Advancing Scan-specific, Parameter-free Artifact Reduction in K-space (SPARK) with Gradient-based Optimization

Junfu Cheng and Vladislav Bukshtynov (Florida Institute of Technology)

Introduction

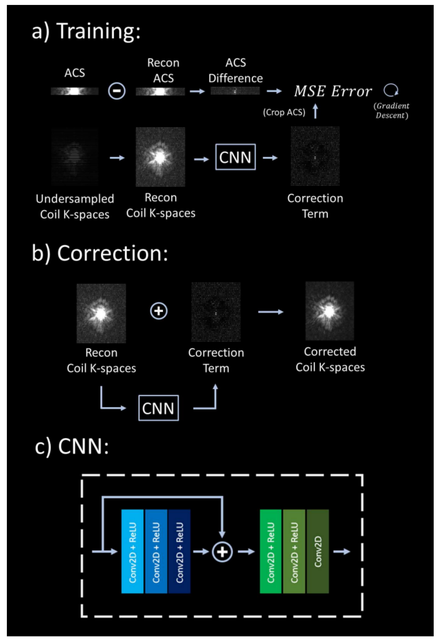
Magnetic resonance imaging (**MRI**) is a modality to collect information inside the organism. Partially Parallel Acquisition (**PPA**) uses spatial information contained in the component coils of an array to replace spatial encoding, typically performed using gradients, thereby reducing imaging time. Parallel imaging reconstruction for accelerated acquisitions of MRI is generally posed as an optimization problem. A convolutional neural network (**CNN**) approach that suppresses construction artifacts at high accelerations is **SPARK** which matches the difference between the acquired autocalibration lines and their erroneous reconstruction and generalizes this error term over the entire K-space [1]. We aim to find the optimal combination of strategies of gradient-based optimization methods for neural networking training to advance SPARK.

SPARK: Training CNN

Optimization: Each SPARK network S_c for coil c is trained by minimizing **objective (loss) function**

$$\min_{w_c} \|crop_{ACS}[S_c(w_c;k)] - d_c\|_2,$$

where w_c are kernel weights, k are reconstructed K-spaces across all coils, $d_c = ACS_c - crop_{ACS}[k_c]$ is the difference between the ACS lines of coil c and the corresponding lines in the reconstructed K-space for the same coil (where reconstruction is performed without substituting the ACS back).



Gradient-based Optimization

Gradient-based optimization is an iterative approach for improving solution $\vec{x}$ starting from its initial guess $\vec{x}_0$, i.e.,

$$\vec{x}^{k+1} = \vec{x}^k + \alpha^k \vec{d}^k,$$
$$\vec{x}^0 = \vec{x}_0,$$

where $\vec{d}^k = -D^k \nabla f(\vec{x}^k)$ is gradient-based search direction to minimize $f(\vec{x}^k)$.

- **Steepest Descent (SD) method**:

$$D^k = D^k_{SD} = I^{n\times n},$$

where I is an identity matrix (symmetric and positive definite).

- **Conjugate Gradient (CG) method** generates Q-conjugate directions using the procedure based only on search directions and gradients computed at the current step k and stored from the previous iteration $k-1$:

$$d^0 = -\nabla f(\vec{x}^0),$$
$$d^k = -\nabla f(\vec{x}^k) + \beta^k \vec{d}^{k-1},$$
$$\beta^k = \frac{\nabla f(\vec{x}^k)^T \nabla f(\vec{x}^k)}{\nabla f(\vec{x}^{k-1})^T \nabla f(\vec{x}^{k-1})}.$$

- **Adaptive Moment Optimization (ADAM)**: combines the **heuristics** of both Momentum and RMSProp [2]:

$$v^{k+1} = \beta_1 v^k - (1-\beta_1)\nabla f(\vec{x}^k),$$
$$s^{k+1} = \beta_2 s^k - (1-\beta_2)\|(\nabla f(\vec{x}^k))\|^2,$$
$$(\Delta\vec{x})^k = -\eta \frac{v^k}{\sqrt{s^k+\epsilon}} \nabla f(\vec{x}^k),$$
$$\vec{x}^{k+1} = \vec{x}^k + (\Delta\vec{x})^k,$$

where η is initial learning rate, v is exponential average of gradients, s is exponential average of square of gradients (magnitudes), and β_1, β_2 are hyperparameters.

