Subspace Recycling in Diffuse Optical Tomography

Misha E. Kilmer[1], Eric de Sturler[2]

misha.kilmer@tufts.edu

[1] Dept. of Mathematics, Tufts University, Medford, MA

[2] Dept. of Mathematics, Virginia Tech, Blacksburg, VA

Overview

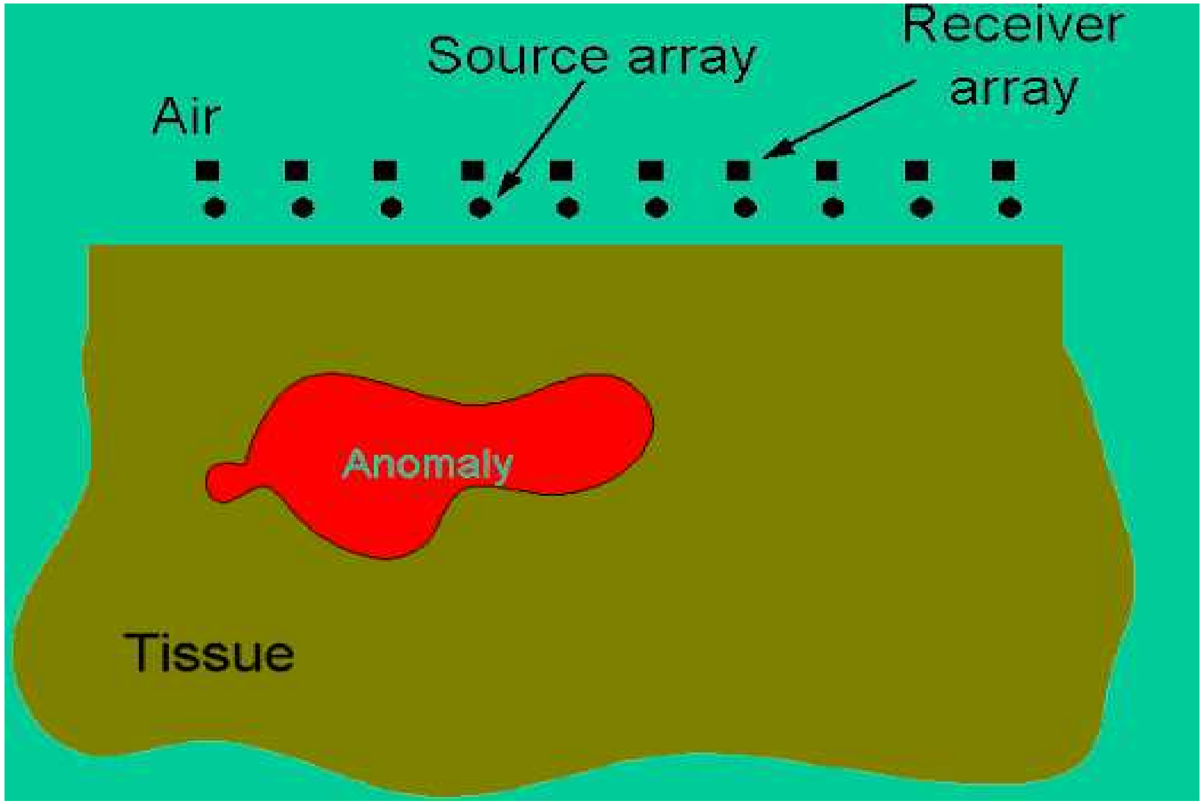
- The DOT forward model
- Parametric inversion for diffuse optical tomography
- Subspace Recycling
- Specifics that can be exploited for recycling
- Theory
- Algorithm tuned to parametric inversion
- Numerical results
- Conclusions and future work

Background

- tissue illuminated by near-infrared, frequency modulated light
- light detected in array(s)
- model of physics used to infer optical properties of (breast) tissue
- tumors, different optical properties than surrounding tissue $\Rightarrow$ 3D images of optical properties show anomalies.

Geometry

- Box
- Sources top, detectors on top and bottom
- Limited data, large number voxels



Diffusion Forward Model

Photon fluence/flux $\phi_s(r)$ given input $f_s(r)$:

$$-\nabla \cdot D(r)\nabla\phi_{s,\omega}(r) + \mu_a(r)\phi_{s,\omega}(r) + i\frac{\omega}{\nu}\phi_{s,\omega}(r) = f_s(r),$$

for r in Ω

$$.25\phi_{s,\omega}(r) + \frac{D(r)}{2}\frac{\partial\phi_{s,\omega}(r)}{\partial\eta} = 0 \text{ top,bottom}$$

$$\phi_{s,\omega}(r) = 0 \text{ on sides.}$$

Discrete Forward Model

$$h_{\omega_i,s_j}(c) \approx \tilde{d}_{\omega_i,s_j} + \eta = d_{\omega_i,s_j},$$

where

- c denotes quantity ('image') of interest (e.g. light diffusion coefficient) evaluated at the grid points in 3D
- η is additive noise
- d_{ω_i,s_j} data due to source s_j at frequency ω_i
- $h_{\omega_i,s_j}(c)$ is computed estimate of measured data given c
    - * Requires numerical soln. of the PDE
    - * Algorithm solving for c very expensive!!

The Inverse Problem

Concatenating data over all sources and frequencies,

$$d = \begin{bmatrix} \text{real}(d_{\omega_1,s(:)}) \\ \text{imag}(d_{\omega_1,s(:)}) \\ \vdots \\ \text{real}(d_{\omega_k,s(:)}) \\ \text{imag}(d_{\omega_k,s(:)}) \end{bmatrix},$$

with a similar definition for $h(c)$.

Inverse problem:

$$\min_c \|W(h(c) - d)\|_2 + \lambda^2 \|\Omega(c)\|$$

The Parametric Inverse Problem

Of special interest:

- c is defined by a (relatively) small set of parameters.
- image is characterized by anomalous regions (jumps).

Simple Example:

$$c(i,j,k) = \alpha S(i,j,k) + \beta(1 - S(i,j,k))$$

where

$$S(i,j,k) \approx \begin{cases} 1 & \text{if } (i,j,k) \in \text{anomaly} \\ 0 & \text{otherwise} \end{cases}$$

Parametric Inverse Problem

Key: A small set of parameters defines the boundaries

Two possibilities in 3D DOT:

- Ellipsoids [K., et al, 2003]
  unknowns are axis lengths, center, rotation angles
- Parametric level sets
  $S(r)$ is (approx) 1 inside the region bounded by the zero level set of a function described by a small number of parameters.

Optimization

$$\min_p \|W\left(h(c(p)) - d\right)\|_2$$

We will be solving using a damped Gauss-Newton method (ideas still valid for other approaches)

Given initial p_0
For $k = 0, ...$ to convergence

- $\epsilon^{(k)} = W(h(p^{(k)}) - d)$
- Solve $(J^{(k)})^T J^{(k)} s = -(J^{(k)})^T \epsilon^{(k)}$
- $p^{(k+1)} = p^{(k)} + \xi s$ where ξ is chosen using a linesearch.

