Momentum optimization for iterative shrinkage algorithms in parallel MRI with sparsity-promoting regularization

Matthew J. Muckley[1], Douglas C. Noll[1], and Jeffrey A. Fessler[2]
[1]Biomedical Engineering, University of Michigan, Ann Arbor, MI, United States, [2]Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, United States

Target Audience: MR physicists and engineers working on accelerated imaging methods.

Purpose: Combinations of parallel MRI and compressed sensing have been proposed for reducing MRI scan time.[1,2] One can combine parallel MRI with compressed sensing by minimizing a cost function of the form, $\Psi(x) = \frac{1}{2}\|y - Ax\|_2^2 + \beta\|Rx\|_1$, where A is a SENSE system matrix and R is a sparsity-promoting transform (e.g., orthogonal wavelets). This cost function is difficult to minimize since the ℓ_1 term is nondifferentiable. Variable-splitting methods are one option for minimizing this function, but these require tuning of penalty parameters.[2] Majorize-minimize methods are an alternative that do not have these same penalty parameters, but they do require a tight bound for the behavior of $RA'AR'$.[3] BARISTA is an algorithm that gives a procedure for computing these bounds and combines it with momentum and adaptive momentum restarting.[4] Here, we review the BARISTA approach for orthogonal wavelets and propose a new momentum update that has a faster convergence rate. We also compare BARISTA to the AL-P2 algorithm,[2] a comparison that had not been made previously.

Methods: For brevity, we only describe the procedure for orthogonal wavelets. When the regularizer uses orthogonal wavelets we can instead optimize $\widetilde{\Psi}(z) = \frac{1}{2}\|y - AR'z\|_2^2 + \beta\|z\|_1$, where $x = R'z$. Majorize-minimize methods require finding a surrogate function and then minimizing the surrogate. One such surrogate is to replace the quadratic term with $\phi_k(z) = \frac{1}{2}\left\|z - \left(z^{(k)} - D_R^{-1}RA'\left(AR'z^{(k)} - y\right)\right)\right\|_{D_R}^2$, where $z^{(k)}$ is the estimate of z at the kth iteration and D_R is a diagonal matrix such that $D_R \succcurlyeq RA'AR'$. The minimum of $\phi_k(z) + \beta\|z\|_1$ is calculated via the ℓ_1-shrinkage operator. In Cartesian SENSE MRI, $A'A \preccurlyeq S'S$, where S is a block-column matrix of sensitivity coil profiles. If R is an orthogonal wavelet transform, then we construct D_R by taking maximums over patches of the sum-of-squares of the sensitivity maps corresponding to the support size of the wavelet coefficient of interest.[4,5] To minimize $\widetilde{\Psi}(z)$ we iteratively apply shrinkage to $\phi_k(z) + \beta\|z\|_1$ with D_R constructed in this manner. This approach extends to analysis regularizers such as total variation.[4] Accelerating the method with momentum[3] and adaptive restarting[6] gives BARISTA.[4] We propose to further accelerate BARISTA by using a new momentum update: $u^{(k+1)} = z^{(k+1)} + \frac{\tau^{(k)}-1}{\tau^{(k+1)}}\left(z^{(k+1)} - z^{(k)}\right) + \frac{\tau^{(k)}}{\tau^{(k+1)}}\left(z^{(k+1)} - u^{(k)}\right)$, which gives a theoretical factor of 2 increase in convergence speed of the cost function.[7] We applied BARISTA with this new momentum term and compared convergence speed to previous methods. Our experiments consisted of collecting a 144 by 256 by 128 sample 3D data set on a GE 3T scanner with an 8-channel head coil. One slice was selected for experiments. The data were retrospectively downsampled with a Poisson-disk sampling pattern with a densely-sampled center. We then minimized $\Psi(x)$ with BARISTA, split Bregman (with optimized parameters for this data set, denoted SB), AL-P2 with parameters based on heuristics (denoted AL-P2),[2] AL-P2 with optimized parameters (denoted Al-P2, opt), and our proposed optimized momentum BARISTA (OMBARISTA) to compare convergence speed.

Results: We plot $\xi(k) = \frac{\|x^{(k)} - x^{(\infty)}\|}{\|x^{(\infty)}\|}$, the norm-residual to convergence, vs. time in all figures. $x^{(\infty)}$ was calculated by running many thousands of iterations. Figure 2 compares the convergence speed of the algorithms in the orthogonal Haar wavelet case, showing the $\sqrt{2}$-factor increase in norm-residual convergence speed. Figure 3 compares the convergence speed with undecimated Haar wavelets. The difference is not as large as with the orthogonal Haar case since the undecimated Haar algorithm requires solving a denoising subproblem,[4] but OMBARISTA is still the fastest method.

Discussion: In addition to converging rapidly, the methods presented use parameters that are easier to tune than the penalty parameters used by variable-splitting methods, making them more robust to use in a clinical setting. We also observed the theoretically-predicted increase in convergence speed with the new momentum term. In conclusion, we have made an improvement to a fast algorithm and observed faster convergence with the new algorithm than current state-of-the-art methods.

References: 1. Lustig et. al, MRM 2007, 2. Ramani et. al, IEEE-TMI 2011, 3. Beck et. al, SIAM-JIS 2009, 4. Muckley et. al, IEEE-TMI to appear, 5. Muckley et. al, ICIP 2014, 6. O'Donoghue et. al, FCM, 7. Kim et. al, arXiv 2014.

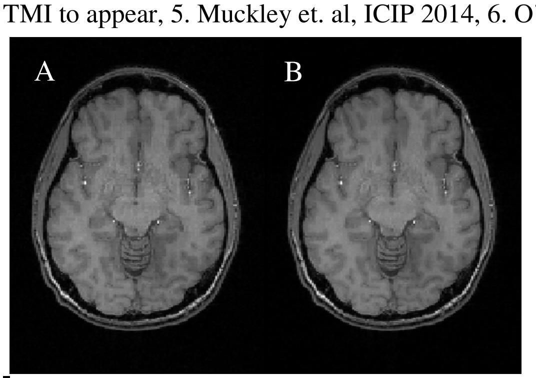

Figure 1: A) Image estimated with orthogonal Haar wavelets. B) Image estimated by minimizing $\Psi(x)$ with 2-level undecimated Haar wavelets

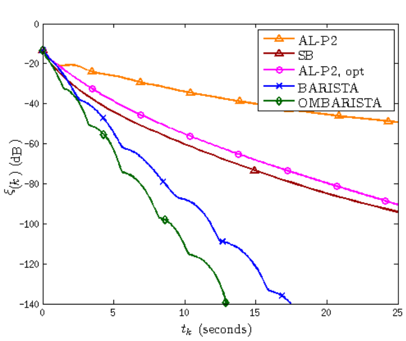

Figure 2: Convergence plot using orthogonal Haar regularizer, showing $\sqrt{2}$ increase in speed.

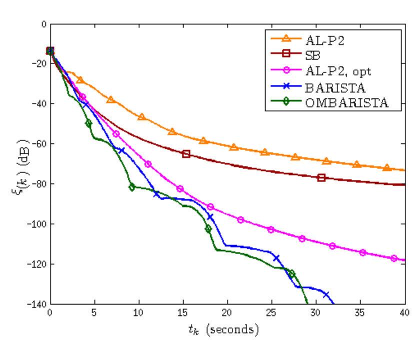

Figure 3: Convergence plot for minimizing $\Psi(x)$ with 2-level undecimated Haar wavelets. AL-P2 requires parameter optimization for this data set to have comparable speed.