}_{k+1} - \bar{p}_k) = \rho\, B\bar{u}_{\mathrm{k+1}}$$

- Uzawa scheme is just a special implementations of
  the preconditioned Richardson method applied to $S\,\bar{p} = M^{-1}\bar{f}$

36

Solution Methods – 3

Block preconditioned iterative solvers

$\mathcal{A} = \begin{pmatrix} M & B^T \\ B & 0 \end{pmatrix}$ $\qquad\qquad$ $\mathcal{B} = \begin{pmatrix} M & 0 \\ 0 & S \end{pmatrix}$, $\quad S = BM^{-1}B^{\mathrm{T}}$

Claim The eigenvalue problem $\mathcal{B}^{-1}\mathcal{A}\, w = \lambda w$ has 3 distinct eigenvalues

$$1, \frac{1+\sqrt{5}}{2}, \frac{1-\sqrt{5}}{2}.$$

Consequently, preconditioned GMRes will terminate in at most 3 iterations.

Proof (exercise)

Part 6
Mixed Finite Element Method (MFEM)

Mixed and Weak Formulations

$-div\ (K\ \nabla p) = f$ in Ω
$p = 0$ on Γ

- What if we are more interested in gradient/flux, $\boldsymbol{u} = -K\nabla p$, rather than p?

$K^{-1}\,\boldsymbol{u} + \nabla p = 0$
$-div\ \boldsymbol{u} = -f$

System of two first order equations

$V = H(div, \Omega)$
$Q = L^2\ (\Omega)$

$$\int_\Omega K^{-1}\boldsymbol{u}\cdot\boldsymbol{v}\,dx - \int_\Omega p\ div\ \boldsymbol{v}\,dx = 0 \quad \forall\ \boldsymbol{v}\in V$$

$$-\int_\Omega q\ div\ \boldsymbol{u}\,dx = -\int_\Omega q\ f\ dx \quad \forall\ q\in Q$$

$$||\boldsymbol{v}||^2 = \int_\Omega \boldsymbol{v}^2\,dx + \int_\Omega div\ \boldsymbol{v}^{\ 2}\,dx$$

$H(div, \Omega)$ Square-integrable vector-functions with square-integrable divergence

Mixed and Weak Formulations

Equivalently:

$$\min_{v\in V}\frac{1}{2}\int_\Omega K^{-1}\boldsymbol{v}\cdot\boldsymbol{v}dx \quad \text{such that } div\ \boldsymbol{v} = f$$ Dual Formulation

Or by KTT theorem

$$L(\boldsymbol{v}, q) = \frac{1}{2}\int_\Omega K^{-1}\boldsymbol{v}\cdot\boldsymbol{v}\,dx - \int_\Omega q\ (div\ v - f)\ dx$$ Mixed Formulation

<u>Claim</u> $(\boldsymbol{u}, p)$ is a saddle-point of the lagrangian $L(\boldsymbol{v}, q)$, i.e.

$$L(\boldsymbol{u}, p) = \min_{v\in V}\ \max_{q\in Q} L(\boldsymbol{v}, q)$$

In this context, p is a Lagrange multiplier corresponding to the constraint
$div\ \boldsymbol{u} = f$

Mixed FEM Discretization

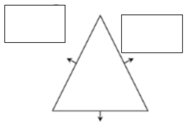
$V_h \subset V$
$Q_h \subset Q$

$$\int_\Omega K^{-1} \boldsymbol{u}_h \cdot \boldsymbol{v}_h \, dx - \int_\Omega p_h \, div \, \boldsymbol{v}_h \, dx = 0 \quad \forall \, \boldsymbol{v}_h \in V_h$$

$$-\int_\Omega q_h \, div \, \boldsymbol{u}_h \, dx = -\int_\Omega q_h \, f \, dx \quad \forall \, q_h \in Q_h$$

- Again, V_h and Q_h should obey the compatibility/inf-sup/LBB condition
- $H(div, \Omega)$ requires a special finite-element space
  The most common choice is Raviart-Thomas elements

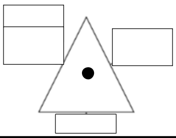
V_h



RT_0
3 constant normal fluxes

$$\boldsymbol{v}_h = \begin{pmatrix} a \\ b \end{pmatrix} + c \begin{pmatrix} x \\ y \end{pmatrix}$$

Q_h



P_0
1 mean value

$\boldsymbol{q}_h = 1$

41

System Properties

$\boldsymbol{u}_h = \sum \bar{u}_i \, \boldsymbol{\eta}_i(x, y)$
$p_h = \sum \bar{p}_i \, \varphi_i(x, y)$

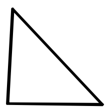
$$\bar{f}_i = \int_\Omega f \phi_i \, dx$$

$$\begin{pmatrix} M & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} \bar{u} \\ \bar{p} \end{pmatrix} = \begin{pmatrix} 0 \\ \bar{f} \end{pmatrix} \qquad \mathcal{A} = \begin{pmatrix} M & B^T \\ B & 0 \end{pmatrix}$$

- $\mathcal{A}$ indefinite symmetric
  very typical for constrained optimization
  real spectrum with both positive and negative eigenvalues
- M symmetric positive definite, mass matrix for RT_0
- B rectangular, discrete *div*
- B^T discrete *grad*
- LLB condition implies $Ker \, B^T = 0$, making $\mathcal{A}$ nonsinguar.