Parametric Inverse Problem

$$\min_p \|W\left(h(c(p)) - d\right)\|_2$$

Difficulty is that each optimization step requires evaluation of $h(c(p^{(k)}))$ and construction of $J^{(k)}$. Linesearch also requires function evals.

- One solve of the discretized PDE for
  - * each source
  - * each frequency
- One “adjoint” (actually, transpose) solve to fill the Jacobian
  - * each detector
  - * each frequency

Summary

After “eliminating” boundary points, the systems to solve are:

$$(A^{(k)} + i\gamma I)x_j^{(k)} = b_j, \qquad j = 1 \ldots \text{num srcs}$$

$$(A^{(k)} + i\gamma I)y_j^{(k)} = m_j, \qquad j = 1 \ldots \text{num dets}$$

where b_j and m_j contain (different) columns of the identity matrix and $A^{(k)}$ is SPD.

Krylov Subspaces and MINRES

$$K_m(A,b) = \text{span}\{b, Ab, A^2b, \ldots, A^{m-1}b\}.$$

Approximate soln. x to $Ax = b$ by solving (MINRES)

$$\min_{x \in K_m(A,b)} \|b - Ax\|_2.$$

- Hope that $\|b - Ax_m\|_2 / \|b\|$ is small for m small.
- For A real, symmetric, short-term recurrences to compute solution.
- Major cost per step is matrix-vector product with A.
- Total cost roughly $O(mM)$ flops, where it costs M flops to do a mat-vec.

Key Observations

- Our mat-vecs are cheap, since matrices are sparse.
- We must keep the number of iterations small.
- It is well known that this will be small if the spectrum of the matrix is clustered around 1, and small eigenvalues close to zero slow convergence.
- We can try to “remove” small eigenvalues.
- We can try to modify the problem so the right-hand side of the new problem has only very small components in the directions associated with smallest and largest eigenvalues.

What is Recycling?

Assume solving $Ax = b$, and have tall-skinny U such that $AU = C$, with $C^T C = I$. Range(U) is the recycled space.

Step 1: Look for best solution in Range$(U) \Rightarrow x = UC^T b$.

Step 2: Compute $r = b - CC^T b$. Consider $K_m((I - CC^T)A(I - CC^T), (I - CC^T)b)$. Look for best solution in Range$(U) \oplus$ Range(K_m).

To accomplish step 2, iteration is on a different system – MINRES applied to

$$(I - CC^T)A(I - CC^T)y = (I - CC^T)b$$

Recycling

We are applying a Krylov solver to solve a different problem,

$$PAP^T y = Pb$$

where $P = (I - CC^T)$ is an orthogonal projector.

If, for example, C contains eigenvectors of A corresponding to small eigenvalues, convergence behaves as if the small eigenvalues of A have been removed. That is, number of iterations m will be small.

What Subspace to Recycle?

The real question is how to choose U.

Recall we want to solve a sequence of systems

$$(A^{(k)} + i\gamma I)x_j^{(k)} = b_j$$

but for each fixed matrix (i.e. k) in the sequence, we have to loop over j and γ. (I.e. systems for each source, detector, frequency)

We should use a slightly different U for each j, k.

Generating an initial U

Consider case k fixed at 0, $\gamma = 0$, and loop over the RHS.

Set $P = I$; For $j = 1..$,

- Run MINRES to solve $PA^{(0)}P^T y = Pb_j$
- Update x
- Compute some approximate eigenvectors if $j < J$, append to U.
- If U changed, $A^{(0)}U = M$, $M = CR$, $U = UR^{-1}$
- $P = I - CC^T$

Since this is the same matrix for each system, we expect this is a good choice for U.

Subspace Recycling

Next consider what happens when k changes.

As explored in [K.& de Sturler, 2006], features to exploit:

- Background captured quickly, shape slowly changing
  - Observe small cluster of "small" eigenvalues disjoint from the rest
  - Corresponding invariant subspaces remain close
- The right-hand sides are only functions of j, not $k \rightarrow$ previous solutions $x_j^{(k)}$ should be added to the U corresponding to right-hand-side j.
- Where you are in the linesearch plays a role in what previous solutions, approximate invariant subspaces, can be kept in U.

Summary

At this point, we are suggesting (real systems only)

$$U^{(k,j)} = [W, x_j^{(k-l)}, x_j^{(k-n)}]$$

where W represents some approximate invariant subspace corresponding to small eigenvalues of $A^{(0)}$, and what l, n are depends on where system k is in the line search.

W may need to be periodically refreshed as k increases.

Justification for Recycling W

Main theoretical result shows that an invariant subspace whose associated eigenvalues are not well-separated from the remaining eigenvalues is still insensitive to perturbations that are concentrated in an invariant subspace whose eigenvalues are sufficiently far removed.

If SPD $A = Q\Lambda Q^T$, and $A + E = \hat{Q}\hat{\Lambda}\hat{Q}^T$ where eigenvalues can be partitioned as

$$\lambda_1^{(1)} \leq \ldots \leq \lambda_{k1}^{(1)} < \lambda_1^{(2)} \leq \ldots \leq \lambda_{k2}^{(2)} < \lambda_1^{(3)} \leq \ldots \leq \lambda_{k3}^{(3)}.$$

and similarly for $\hat{\Lambda}$. If Q_1 corresponds to (a subset of) the recycled invariant subspace, under some assumptions we show the cannonical angles between Q_1, $\hat{Q}_1$ are small.

Consequences

We argue that due to fast convergence of background params. and changes mostly in shape, system matrix changes are primarily concentrated in the high frequency (large ev) components and not in the low frequency.

Implies that one may be able to keep (estimates of) Q_1 around in the recycled space for a long period of time. Estimates of Q_1 are those obtained by initial MINRES runs.

Tuning Based on the Search

Now we return to the issues of how to augment U, for a fixed right-hand side b_j, with solutions to previous systems.

Key is to consider the effect of the linesearch in the damped GN solver.

