ACCELERATED METHODS FOR LOW-RANK PLUS SPARSE IMAGE RECONSTRUCTION

Claire Yilin Lin, Jeffrey A. Fessler

University of Michigan, Mathematics and EECS Departments, Ann Arbor, MI, USA

ABSTRACT

The low-rank plus sparse (L+S) decomposition model enables the reconstruction of undersampled dynamic magnetic resonance imaging (MRI) data. Solving for the L and S components is a nonsmooth composite convex minimization problem. While current techniques for this model are based on the classical iterative soft thresholding algorithm (ISTA), accelerated methods can be applied to obtain faster rate of convergence of the algorithm. This paper proposes two alternative methods for solving the L+S problem, one based on the fast iterative shrinkage-thresholding algorithm (FISTA), and the other based on the recent proximal optimized gradient method (POGM). Numerical results suggest faster convergence than the traditional ISTA, while preserving its computational simplicity.

Index Terms— Low rank plus sparse model, dynamic MRI, accelerated first-order algorithms, proximal gradient methods

1. INTRODUCTION

One method for reconstructing undersampled dynamic magnetic resonance imaging (MRI) data uses the low-rank plus sparse (L+S) matrix decomposition, where the low-rank component L models the temporally correlated background, while the sparse component S models the dynamic information that lies on top of the background. With incoherence between the acquisition space and the representation space, the model offers high compressibility of dynamic MRI data, and has various applications in clinical studies, such as separating contrast enhancement from background, and efficient background suppression [1].

The L+S decomposition can be formulated as a convex optimization problem, in which the nuclear norm and the l_1 norm enforce respectively the low-rankness and the sparsity. Popular minimization techniques for this problem include the iterative soft thresholding (ISTA) [1], and the alternating direction methods of multipliers (ADMM) [2]. This paper considers two algorithms in the category of fast proximal gradient methods (FPGM), and compares convergence with ISTA that has the same convergence rate as the classical proximal gradient method (PGM).

We present an algorithm framework with two alternatives to ISTA, and numerically investigate their performance on the L+S model. The first one is based on the fast iterative shrinkage-thresholding algorithm (FISTA), which has a significantly improved global rate of convergence [3]. The other uses the recent proximal optimized gradient method (POGM), which numerically satisfies a worst-case cost function bound that is about twice better than FISTA [4]. Both methods preserve the computational simplicity of ISTA, thus accelerating the L+S reconstruction. Experiments on the retrospectively undersampled multicoil cardiac perfusion and cardiac cine data in [1] demonstrate that FISTA and POGM converge two to three times faster than ISTA, with the recent POGM approach being the fastest of all three methods. Finally, we briefly discuss the ADMM framework in [2], and its possible extension for improving its computational efficiency.

2. PROBLEM FORMULATION

We consider the problem in the setting of undersampled dynamic MRI reconstruction. Given undersampled k-t data d, an acquisition operator E, the L+S model aims to estimate space-time matrices L and S, where L is the low-rank component, and S has a sparse representation in the temporal frequency domain. The optimization problem is:

$$\min_{L,S} \frac{1}{2}\|E(L+S) - d\|_2^2 + \lambda_L\|L\|_* + \lambda_S\|TS\|_1, \quad (1)$$

where T is a sparsifying transform, which is the (unitary) temporal Fourier transform operator in our results. The regularization parameters λ_L, λ_S balance the contributions between data consistency, low-rankness, and sparsity [1]. Here the data consistency is enforced by the l_2-norm term, the low-rankness of L by the nuclear norm, and the sparsity of the transformed S by the l_1 norm.

To solve this problem using PGM, we reformulate (1) by concatenating L and S to form a single variable $X = \begin{bmatrix} L \\ S \end{bmatrix}$. Then an equivalent expression is

$$\min_X g(X) + h_1(X) + h_2(X), \quad \text{where} \quad (2)$$
$$g(X) = \frac{1}{2}\|[E \ E]X - d\|_2^2,$$
$$h_1(X) = \lambda_L\|[I \ 0]X\|_*, \quad \text{and} \quad h_2(X) = \lambda_S\|[0 \ T]X\|_1.$$

Here I and 0 are respectively the identity and the zero matrices, with the same size as L and S. Since $g(X)$ is smooth, convex, and continuously differentiable with Lipschitz continuous gradient, and $h_1(X)$, $h_2(X)$ are continuous, convex but nonsmooth, problem (2) can be solved by proximal gradient methods. In particular, the classical PGM computes iterates

$$X_k = \text{prox}_{h_i}\big(X_{k-1} - t\nabla g(X_{k-1})\big), \quad i = 1, 2, \quad \text{where} \quad (3)$$

$$\text{prox}_{h_i}(Y) = \underset{X}{\arg\min}\, h_i(X) + \frac{1}{2}\|X - Y\|_2^2$$

is the proximal operator of $h_i(X)$, and t is the stepsize, which depends on the Lipschitz constant $l(g)$.

