Fast algorithms for Hyperspectral Diffuse Optical Tomography

Arvind K. Saibaba\(^1\)
Joint work with Misha E. Kilmer\(^2\), Eric L. Miller\(^1\)

\(^1\)Department of Electrical and Computer Engineering
\(^2\)Department of Mathematics
Tufts University

April 8, 2014
Diffuse Optical Tomography

Image courtesy of Boas, et al., “Imaging the body with Diffuse Optical Tomography” IEEE Signal Processing magazine vol. 18 issue 6
Modeling assumptions

The photon fluence satisfies

\[-\nabla D(r, \lambda) \cdot \nabla \phi(r, \lambda) + \nu \mu_a(r, \lambda) \phi(r, \lambda) = S(r, \lambda) \quad r \in \Omega\]

\[\phi = 0 \quad r \in \partial \Omega_D\]

\[\phi(r, \lambda) + 2\eta \frac{\partial \phi(r, \lambda)}{\partial n} = 0 \quad r \in \partial \Omega_R\]

Decompose absorption coefficient

\[\mu_a(r, \lambda) = \mu(\lambda) + \Delta \mu(r, \lambda)\]

\(\mu(\lambda)\) Background \quad \Delta \mu(r, \lambda) Perturbation

Split \(\phi\) into an incident field \(\phi_i\) and scattered field \(\phi_s\)

\[-\nabla^2 \phi_i + \frac{\nu \mu(\lambda)}{D(\lambda)} \phi_i = \frac{S_0(\lambda)}{D(\lambda)} \delta(r - r_s)\]

\[-\nabla^2 \phi_s + \frac{\nu \mu(\lambda)}{D(\lambda)} \phi_s = -\frac{1}{D(\lambda)} \Delta \mu(r, \lambda)(\phi_s + \phi_i)\]
An approximation is Born

Born approximation: if $|\Delta \mu(x, \lambda)| \ll \mu(\lambda)$

$$-\nabla^2 \phi_s + \frac{\nu \mu(\lambda)}{D(\lambda)} \phi_s = -\frac{1}{D(\lambda)} \Delta \mu(\mathbf{r}, \lambda)(\phi_s + \phi_i)$$

Formally, we can write measurements as the Lippmann-Schwinger integral

$$\phi(\mathbf{r}_d; \lambda) \approx \int_{\Omega} G(\mathbf{r}_d, \mathbf{r}'; \lambda) \Delta \mu_a(\mathbf{r}; \lambda) G(\mathbf{r}', \mathbf{r}_s; \lambda) d\mathbf{r}'$$

We choose to retain the PDE formulation

- Allows flexibility in modeling complex geometries
- Allows for nonhomogenous background diffusion and absorption operators.
- Linearizes the parameter-to-observation mapping
- We can develop fast solvers for multiple wavelengths $(K + \sigma_j M)x_j = b$ for $j = 1, \ldots, N_{\lambda}$. 
An approximation is Born

Born approximation: if $|\Delta \mu(x, \lambda)| \ll \mu(\lambda)$

$$-\nabla^2 \phi_s + \frac{\nu \mu(\lambda)}{D(\lambda)} \phi_s \approx -\frac{1}{D(\lambda)} \Delta \mu(r, \lambda) \phi_i$$

Formally, we can write measurements as the Lippmann-Schwinger integral

$$\phi(r_d; \lambda) \approx \int_{\Omega} G(r_d, r'; \lambda) \Delta \mu_a(r; \lambda) G(r', r_s; \lambda) dr'$$

We choose to retain the PDE formulation

- Allows flexibility in modeling complex geometries
- Allows for nonhomogenous background diffusion and absorption operators.
- Linearizes the parameter-to-observation mapping
- We can develop fast solvers for multiple wavelengths $(K + \sigma_j M)x_j = b$ for $j = 1, \ldots, N_\lambda$. 
Computational challenges

