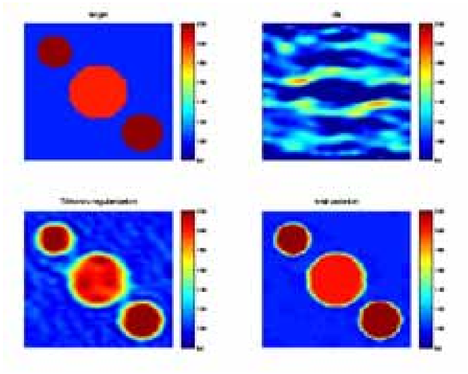
Fast algorithms for inverse problems governed by PDEs

George Biros



$$\min_u \mathcal{J}(u) := \frac{1}{2}\|Ku - d\|^2 + \frac{\beta}{2}\|u\|^2$$

Georgia Institute of Technology

Collaborators and Support

- Santi S. Adavani (Penn), Volkan Akcelik (Stanford), Omar Ghattas (UT Austin)

- National Science Foundation
- National Institutes of Health
- Air Force Office of Scientific Research
- Department of Energy

Georgia Institute of Technology

Outline

- Inverse problems with PDEs
- The input-output map
- Ill-posedness and regularization
- Algorithmic goals
- The Hessian
    - Structure (equilibrium,evolution)
    - Preconditioning and regularization

Key points of the talk

- Hessian
  - Regularization
  - Acceleration
  - Probability
  - “Hard” to compute

Inverse problems in earth sciences

- Weather/Climate/Cloud
- Space
- Ocean/Ice
- Carbon
- Ecosystems
- Subsurface imaging
  - o Gravity, electromagnetics, porous media, seismic

- References for inverse problems theory
  - o Tarantola, Kaipio & Somersalo, Vogel

General problem statement

observation operator

data

state

inversion params

$$d = Q(y)$$

$$\frac{dy}{dt} = m(y, u, t) + n(y, t), \qquad y(0) = y_0$$

model

noise

$$y = \mathcal{M}(u)$$

- Model encapsulates conservation and constitutive laws

Linear case: input-output operator and the Hessian

- Forward operator $$y = \mathcal{M}u$$
- Observation operator $$d = Qy$$
- Input-output operator $$d = Ku$$
- Hessian $$\boldsymbol{H = K^{*}K}$$
- Adjoint $$K^{*} = M^{*}Q^{*}$$

Inverse problem, Linear case

- Inverse problem $$\min_u \mathcal{S}(Ku, d) + \mathcal{R}(u)$$
- Least squares $$\min_u \frac{1}{2}\|Ku - d\|_2^2 + \frac{1}{2}\|u\|_R^2$$
  - Linear system for *u*

$$(R + K^*K)u = K^*d$$

- How do you choose *R*?
- Scalable algorithms
  - Relation to probabilistic/Bayesian approaches

    Maximum likelihood

Inverse scattering

- Forward problem
  Given tissue material and excitation parameters, compute acoustic response

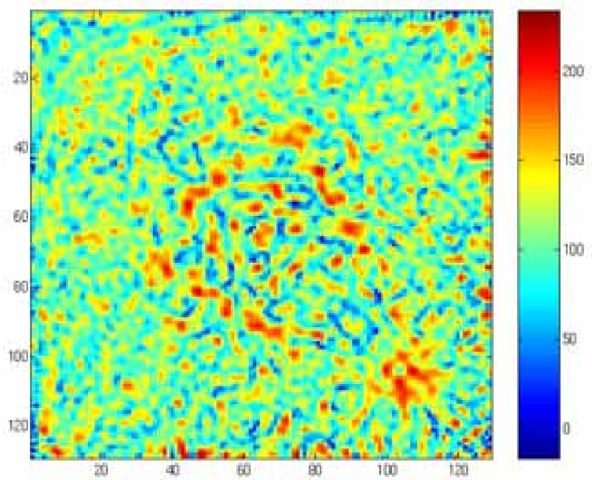
- Inverse problem
  Given observations (and source), estimate material parameters
  - o Constrain w/deviation with prior models
  - o Penalize w/total variation

$$\min_{u,y} \frac{1}{2} \sum_{j=1}^{N} \int_0^T (d_j - y(x_j))^2 \, dt + \frac{\beta}{2} \int_\Omega |\nabla u|^2 \, d\Omega$$

