Optimization methods in cardiac MRI

Chong Chen, Department of Biomedical Engineering

Introduction

Magnetic Resonance Imaging (MRI) is a non-invasive tool that provides a comprehensive assessment of the soft tissues. The measured signal (k-space) in MRI is the Foulier Transformation (FT) of the image. Once a sufficient number of k-space samples is acquired (fully sampled, based on Nyquist-Shannnon sampling theorem), the image can be reconstructed by applying an inverse discrete FT. However, due to the low inefficient acquisition of MRI, acceleration techniques are usually applied on the modern MR scanner, such as parallel imaging (MR signal received by multiple coils distributed around the anatomy of interest) and compressed sensing (CS).

Parallel imaging alone can enable 2 to 3 fold acceleration. The reconstruction of parallel imaging is a least square problem. In this project, I implemented a class of first-order methods which include gradient method (GM), fast gradient method (FGM) and optimized gradient method (OGM) (1) to solve the least square problem. The performances of these methods were evaluated in the retrospectively undersampled 2D+t cardiac cine data with uniform sampling pattern (acceleration rate 2, 4 and 6).

Combining the parallel imaging with the CS recovery, several studies have demonstrated the feasibility of dynamic MRI with improved resolution and reduced acquisition time using a high acceleration rate ($\sim$ 8). CS utilizes the prior knowledge that the image is sparse in a specific domain. If the sparsifying transform is invertable and the l_1 penalty is used, the reconstruction of CS can be modeled as the l_1 penalized linear regression, which is also known as the Lasso problem. I used three methods for the reconstruction of compressive sensing: (I) iterative soft-thresholding algorithm/proximal gradient method (ISTA/PGM)(2, 3), (II) the fast iterative softthresholding algorithm (FISTA/FPGM) (4, 5) and (III) the proximal optimized gradient method (POGM)(6). Their performances were compared using retrospectively undersampled 2D+t cardiac cine data with pesudo random sampling pattern (acceleration rate 4, 6 and 8).

The rest of this report is organized as below. First the MRI forward model is introduced briefly. Then I reviewed the optimization methods implemented in this project, and presented the numerical results for parallel imaging (PI) and compressive sensing (CS) respectively. Finally a short summary

was given.

MRI forward model

The measured MRI signal (k-space data) are samples of the continuous Fourier transform of the underlying 2D or 3D image. In all modern MRI scanners, the k-space data are collected in parallel using multiple receive coils, and can be expressed as:

$$b = \underbrace{DFS}_{A} x + n \qquad [1]$$

where $x \in \mathbb{C}^{N\times 1}$ is a vectorized spatiotemporal image; S contains the coil sensitivity maps [1], F is the discrete Fourier transform, D is the undersampling matrix, $b \in \mathbb{C}^{M\times 1}$ is measured noisy k-space and $n \in \mathbb{C}^{M\times 1}$ is the Gaussian noise. MRI reconstruction is to recover the image $\widehat{x}$ from

