Fast Regularized Reconstruction of Non-uniformly Subsampled Parallel MRI data

W. Scott Hoge\textsuperscript{1} Misha E. Kilmer\textsuperscript{2}, Steve J. Haker\textsuperscript{1}, Dana H. Brooks\textsuperscript{3}, Walid E. Kyriakos\textsuperscript{1}

misha.kilmer@tufts.edu

\textsuperscript{1} Dept. of Radiology, Brigham and Women’s Hospital and Harvard Medical School, Boston, MA
\textsuperscript{2} Dept. of Mathematics, Tufts Univ., Medford, MA
\textsuperscript{3} ECE Dept., Northeastern Univ., Boston, MA
Overview

• Background
  • Basics of MRI
  • Basics of accelerated parallel MR imaging
• Details of fast reconstruction algorithm:
  • particular structure of reconstruction problem
  • solution of projected problem
  • fast regularization parameter selection
• Results
• Conclusions and future work
Basics of MRI

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

Encoding typically corresponds to Fourier transform (data is acquired in $k$–space domain)

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

Reduce number of lines $\leftrightarrow$ Reduce acquisition time
Basics of Parallel MRI

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

Parallel MR imaging (pMRI) employs spatial domain encoding \((\text{via multiple receiver coils})\) in tandem with Fourier encoding

Remove aliasing using estimates of coil sensitivities
Basics of pMRI, con’t

• Get data from multiple coils
• Fewer than $n_y$ lines acquired (simultaneously) by each coil
• Fewer lines acquired translates to “acceleration”
• **Question**: Which lines of data should we acquire? (i.e. how to subsample k-space?)
Advantages of non-uniform subsampling

• Uniform sub-sampling produces a very computationally tractable reconstruction problem. (Cartesian SENSE)

• Typical images have higher signal energy at low frequencies

• Acquiring more densely at center of $k$–space will
  • improve SNR
  • reduce visible artifacts
  • *but*, substantially increase the computational cost (until now)
Data acquisition model

Parallel MRI signal model

\[ s_l(k) = \int_V W_l(r) \rho(r) e^{j2\pi k \cdot r} dr. \]

- \( \rho(r) \): distribution of excited spins in field of view (FOV)
- \( W_l(r) \): coil sensitivity within FOV
- \( k \): Fourier encoding domain coordinates
- \( r \): spatial domain coordinates
- \( s_l(k) \): acquired signal data for coil \( l \)
- \( L \) coils: \( L > 1 \) for parallel MRI
2D Discrete Model

Discrete signal in so-called \( k \)-space:

\[
s_l(k_y, k_x) = \sum_x \sum_y W_l(y, x) \rho(y, x) e^{-i2\pi(k_y y + k_x x)},
\]

number of discrete points \( x, y, k_x \) is the same. (i.e. 256, 512).

In matrix form:

\[
S_l = F_y (W_l \circ R) F_x,
\]

where \( F_x, F_y \) are (subsampled) versions of DFT matrices, \( R \) is the desired image.
Model Used for Reconstruction

Preprocessing: By orthogonality of the rows of $F_x$,

$$
\hat{S}_l = \left( \frac{1}{n_x} \right) S_l F_x^* = F_y (W_l \odot R),
$$

Matching columns on the right with columns on the left:

$$
\hat{S}_l(:, m) = F_y \text{diag}(W_l(:, m)) R(:, m).
$$

We have one such equation for each coil $l$. Hence,

$$
\begin{bmatrix}
\hat{S}_1(:, m) \\
\hat{S}_2(:, m) \\
\vdots \\
\hat{S}_L(:, m)
\end{bmatrix} =
\begin{bmatrix}
F_y \text{diag}(W_1(:, m)) \\
F_y \text{diag}(W_2(:, m)) \\
\vdots \\
F_y \text{diag}(W_L(:, m))
\end{bmatrix} R(:, m)
$$
Linear System Notation

\[
\begin{bmatrix}
\hat{S}_1(:, m) \\
\hat{S}_2(:, m) \\
\vdots \\
\hat{S}_L(:, m)
\end{bmatrix}
= 
\begin{bmatrix}
F_y\text{diag}(W_1(:, m)) \\
F_y\text{diag}(W_2(:, m)) \\
\vdots \\
F_y\text{diag}(W_L(:, m))
\end{bmatrix}R(:, m)
\]