Typical problem specifications

<table>
<thead>
<tr>
<th>Number of</th>
<th>Symbol</th>
<th>Typical number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sources</td>
<td>$N_s$</td>
<td>$10 - 100$</td>
</tr>
<tr>
<td>Detectors / source</td>
<td>$N_{ds}$</td>
<td>$5 - 10$</td>
</tr>
<tr>
<td>Wavelengths</td>
<td>$N_\lambda$</td>
<td>$10 - 200$</td>
</tr>
<tr>
<td>Grid size</td>
<td>$N$</td>
<td>$32^3 - 100^3$</td>
</tr>
</tbody>
</table>

Assuming the true field is known, to reproduce measurements costs

$$\text{Cost: } N_s (1 + N_{ds}) N_\lambda O(N)$$
Computational challenges

Typical problem specifications

<table>
<thead>
<tr>
<th>Number of</th>
<th>Symbol</th>
<th>Typical number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sources</td>
<td>(N_s)</td>
<td>10 – 100</td>
</tr>
<tr>
<td>Detectors / source</td>
<td>(N_{ds})</td>
<td>5 – 10</td>
</tr>
<tr>
<td>Wavelengths</td>
<td>(N_\lambda)</td>
<td>10 – 200</td>
</tr>
<tr>
<td>Grid size</td>
<td>(N)</td>
<td>(32^3 – 100^3)</td>
</tr>
</tbody>
</table>

Assuming the true field is known, to reproduce measurements costs

\[
\text{Cost: } \quad N_s (1 + N_{ds}) N_\lambda \mathcal{O}(N)
\]

In other words, have to solve 200,000 PDE’s and store 200 GB in memory!
Outline

1. The Forward Problem

2. The Inverse Problem

3. Conclusions and future work
Krylov subspace methods for shifted systems

Previous works include

1. Symmetric Lanczos based methods
   - Conjugate Gradient - Freund 1990
   - Meerbergen 2003, Meerbergen and Bai 2010

2. Unsymmetric Lanczos based methods
   - Quasi Minimum Residual (QMR), TFQMR - Freund 1993
   - BQMR - Freund and Malhotra 1997

3. Arnoldi based methods
   - Restarted Generalized Minimum RESidual (GMRES) - Frommer and Glassner 1996
   - Restarted Full Orthogonalization Method (FOM) - Simoncini 2003
   - Flexible Arnoldi - Gu et al. 2007
   - GMRES-DR Darnell, Morgan and Wilcox 2008

and several other block variants for multiple right hand sides.
Shift invariance

Krylov subspaces are shift invariant

\[ \mathcal{K}_m(A + \sigma I, b) = \mathcal{K}_m(A, b) \]
Shift invariance

Krylov subspaces are shift invariant

\[ \mathcal{K}_m(A + \sigma I, b) = \mathcal{K}_m(A, b) \]

To see this

\[ AV_m = V_{m+1} \bar{H}_m \]
\[ IV_m = V_m \]

\[ (A + \sigma I)V_m = V_{m+1} \left( \begin{bmatrix} I \\ 0 \end{bmatrix} + \sigma \bar{H}_m \right) \]
\[ \overset{\text{def}}{=} \bar{H}_m(\sigma) \]
Shift invariance

Krylov subspaces are shift invariant

\[ \mathcal{K}_m(A + \sigma I, b) = \mathcal{K}_m(A, b) = \text{Span}\{V_m\} \]

In summary,

\[ (A + \sigma I)V_m = V_{m+1}\bar{H}_m(\sigma) \]
Shift invariance

Krylov subspaces are shift invariant

\[ \mathcal{K}_m(A + \sigma I, b) = \mathcal{K}_m(A, b) = \text{Span}\{V_m\} \]

In summary,

\[ (A + \sigma I)V_m = V_{m+1}\bar{H}_m(\sigma) \]

Our general strategy is the following,

- Build a shift-invariant Krylov basis \( V_m \)
- Solve the smaller-dimensional subproblem for \( y_m(\sigma) \)
- The approximate solution is \( x_m(\sigma) = V_m y_m(\sigma) \)
Preconditioners

Practical use of shift-invariant method is challenging because

- An initial guess has to satisfy collinearity property.
- Preconditioners need to work across entire range of shifts.