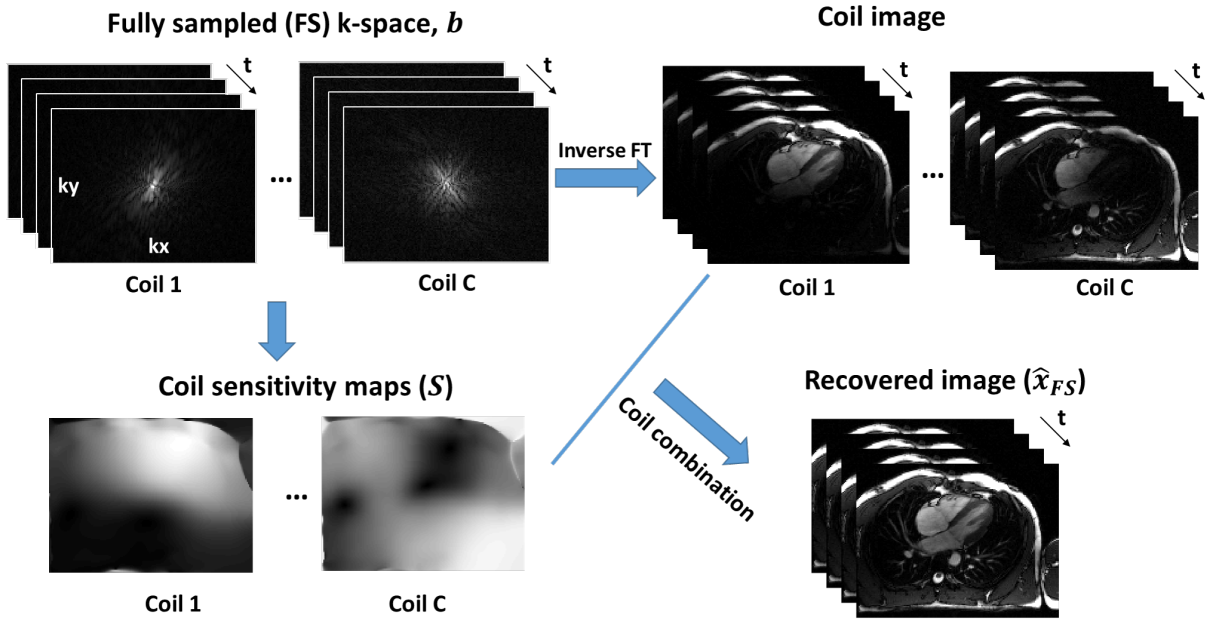

Figure 1: Reconstruction of cardic cine image from the fully sampled (FS) k-space ($\widehat{x}_{FS}$ as the reference)

[1] In this work, I use the same coil sensitivity maps extracted from the fully sampled dataset in all retrospectively undersampling studies. It is reasonable since this project only focuses on the convergence speed of different methods.

the sampled noisy k-space data b. If the MR signal is fully sampled ($M = NC > N$),

$$D = I, A^H A = I \quad [2]$$

$$\Longrightarrow \widehat{x} = \underset{x}{\operatorname{argmin}} ||Ax - b||_2^2 = (A^H A)^{-1} A^H b = S^H F^H b \quad [3]$$

i.e. the image $\widehat{x}$ can be obtained using the inverse FT followed by the adaptive coil combination, as shown in Fig.1. In this work, the image $\widehat{x}_{FS}$ recovered from a fully sampled 2D+t cardiac cine MRI dataset was chosen as the reference, and the normalized MRSE $||\widehat{x} - \widehat{x}_{FS}||_2 / ||\widehat{x}_{FS}||_2$ (NMRSE) was used to evaluate the quality of the image reconstructed from the retrospectively undersampled data.

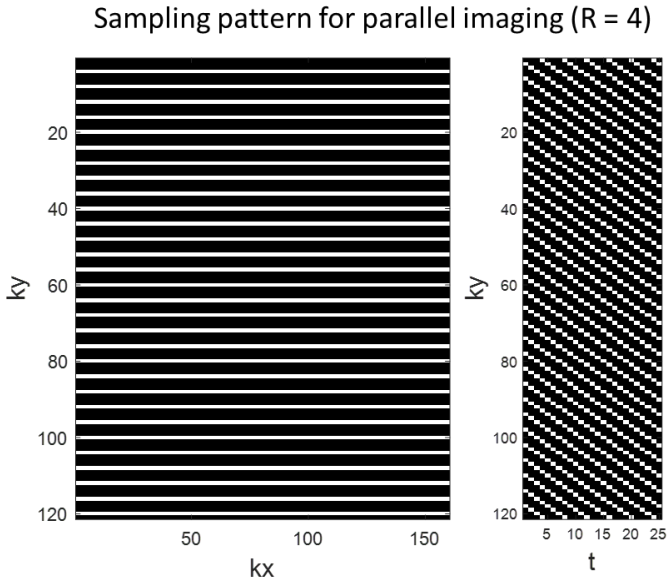

Figure 2: Time interleaved uniform sampling pattern with acceleration rate 4

Parallel imaging (PI)

In parallel imaging, the k-pace are undersampled uniformly. Fig.2 shows an example of the time interleaved uniform sampling pattern used in this project with acceleration rate 4. Due to the nature of MRI acquisition, there is no undersampling along the readout (k_x) dimension and the sampling pattern is the same for all the receive coils. The reconstruction of parallel imaging is to solve the

following optimization problem:

$$\widehat{x} = \underset{x}{\operatorname{argmin}} \; ||Ax - b||_2^2 \qquad [4]$$

where $x \in \mathbb{C}^{N\times 1}$ is a vectorized spatiotemporal image; $b \in \mathbb{C}^{M\times 1}$ is measured noisy k-space; $A = DFS \in \mathbb{C}^{M\times N}$ is the MRI forward operator.