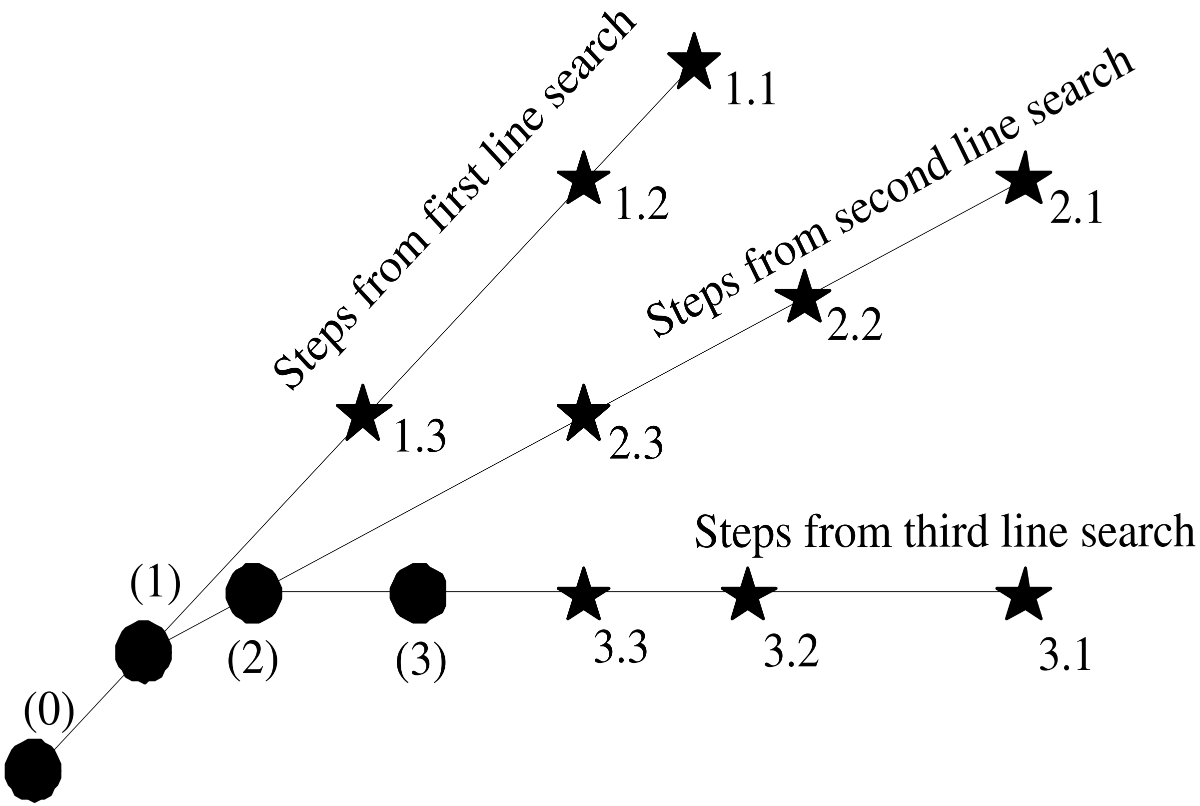
DGN

$$\min_p \|W\left(h(c(p)) - d\right)\|_2$$

Given initial p_0
For $k = 0, ...$ to convergence

- $\epsilon^{(k)} = W(h(p^{(k)}) - d)$
- Solve $(J^{(k)})^T J^{(k)} s = -(J^{(k)})^T \epsilon^{(k)}$
- $p^{(k+1)} = p^{(k)} + \xi s$ where ξ is chosen using a linesearch (minimize residual as a function of ξ).

Utilizing Search Information



Strategy, all right-hand-sides, no shift

$$A^{(k)} x_j^{(k)} = b_j$$

1. Initial system & RHS, iterate to complete, save some Ritz vectors in W
   * Other RHS, use current W, add some new vectors to W
2. For each other system k, loop over RHS:
   - If in line search, U contains some vectors in W, beginning solution (in LS), prev. accepted step

     i.e. $U = [W, x_j^{(k-l)}, x_j^{(k-n)}]$

   - If at beginning, U contains W, previous solution (end of line search) i.e. $U = [W, x_j^{(k-1)}]$

Shifted Systems

For fixed (real) system, $Ax = b$, we have $AU = C, C^T C = I$, and recall we look for a solution in $K_m(PAP^T, Pb) \oplus \text{Range}(U)$ where $P = I - CC^T$

The Krylov subspace of the shifted matrix $P(A + i\gamma I)P^T$ is the same, although $(A + i\gamma I)U = C + i\gamma U$.

Still, we look for solutions to the shifted system in $K_m(PAP^T, Pb) \oplus \text{Range}(U)$.

Result: As long as we use the same U for the real and complex system, we obtain solution to the complex system with a small amount of overhead, but no extra mat-vecs.

Shifted Systems

Observations:

- We need to use a U that has only real entries, so that we don't introduce unnecessary complex arithmetic when solving the real systems.
- The the real part of the solution to the shifted system and the solution to the real system for a fixed b_j are close.

Idea: augment U for a fixed b_j by the previous solution to the real (unshifted) system and the imaginary part of the shifted system.

Algorithm

1. Initial system & RHS, iterate to complete, save some Ritz vectors in W
   - Set $U = W$, update x for the real and shifted systems for initial RHS.
   - Loop over other RHS, solve for x on both the real and shifted systems, expanding $U = W$.
2. For every other system, looping over RHS:
   - If in line search,

$$U = [W, x_{j,0}^{(k-l)}, imag(x_{j,\gamma}^{(k-l)}), x_{j,0}^{(k-n)}, imag(x_{j,0}^{(k-n)})]$$

   - If at beginning, $U = [W, x_{j,0}^{(k-1)}, imag(x_{j,\gamma}^{(k-l)})]$
   - Update solns. for the corresponding real and shifted systems

