Two-Parameter Selection Techniques for Projection-based Regularization Methods: Application to Partial-Fourier pMRI

Misha E. Kilmer[1]

Scott Hoge[2]

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

[2] Dept. of Radiology, Brigham and Women's Hospital and
Harvard Medical School, Boston, MA

Overview

- Hybrid Methods for single parameter Tikhonov regularization
- Generalization to specialized, two parameter case
- Parallel MRI background
- Numerical results
- Conclusions and future work

Motivation

Forward model is a (real) linear system

$$Ax - \eta = b_{ex},$$

where

- A is
  - $m \times n$, (e.g. $n \geq 100,000$,$m \geq n$)
  - not available explicitly (fast matvecs)
  - ill-conditioned
- Only $b = b_{ex} + \eta$ measured (known)
- η white (or close)

Regularized Problem

$$\min_x \|Ax - b\|_2^2 + \lambda^2 \|L_1 x\|_2^2 + \mu^2 \|L_2 x\|_2^2$$

where λ, μ are not known a priori.

Our Motivation: Image reconstruction in partial Fourier, parallel (i.e. multiple coil) MRI.

Assumption*: $L_1, L_2, n/2 \times n$, $\begin{bmatrix} L_1 \\ L_2 \end{bmatrix}$ invertible.

Issue: Choosing appropriate (λ, μ) without naive solution over a grid of possible choices.

*can be relaxed

Single Parameter Case

First consider

$$\min_x \|Ax - b\|_2^2 + \lambda^2 \|Lx\|_2^2,$$

where L is (cheaply) invertible. A change of variables $y = Lx$ gives

$$\min_y \left\| \begin{bmatrix} AL^{-1} \\ \lambda I \end{bmatrix} y - \begin{bmatrix} b \\ 0 \end{bmatrix} \right\|_2^2 .$$

Naive Approach: for fixed set of λ's, repeatedly solve and use a heuristic (e.g. L-curve, methods from previous talk) to approximate the best one.

Single Parameter Case, cont

Since cheap to compute Av and $L^{-1}v$, use LSQR to solve.

Solve instead

$$\min_{y \in \mathcal{K}_k} \left\| \begin{bmatrix} AL^{-1} \\ \lambda I \end{bmatrix} - \begin{bmatrix} b \\ 0 \end{bmatrix} \right\|_2^2$$

where, with $C = AL^{-1}$,
$\mathcal{K}_k =$ span$\{C^T b, (C^T C) C^T b, (C^T C)^2 C^T b, \ldots, (C^T C)^{k-1} C^T b\}$

Soln. cost $\approx k$ times sum of matvec cost with A and cost of $L^{-1}v$. If we wanted to solve this “accurately” for each specific λ, k could change and be large. Too expensive!

Projected Problem

In LSQR we have the relations

$$AL^{-1}V_k = U_{k+1}B_k, \qquad u_1 = U_{k+1}e_1 = \beta b$$

where V_k is $n \times k$, U_k is $m \times k$ each with orthogonal columns and B_k is $k+1 \times k$ bidiagonal.

$$\min_{y_k \in \mathcal{K}_k} \left\| \begin{bmatrix} AL^{-1} \\ \lambda I \end{bmatrix} y_k - \begin{bmatrix} b \\ 0 \end{bmatrix} \right\|_2^2$$

becomes, with $x_k = V_k y_k$:

$$\min_{z_k} \left\| \begin{bmatrix} B_k \\ \lambda I_k \end{bmatrix} z_k - \beta e_1 \right\|_2^2$$

Regularized, Projected Problem

$$\min_{z_k} \left\| \begin{bmatrix} B_k \\ \lambda I_k \end{bmatrix} z_k - \beta e_1 \right\|_2^2$$

This is a size k (small), Tikhonov-regularized, projected problem.

KEY: Choose λ "optimally" for this problem. Then, the regularized solution to the original equation is set as $y_k^{(\lambda^*)} = V_k z_k^{(\lambda^*)}$.

Benefits

$$\min_{z_k} \left\| \begin{bmatrix} B_k \\ \lambda I_k \end{bmatrix} z_k - \beta e_1 \right\|_2^2$$

- Can compute $y_k^{(\lambda)} = L^{-1} x_k^{(\lambda)}$ with short-term recurrences for multiple λ simultaneously.
- Try to choose optimal λ *for projected problem* using appropriate heuristic [K. and O'Leary, '01]
  - $\|Lx_k^{(\lambda)}\| = \|y_k^{(\lambda)}\| = \|z_k^{(\lambda)}\|$ virtually free
  - $\|Ax_k^{(\lambda)} - b\| = \|AL^{-1}y_k^{(\lambda)} - b\| = \|B_{k+1} z_k^{(\lambda)} - \beta e_1\|$ virtually free
- If k not too large, other options possible (e.g. WGCV, [Chung, Nagy, O'Leary '08])