Optimization methods for PI

This work focused on the first-order methods of the form shown in Algorithm 1 for solving the optimization problem 4. Different choices of the parameters $(\alpha, \beta_k, \gamma_k)$ lead to different methods. The fast gradient method (FGM) (with $\gamma_k = 0$ in Algorithm 1) accelerates the gradient method (GM) (with $\beta_k = \gamma_k = 0$) using the momentum term $\beta_k(y_{k+1} - y_k)$ with negligible additional computation. The optimized gradient method (OGM) uses an *over-relaxation* term $\gamma_k(y_{k+1} - x_k) = -\gamma_k \alpha \nabla f(x_k)$ for further acceleration (1). Table. 1 summarizes the standard choice of $(\alpha, \beta_k, \gamma_k)$ for GM, FGM, OGM and their worst-case rates for smooth convex function $f(x)$, which has Lipschitz continuous gradient with Lipschitz constant L.

Algorithm 1 Accelerated First-order Methods

Input: cost function f(x) = $||Ax - b||_2^2$, $y_0 = x_0$.
1: **for** $k = 0$ to K **do**
2: $y_{k+1} = x_k - \alpha \nabla f(x_k) = x_k - \alpha(2A^H(Ax_k - b))$
3: $x_{k+1} = y_{k+1} + \beta_k(y_{k+1} - y_k) + \gamma_k(y_{k+1} - x_k)$
4: **end for**
Output: Reconstructed image, $\widehat{x} = x_{K+1}$

Retrospectively undersampling study for PI

The cardiac cine dataset (fully sampled) was retrospectively undersampled using uniform sampling pattern with acceleration rate 2, 4 and 6. Then, the Lipschitz constant of $\nabla f(x) = 2A^H(Ax - b)$ is calculated $L = 2\lambda_{max}(A^H A)$, where $\lambda_{max}(A^H A)$ is the largest eigenvalue of matrix $A^H A$ and was estimated using power iteration with a random start. The optimization problem 4 is solved using GM, FGM and OGM respectively, with the stepsize $\alpha = 1/L$ and the total number of iteration $K = 150$.

Table 1: Accelerated First-order Methods for Smooth Convex Problems

Method	α	β_k	γ_k	Worst-case Rate
GM	$\frac{1}{L}$	0	0	$f(y_k)-f(x_*)\leq\frac{L\|\|x_0-x_*\|\|^2}{4k+2}$
FGM	$\frac{1}{L}$	$\frac{t_k-1}{t_{k+1}}$	0	$f(y_k)-f(x_*)\leq\frac{L\|\|x_0-x_*\|\|^2}{2t_{k-1}^2}\leq\frac{2L\|\|x_0-x_*\|\|^2}{(k+1)^2}$ $f(x_k)-f(x_*)\leq\frac{L\|\|x_0-x_*\|\|^2}{2t_k^2}\leq\frac{2L\|\|x_0-x_*\|\|^2}{(k+2)^2}$
OGM	$\frac{1}{L}$	$\frac{t_k-1}{t_{k+1}}$	$\frac{t_k}{t_{k+1}}$	$f(y_k)-f(x_*)\leq\frac{L\|\|x_0-x_*\|\|^2}{4t_{k-1}^2}\leq\frac{L\|\|x_0-x_*\|\|^2}{(k+1)^2}$
Parameters				
$t_0=1,\ \ t_k=\frac{1}{2}\left(1+\sqrt{1+4t_{k-1}^2}\right),\ k=1,...,K$ L is the Lipschitz constant of gradient $\nabla f(x)$				

Fig. 3 demonstrates the reconstructed image, the error map, and the convergent speeds with different optimization methods and acceleration rates for parallel imaging. As expected, the convergence speed of FGM and OGM are much faster than GM, and OGM has the best convergence performance. For $R = 4$ and 6, GM didn't converge within 150 iterations due to its slow convergence rate. For $R = 6$, FGM and OGM diverged due to the high acceleration rate (the optimization problem 4 was ill posed).