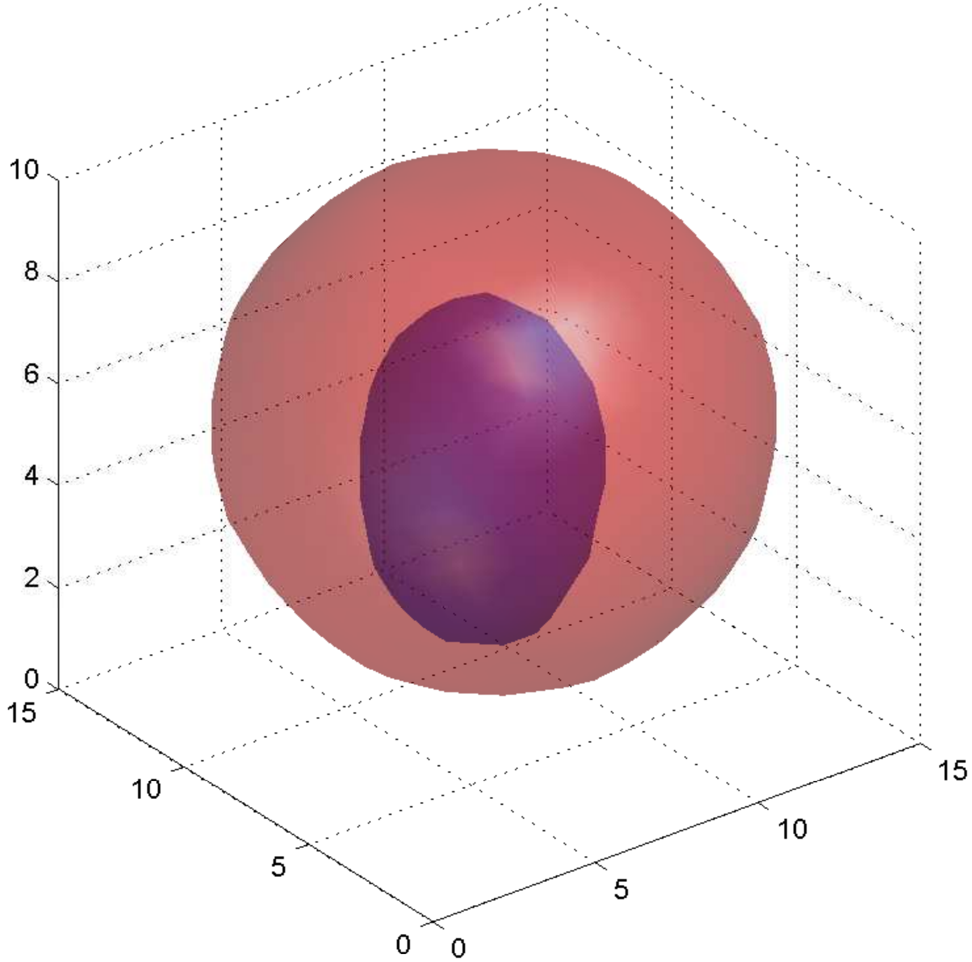
Numerical Results

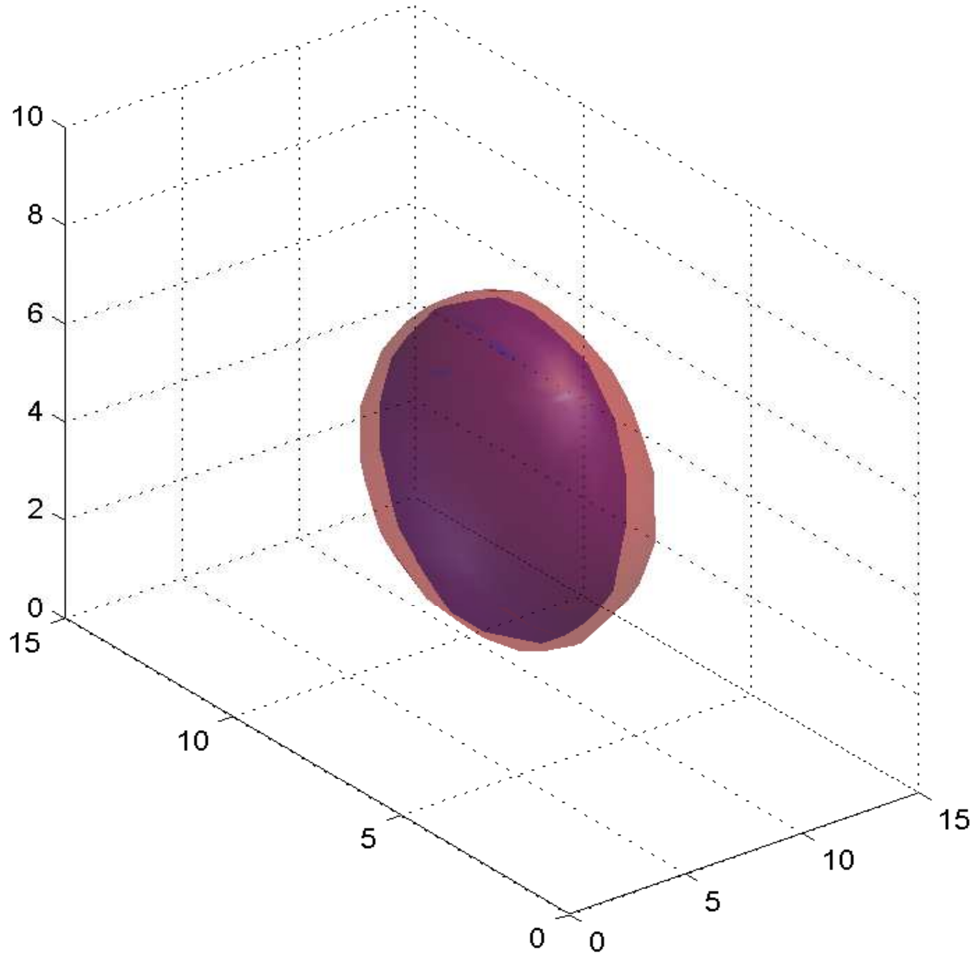
All experiments carried out in Matlab

Example 1: PaLS reconstruction for diffusion

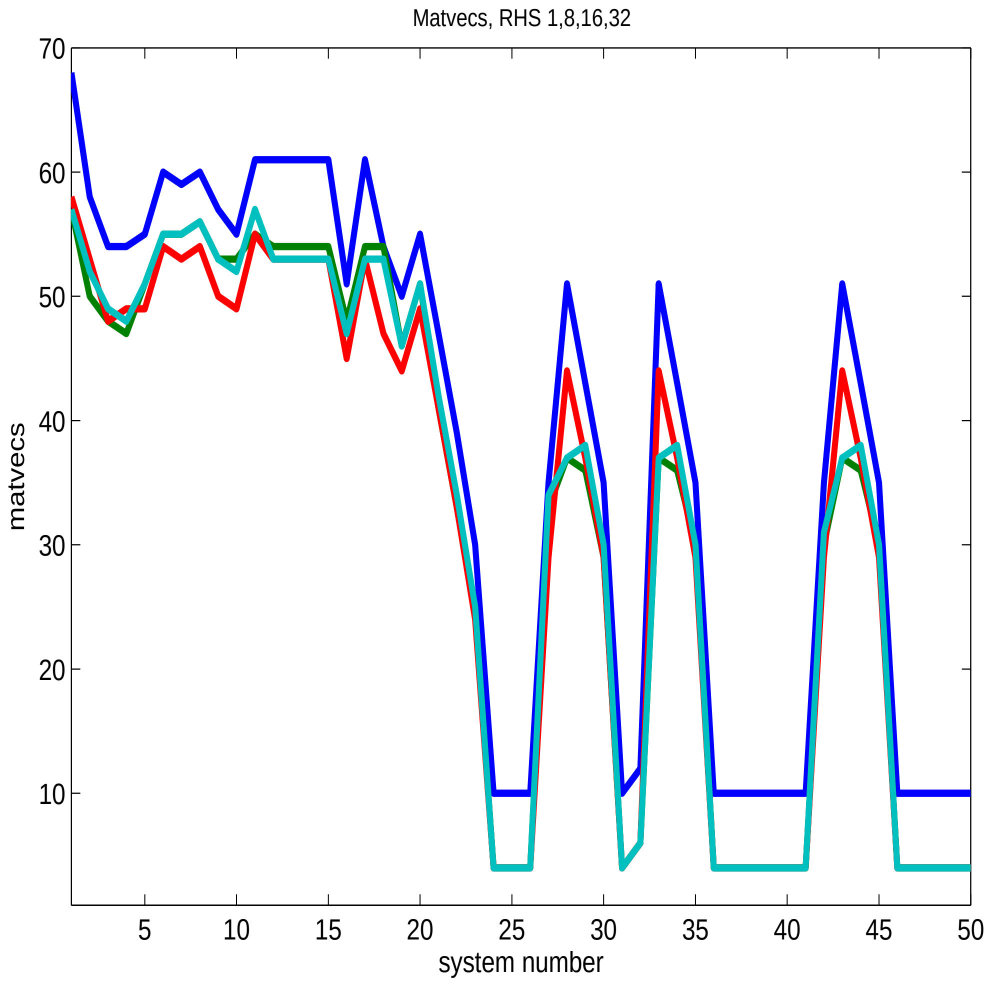
- constant background, shape boundary described by the (near) zero level set of a linear combination of polynomial basis functions.
- $15 \times 15 \times 10$ grid, 8cm $\times$ 8cm $\times$ 4cm
- 16 sources top/bottom; 16 detectors top/bottom
  - * 32 RHS for forward problem; 32 for “adjoint”
- 2 Harmonic Ritz vectors added for each of the first 3 systems

