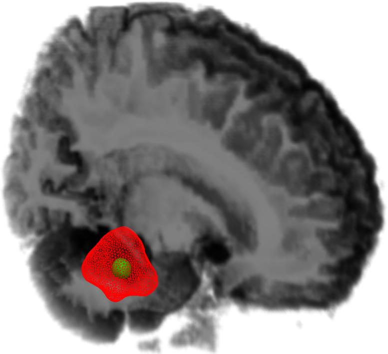
On Preconditioning Newton Method for PDE Constrained Optimization Problems

Amir Gholami, George Biros

INSTITUTE FOR COMPUTATIONAL ENGINEERING & SCIENCES

THE UNIVERSITY OF TEXAS AT AUSTIN

May 26, 2016



Outline

Introduction - Gliomas

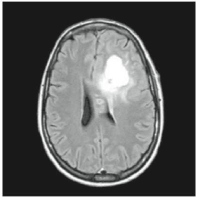
Gliomas:

- Account for **80%** of all malignant brain tumors.
- More than **60,000** new cases each year in US.

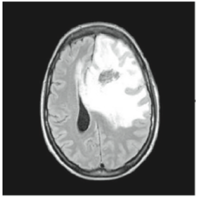
Goals of biophysical tumor growth model:

- Determining the extent of tumor infiltration
- Pre- and post-operative planning

This requires a patient specific method to approximate a tumor growth model parameters.



Initial FLAIR Scan

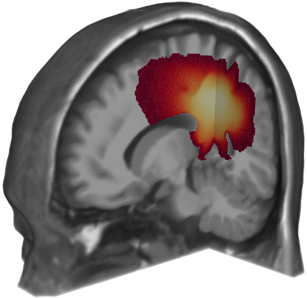


Time2 FLAIR Scan

Image from (P. Mosayebi et. al.)

Outline

Forward Model

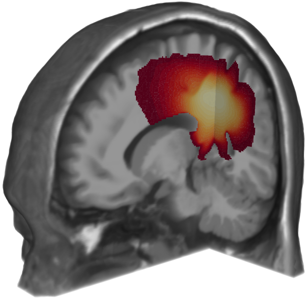


Phenomenological model:

$$\frac{\partial c}{\partial t} - \nabla.(k(x)\nabla c) - \rho c(1-c) = 0 \text{ in } U = \Omega \times (0, T],$$

$$\frac{\partial c}{\partial n} = 0 \text{ on } \Gamma \times (0,1),$$

Forward Model



Phenomenological model:

$$\frac{\partial c}{\partial t} - \nabla.(k(x)\nabla c) - \rho c(1-c) = 0 \text{ in } U = \Omega \times (0,T],$$

$$\frac{\partial c}{\partial t} - Dc - r(c) = 0 \text{ in } U = \Omega \times (0,T],$$

$$\frac{\partial c}{\partial n} = 0 \text{ on } \Gamma \times (0,1),$$

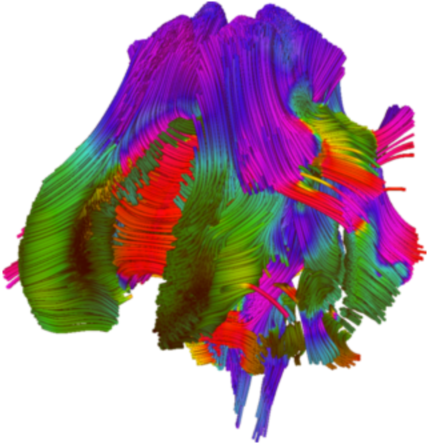
Forward Model - Diffusion

$$Dc = \nabla.(k(x)\nabla c),$$

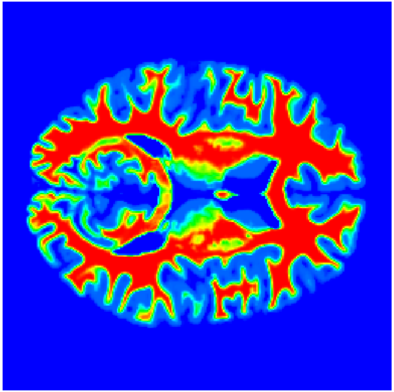
$$k(x) = k_0(x)I + k_f T(x)$$

- k_0: Inhomogeneous diffusion part
- $T(x)$: Anisotropic diffusion part
- k_f: Anisotropic diffusion coefficient

Forward Model - Diffusion



Principle direction extracted from DTI
(raw data provided by LONI lab of USC).