Compressive sensing (CS)

One of the requirements of compressive sensing is the incoherence of undersampling artifacts, which is generally achieved by employing Cartesian patterns with random or pseudo-random sampling. In MRI, the center of k-space has higher SNR. So pseudo-random sampling with the central region of k-space sampled more frequently is a popular scheme. Fig. 4 shows the pseudo-random VISTA sampling pattern (7) used in this work with acceleration rate 6. The reconstruction of dynamic MRI image from the highly undersampled k-space can by modeled as the following ℓ_2-ℓ_1 optimization problem:

$$\widehat{x} = \underset{x}{\operatorname{argmin}} \frac{1}{\sigma^2} \|Ax - b\|_2^2 + \lambda \|\Psi x\|_1 , \qquad [5]$$

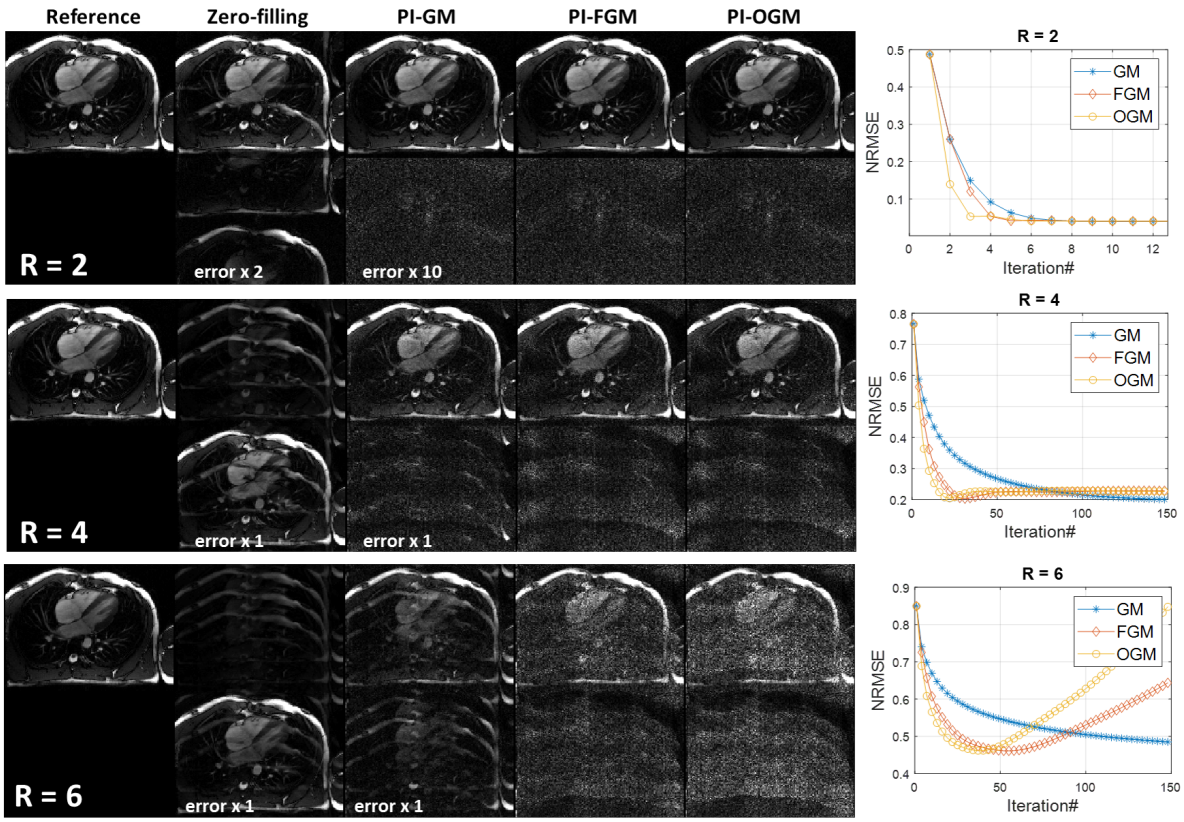

Figure 3: Reconstructed cardiac images (systolic frame), the error maps and the convergent speeds using different optimization methods (GM, FGM, OGM) and acceleration rates (R = 2, 4, 6) for parallel imaging (PI). The stepsize $\alpha = 1/L$ and the total number of iteration $N = 150$. As expected, the convergence speed of FGM and OGM are much faster than GM, and OGM has the best performance. For $R = 4, 6$, GM didn't converge within 150 iterations. For $R = 6$, FGM and OGM diverged due to the high acceleration rate.