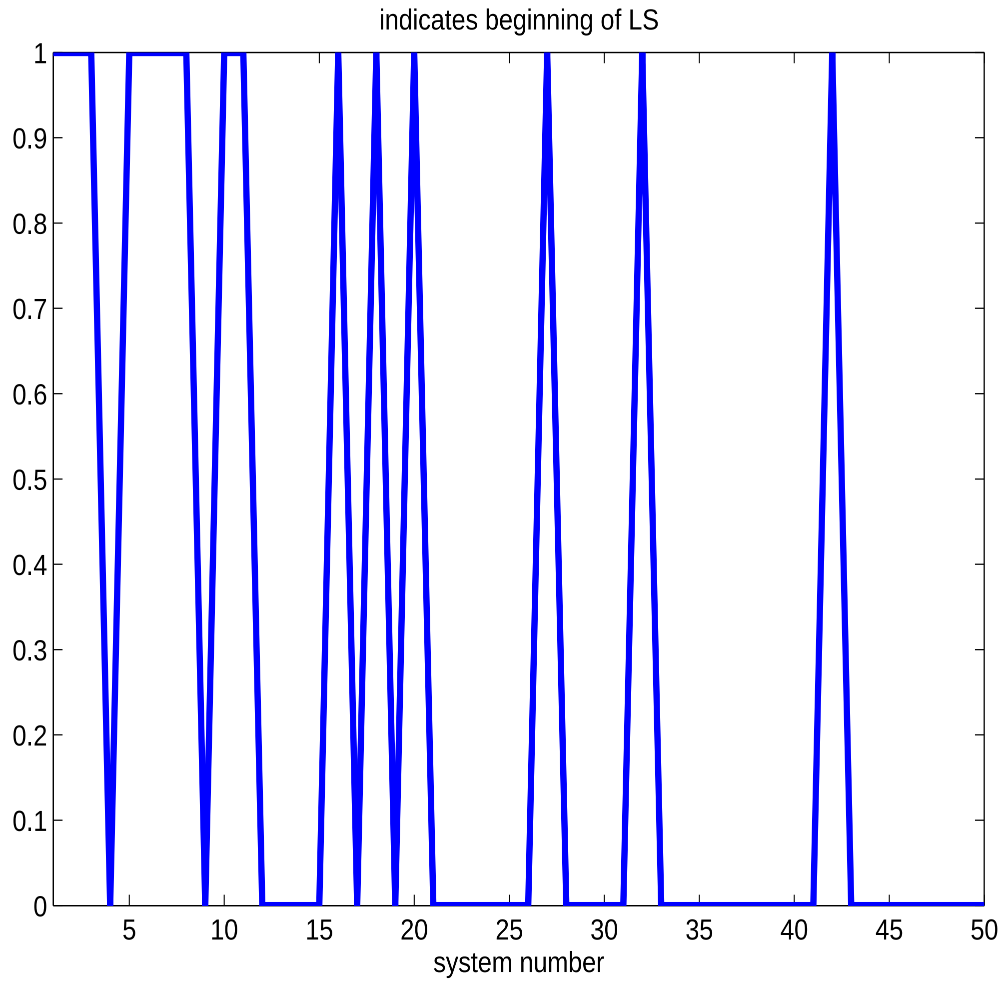
Exp 1



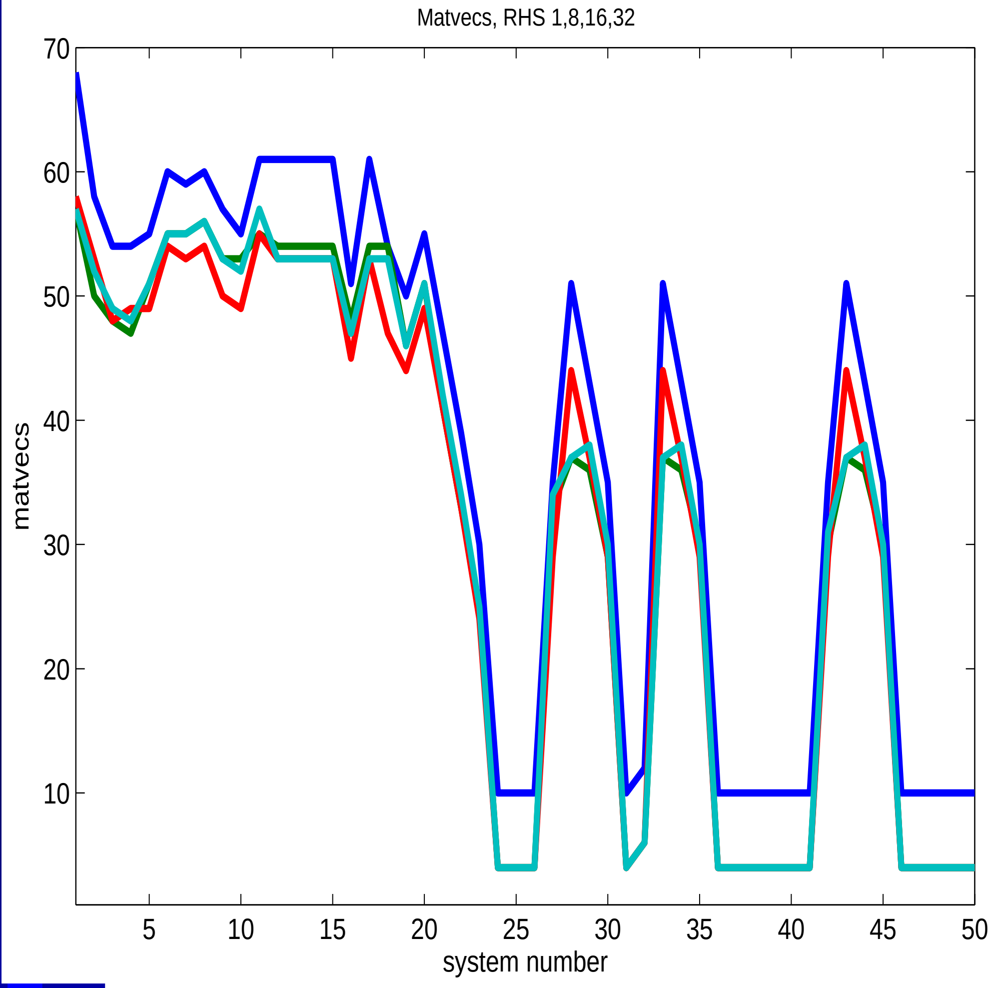


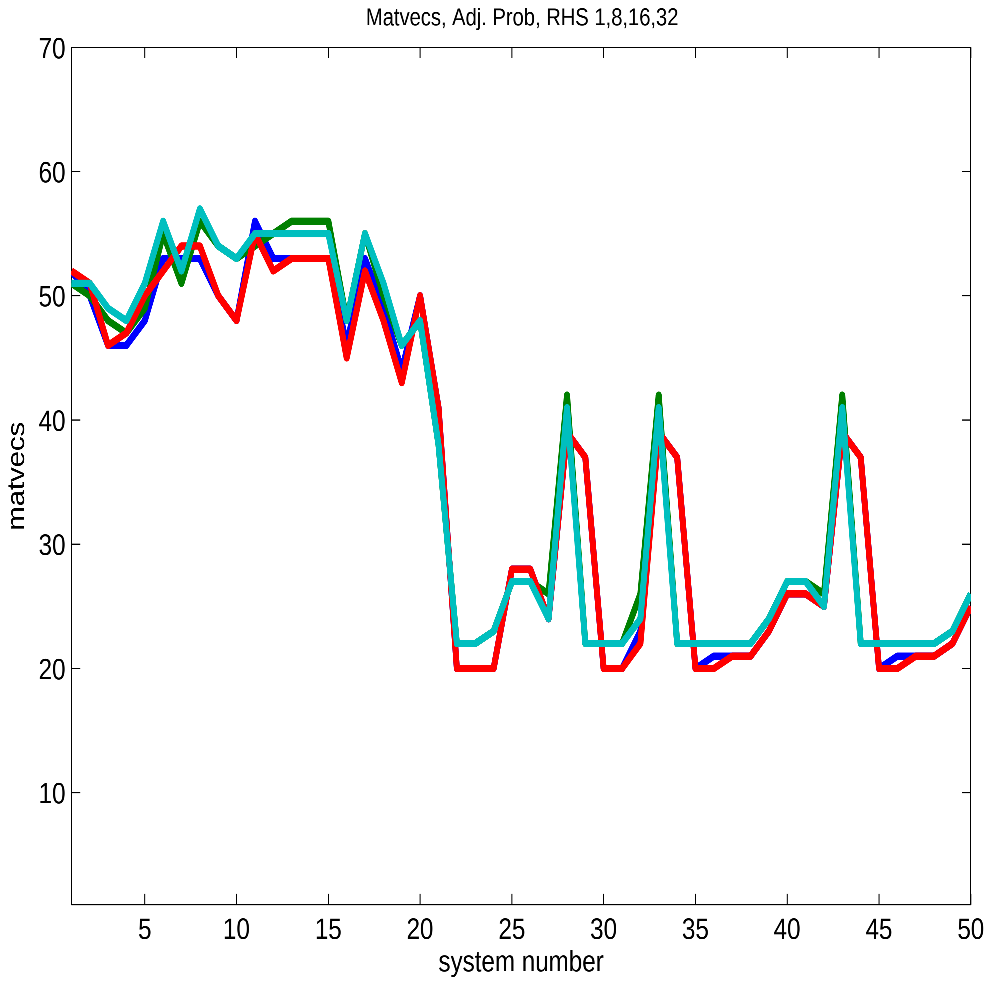
Exp 1


Matvecs, RHS 1,8,16,32
matvecs
system number
70
60
50
40
30
20
10
5
10
15
20
25
30
35
40
45
50


indicates beginning of LS