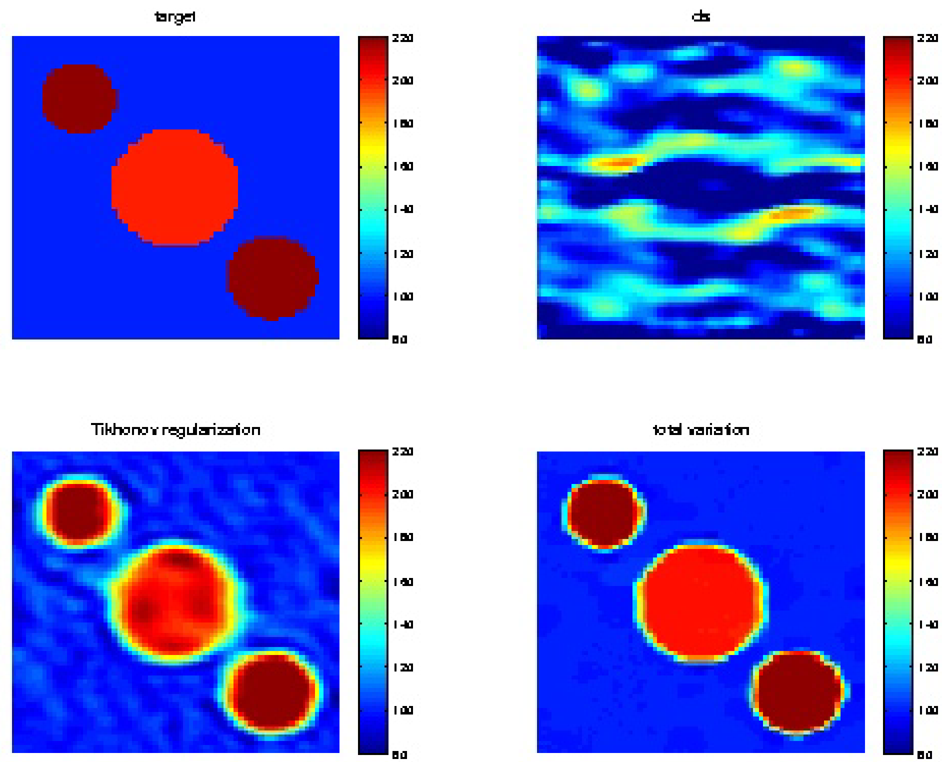
$$\text{subject to} \quad \frac{\partial^2 y}{\partial t^2} - \nabla \cdot u \nabla y = f \quad \text{in} \quad \Omega \times [0, T] \qquad + B.C., \; I.C.$$