So for each of the \( m = 1, \ldots, n_y \) columns of the image we have a (rectangular, since number of coils usually exceeds the acceleration factor) linear system

\[
s_m = P_m \rho_m.
\]

For convenience, we drop the subscript and consider

\[
s = P \rho.
\]
Goals of 2D PMRI

We subsample in $k_y$ space. That is, $F_y$ contains only a few rows of a square DFT matrix.

- All coils acquire data at the same time.
- The time for a coil to obtain its data is proportional to the number of k-space lines sampled (i.e. to the number of rows in $F_y$).

Example: For 256 k-space lines, if we want to accelerate by a factor of 4, we could subsample every 4th line, so $F_y$ is $64 \times 256$. A scan that took 16 min. now takes about 4 min. We will sample non-uniformly, however.
Discrete System

So, \( s = P\rho \) is the discrete model.

- \( P \) is ill-conditioned, rectangular, complex, dense.
- \( P \) can be large (2D, \( O(n_x) \times n_x \); 3D, \( O(n_x^2) \times n_x^2 \)).
- \( s \) contains noise \( \Rightarrow \) regularization.
- Iterative algorithms good candidates because:
  - fast matrix-vector products are possible.
  - \( P \) has tightly clustered spectrum.
Matrix Structure

Recall for a particular column $m$ in the 2D recon problem:

$$P = \begin{bmatrix}
F_y\text{diag}(W_1(:, m)) \\
F_y\text{diag}(W_2(:, m)) \\
\vdots \\
F_y\text{diag}(W_L(:, m))
\end{bmatrix}$$

Observations:

- Matvecs with diagonal matrices, fast.
- Matvecs with DFT matrices, fast via FFT.
- $\Rightarrow$ matvecs with $P (P^H)$ done quickly, without explicitly storing entires of $P$. 
Ensures the number of iterations needed by Krylov subspace iterative method will be relatively small
Regularization

One can recast image reconstruction as a minimization problem

\[
\min_{\rho} \| s - P\rho \|_2^2.
\]

Adding regularization terms allows one to balance between noise reduction and artifact suppression.

\[
\min_{\rho} \| s - P\rho \|_2^2 + \lambda^2 \| \rho \|_2^2
\]

Iterative algorithms, such as Conjugate Gradient (CG), are employed to provide a reasonable reconstruction time.
Iterative Options

Assume $\lambda$ is fixed. It is a simple exercise to show solving

$$\min_{\rho} \| P\rho - s \|_2^2 + \lambda^2 \| \rho \|_2^2$$

is equivalent to solving

$$\min_{\rho} \left\| \begin{bmatrix} P \\ \lambda I \end{bmatrix} \rho - \begin{bmatrix} s \\ 0 \end{bmatrix} \right\|.$$

Naive (traditionally used in MR literature) approach is to solve the normal equations for this using CG...
The Normal Equations

The associated normal equations are

\[(PHP + \lambda^2 I)\rho = PHs.\]

At each CG iteration,

- \(\rho_k^\lambda\) is an approximate solution to the system,
- and \(\rho_k^\lambda\) lives in the k-dimensional Krylov subspace

\[\text{span}\{PHs, (PHP + \lambda^2 I)PHs, \ldots, (PHP + \lambda^2 I)^{k-1}PHs\} \]
An Alternative

Projecting the large-scale problem to a smaller one, using Krylov vectors, the multiple $\lambda$ Tikhonov problem is more computationally tractable. To see this, consider the following...

Rather than solving

$$(P^HP + \lambda^2 I)\rho = PHs$$

with CG, solve the LS problem

$$\min_{\rho} \| \begin{bmatrix} P \\ \lambda I \end{bmatrix} \rho - \begin{bmatrix} s \\ 0 \end{bmatrix} \|$$

with LSQR$^2$.

