Fast algorithms for inverse problems governed by PDEs

George Biros

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

• Inverse problems with PDEs
• The input-output map
• Ill-posedness and regularization
• Algorithmic goals
• The Hessian
  o Structure (equilibrium, evolution)
  o 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
  - Gravity, electromagnetics, porous media, seismic

- References for inverse problems theory
  - Tarantola, Kaipio & Somersalo, Vogel
General problem statement

- Model encapsulates conservation and constitutive laws

\[
\frac{dy}{dt} = m(y, u, t) + n(y, t), \quad y(0) = y_0
\]

\[
d = Q(y)
\]

\[
y = M(u)
\]
Linear case: input-output operator and the Hessian

- Forward operator: $y = Mu$
- Observation operator: $d = Qy$
- Input-output operator: $d = Ku$
- Hessian: $H = K^*K$
- Adjoint: $K^* = M^*Q^*$
Inverse problem, Linear case

• Inverse problem
  \[
  \min_u S(Ku, d) + \mathcal{R}(u)
  \]

• Least squares
  \[
  \min_u \frac{1}{2} \| Ku - d \|_2^2 + \frac{1}{2} \| u \|_R^2
  \]
  o Linear system for \( u \)
  \[
  (R + K^* K)u = K^* d
  \]

• How do you choose \( R \)?

• Scalable algorithms
  o 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
  - Constrain w/deviation with prior models
  - Penalize w/total variation

\[
\begin{align*}
\min_{u,y} \frac{1}{2} \sum_{j=1}^{N} & \int_{0}^{T} (d_j - y(x_j))^2 \, dt + \frac{\beta}{2} \int_{\Omega} |\nabla u|^2 \, d\Omega \\
\text{subject to} \quad & \frac{\partial^2 y}{\partial t^2} - \nabla \cdot u \nabla y = f \quad \text{in} \quad \Omega \times [0,T] + B.C., \ I.C.
\end{align*}
\]
2D acoustics  Akcelik et al 02

No regularization

Tikhonov and total variation

Tikhonov regularization

Total variation
3D acoustics
256 cores at PSC
Akcelik et al 02

<table>
<thead>
<tr>
<th>grid size</th>
<th>material parameters</th>
<th>L-BFGS its</th>
<th>GNK its</th>
<th>PGNK its</th>
</tr>
</thead>
<tbody>
<tr>
<td>$65^3$</td>
<td>8</td>
<td>16</td>
<td>17 (5)</td>
<td>10 (5)</td>
</tr>
<tr>
<td>$65^3$</td>
<td>27</td>
<td>36</td>
<td>57 (6)</td>
<td>20 (6)</td>
</tr>
<tr>
<td>$65^3$</td>
<td>125</td>
<td>144</td>
<td>131 (7)</td>
<td>33 (6)</td>
</tr>
<tr>
<td>$65^3$</td>
<td>729</td>
<td>156</td>
<td>128 (5)</td>
<td>85 (4)</td>
</tr>
<tr>
<td>$65^3$</td>
<td>4913</td>
<td>*</td>
<td>144 (4)</td>
<td>161 (4)</td>
</tr>
<tr>
<td>$65^3$</td>
<td>35, 937</td>
<td>*</td>
<td>177 (4)</td>
<td>159 (6)</td>
</tr>
<tr>
<td>$65^3$</td>
<td>274, 625</td>
<td>—</td>
<td>350 (7)</td>
<td>197 (6)</td>
</tr>
<tr>
<td>$129^3$</td>
<td>2, 146, 689</td>
<td>—</td>
<td>1470† (22)</td>
<td>409 (16)</td>
</tr>
</tbody>
</table>
Structure of the Hessian, parabolic PDE

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

subject to

\[ \frac{\partial y}{\partial t} - \nu \Delta y + a(x)y + u(x)1(t) = 0 \]
\[ y(x, 0) = 0, \; x \in \Omega, \quad y(0, t) = y(1, t) = 0. \]
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
  \[ H u := \beta u(x) + \int_T \int_T \int_\Omega \int_\Omega 1(t)G(x, y; T-t-\tau)G(y, z; \tau-\sigma)1(\sigma)u(z) \ d\Omega \ d\sigma \ d\Omega \ d\tau \ d\tau \]

- Reduced Hessian
  \[ h_k = \beta + \frac{3e^{-2\lambda_k^T} - 4e^{-\lambda_k^T} + 2\lambda_k^T}{2\lambda_k^3} = \beta + O\left(\frac{1}{\lambda_k^2}\right) \]
Use CG as a solver?
CG for Hessian

<table>
<thead>
<tr>
<th>$N_s$</th>
<th>$\beta$ ($\sigma &gt; \beta$)</th>
<th>iters</th>
</tr>
</thead>
<tbody>
<tr>
<td>512</td>
<td>6e-08 (19)</td>
<td>69</td>
</tr>
<tr>
<td>1024</td>
<td>6e-08 (19)</td>
<td>70</td>
</tr>
<tr>
<td>2048</td>
<td>6e-08 (19)</td>
<td>68</td>
</tr>
<tr>
<td>4096</td>
<td>6e-08 (19)</td>
<td>71</td>
</tr>
</tbody>
</table>