system number
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
5
10
15
20
25
30
35
40
45
50

Exp 1

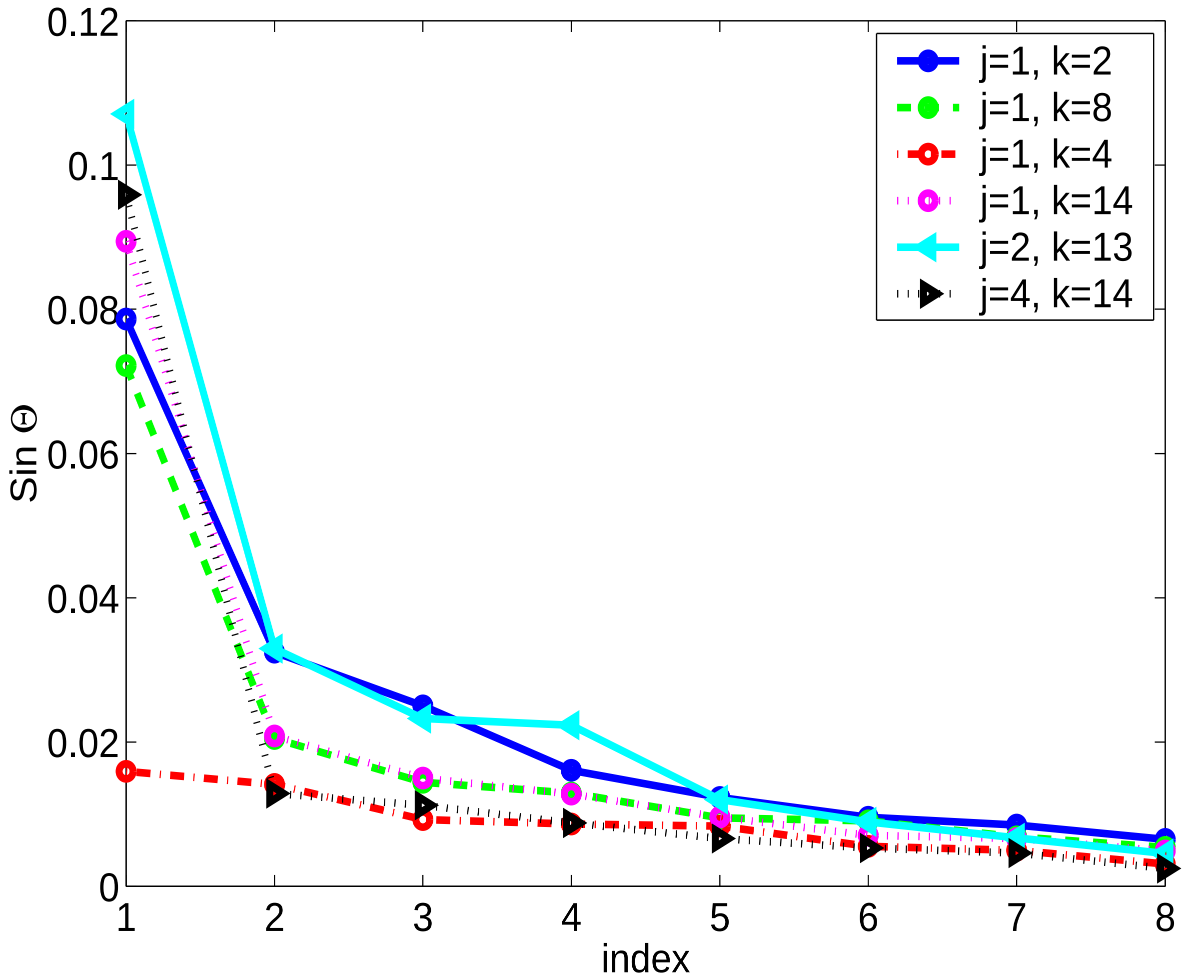

Matvecs, RHS 1,8,16,32
matvecs
system number
70
60
50
40
30
20
10
5
10
15
20
25
30
35
40
45
50