2D acoustics Akcelik et al 02

No regularization



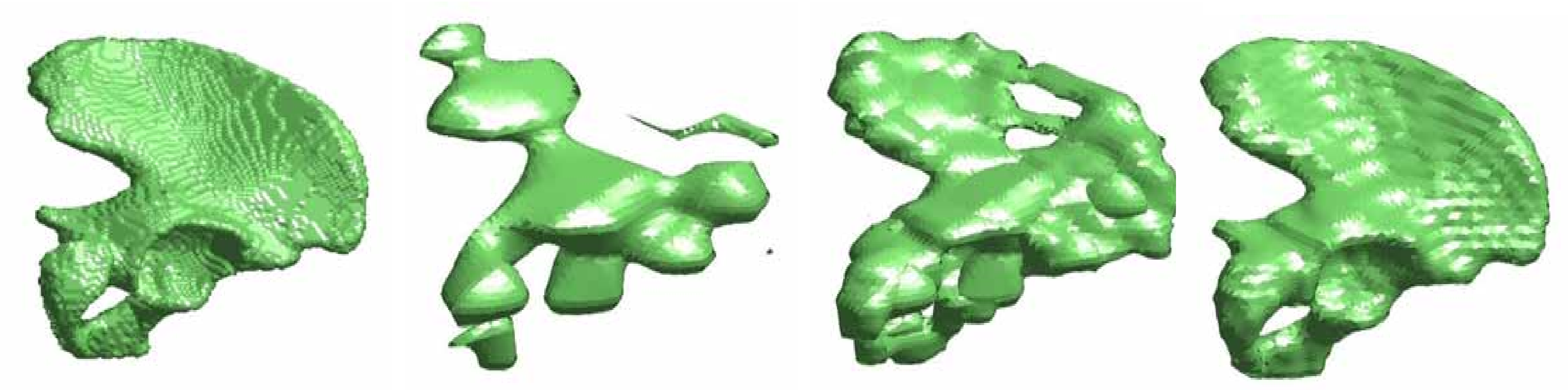
Tikhonov and total variation



3D acoustics

256 cores at PSC

Akcelik et al 02



grid size	material parameters	L-BFGS its	GNK its	PGNK its
65^3	8	16	17 (5)	10 (5)
65^3	27	36	57 (6)	20 (6)
65^3	125	144	131 (7)	33 (6)
65^3	729	156	128 (5)	85 (4)
65^3	4913	*	144 (4)	161 (4)
65^3	35,937	*	177 (4)	159 (6)
65^3	274,625	—	350 (7)	197 (6)
129^3	2,146,689	—	$1470^{\dagger}$ (22)	409 (16)

Structure of the Hessian, parabolic PDE

$$\text{minimize} \quad \frac{1}{2}\int_0^T \int_\Omega (y-d)^2 \, d\Omega \, dt + \frac{\beta}{2}\int_\Omega u^2 \, d\Omega$$