where $x \in \mathbb{C}^{N\times 1}$ is a vectorized spatiotemporal image; $b \in \mathbb{C}^{M\times 1}$ is measured noisy k-space; σ^2 is noise variance in the measured data; λ is the regularization weight; $A = DFS \in \mathbb{C}^{M\times N}$ is the MRI forward operator; and $\Psi \in \mathbb{C}^{L\times N}$ is a spatiotemporal sparsifying transform. In this work, Ψ was chosen as Temporal Fourier Transform, which is invertable. Then Eq. 5 can be rewritten as the l_1 penalized liner regression problem:

$$\widehat{z} = \underset{z}{\operatorname{argmin}} \left\| \tilde{A}z - b \right\|_2^2 + \tilde{\lambda} \left\| z \right\|_1 \qquad [6]$$

where $\tilde{A} = A\Psi^{-1}$, $\tilde{\lambda} = \lambda\sigma^2$ and the reconstructed image $\widehat{x} = \Psi^{-1}\widehat{z}$.

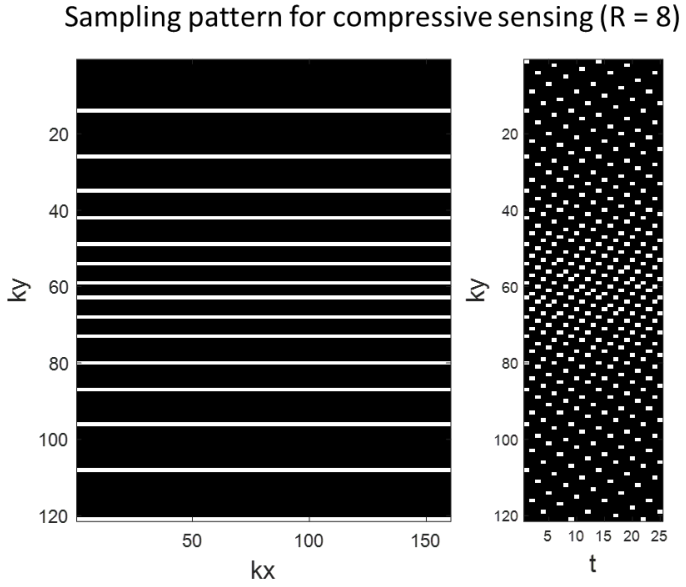

Figure 4: Pseudo-random VISTA sampling pattern with acceleration rate 6

Optimization methods for CS

To solve the optimization problem 6, three methods were implemented in Matlab: (I) the iterative soft-thresholding algorithm (ISTA/PGM) shown in Alogrithm 2, which is very simple, but has an undesirably slow convergence bound $\mathbb{O}(1/k)$, with k denoting the number of iterations. (II) the fast iterative softthresholding algorithm (FISTA/FPGM) as shown in Algorithm 3, with an $\mathbb{O}(1/k^2)$ convergence bound; and (III) the proximal optimized gradient method (POGM) (Algorithm 4), which has a worst-case convergence bound about twice as good as that of the FISTA.

Algorithm 2 the iterative soft-thresholding algorithm (ISTA)

Input: $f(z) = ||\tilde{A}z - b||_2^2$, $g(z) = \tilde{\lambda}|z|_1$, L–Lipschitz constant of $\nabla f(z)$

1: $\alpha = 1/L$.