Two Parameter Case

$$\min_x \|Ax - b\|_2^2 + \lambda^2 \|L_1 x\|_2^2 + \mu^2 \|L_2 x\|_2^2$$

Recall the assumption $L_1, L_2, n/2 \times n$, $\begin{bmatrix} L_1 \\ L_2 \end{bmatrix}$ invertible.

$$\min_x \|Ax - b\|_2^2 + \lambda^2 \left\| \begin{bmatrix} L_1 \\ \frac{\mu}{\lambda} L_2 \end{bmatrix} x \right\|_2^2$$

Fix $c = \mu/\lambda$, define $L_c = \begin{bmatrix} L_1 \\ cL_2 \end{bmatrix}$ and $y^{(\lambda,\mu)} = L_c x^{(\lambda,\mu)}$:

$$\min_y \left\| \begin{bmatrix} AL_c^{-1} \\ \lambda I \end{bmatrix} y - \begin{bmatrix} b \\ 0 \end{bmatrix} \right\|_2^2$$

Two Parameter Case

$$\min_{z_k} \left\| \begin{bmatrix} B_{k,c} \\ \lambda I \end{bmatrix} z_{k,c} - \beta e_1 \right\|_2^2$$

- For a fixed value of c, a different projected problem, regularized using Tikhonov.
- Question: Which values of λ need to be tested for fixed c?
- Question: What information about the projected problems do we retain to make a decision about **both** λ and μ?

Grid

Typically, choose a set of ℓ_1 values for λ equally spaced in log space. Likewise, ℓ_2 log-equispaced points for μ. Then "search" over the $\ell_1\ell_2$ possible pairs in the grid.

Thus, in logspace, each pair (λ, μ) lies on one of the $\ell_1 + \ell_2 - 1$ lines of slope 1 in this grid.

Using $\mu = c\lambda$, each line corresponds to one value of c. For each fixed c, we need only take the λ values on this line. For each projected problem, at most $\min(\ell_1, \ell_2)$ λ values are tested, at best, 1. Work could be done in parallel.

Summary

Original:

$$y^{(\lambda,\mu)} = \arg\min_y \left\| \begin{bmatrix} AL_c^{-1} \\ \lambda I \end{bmatrix} y - \begin{bmatrix} b \\ 0 \end{bmatrix} \right\|_2^2,$$

$$x^{(\lambda,\mu)} = L_c^{-1} y^{(\lambda,\mu)}, \qquad \mu = c\lambda.$$

Apply k steps of LSQR to approximately it, equivalent to:

$$z_{k,c}^{(\lambda)} = \arg\min \left\| \begin{bmatrix} B_{k,c} \\ \lambda I_k \end{bmatrix} z - \beta e_1 \right\|_2^2$$

Summary, Cont

We are able to compute the following with short-term recurrences, for all appropriate values of c, λ_1 by considering multiple projected problems:

- Solutions* $y_k^{(\lambda,\mu)} = L_c x_k^{(\lambda,\mu)}$
- $\|y_k^{(\lambda,\mu)}\| = \|z_{k,c}^{(\lambda)}\|$
- $\|r_k^{(\lambda,\mu)}\| = \|Ax_k^{(\lambda,\mu)} - b\| = \|B_{k+1,c} z_{k,c}^{(\lambda)} - \beta e_1\|$

* Not needed to obtain items 2 and 3.

Goal

Compute near "optimal" values of λ, μ. Would like to do this using only information that was cheaply computed for each projected problem.

Following single-parameter case logic, knowing an "optimal" value of λ for each fixed-c line might be useful.

Added difficulty: each projected problem depends on a fixed choice of c, but need whole picture. In particular, $\|L_c x_k^{(\lambda,\mu)}\|$ is what is returned, not $\|L_1 x_k^{(\lambda,\mu)}\|$, $\|L_2 x_k^{(\lambda,\mu)}\|$.

Regularization Parameter Selection

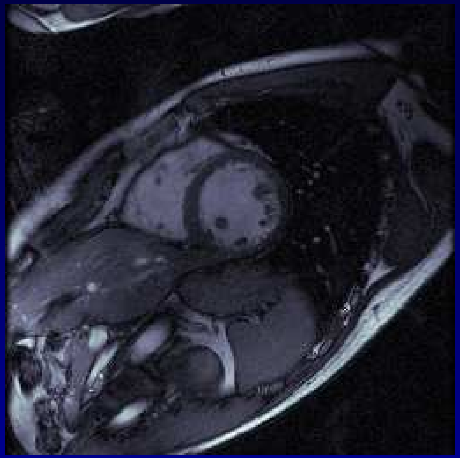
Scheme somewhat problem specific, main idea but may be useful in other applications as well.

For each c, select regularization parameter λ for the corresponding projected problem. Using $\mu = c\lambda$, gives us $\ell_1 + \ell_2 - 1$ possible choices. Next use other (problem dependent) a priori information to select from among these.