$^2$ Paige and Saunders, ACM Trans Math Software, 1982
For the moment, assume that $\lambda$ is fixed and let $\rho_0 = 0$.

LSQR produces iterates $\rho_k^\lambda$ such that

$$
\rho_k^\lambda = \arg \min_{\rho \in K_k} \left\| \begin{bmatrix} P \\ \lambda I \end{bmatrix} \rho - \begin{bmatrix} s \\ 0 \end{bmatrix} \right\|
$$

where

$$
K_k = \text{span}\{ P^H s, (P^H P + \lambda^2 I) P^H s, \ldots, (P^H P + \lambda^2 I)^{k-1} P^H s \} = \text{span}\{ P^H s, (P^H P) P^H s, \ldots, (P^H P)^{k-1} P^H s \}
$$

is a $k$-dimensional Krylov subspace. Note the Krylov subspace is independent of $\lambda$. 
LSQR - CG connection

Mathematically, this is the same $\rho_k^\lambda$ after $k$ steps of CG applied to the regularized normal equations for this fixed $\lambda$. 
The Projected Problem

Through a few linear algebra steps, we convert our problem to

\[
\min_{\rho \in K_k} \left\| \begin{bmatrix} P & \lambda I \end{bmatrix} \rho - \begin{bmatrix} s \\ 0 \end{bmatrix} \right\| \quad \rightarrow \quad \min_{z} \left\| \begin{bmatrix} B_k & \lambda I \end{bmatrix} \begin{bmatrix} z \\ \beta e_1 \end{bmatrix} \right\|
\]

where \( B_k \) is \((k + 1) \times k\) lower bidiagonal. Example,

\[
B_2 = \begin{bmatrix} a_1 & 0 \\ b_1 & a_2 \\ 0 & b_2 \end{bmatrix}.
\]

Note independence of \( \lambda \): \( B_k \) is independent of \( \lambda \), \( z \) is not.
The Projected Problem

Through a few linear algebra steps, we convert our problem to

$$\min_{\rho \in K_k} \left\| \begin{bmatrix} P & \lambda I \end{bmatrix} \rho - \begin{bmatrix} s & 0 \end{bmatrix} \right\| \rightarrow \min_{z} \left\| \begin{bmatrix} B_k & \lambda I \end{bmatrix} z - \beta e_1 \right\|$$

where $B_k$ is $(k + 1) \times k$ lower bidiagonal. Example,

$$B_3 = \begin{bmatrix} a_1 & 0 & 0 \\ b_1 & a_2 & 0 \\ 0 & b_2 & a_3 \\ 0 & 0 & b_3 \end{bmatrix}.$$

*Note independence of $\lambda$: $B_k$ is independent of $\lambda$, $z$ is not.*
Advantage of Projected Problem

\[
\min_z \left\| \begin{bmatrix} B_k & \lambda I \\ \end{bmatrix} z - \beta e_1 \right\|
\]

is a *smaller* problem than the original if \( k \) is small.

Specifically,

\[2N \times N \text{ LS problem} \rightarrow (2k + 1) \times k \text{ LS problem.}\]

In the worst case, due to the structure of \( B_k \), cost of solving this is \( O\{k\} \) flops. Determining \( \rho_k^\lambda \) naively is \( O\{Nk\} \) flops.

Fortunately, we can do the latter by *short term recurrences.*
Summary of Projected Problem

Solving

$$\min_z \left\| \begin{bmatrix} B_k \\ \lambda I \end{bmatrix} z - \beta e_1 \right\|$$

is equivalent to the Tikhonov regularized problem

$$\min_z \|B_k z - \beta e_1\|^2 + \lambda^2 \|z\|^2.$$ 

For large enough $k$, $B_k$ inherits the ill-conditioning of $P$, so the projected problem is itself a regularized problem.

We still need to find $\lambda$ ...
Choosing $\lambda$

*Finding the best $\lambda$ can be expensive*

One option: use the L-curve\[^3\] approach

- **plot** $\log(||P\rho^{(\lambda)} - s||)$ vs. $\log(||\rho^{(\lambda)}||)$

- Point of maximum (positive) curvature corresponds to a good tradeoff between the residual error and noise in the solution.