Shift-and-invert preconditioners

- Use preconditioners of the form \((K + \tau M)^{-1}\)

\[
(K + \tau M)^{-1}(K + \sigma M) = I + (\sigma - \tau)(K + \tau M)^{-1}M \\
\text{def} = A
\]

and the shift invariance \(K_m\{A, b\} = K_m\{I + (\sigma - \tau)A, b\}\)

- Rule of thumb - systems with \(\sigma\) near \(\tau\) converge first.
- Inexactness issues due to use of iterative solver to apply \(A \triangleq (K + \tau M)^{-1}M\)
Speedup for DOT application

- System size $35 \times 35 \times 32$ with 200 shifts
- AMG solver used to apply $(K + \tau M)^{-1} M$
- Stopping tolerance used for all systems $\|r_k(\sigma_j)\|_2/\|r_0\|_2 < 10^{-10}$
Block Krylov methods

- Most iterative methods are generalized easily to block methods,
- But the stability of block methods requires extra effort.
- Block methods may be, but need not be much faster than solving the $N_s$ systems separately
Block Krylov methods

- Most iterative methods are generalized easily to block methods,
- But the stability of block methods requires extra effort.
- Block methods may be, but need not be much faster than solving the $N_s$ systems separately

Reasons for using Block Krylov methods

- Search space is bigger, since using information from different rhs
- Sometimes, block matrix-vector products can be computed at once

Search for solutions in the space

$$\mathcal{K}_m(A, B) \overset{\text{def}}{=} \text{Span}\{B, AB, \ldots, A^{m-1}B\}$$
Results: Multiple Shifts Multiple RHS

- System size $35 \times 35 \times 32$ with 200 shifts and 16 rhs = 3200 systems
- AMG solver used to apply $(K + \tau M)^{-1} M$
- Stopping tolerance used for all systems $\|r_k(\sigma_j)\|_2/\|r_0\|_2 < 10^{-10}$
Breakup of computational costs

- Computational cost is dominated by the application of $(K + \tau M)^{-1} M$.
- Reduction in max. number of iterations reduces runtime costs.

<table>
<thead>
<tr>
<th>Time</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_1$</td>
<td>Matvec with $(K + \tau M)^{-1} M$</td>
</tr>
<tr>
<td>$t_2$</td>
<td>Arnoldi orthogonalization</td>
</tr>
<tr>
<td>$t_3$</td>
<td>Projected shifted systems</td>
</tr>
<tr>
<td>$t_4$</td>
<td>Checking convergence</td>
</tr>
<tr>
<td>$t_5$</td>
<td>Computing approx. soln.</td>
</tr>
</tbody>
</table>

Reduction in max. number of iterations reduces runtime costs.
Overall speedup

- System size $35 \times 35 \times 32$ with 200 shifts and 16 rhs = 3200 systems
- AMG solver used to apply $(K + \tau M)^{-1} M$
- Stopping tolerance used for all systems $\|r_k(\sigma_j)\|_2/\|r_0\|_2 < 10^{-10}$
Outline

1. The Forward Problem
2. The Inverse Problem
3. Conclusions and future work
Compression of measurement operator

The measurement operator is

\[ y = H\mu + \eta \quad \eta \sim \mathcal{N}(0, W^{-2}) \]

- Measurement operator is large, dense.
- High levels of compression is possible across wavelengths.
Compression of measurement operator

The measurement operator is

\[ y = H\mu + \eta \quad \eta \sim \mathcal{N}(0, W^{-2}) \]

- Measurement operator is large, dense.
- High levels of compression is possible across wavelengths.

Consider the measurement operator from 2 different sources.

\[ \begin{align*}
N_{ds} N_\lambda & \quad H_1 \\
N_{ds} N_\lambda & \quad H_2
\end{align*} \]
Compression of measurement operator

The measurement operator is

\[ y = H\mu + \eta \quad \eta \sim \mathcal{N}(0, W^{-2}) \]

- Measurement operator is large, dense.
- High levels of compression is possible across wavelengths.

\[
\begin{align*}
U_1 & \quad \Sigma_1 V_1^* \\
U_2 & \quad \Sigma_2 V_2^*
\end{align*}
\]
Compression of measurement operator

The measurement operator is

\[ y = H\mu + \eta \quad \eta \sim \mathcal{N}(0, W^{-2}) \]

- Measurement operator is large, dense.
- High levels of compression is possible across wavelengths.
Compression of measurement operator