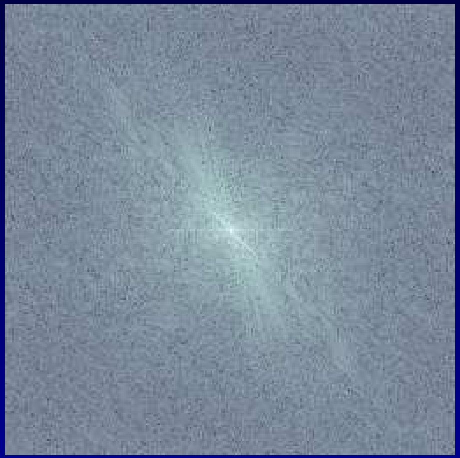
For our application, enough to *monitor sharp transitions in residual norms* (cheap, available).

Background: pMRI (2D)

MRI uses magnetic field gradients and RF signals to encode field-of-view



$\longleftarrow \mathcal{F} \longrightarrow$



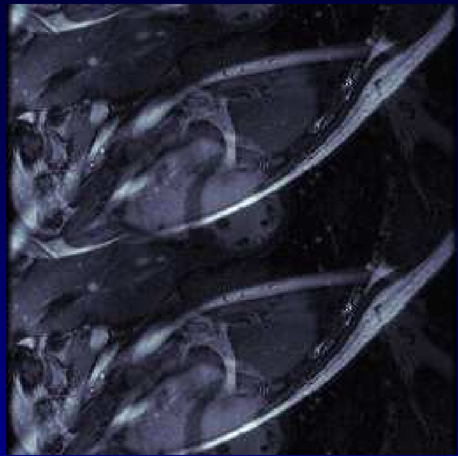
Encoding typically corresponds to DFT, both sides $\longrightarrow$ data is acquired in $k-$space domain

$k-$space is sampled in line-by-line fashion.

Reduce number of lines $\leftrightarrow$ Reduce acquisition time

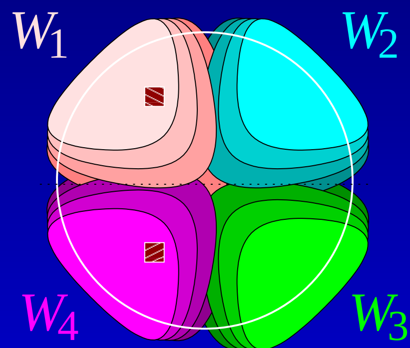
2D pMRI

Sub-sampling k-space produces aliasing in spatial domain.



$\longleftarrow \mathcal{F} \longrightarrow$

Use *multiple receiver coils*
and each coil subsamples in parallel. 4 coils, each subsampling by 4 $\rightarrow$ 16 min. scan now takes 4 min.



Reconstruct image one column at a time (regularized soln. to $W\rho = s$).
Similarly, 3D, reconstruct volume one 2D image slice at a time.

Fast imaging using partial-Fourier encoding

Strategy:

- Acquire one half of k-space (top 1/2)
- Use conjugate-symmetry assumption to reconstruct the other half

Issues:

- Conjugate-symmetry implies a real-valued image
- Field inhomogeneity and gradient field errors prevent exact conjugate-symmetry in k-space encoding.

Partial-Fourier Problem Formulation

Want to constrain solution to be 'nearly' real.

We use a two-parameter minimization, to constrain real and imaginary components separately.

$$\min_{\rho} \left\{ \|W\rho - s\|_2^2 + \lambda_{re}\|\Re\{\rho\}\|_2^2 + \lambda_{im}\|\Im\{\rho\}\|_2^2 \right\}$$

which is equivalent to solving

$$\min_{x} \|Ax - b\|_2^2 + \lambda^2\|L_1 x\|_2^2 + \mu^2\|L_2 x\|_2^2$$

with

$$A = \begin{bmatrix} \Re\{W\} & -\Im\{W\} \\ \Im\{W\} & \Re\{W\} \end{bmatrix}, x = \begin{bmatrix} \Re\{\rho\} \\ \Im\{\rho\} \end{bmatrix}, L_1 = [I_n, 0], L_2 = [0, I_n]$$

Parameter Selection

STEP 1: For each c (line) do:

- Compute the residual norms, as function of λ_1, for the projected problem (cheap!). Note these are the same as residual norms corresponding to the large problem.
- Compute relative difference between neighboring terms on that line.
- Record λ value corresponding the sharpest transition.