subject to

$$\frac{\partial y}{\partial t} - \nu \Delta y + a(x)y + u(x)1(t) = 0$$

$$y(x,0) = 0, \;\; x \in \Omega, \qquad y(0,t) = y(1,t) = 0.$$

Computing g = Hu

Forward

$$\frac{\partial y}{\partial t} - \nu \Delta y + u(x) 1(t) = 0$$

$$y(x,0) = 0, \; x \in \Omega, \quad y(0,t) = y(1,t) = 0$$

Adjoint

$$-\frac{\partial p}{\partial t} - \nu \Delta p + y = 0$$

$$p(x,T) = 0, \; x \in \Omega, \quad p(0,t) = p(1,t) = 0$$

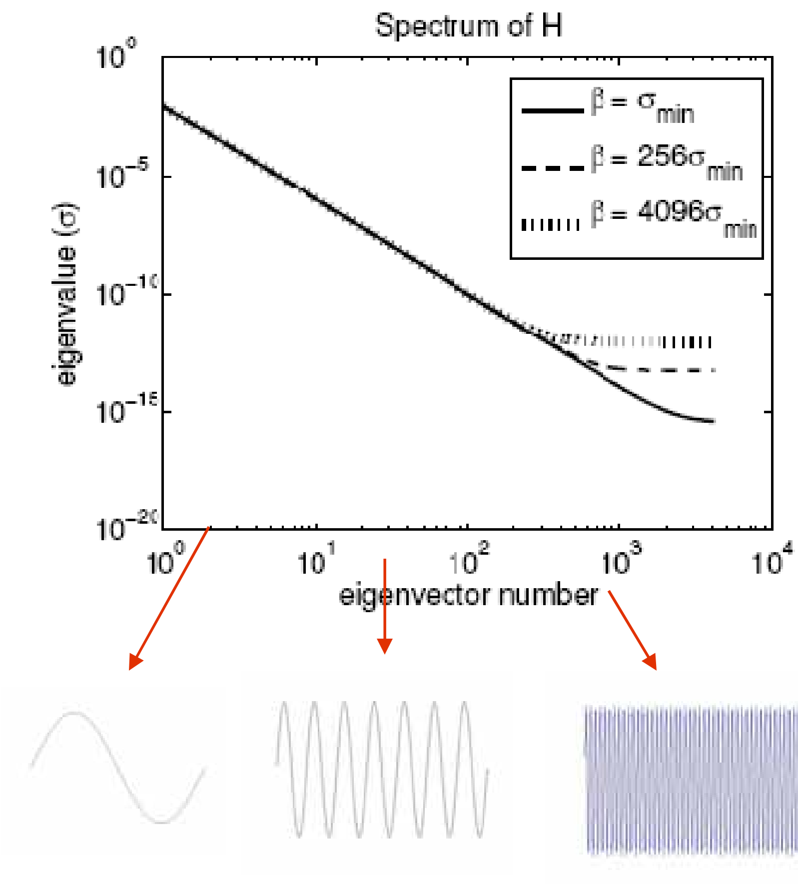
Gradient

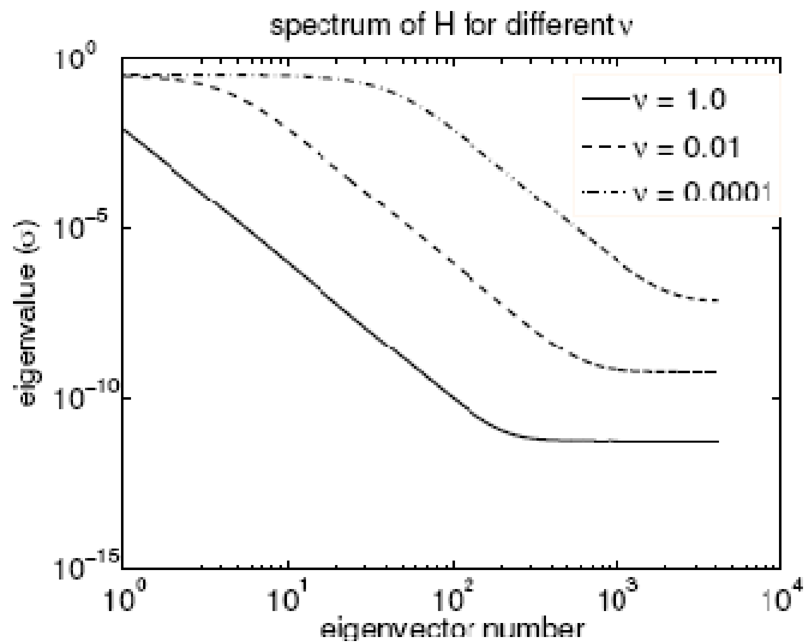
$$g = \beta u + \int_0^T p \, dt$$

Spectral analysis of the Hessian

- Green's function
  $G(x, y; t) := \sum_{k=1}^{\infty} e^{-\lambda_k t}\ 2\sin(k\pi x)\sin(k\pi y)$
- Laplacian
  $\lambda_k = \nu k^2 \pi^2$
- Reduced Hessian operator
  $Hu :=$
  $\beta u(x) + \int_T \int_T \int_\Omega \int_T \int_\Omega 1(t) G(x, y; T-t-\tau) G(y, z; \tau-\sigma) 1(\sigma) u(z)\, d\Omega\, d\sigma\, d\Omega\, d\tau\, dt$
- Reduced Hessian
  $h_k = \beta + \frac{3+e^{-2\lambda_k T} - 4e^{-\lambda_k T} + 2\lambda_k T}{2\lambda_k^3} = \beta + \mathcal{O}(\frac{1}{\lambda_k^2})$

Spectrum





Use CG as a solver?