Inhomogeneous part of the diffusion

Outline

Inverse Problem - Setup

PDE-constrained optimization:

$$\min_{\color{blue}{p}} \mathcal{J} := \frac{1}{2}\|O_0 c_0 - d_0\|^2_{L^2(\Omega)} + \frac{1}{2}\|O_1 c_1 - d_1\|^2_{L^2(\Omega)} + \frac{\beta_p}{2}\|p\|^2_{\mathcal{R}^{N_k}},$$

subject to:

$$\frac{\partial c}{\partial t} - Dc - r(c) = 0 \text{ in } U$$

$$c_0 - \Phi \color{blue}{p} = 0.$$

Inverse Problem - KKT Optimality Conditions

THE UNIVERSITY OF TEXAS AT AUSTIN

By requiring stationarity of the Lagrangian with respect to the adjoint, state, and inversion variable, we obtain the so called KKT optimality conditions:

$$\frac{\partial \mathcal{L}}{\partial c} = 0 \Rightarrow \text{ adjoint equation}$$

$$\frac{\partial \mathcal{L}}{\partial \alpha} = 0 \Rightarrow \text{ state equation}$$

$$\frac{\partial \mathcal{L}}{\partial p} = 0 \Rightarrow \text{ inversion equation}$$

Inverse Problem - First Optimality Conditions

THE UNIVERSITY OF TEXAS AT AUSTIN

Find c, α, p such that:

$$\begin{cases} -\dfrac{\partial \alpha}{\partial t} - D\alpha - \dfrac{\partial r(c)}{\partial c}\alpha = 0 \\ \alpha_1 + O_1^T\left(O_1 c_1 - d_1\right) = 0 \\ \dfrac{\partial c}{\partial t} - Dc - r(c) = 0 \\ c_0 - \Phi p = 0 \\ \beta p + O_0^T\left(O_0 p - d_0\right) - \Phi^T \alpha_0 = 0 \end{cases}$$

Reduced Space Method

THE UNIVERSITY OF TEXAS AT AUSTIN

$$H\tilde{p} = -\partial_p \mathcal{L}|_{p^0}.$$

where H is the reduced Hessian and $H\tilde{p}$ is computed by:

1. Given p^0, compute c^0 and α^0 from the state and adjoint equations.
2. Solve for $\tilde{c}$ and $\tilde{\alpha}$:

$$\frac{\partial \tilde{c}}{\partial t} - \nabla \cdot (k\nabla\tilde{c}) - \rho(1 - 2c^0)\tilde{c} = 0$$

$$\tilde{c}(0) - \Phi\tilde{p} = 0$$

$$-\frac{\partial \tilde{\alpha}}{\partial t} - \nabla \cdot (k\nabla\alpha) - \rho(1 - 2c^0)\tilde{\alpha} + 2\rho\tilde{c}\alpha^0 = 0$$

$$\tilde{\alpha}(T) + O^T O\tilde{c}(T) = 0$$

Reduced Space Method

$$H\tilde{p} = -\partial_p \mathcal{L}|_{p^0}.$$

where H is the reduced Hessian and $H\tilde{p}$ is computed by:

1. Given p^0, compute c^0 and α^0 from the state and adjoint equations.
2. Solve for $\tilde{c}$ and $\tilde{\alpha}$:

$$\frac{\partial \tilde{c}}{\partial t} - \nabla \cdot (k \nabla \tilde{c}) - \rho(1 - 2c^0)\tilde{c} = 0$$

$$\tilde{c}(0) - \Phi \tilde{p} = 0$$

$$-\frac{\partial \tilde{\alpha}}{\partial t} - \nabla \cdot (k \nabla \alpha) - \rho(1 - 2c^0)\tilde{\alpha} + 2\rho \tilde{c} \alpha^0 = 0$$

$$\tilde{\alpha}(T) + O^T O \tilde{c}(T) = 0$$

Straing Splitting: Forward Problem

THE UNIVERSITY OF TEXAS AT AUSTIN

1. Solve $\frac{\partial c}{\partial t} = Dc$ over time $\Delta t/2$ with c^n as initial condition, to obtain $c^\dagger$.
2. Solve $\frac{\partial c}{\partial t} = R(c)$ over time Δt with $c^\dagger$ as initial condition, to obtain $c^{\dagger\dagger}$.
3. Solve $\frac{\partial c}{\partial t} = Dc$ over time $\Delta t/2$ with $c^{\dagger\dagger}$ as initial condition, to obtain c^{n+1}.