STEP 2:

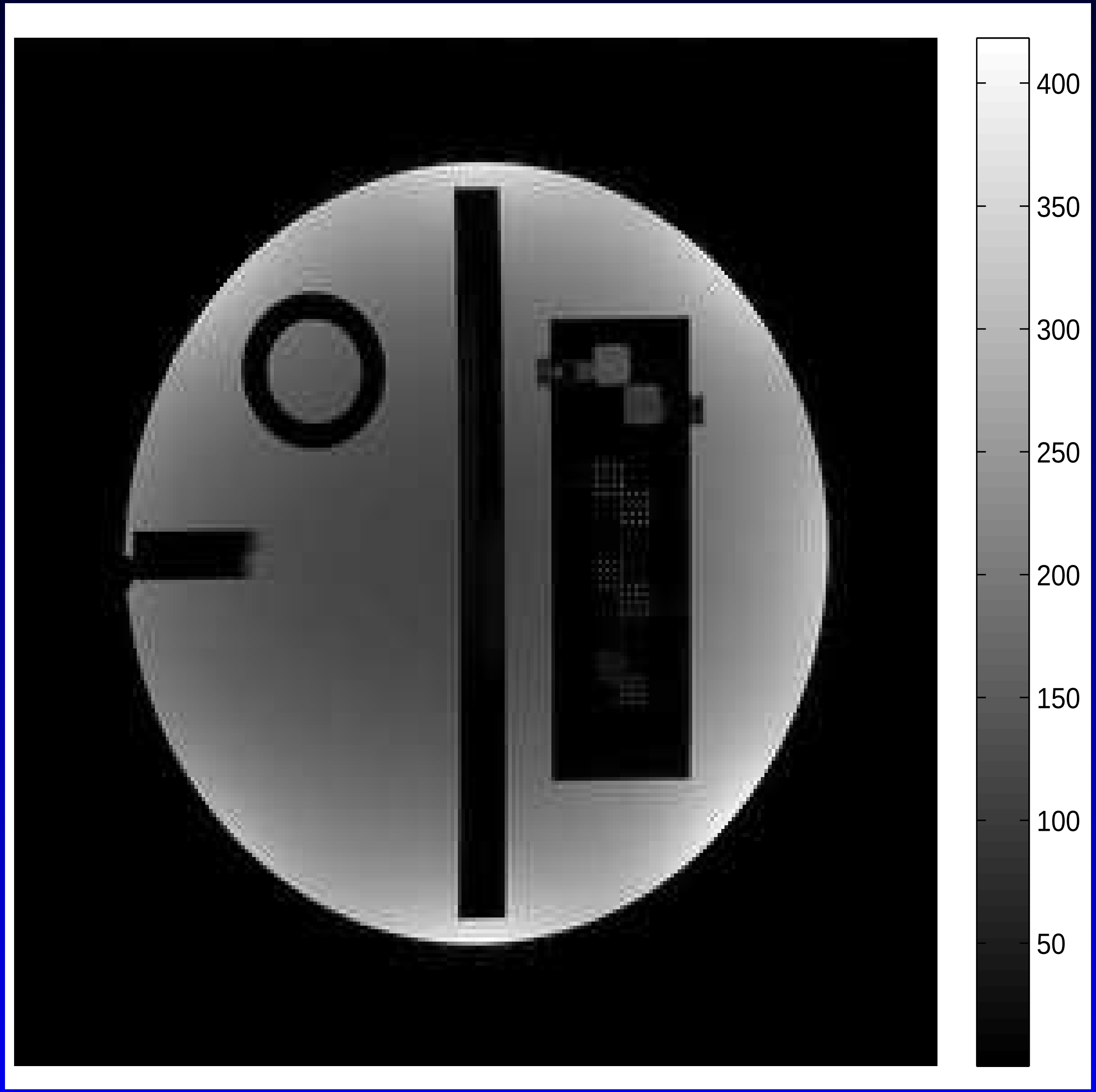
- If haven't already, compute the $x_k^{\lambda,\mu}$'s for these pairs.
- Throw out any "non-physical" solutions (e.g. ratio of imaginary part to real part too large).
- Choose the remaining term with smallest residual norm.