CG for Hessian

N_s	β $(\sigma > \beta)$		iters	
512	6e-08 (19)	1e-10 (99)	69	725
1024	6e-08 (19)	1e-10 (99)	70	781
2048	6e-08 (19)	1e-10 (99)	68	763
4096	6e-08 (19)	1e-10 (99)	71	713

- Fixed β : CG mesh-independent
- Fixed mesh : CG β-dependent
- β depends on frequency information that we need to recover
  - Truncation noise $\rightarrow \beta \geq h^2$

Difficulties

- For constant coefficients we can construct analytic representation of Hessian
  - Algebraic ill-posedeness
  - For partial-observations we have singular Hessian
- Matrix-free iterations
- 1 forward + 1 adjoint per Hessian matvec
- Hessian ill-conditioned
- Precondition
  - multigrid
  - analytic Hessian

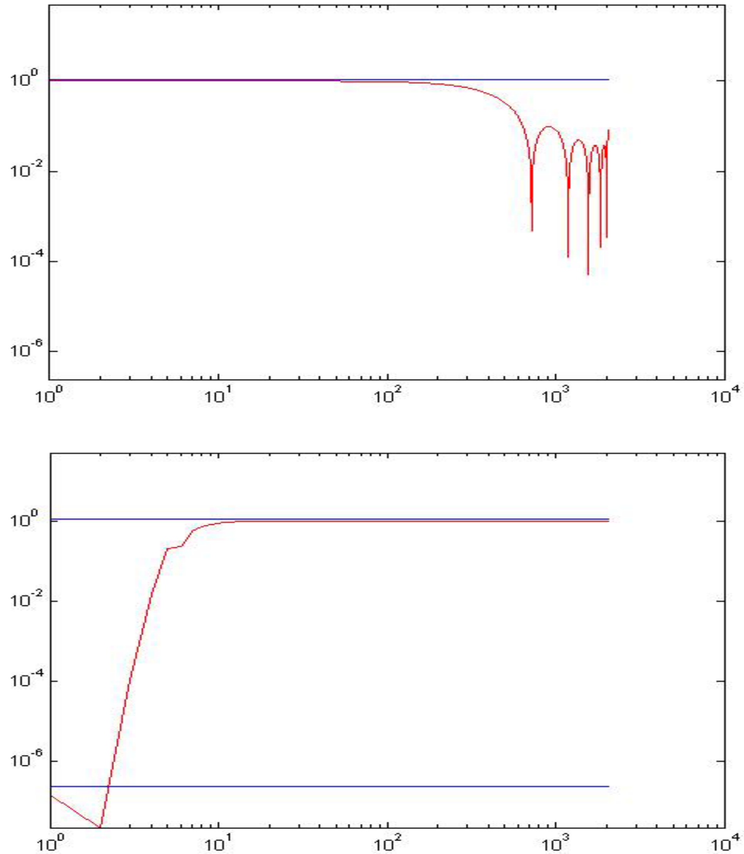
Multigrid

- Multigrid - elliptic PDEs
  - *Brandt, Braess, Bramble, Hackbusch*
- Multigrid – second kind Fredholm
  - *Hackbusch, Hemker & Schippers*
- Multigrid for optimization/inverse problems
  - *Ascher & Haber & Oldenburg, Borzi, Borzi & Kunisch, Borzi & Griesse, Chavent, Dreyer & Maar & Schultz, Draganescu, Hanke & Vogel, Lewis & Nash, Kaltenbacher, King, Kunoth, Ta'asan, Tau & Xu, Vogel, Toint*

Multigrid for Hessians: challenges

- Typically “dense”
- Typically only “MatVec” available
  - Differential
  - I + Compact
  - **Compact**
    - Smoothers
    - Hessian approximation
    - Coarse grid operator (Galerkin vs Non-Galerkin)