This scheme can more compactly be written as:

$$c^{n+1} = S_D^{\frac{\Delta t}{2}} S_R^{\Delta t} S_D^{\frac{\Delta t}{2}} c^n,$$

Outline

Hessian Preconditioners

THE UNIVERSITY OF TEXAS AT AUSTIN

Extensive work on Hessian preconditioners on stationary problems but much less work on solvers for nonlinear parabolic systems

General ideas:

- Low rank approximation,
- Domain-decomposition,
- Analytic preconditioners,
- Multilevel preconditioners,

Our approach fits in the last two categories with some elements on multilevel.

Hessian Preconditioner

1. Given p^0, compute c^0 and α^0 from the state and adjoint equations.
2. Solve for $\tilde{c}$ and $\tilde{\alpha}$:

$$\frac{\partial \tilde{c}}{\partial t} - \nabla \cdot (k \nabla \tilde{c}) - \rho(1 - 2c^0)\tilde{c} = 0$$

$$\tilde{c}(0) - \Phi \tilde{p} = 0$$

$$-\frac{\partial \tilde{\alpha}}{\partial t} - \nabla \cdot (k \nabla \alpha) - \rho(1 - 2c^0)\tilde{\alpha} + 2\rho \tilde{c} \alpha^0 = 0$$

$$\tilde{\alpha}(T) + O^T O \tilde{c}(T) = 0$$

3. $H\tilde{p} = \beta \tilde{p} - \Phi^T \tilde{\alpha}(0)$.

$\star$ Approximate J^{-1} by including the nonlinear reaction term, and using average diffusion coefficient $(\bar{k})$

Hessian Preconditioner

1. Given p^0, compute c^0 and α^0 from the state and adjoint equations.
2. Solve for $\tilde{c}$ and $\tilde{\alpha}$:

$$\frac{\partial \tilde{c}}{\partial t} - \nabla \cdot (\bar{k} \nabla \tilde{c}) - \rho(1 - 2c^0)\tilde{c} = 0$$
$$\tilde{c}(0) - \Phi \tilde{p} = 0$$
$$-\frac{\partial \tilde{\alpha}}{\partial t} - \nabla \cdot (\bar{k} \nabla \alpha) - \rho(1 - 2c^0)\tilde{\alpha} + 2\rho\tilde{c}\alpha^0 = 0$$
$$\tilde{\alpha}(T) + O^T O \tilde{c}(T) = 0$$

3. $P_H \tilde{p} = \beta \tilde{p} - \Phi^T \tilde{\alpha}(0)$.

Analytical Hessian Preconditioner

THE UNIVERSITY OF TEXAS AT AUSTIN

- AP_0: Use $\bar{k}$ for the diffusion term and ignore the rest
  - Computational cost independent of the number of time steps.
- AP_1: Use $\bar{k}$ for the diffusion term and numerically solve the rest
  - Computational cost scales linearly with the number of time steps.

OR

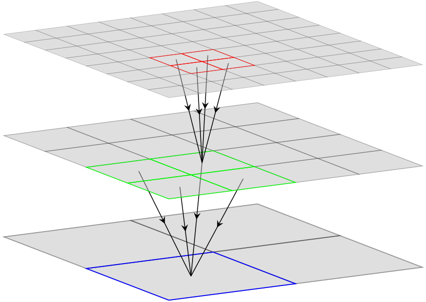
Compute the true H^{-1}, on a coarser grid $\rightarrow MLP_i$

Multi Level Preconditioner

$$\tilde{H} = \tilde{J}^{-T} O^T O \tilde{J}^{-1} + \beta I.$$

Use an iterative solver to compute the matvec of $\tilde{H}^{-1}$

- Higher computational cost compared to APs.



Reduced Space Method - Compact Form

(1) Start with some p^0.

(2) $P_H^{-1} H\tilde{p} = -P_H^{-1}\frac{\partial J}{\partial p}$.