2: **for** $k = 0$ to K **do**

3: $z_{k+1} = \text{prox}_{\alpha g(z)}\{z_k - \alpha \nabla f(z_k)\} = \text{prox}_{\alpha g(z)}\left\{z_k - \alpha(2\tilde{A}^H(\tilde{A}z_k - b))\right\}$

4: **end for**

Output: Reconstructed image, $\widehat{x} = \Psi^{-1}\widehat{z} = \Psi^{-1}z_{K+1}$

Algorithm 3 fast iterative soft-thresholding algorithm (FISTA)

Input: $f(z) = ||\tilde{A}z - b||_2^2$, $g(z) = \tilde{\lambda}|z|_1$, L–Lipschitz constant of $\nabla f(z)$

1: $y_1 = z_0; t_1 = 1, \alpha = 1/L$

2: **for** $k = 1$ to K **do**

3: $z_k = \text{prox}_{\alpha g(z)}\{y_k - \alpha \nabla f(y_k)\} = \text{prox}_{\alpha g(z)}\left\{y_k - \alpha(2\tilde{A}^H(\tilde{A}y_k - b))\right\}$

4: $t_{k+1} = \frac{1+\sqrt{1+4t_k^2}}{2}$

5: $y_{k+1} = z_k + \frac{t_k - 1}{t_{k+1}}(z_k - z_{k-1})$

6: **end for**

Output: Reconstructed image, $\widehat{x} = \Psi^{-1}\widehat{z} = \Psi^{-1}z_{K+1}$

Algorithm 4: POGM

initialize k = 0, $\tau_0 = 1$, y_0, z_0; $\beta = L$

while $k \leq K$ **do**

  if $k < K - 1$ **then**

    $\tau_{k+1} \leftarrow \frac{1+\sqrt{1+4\tau_k^2}}{2}$

  else

    $\tau_{k+1} \leftarrow \frac{1+\sqrt{1+8\tau_k^2}}{2}$

  end

  $\gamma_{k+1} \leftarrow \frac{1}{\beta}\frac{2\tau_k + \tau_{k+1} - 1}{\tau_{k+1}}$;

  $x_{k+1} \leftarrow y_k - \frac{1}{\beta}\nabla f(y_k)$;

  $z_{k+1} \leftarrow x_{k+1} + \frac{\tau_k - 1}{\tau_{k+1}}(x_{k+1} - x_k) + \frac{\tau_k}{\tau_{k+1}}(x_{k+1} - y_k) + \frac{\tau_k - 1}{\beta\gamma_k\tau_{k+1}}(z_k - y_k)$;

  $y_{k+1} \leftarrow \text{prox}_{\gamma_{k+1}g}(z_{k+1})$

end

Retrospectively undersampling study for CS

The cardiac cine dataset was retrospectively undersampled using pseudo-random sampling pattern with acceleration rate 4, 6 and 8. Then the Lipschitz constant of $\nabla f(z)$ is calculated $L = 2\lambda_{max}(\tilde{A}^H\tilde{A})$ using power iteration with a random start. Finally, the optimization problem 6 is solved using ISTA, FISTA and POGM respectively, with the stepsize $\alpha = 1/L$ and total number of iteration $K = 150$. The regularization strength $\tilde{\lambda}$ in the optimization problem 6 was manually tuned for $R = 6$, and the same $\tilde{\lambda}$ was used for $R = 4$ and 8.

Fig. 5 demonstrates the reconstructed image, the error map, and the convergent speeds with different optimization methods and acceleration rates for CS. As expected, the convergence speed of FISTA and POGM are much faster than ISTA, and POGM has the best convergence performance.