Matvecs, Adj. Prob, RHS 1,8,16,32
matvecs
system number
70
60
50
40
30
20
10
5
10
15
20
25
30
35
40
45
50

Exp 2

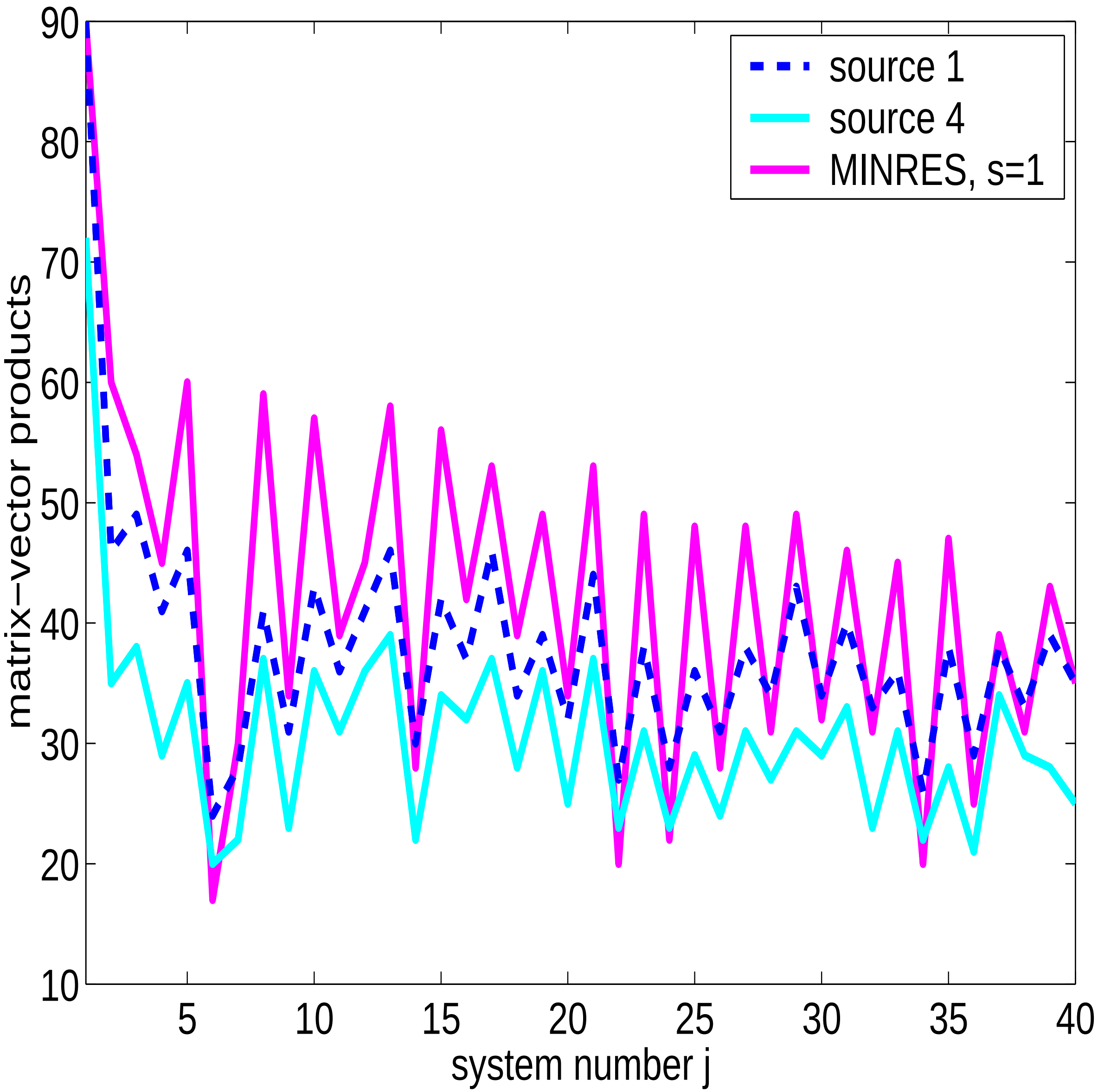
- 31x31x21 grid, ellipsoidal reconstruction for diffusion and absorption.
- GN steps = 24 and the total no. of system matrices = 51. Looked at the first 40 systems: indices corresponding to the beginning of a line search are 2,5,8,10,13-39 (odd)
- two vectors to W corresponding to the smallest harmonic Ritz values for each of the first 4 sources.