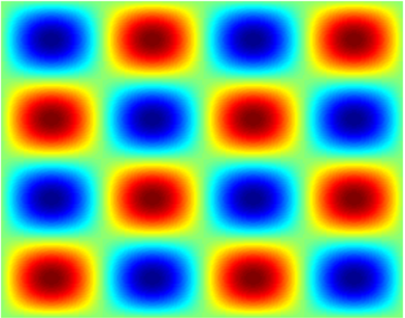
(3) Set $p^0 = p^0 + \gamma\tilde{p}$ (with line search).

(4) If tolerance is reached break, otherwise go back to 2.

Preconditioner Performance: Test Case 1

THE UNIVERSITY OF TEXAS AT AUSTIN

$$\frac{\partial c}{\partial t} - \nabla.(k(x)\nabla c) - \rho c(1-c) = 0$$



$k = 1 + sin(2x)$ and ρ constant

Preconditioner Performance: Test Case 1

THE UNIVERSITY OF TEXAS AT AUSTIN

The number of Hessian matvecs to solve the optimality conditions for one iteration (tol=1E-3, $N = 128^2$). AP: Analytical Preconditioner, MLP_i Coarse Grid Preconditioner (level i coarsening).

ρ	No Prec	P_0	P_1	MLP_1	MLP_2	MLP_3
0	**6**	**3+0.00**	3+0.07	3+1.50	3+0.29	4+0.07
0.1	**6**	**3+0.00**	3+0.07	3+1.50	3+0.29	4+0.07
1	**8**	5+0.01	**4+0.11**	3+1.82	4+0.51	6+0.11
5	**13**	15+0.02	**6+0.40**	5+9.00	6+2.28	10+0.71

Hybrid inexact Newton method

THE UNIVERSITY OF TEXAS AT AUSTIN

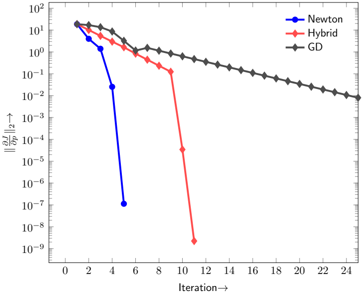
💡 Since the preconditioners perform very well, we can use them at the early stages of the Newton solves

$\longrightarrow$ hybrid inexact Newton Method

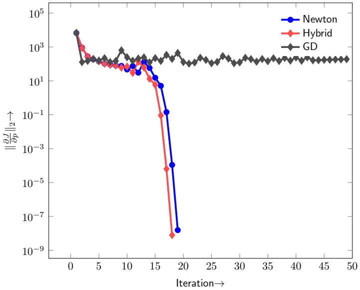
Preonditioner Performance: Test Case 1

The total of Hessian matvecs to required to reduce the L2 norm of the gradient below 1E-3.

Method	$\rho = 0$	$\rho = 0.1$	$\rho = 1.0$	$\rho = 5.0$
UnPrec. Newton	**12**	**12**	**22**	**85**
Prec. Newton	8.90	8.90	15.00	44.43
Hybrid Newton	**2.34**	**2.35**	**5.69**	**32.59**



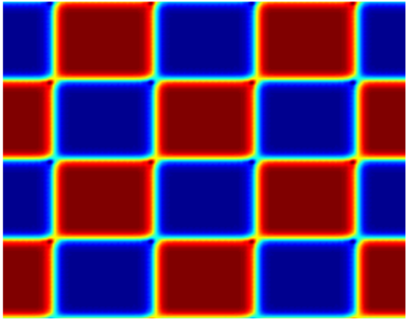
(a) $\rho = 0.$



(b) $\rho = 5.0$

Preconditioner Performance: Test Case 2

$$\frac{\partial c}{\partial t} - \nabla.(k(x)\nabla c) - \rho c(1-c) = 0$$



Checkerboard diffusion coefficient, and $\rho = 1$ constant

Preconditioner Performance: Test Case 2

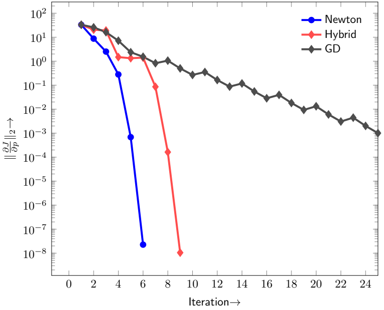
The number of Hessian matvecs to solve the optimality conditions for one iteration (tol=1E-3, $N = 128^2$). AP: Analytical Preconditioner, MLP_i Coarse Grid Preconditioner (level i coarsening).