<u>Exercise</u>
Derive $RT_0 - P_0$ element system matrix for the rectangular triangle.

h



h

42

Solution Methods

Solution approaches are quite similar to those in Stokes problem:

- Pressure-matrix method
- Uzawa method
- Block preconditioned iterative solvers

Yet there is one approach that allows to construct a symmetric-positive definite matrix of smaller size.

Hybridization

$$K^{-1}\,\boldsymbol{u} + \nabla p = 0$$
$$-div\,\boldsymbol{u} = -f$$

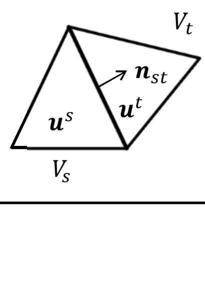
Integrate over each element area

$$\int_{V_s} K^{-1}\boldsymbol{u}\cdot\boldsymbol{v}\,dx - \int_{V_s} p\,div\,\boldsymbol{v}\,dx + \int_{\partial V_s} \overset{\lambda}{p}\,\boldsymbol{v}\cdot\boldsymbol{n}\,dl = 0 \quad \forall\,\boldsymbol{v}\in V$$

$$-\int_{V_s} q\,div\,\boldsymbol{u}\,dx = -\int_{V_s} q\,f\,dx \quad \forall\,q\in Q$$

- Assume $\boldsymbol{u}$ and $\boldsymbol{v}$ are discontinuous across the element boundaries and enforce continuity of the normal flux separately

$$\int_{\Gamma_{st}} \mu\,(\boldsymbol{u}^s - \boldsymbol{u}^t)\cdot\boldsymbol{n}_{st}\,dl = 0 \quad \forall\,\mu\in\Lambda$$



Hybridization

$$\begin{pmatrix} A & -B^T & C^T \\ -B & & \\ C & & \end{pmatrix}\begin{pmatrix} u \\ p \\ \lambda \end{pmatrix} = \begin{pmatrix} 0 \\ f \\ 0 \end{pmatrix}$$

A block-diagonal
C assembled from diagonal element matrices

$\mathcal{S}\,\lambda = g$

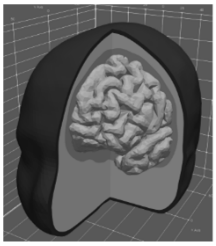
$\mathcal{S} = C^T[A^{-1} - A^{-1}B(B^TA^{-1}B)^{-1}B^TA^{-1}]C$

$\mathcal{S}$ is sparse symmetric positive-definite!

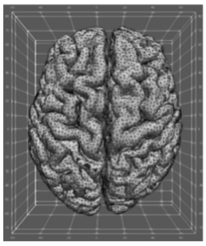
Multigrid can be applied here.

45

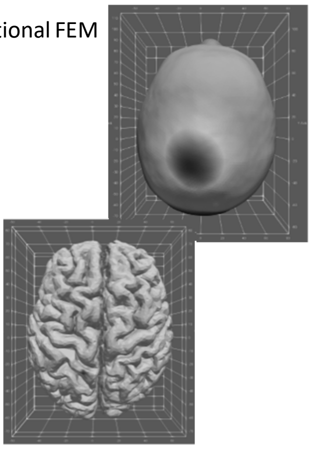
Example



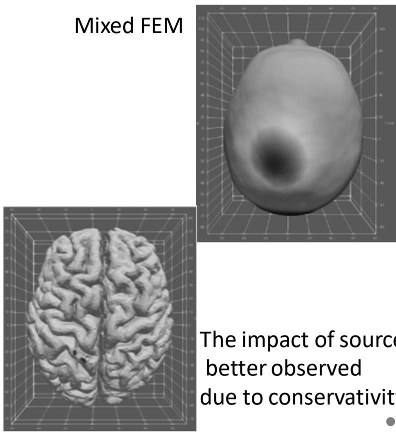
Neural currents propagation



Conventional FEM



Mixed FEM



The impact of sources is better observed due to conservativity

46

Summary on Mixed FEM

- Better approximation to the gradient
- Conservative
  Good for localized sources, rough grids, highly heterogeneous problems
- More computationally demanding than the conventional FEM