Numerical Results

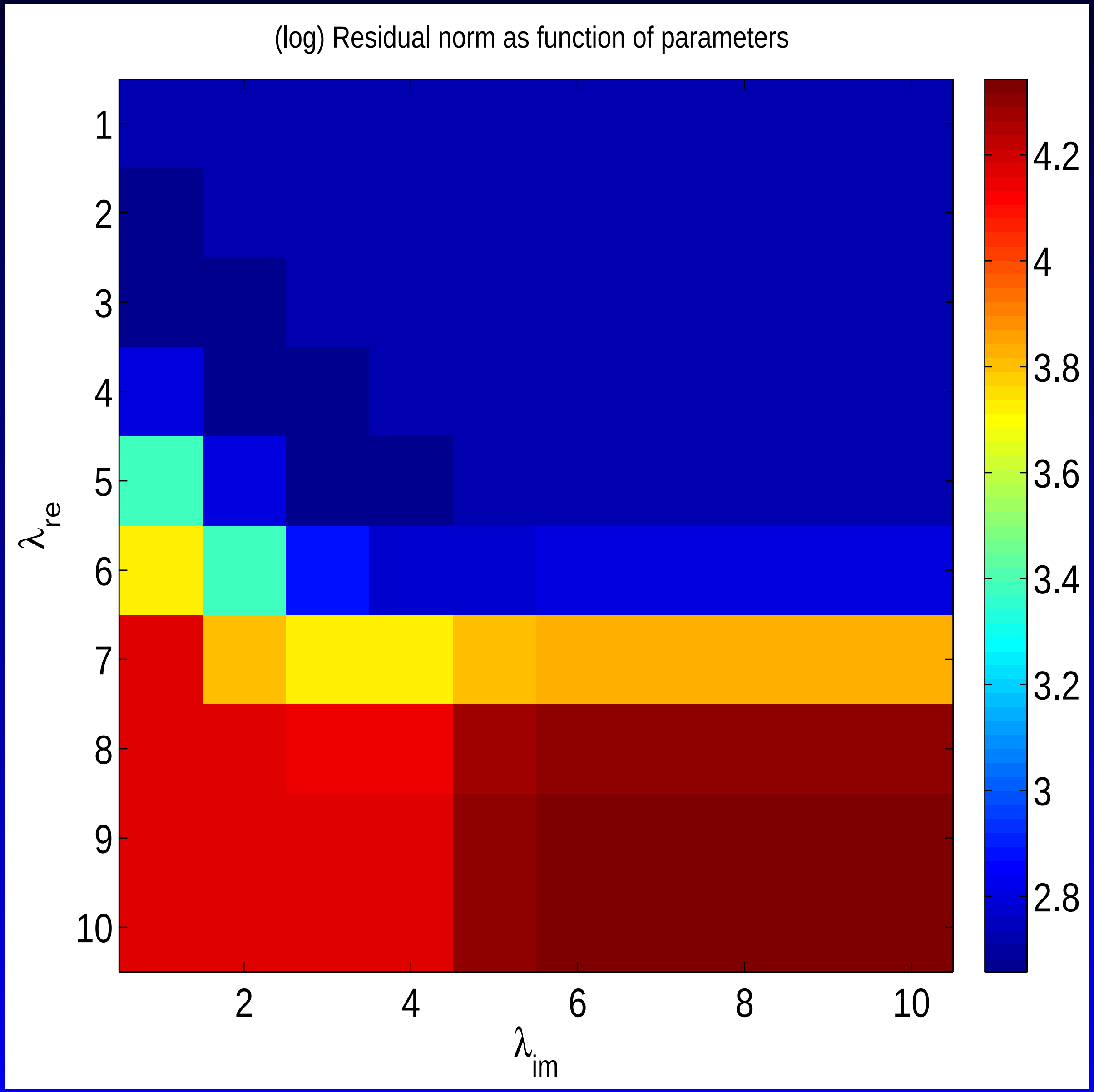
High resolution phantom, 8 coil GE Scanner at BWH, single 256x256 slice of 3D data set

- Sampled (partial Fourier) in ky-space 73 lines (at or above 128); Sampled in kx-space 114 lines, nonuniformly.
- Acceleration factor ≈ 8
- A is 133,152 x 131,072
- λ_1 = `logspace`(-5,2,10); λ_2 = `logspace`(-3,4,10)
- k fixed at 30
- Simple thresholding on aliased image to throw out non-physical solutions.