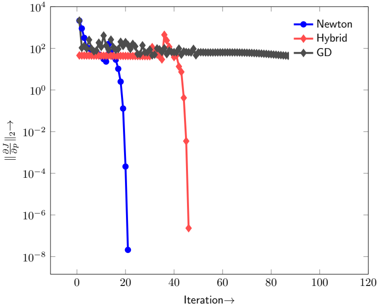
T	No Prec	P_0	P_1	MLP_1	MLP_2	MLP_3
0.2	**6**	**3+0.01**	3+0.22	2+1.31	3+0.37	3+0.08
0.5	**7**	4+0.01	4+0.14	3+2.19	**3+0.39**	4+0.12
1	**9**	6+0.00	5+0.12	3+2.84	**4+0.61**	6+0.18
2	**15**	12+0.01	8+0.15	5+7.77	**6+1.46**	10+0.39
4	**22**	23+0.00	16+0.20	10+23.84	**9+3.79**	14+0.05

Preconditioner Performance: Test Case 2

The total of Hessian matvecs to required to reduce the L2 norm of the gradient below 1E-3.

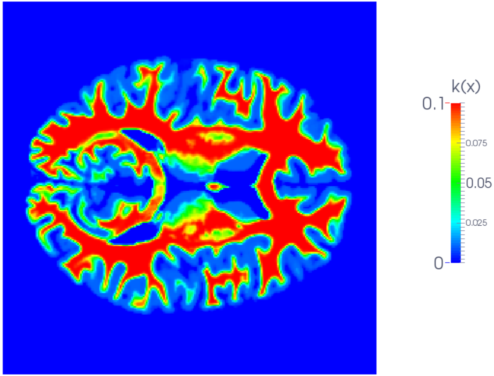
Method	$T = 0.2$	$T = 0.5$	$T = 1.0$	$T = 2.0$	$T = 4.0$
UnPrec. Newton	**12**	**12**	**27**	**43**	**96**
Prec. Newton	8.37	8.37	16.63	27.06	**54.41**
Hybrid Newton	**2.63**	**5.90**	**11.81**	**22.13**	54.60





Preonditioner Performance: Test Case 3

$$\frac{\partial c}{\partial t} - \nabla.(k(x)\nabla c) - \rho c(1-c) = 0$$



Diffusion coefficient for test case 4.

Preconditioner Performance: Test Case 3

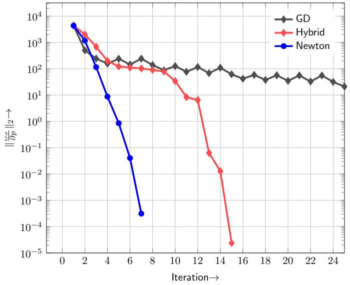
The number of Hessian matvecs to solve the optimality conditions for one iteration (tol=1E-3, $N = 64^3$). AP: Analytical Preconditioner, MLP_i Coarse Grid Preconditioner (level i coarsening).

ρ	No Prec	AP_0	AP_1	MLP_1	MLP_2
1.0	**7**	**3.00**	3+0.05	3+0.25	3+0.02
2.0	**8**	5+0.01	**3+0.06**	5+0.63	6+0.10
4.0	**9**	10+0.02	**3+0.06**	12+0.06	12+0.24

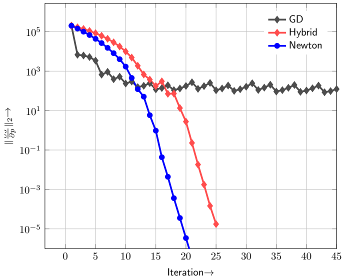
Preonditioner Performance: Test Case 3

Total number of Hessian matvecs necessary to reach convergence (i.e. $\|\frac{\partial J}{\partial p}\|_2 < 1$E-3)

Method	$\rho = 1.0$	$\rho = 2.0$	$\rho = 4.0$
UnPrec. Inexact Newton	**69**	**74**	**137**
Prec. Inexact Newton	34.34	38.42	47.66
Hybrid Inexact Newton	**28.06**	**28.01**	**42.6**



(a) $\rho = 1.0$



(b) $\rho = 4.0$

Acknowledgment

THE UNIVERSITY OF TEXAS AT AUSTIN

Thank You

- Andreas Mang
- Florian Tramnitzke
- Hari Sundar
- Dhairya Malhotra
- Nick Alger