Cannonical Angles



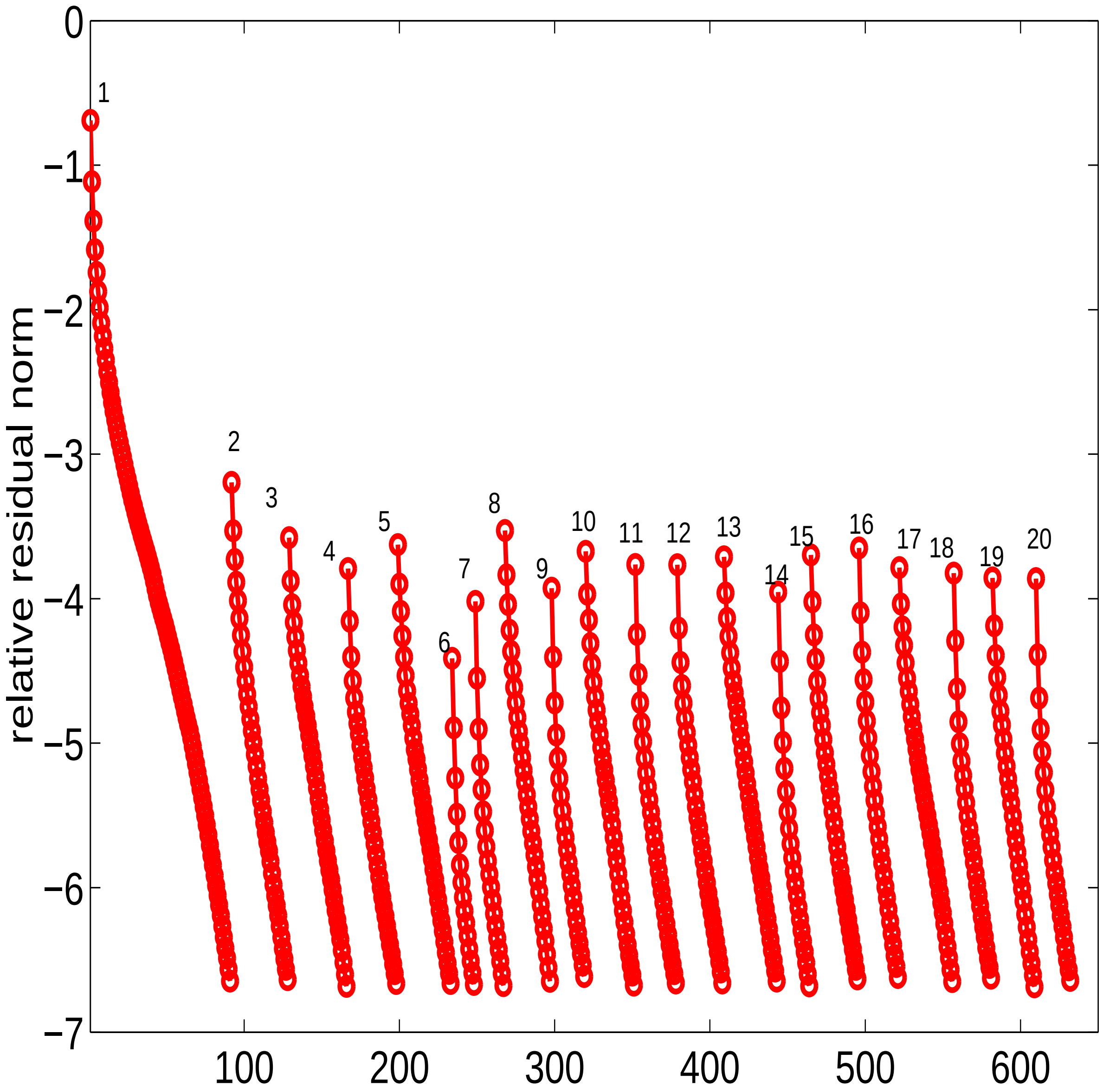
Plots of $\sin\Theta[\mathrm{range}(W^{(j)}), \mathrm{range}(W^{(k)})]$ for various (j, k) assuming a subspace dimension of 8.

Exp 2

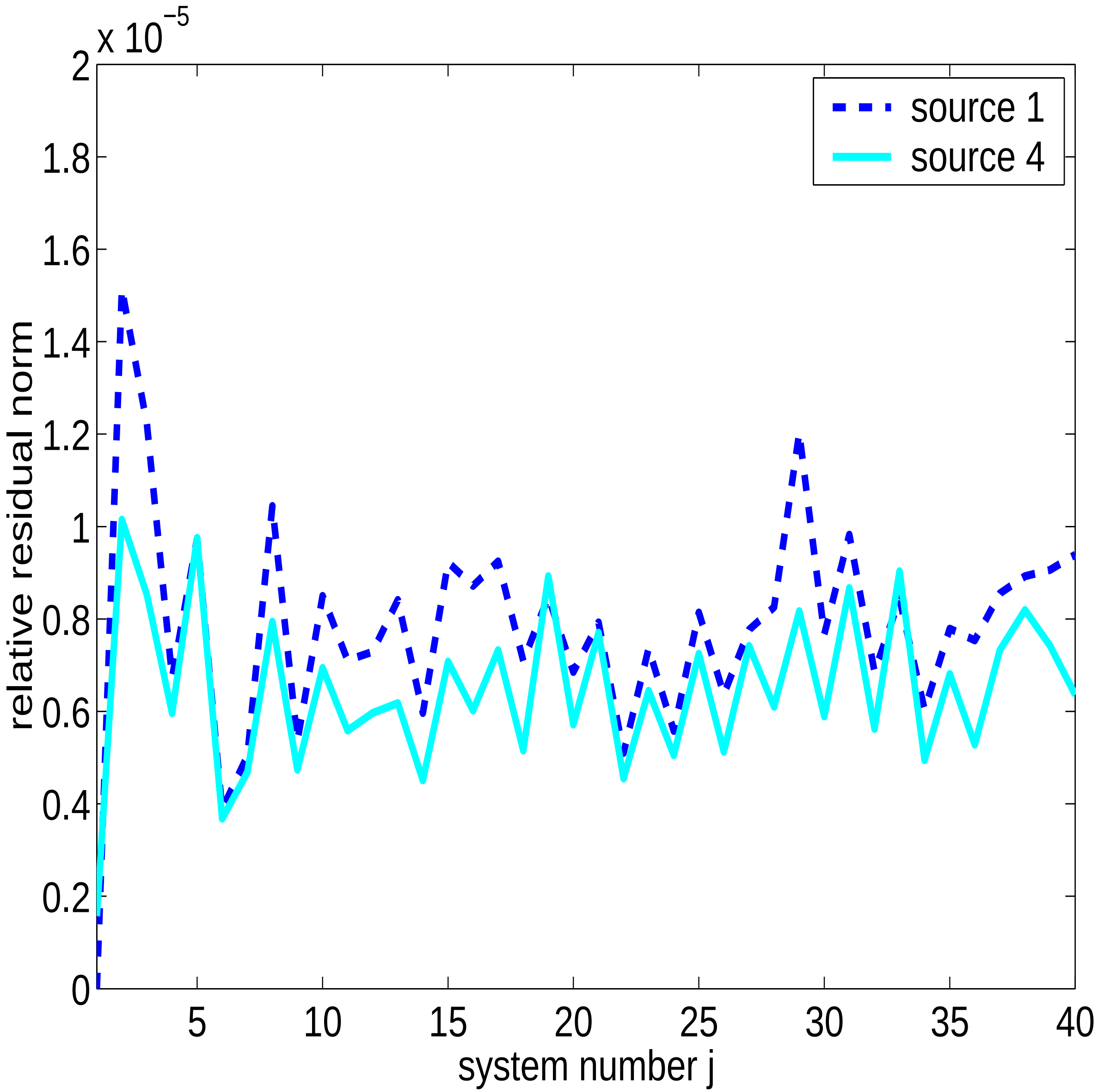


source 1
source 4
MINRES, s=1
matrix−vector products
system number j
90
80
70
60
50
40
30
20
10
5
10
15
20
25
30
35
40

Conv. Plots for (Real) Systems 1:20, src 1



Conv. Plots for Shifted Case



Conclusions and Future Work

- Combined strategies of recycling approx. invariant subspaces with those for subspaces from previous solutions.
- Careful analysis (using application, matrix symmetry) of which strategy is most useful at ea. stage of the optimization.
- Expanded algorithm combines subspace recycling with solving for multiple shifted systems using a single Krylov subspace.
- Optimization algs. that don't require line searches.
- Automating recycle space updating process using GCROT-like techniques to measure the subspace effectiveness.
- Other applications (e.g. linear pMRI).