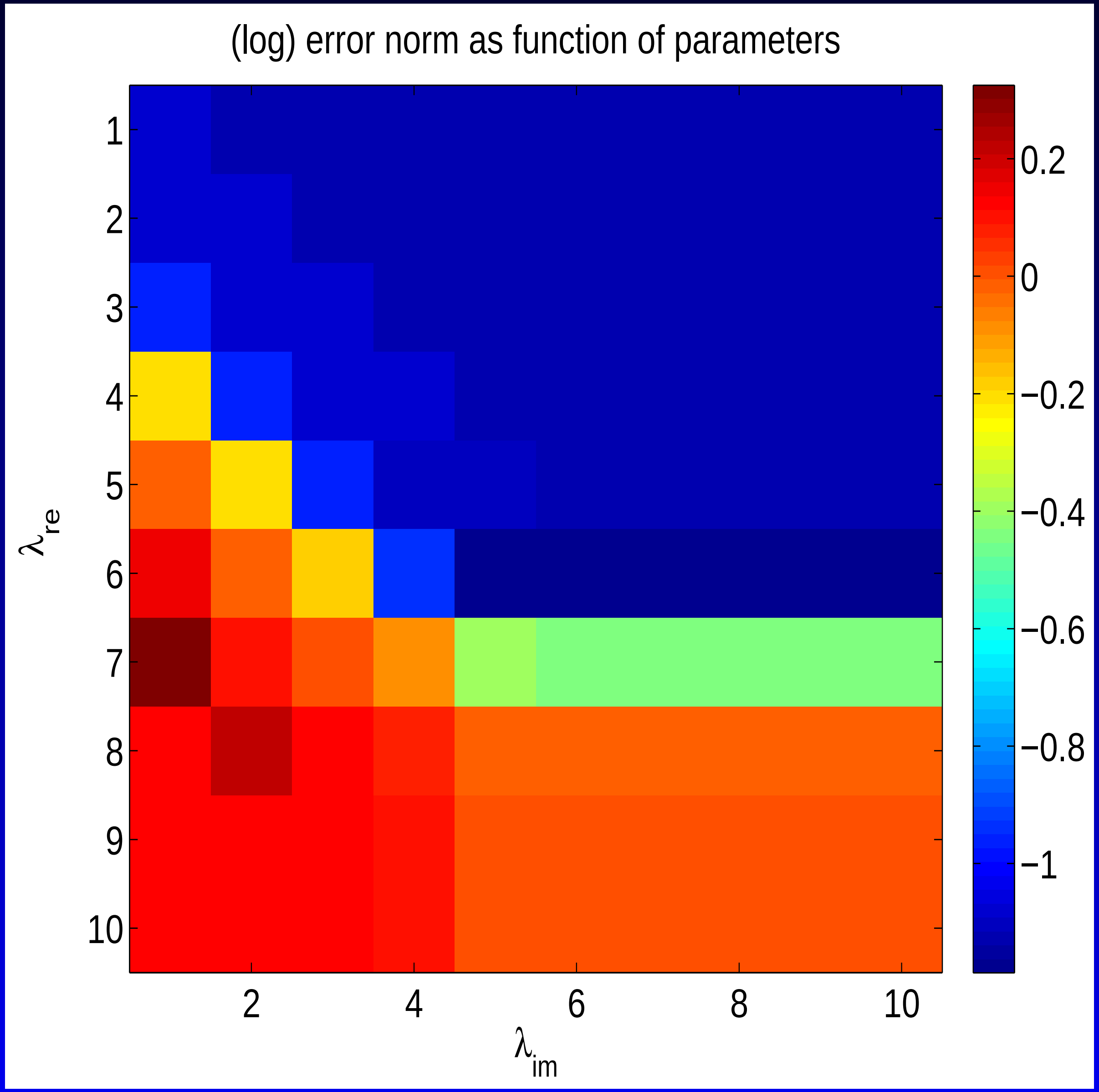
Full Data Reconstruction



Plot of Residual Norms

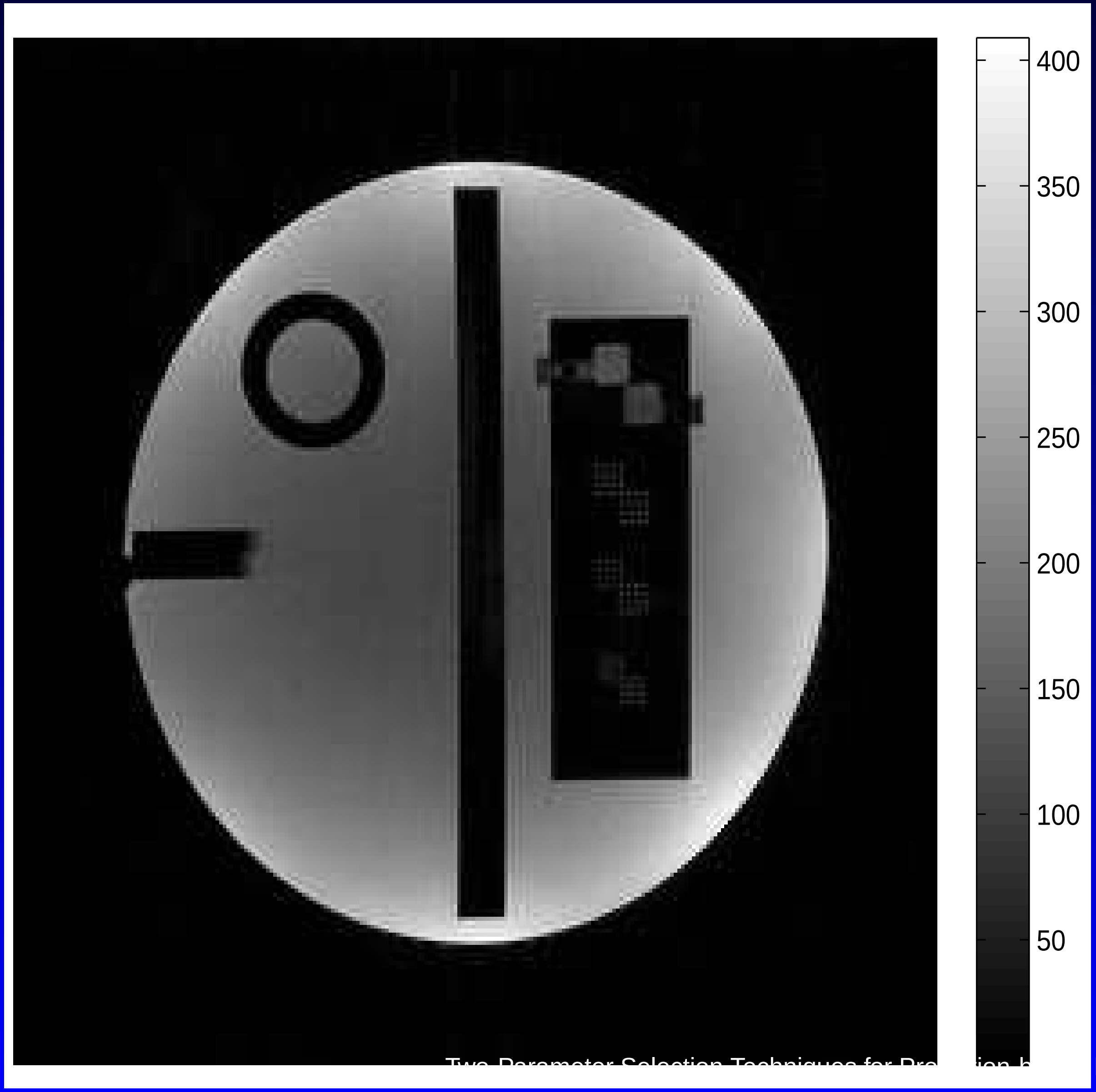


Plot of Relative Error

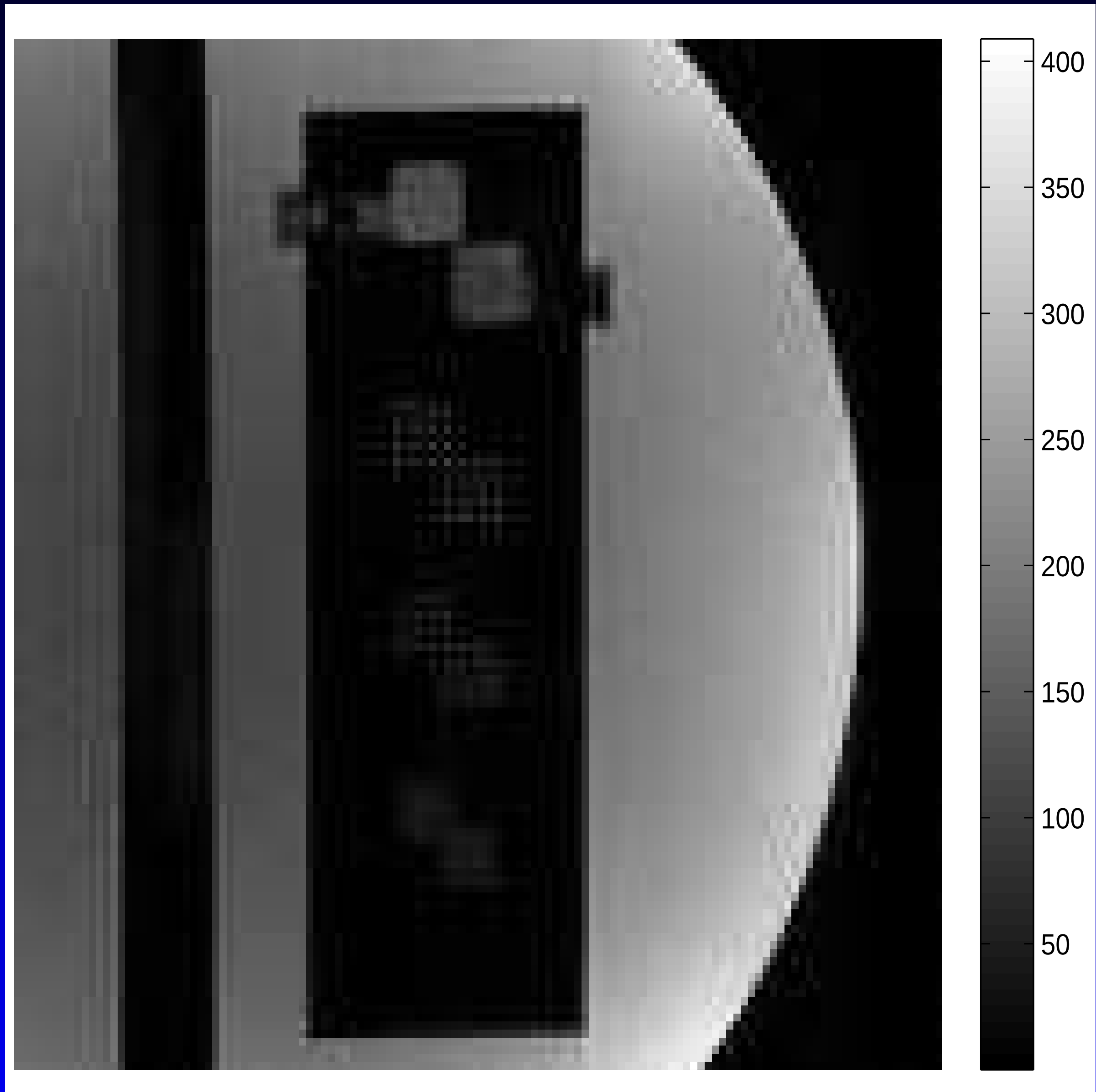


Reconstruction

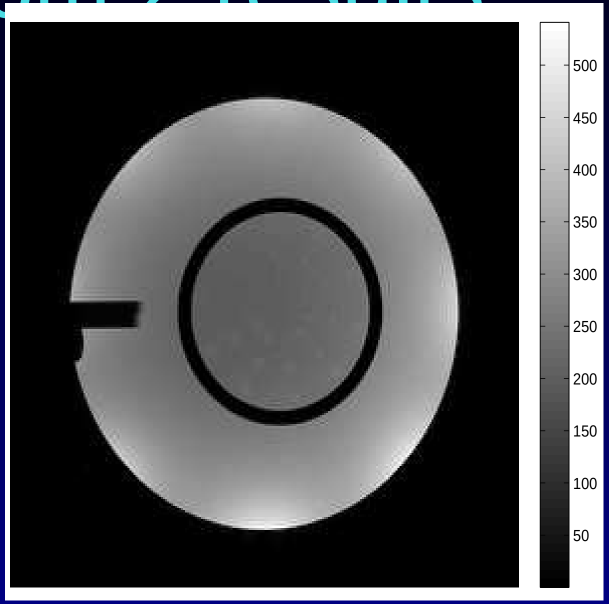
$\lambda_1(6) = 7.7e^{-2}$, $\lambda_2(5) = 1.29$