The measurement operator is

\[ y = H\mu + \eta \quad \eta \sim \mathcal{N}(0, W^{-2}) \]

- Measurement operator is large, dense.
- High levels of compression is possible across wavelengths.
Compression of measurement operator

The measurement operator is

\[ y = H\mu + \eta \quad \eta \sim \mathcal{N}(0, W^{-2}) \]

- Measurement operator is large, dense.
- High levels of compression is possible across wavelengths.
Compression of measurement operator

The measurement operator is

\[ y = H\mu + \eta \quad \eta \sim \mathcal{N}(0, W^{-2}) \]

- Measurement operator is large, dense.
- High levels of compression is possible across wavelengths.

Low-rank factorization can be done using
- Singular Value Decomposition
- Rank revealing factorizations - Interpolative Decomposition, Randomized SVD.
Recursive Compression

- \( N = 35 \times 35 \times 32 \)
- \( N_s = 16 \) and \( N_{ds} = 9 \) - Total number of systems = 12960.
- **Accuracy**: \( \| H - U \Sigma V^* \|_2 \leq (\log_2 N_s + 1)\sigma_{R^\lambda + 1}(H_i) + \sigma_{R^\lambda + 1}(H_\epsilon) \)
Shape based approach

Model unknowns as piecewise constant

\[ f(x) = f_i(x) \chi_D(x) + f_o(x)(1 - \chi_D(x)) \]

\[ \chi_D(x) = \begin{cases} 
1 & x \in D \\
0 & x \in \Omega \end{cases} \]

Level-set

\[ \Delta \mu(r) = c_p H(\phi(r) - \tau) + c_b (1 - H(\phi(r) - \tau)) \]

\[ H(x) = \frac{1}{2} (1 + \text{sign}(x)) \]

Figure 1: Image courtesy of Wikipedia

- Topologically flexible - able to recover multiple connected components
- Evolve the shape by the minimizing an objective function.
Parametric level-sets (PaLS)\(^3\)

Parametric representation of surface

\[
\phi(x; p) = \sum_{k=0}^{n_p} \alpha_k \psi \left( \left\| \beta_k (x - \chi^{(k)}) \right\|^{\dagger} \right)
\]

where,

\[
\begin{align*}
\alpha_k & := \text{expansion coefficients} \\
\beta_k & := \text{dilation factors} \\
\chi^{(k)} & := \text{centers} \\
\left\| x \right\|^{\dagger} & := \sqrt{\left\| x \right\|_2^2 + \nu^2}
\end{align*}
\]

PaLS parameters

\[
p = [\alpha, \beta, \chi_x, \chi_y, \chi_z]^T
\]

\(^3\)A. Aghasi, M. Kilmer and E. Miller, Parametric Level Set Methods for Inverse Problems, SIAM J. Imaging Science 2012
Trust region approach

Optimization problem

\[
\min_p \quad \frac{1}{2} \|r\|_2^2 = \frac{1}{2} \| W (y - H\mu(p)) \|_2^2
\]

Solution update \( p_+ = p + s \),

\[
\min_s \quad r^T Js + \frac{1}{2} s^T J^T Js \quad \text{s.t.} \quad \|s\|_2 \leq \Delta
\]

\[
\begin{align*}
r &= W (y - H\mu(p)) \\
J &= -WH\nabla p\mu(p)
\end{align*}
\]

- Radius of the trust region controlled by improvement in objective function
- Stopping tolerance based on discrepancy principle.
Spherical inclusion

- 40k unknowns, 900 measurements
- 1% additive gaussian noise
- 22.1% relative $L^2$ error
Reconstructed model

- 40k unknowns, 900 measurements
- 1% additive gaussian noise
- 35.6% relative $L^2$ error
Outline

1. The Forward Problem

2. The Inverse Problem

3. Conclusions and future work
Contributions

Solution of Born approximation can be significantly accelerated.

Large, dense measurement operator can be efficiently compressed with little loss of accuracy.

Showed reconstruction of chromophore shapes using different methods.

Future Work

Compare reconstruction with experimental setup.

Relaxing the Born approximation - fully nonlinear inverse problem.

Combining mesh adaptivity with Level Set framework.