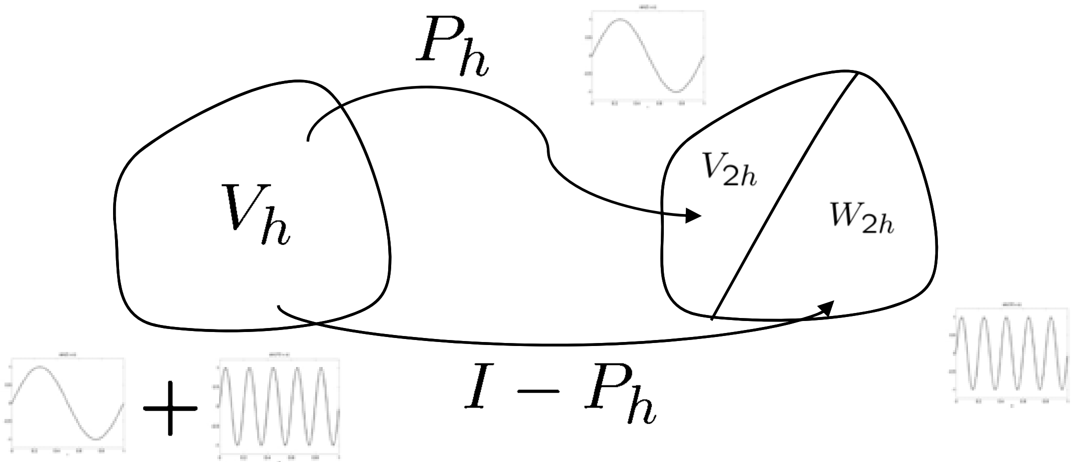
CG as a smoother



- Laplacian
- Hessian

Smoother for MG

- $u = u_s + u_o$ where $u_s \in V_{2h}, u_o \in W_{2h}$



$$(I - P_h + P_h)H^h(I - P_h + P_h)(u) = g$$

Smoothing equation

$$(I - P_h)H^h u_o = (I - P_h)g$$

Coarse-grid equation

$$P_h H^h u_s = P_h g$$

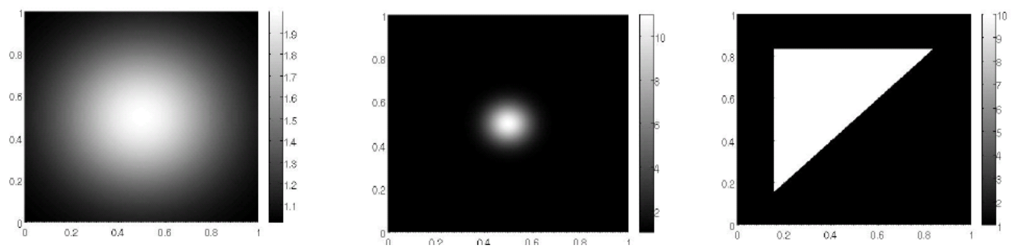
Results Adavani & Biros SISC 08

$\nu = 1.0$

N_s	β	M_{sec}^{-1}	$\tilde{M}_{\text{sec}}^{-1}$	M_{stc}^{-1}	$\tilde{M}_{\text{stc}}^{-1}$
31	5e-07 (31)	10	10	10	10
63	1e-07 (44)	13	13	13	17
127	3e-08 (63)	13	14	14	16
255	7e-09 (89)	13	19	13	16
511	2e-09 (127)	15	18	15	17
1023	5e-10 (180)	15	17	15	17

- **CG precondioned by multigrid**
- V(2,2) cycles
  - mesh β-independent
  - Time coarsening (sec vs stc)

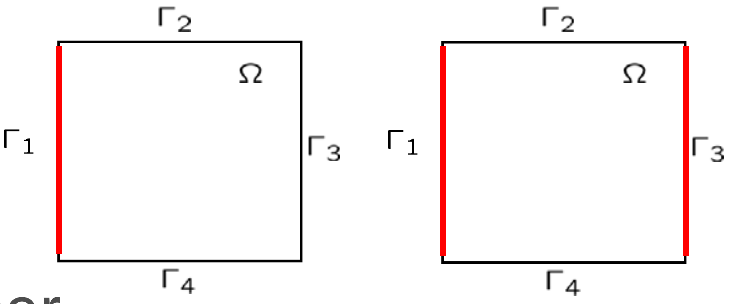
2D, full domain observations



	α_1		α_2		α_3	
$n \times n$	**none**	**preco**	**none**	**preco**	**none**	**preco**
17×17	8	3	8	3	11	6
33×33	15	3	14	4	21	7
65×65	36	4	35	4	80	8
129×129	169	4	136	5	248	12
257×257	639	5	647	6	-	13