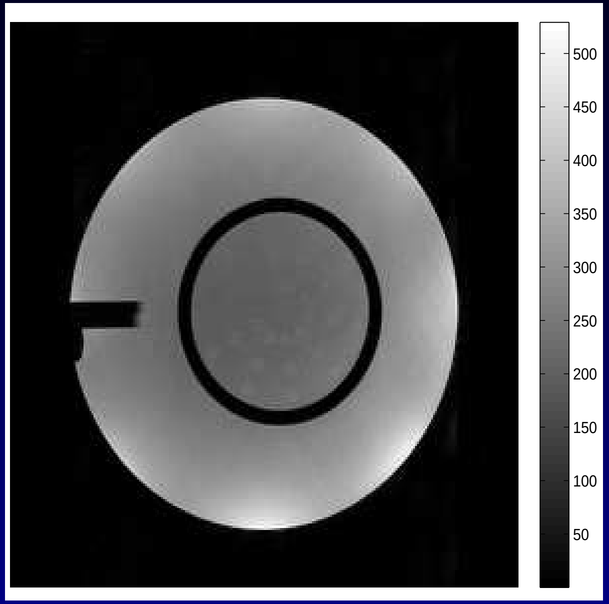


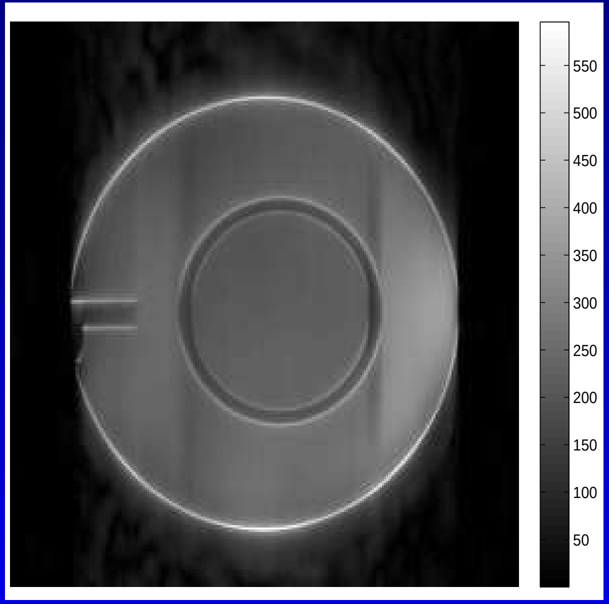
Reconstruction, zoom



Phantom 2 results







Conclusions and Future Work

- Projection approaches can be very computationally efficient – choose the regularization parameter for the smaller, projected problem (cheaper).
- For 2D, we select first for individual projected problems, then over the whole.
- Best selection methods may be problem dependent.
- Basic idea valid when L_1, L_2 not this special: Transform to standard form or use hybrid approach of [K., Hansen, Espanol, '07].
- Issue of choosing k [Chung, Nagy, O'Leary '08]. Not a factor for our application.