We apply different gradient-based optimization schemes in the neural network single channel training to minimize the mean squared error (**MSE**) as a loss function value: e.g., steepest descent, conjugate gradient, and ADAM optimization algorithms. These methods are supplied with various methodologies for the optimal step size search: e.g., simple "correction"

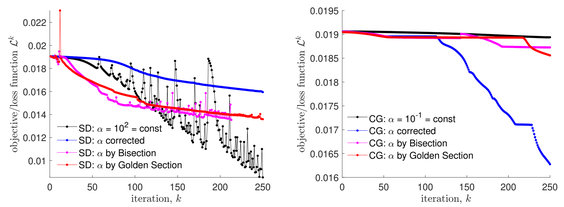
$$\text{if} \quad f(\vec{x}^{k+1}) > f(\vec{x}^k): \ \alpha \leftarrow \frac{\alpha}{2} \quad \text{or} \quad \eta \leftarrow \frac{\eta}{2},$$

and also bisection and golden section methods.

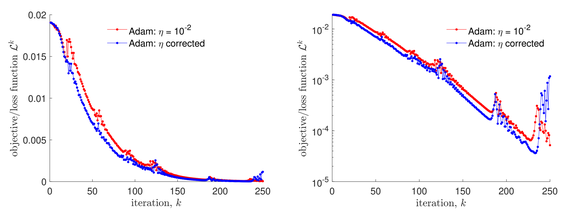
Results

Loss function value changes: the channel 0i trained loss function value (MSE) changes in the 250 or 350 iterations when combinations of different gradient-based optimization schemes are used.

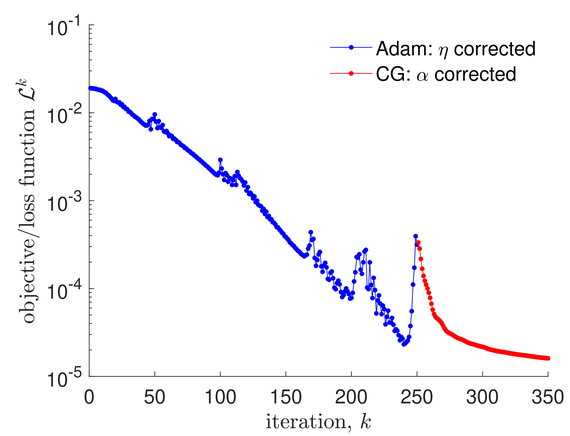
- MSE by Steepest Descent vs. Conjugate Gradient



- MSE by ADAM

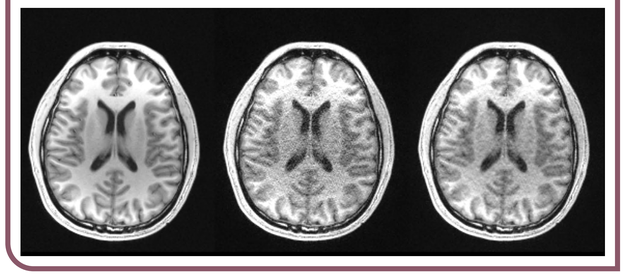


- MSE by ADAM combined with Conjugate Gradient



Conclusion

Promising strategy: gradient-based optimization when ADAM is combined with the Conjugate Gradient method using corrected learning rate η and step size α, respectively. According to channel 0i training test results, the gradient-base optimization using ADAM (initial learning rate is 10^{-2}) and CG (max step size is 10^{-1}) in the first 250 and last 100 iterations, respectively, has the best performance to make the MSE convergence to the minimal value. The AdAM + CG algorithm with corrected learning rate/step size should train all 31 channels of neural network connections between inputs and outputs.



References

[1] **O. Beker, C. Liao, J. Cho, Z. Zhang, K. Setsompop, and B. Bilgic.** *Scan-specific, Parameter-free Artifact Reduction in K-space (SPARK)*, arXiv:1911.07219, 2019.

[2] **D. P. Kingma, J. Ba.** *Adam: A method for Stochastic Optimization*, arXiv:1412.6980, 2014.