Hessian for boundary observations

- Construct Hessian analytically for const coefficients
- Construct inverse

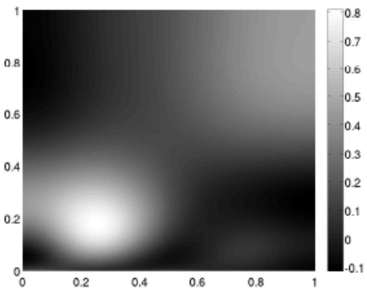


- Not an exact preconditioner
- Use to precondition variable-coefficient case

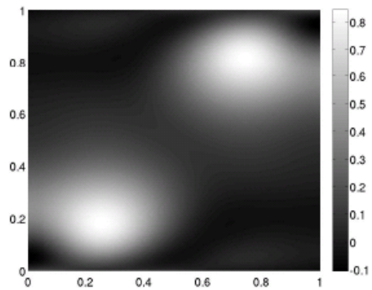
$$H^p(i,j) = \frac{T - E_{ip0}(T) - E_{jp0}(T) + E_{ijp}(T)}{\nu^2(i^2+p^2)(j^2+p^2)\pi^4}$$

$$E_{lmr}(t) = \frac{1 - \exp(-(l^2+m^2+2r^2)\nu\pi^2 t)}{\nu(l^2+m^2+2r^2)\pi^2}$$

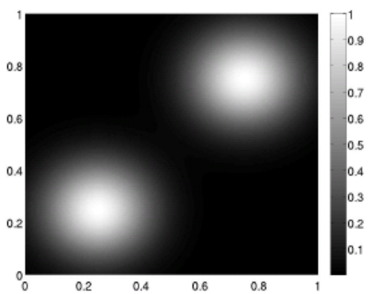
Reconstructions



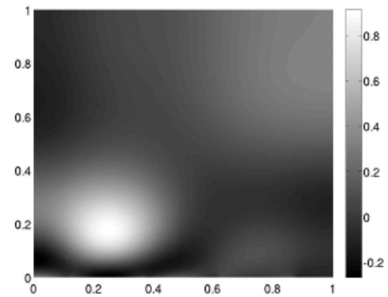
(a) Measurements on the left boundary



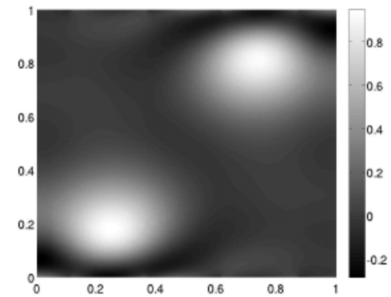
(b) Measurements on the left and right boundaries



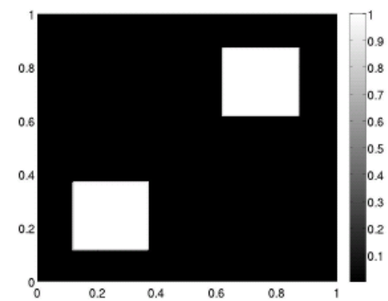
(c) Exact source



(a) Measurements on the left boundary



(b) Measurements on the left and rigl boundaries



(c) Exact source

Scalability

Left boundary data

	u_1				u_2			
	$\beta = 10^{-4}/n^2$		$\beta = 10^{-5}/n^2$		$\beta = 10^{-4}/n^2$		$\beta = 10^{-5}/n^2$	
$n \times n$	**none**	**preco**	**none**	**preco**	**none**	**preco**	**none**	**preco**
17×17	3	3	4	5	5	4	7	6
33×33	3	3	21	5	7	4	33	5
65×65	26	3	-	5	18	4	-	6
129×129	-	4	-	5	-	4	-	6
257×257	-	4	-	5	-	4	-	6