However, naively solving multiple Tikhonov problems for a set of $\lambda$’s, is very expensive!

\[^3\] P. C. Hansen, SIAM Press, 1998
Surprising Results

Theorem:

\[ \| B_k z^{(\lambda)} - \beta e_1 \| = \| P \rho_k^{(\lambda)} - s \| \]
\[ \| z^{(\lambda)} \| = \| \rho_k^{(\lambda)} \| \]

Observations:
- The terms on the right are easily computed from short-term recurrences \textit{without} even computing \( \rho_k^{(\lambda)} \)!
- This correspondence means the L-curve for the projected problem gives an L-curve for the original problem for a fixed \( k \).
LSQR-Hybrid Summary

**Bottom line**\(^4\): For minimal overhead, we can find both a good value for \(\lambda\) and the corresponding regularized solution.

- Run LSQR for \(\lambda = 0\) case.
- For each iteration \(k\):
  - Major cost is matvec with \(P\) and \(PH\).
  - Generating points on L-curve only \(\mathcal{O}\{N\lambda\}\) flops.
  - Can update (all) \(\rho_k^{(\lambda)}\) to use for at \(k + 1\), \(\mathcal{O}\{NN\lambda\}\),
- Stop at \(k\) value chosen after \(\lambda^*\’s\) settle down.
- Return that \(\rho_k^{\lambda^*}\).

**Observation:** \(k\) will not be large due to clustered spectrum.

\(^4\) Kilmer and O’Leary, SIMAX, 2001
Example

Acceleration factor = 3, exponentially-weighted sub-sampling density, 8-channel FSE data acquired on a 1.5T GE Signa EXCITE scanner.

\[ \lambda = 1.0000 \times 10^{-4} \]
Example

Acceleration factor = 3, exponentially-weighted sub-sampling density, 8-channel FSE data acquired on a 1.5T GE Signa EXCITE scanner.

\[ \lambda = 6.2506 \times 10^3 \]
Example

Acceleration factor = 3, exponentially-weighted sub-sampling density, 8-channel FSE data acquired on a 1.5T GE Signa EXCITE scanner.

\[ \lambda = 1.4563 \times 10^4 \]
Example

Acceleration factor = 3, exponentially-weighted sub-sampling density, 8-channel FSE data acquired on a 1.5T GE Signa EXCITE scanner.

\[ \lambda = 7.9060 \times 10^4 \]
Computational Comparison

Reconstruction time comparison for multiple $\lambda$ L-curve generation. (Matlab implementation on a modest PIII)

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Fast FFT mat-vec prod</th>
<th>Embedded L-curve generation</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSQR</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>CG</td>
<td>X</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of $\lambda$ solutions:</th>
<th>1</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>25</th>
<th>40</th>
<th>50</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSQR recon time (sec):</td>
<td>14</td>
<td>16</td>
<td>20</td>
<td>22</td>
<td>30</td>
<td>37</td>
<td>45</td>
</tr>
<tr>
<td>CG recon time (sec):</td>
<td>9</td>
<td>49</td>
<td>97</td>
<td>144</td>
<td>240</td>
<td>417</td>
<td>518</td>
</tr>
</tbody>
</table>
On MR Host Reconstruction Results

C-code version running on 1.5 T GE MR scanner (EXCITE 11m4)

3.5x acceleration (74 of 256 lines), 8 channel cardiac coil, FGRE SSFP (FIESTA)

Self-referenced coil sensitivities

20 λ’s, 10 iterations

128x128 images requires ~ 1.5 sec
256x256 images requires ~ 4 sec
Conclusions and Future Work

- LSQR-Hybrid provides a mechanism for fast regularized reconstruction
- Enables us to subsample non-uniformly (traditionally too expensive relative to SENSE) and still obtain high quality images
- $\lambda$ for each column
- Non-identity regularizers (within reason) possible in this framework
- Move to 3D (larger savings)
- Merge with temporal filtering methods
- Certain other patentable ideas that cannot be mentioned