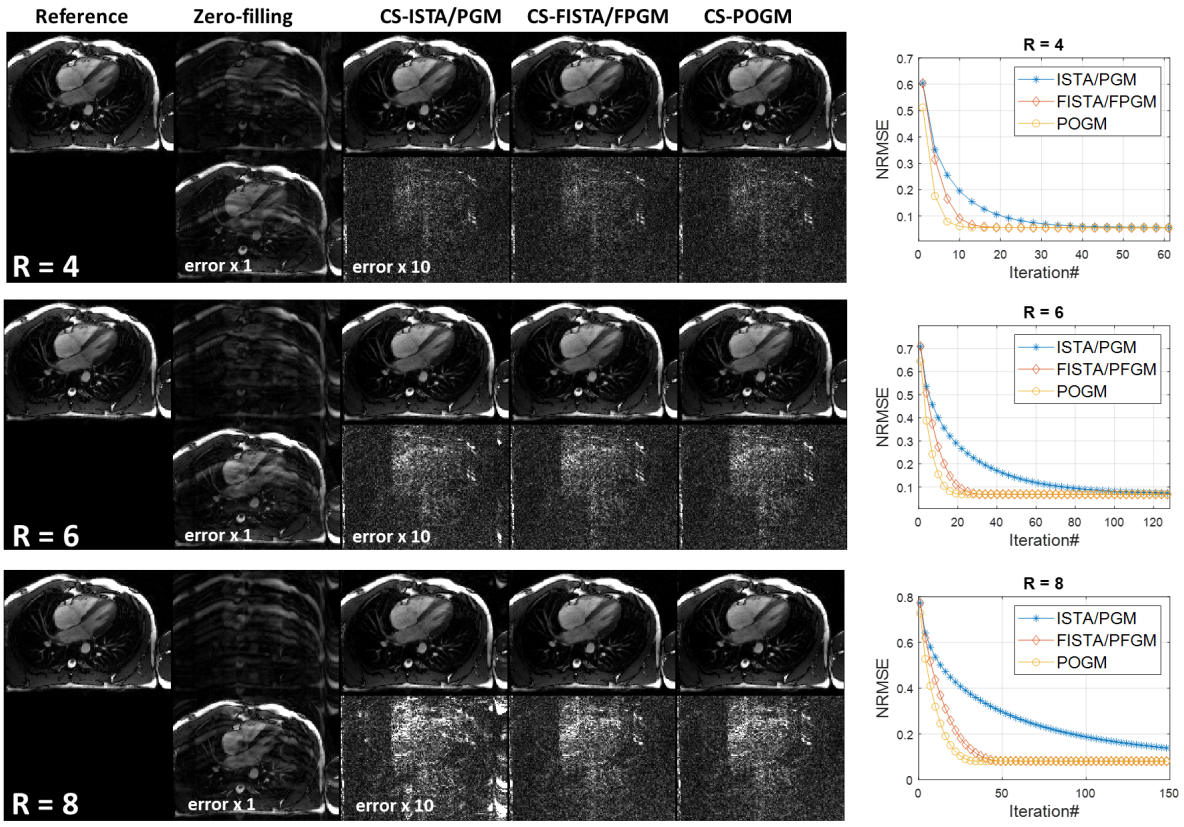

Figure 5: Reconstructed cardiac images (systolic frame), the error maps and the convergent speeds using different optimization methods (ISTA, FISTA, POGM) and acceleration rates (R = 4, 6, 8) for compressive sensing (CS). The stepsize $\alpha = 1/L$ and the total number of iteration $N = 150$. As expected, the convergence speed of FISTA and POGM are much faster than ISTA, and POGM has the best convergence performance. For $R = 8$, ISTA didn't converge within 150 iterations.

For $R = 8$, ISTA didn't converge within 150 iterations. The image quality reconstructed from compressive sensing is consistently better than the parallel imaging.

Summary

In this project, I implemented and compared several optimization method for parallel imaging (least square problem) and compressive sensing (Lasso problem) in cardiac MRI. The numerical results are as expected: the momentum term and over-relaxation term can accelerate the convergence significantly with negligible additional computation. Another fact I noticed was that different optimization methods converged to the same stationary point finally, which is also as expected. Compared to the parallel imaging, the image quality reconstructed from compressive sensing is

consistently better due to the regularization term. The Matlab codes used in this project are avaible in GitHub: `https://github.com/MRIOSU/ECE5759_Project`.

References

[1] Kim D, Fessler JA. Adaptive restart of the optimized gradient method for convex optimization. Journal of Optimization Theory and Applications 2018;178(1):240–263.

[2] Daubechies I, Defrise M, De Mol C. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences 2004;57(11):1413–1457.

[3] Combettes PL, Pesquet JC. Proximal splitting methods in signal processing. In: Fixed-point algorithms for inverse problems in science and engineering Springer; 2011.p. 185–212.

[4] Beck A, Teboulle M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM journal on imaging sciences 2009;2(1):183–202.

[5] Beck A, Teboulle M. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. IEEE transactions on image processing 2009;18(11):2419–2434.

[6] Kim D, Fessler JA. Adaptive restart of the optimized gradient method for convex optimization. Journal of Optimization Theory and Applications 2018;178(1):240–263.

[7] Ahmad R, Xue H, Giri S, Ding Y, Craft J, Simonetti OP. Variable density incoherent spatiotemporal acquisition (VISTA) for highly accelerated cardiac MRI. Magnetic resonance in medicine 2015;74(5):1266–1278.