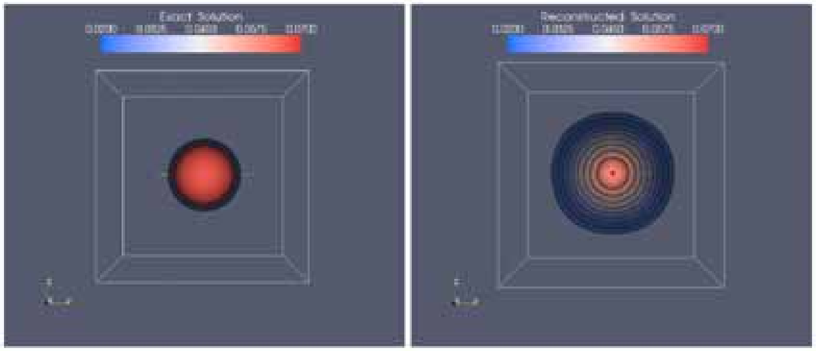
Left and Right boundary data

	u_1				u_2			
	$\beta = 10^{-4}/n^2$		$\beta = 10^{-5}/n^2$		$\beta = 10^{-4}/n^2$		$\beta = 10^{-5}/n^2$	
$n \times n$	**none**	**preco**	**none**	**preco**	**none**	**preco**	**none**	**preco**
17×17	3	3	3	4	5	3	11	4
33×33	3	3	12	4	8	3	10	5
65×65	-	3	77	4	11	3	191	5
129×129	-	3	-	4	-	4	-	4
257×257	-	3	-	5	-	4	-	5

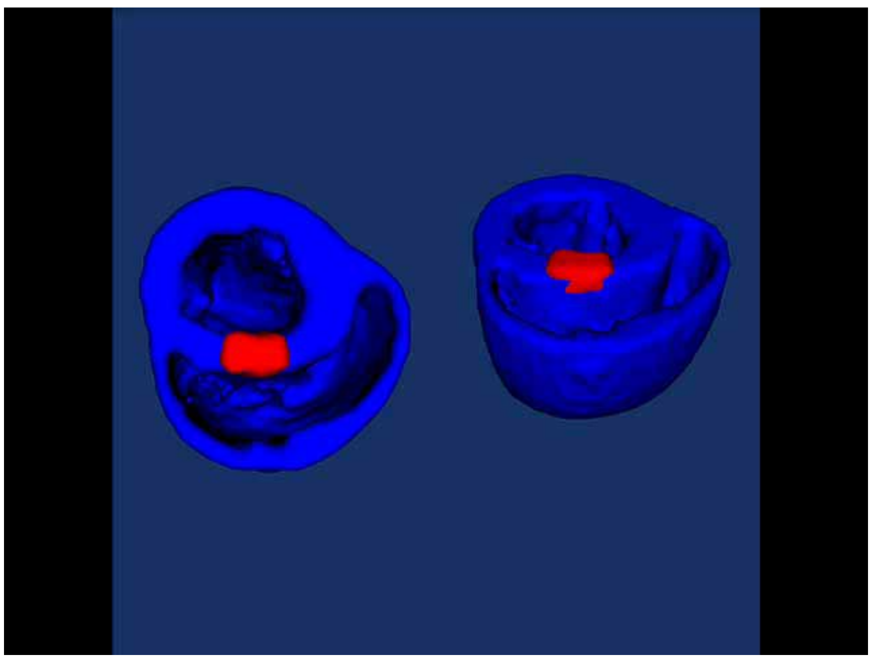
3D parabolic

Adavani & Biros' 08

$n \times n \times n$	**none**	**preco**
$16 \times 16 \times 16$	51	10
$32 \times 32 \times 32$	82	10
$64 \times 64 \times 64$	102	11
$128 \times 128 \times 128$	165	10



Forward: 256^3 x 1024



Summary

- Hessian: Important when we have a lot of data and high-dimensional u
  - Operator, parametrization, observations
- Second derivatives (need for adjoints)
- Regularization
- Case-by-case analysis needed
  - Multigrid
  - Analytic preconditioners
    - Limited on regular geometries, smooth coefficients
  - Orders of magnitude improvement

Georgia Institute of Technology

Not discussed

- Adaptive mesh refinement
- Bayes and probabilistic approaches
- Nonlinear inversion
  - Nonlinear regularization
- Data assimilation
- Parallel scalability
- Model error