- Fixed $\beta$ : CG mesh-independent
- Fixed mesh : CG $\beta$-dependent
- $\beta$ 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
  o Algebraic ill-posedeness
  o 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
  o Differential
  o I + Compact
  o **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

\[
P_h H^h u_s = P_h g
\]
CG preconditioned by multigrid

V(2,2) cycles
  - mesh $\beta$-independent
  - Time coarsening (sec vs stc)
2D, full domain observations

<table>
<thead>
<tr>
<th>$n \times n$</th>
<th>$\alpha_1$</th>
<th>$\alpha_2$</th>
<th>$\alpha_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>none</td>
<td>preco</td>
<td>none</td>
</tr>
<tr>
<td>17 $\times$ 17</td>
<td>8</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>33 $\times$ 33</td>
<td>15</td>
<td>3</td>
<td>14</td>
</tr>
<tr>
<td>65 $\times$ 65</td>
<td>36</td>
<td>4</td>
<td>35</td>
</tr>
<tr>
<td>129 $\times$ 129</td>
<td>169</td>
<td>4</td>
<td>136</td>
</tr>
<tr>
<td>257 $\times$ 257</td>
<td>639</td>
<td>5</td>
<td>647</td>
</tr>
</tbody>
</table>
Hessian for boundary observations

- Construct Hessian analytically for const coefficients
- Construct inverse

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

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

\[ E_{imp}(t) = \frac{1 - \exp(-(l^2 + m^2 + 2r^2)\nu^2 t)}{\nu(l^2 + m^2 + 2r^2)\pi^2} \]
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 right boundaries

(c) Exact source
Scalability

<table>
<thead>
<tr>
<th>$n \times n$</th>
<th>$\beta = 10^{-4}/n^2$</th>
<th>$\beta = 10^{-5}/n^2$</th>
<th>$\beta = 10^{-4}/n^2$</th>
<th>$\beta = 10^{-5}/n^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$u_1$</td>
<td>$u_1$</td>
<td>$u_2$</td>
<td>$u_2$</td>
</tr>
<tr>
<td></td>
<td>none</td>
<td>preco</td>
<td>none</td>
<td>preco</td>
</tr>
<tr>
<td>17 $\times$ 17</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>33 $\times$ 33</td>
<td>3</td>
<td>3</td>
<td>21</td>
<td>5</td>
</tr>
<tr>
<td>65 $\times$ 65</td>
<td>26</td>
<td>3</td>
<td>-</td>
<td>5</td>
</tr>
<tr>
<td>129 $\times$ 129</td>
<td>-</td>
<td>4</td>
<td>-</td>
<td>5</td>
</tr>
<tr>
<td>257 $\times$ 257</td>
<td>-</td>
<td>4</td>
<td>-</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n \times n$</th>
<th>$\beta = 10^{-4}/n^2$</th>
<th>$\beta = 10^{-5}/n^2$</th>
<th>$\beta = 10^{-4}/n^2$</th>
<th>$\beta = 10^{-5}/n^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$u_1$</td>
<td>$u_1$</td>
<td>$u_2$</td>
<td>$u_2$</td>
</tr>
<tr>
<td></td>
<td>none</td>
<td>preco</td>
<td>none</td>
<td>preco</td>
</tr>
<tr>
<td>17 $\times$ 17</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>33 $\times$ 33</td>
<td>3</td>
<td>3</td>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td>65 $\times$ 65</td>
<td>-</td>
<td>3</td>
<td>77</td>
<td>4</td>
</tr>
<tr>
<td>129 $\times$ 129</td>
<td>-</td>
<td>3</td>
<td>-</td>
<td>4</td>
</tr>
<tr>
<td>257 $\times$ 257</td>
<td>-</td>
<td>3</td>
<td>-</td>
<td>5</td>
</tr>
</tbody>
</table>
**3D parabolic**  Adavani & Biros’ 08

<table>
<thead>
<tr>
<th>$n \times n \times n$</th>
<th>none</th>
<th>preco</th>
</tr>
</thead>
<tbody>
<tr>
<td>$16 \times 16 \times 16$</td>
<td>51</td>
<td>10</td>
</tr>
<tr>
<td>$32 \times 32 \times 32$</td>
<td>82</td>
<td>10</td>
</tr>
<tr>
<td>$64 \times 64 \times 64$</td>
<td>102</td>
<td>11</td>
</tr>
<tr>
<td>$128 \times 128 \times 128$</td>
<td>165</td>
<td>10</td>
</tr>
</tbody>
</table>

Forward: $256^3 \times 1024$
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
Not discussed

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