Since h_1 and h_2 act upon the L and S components separately, and are represented by different norms, their proximal operators correspond to different expressions. In particular, prox_{h_1} is given by singular value thresholding, and prox_{h_2} by soft thresholding. Adopting the notations in [1], we define the soft thresholding operator by

$$\Lambda_\lambda(X) = \text{sign}(X)(|X| - \lambda)_+,$$

and the singular value thresholding operator by

$$\text{SVT}_\lambda(X) = U\Lambda_\lambda(\Sigma)V^*,$$

where $U\Sigma V^*$ is a singular value decomposition of X. Then the PGM iteration in (3) updates $X_k = \begin{bmatrix} L_k \\ S_k \end{bmatrix}$ by applying the two operators on L_k and S_k separately:

$$L_k = \text{SVT}_{\lambda_L}\big(L_{k-1} - t\, d(X_{k-1})\big), \quad \text{and}$$

$$S_k = T^{-1}\Big(\Lambda_{\lambda_S}\big[T\big(S_{k-1} - t\, d(X_{k-1})\big)\big]\Big), \quad \text{where} \quad (4)$$

$$d(X) = [I\ \ 0]\nabla g(X) = E^*([E\ \ E]X - d) = [0\ \ I]\nabla g(X).$$

Because the two gradient terms have the same expression, each iteration of the simultaneous update of L and S requires only one such evaluation. The algorithms below take advantage of this fact, and further accelerate it using FPGM.

3. ALGORITHM FRAMEWORK

We present a framework that accommodates three updates for the iterative scheme, using different proximal gradient methods. In each iteration, we update first the unknown X, then the gradient term to enforce data consistency. Written out explicitly in the algorithm, the update of X corresponds to computing the L and S components using singular value thresholding and soft thresholding. Adopting the notations in [1], which uses the ISTA update, we extend the algorithm to allow updates using FISTA and POGM in the proximal gradient step, achieving faster convergence rate. We first discuss the three different updates, then present the general framework in Section 3.4.

3.1. ISTA update

The ISTA update, presented in [1], uses classical PGM to update L_k and S_k in each iteration. Due to the same expression of the gradient term, an equivalent formulation of (4) is used in the update, so that the gradient is only evaluated once, in the data consistency step. Given initialization L_0, S_0 and $M_0 L_0 + S_0$, the updates for the unknown X_k and the data consistency term M_k are:

(3.1_{X_k}) $\quad L_k = \text{SVT}_{\lambda_L}(M_{k-1} - S_{k-1}),$
$\quad\quad\quad\quad S_k = T^{-1}\Big(\Lambda_{\lambda_S}\big[T\big(M_{k-1} - L_{k-1}\big)\big]\Big);$
(3.1_{M_k}) $\quad M_k = L_k + S_k - tE^*\big(E(L_k + S_k) - d\big).$

When E is normalized such that $\|E\| = 1$ for the fully sampled case, where $\|E\|$ is the spectral norm of E, the Lipschitz constant of ∇g satisfies $l(g) = 2\|E\|^2 \le 2$. Convergence of ISTA here requires $0 < t < \frac{2}{2\|E\|^2} = 1$ (see [3]), and we set $t = 0.99$. The convergence analysis of ISTA reduces to that of the classical gradient method (when $h_1(X) = h_2(X) = 0$), where the the sequence of function values converges to the optimal function value at a rate of $O(1/k)$ [3].

3.2. FISTA update

Among first order methods, Nesterov's fast gradient method (FGM) achieves the improved rate of $O(1/k^2)$[5]. [3] then extends this result to the case of proximal gradient methods, showing the same rate of convergence with FISTA. In particular, FISTA uses a secondary sequence $\widetilde{X}_k = \begin{bmatrix} \widetilde{L}_k \\ \widetilde{S}_k \end{bmatrix}$, while the main computational effort of the gradient evaluation remains the same as in ISTA. The FISTA initialization and updates of X_k and M_k are:

$(3.2 : I)$ $\quad \widetilde{X}_0 = X_0,\ \theta_0 = 1;$
$(3.2 : X_k)$ $\quad L_k = \text{SVT}_{\lambda_L}(M_{k-1} - \widetilde{S}_{k-1}),$
$\quad\quad\quad\quad S_k = T^{-1}\Big(\Lambda_{\lambda_S}\big[T\big(M_{k-1} - \widetilde{L}_{k-1}\big)\big]\Big),$
$\quad\quad\quad\quad \theta_k = \frac{1+\sqrt{1+4\theta_{k-1}^2}}{2},$
$\quad\quad\quad\quad \widetilde{X}_k = X_k + \frac{\theta_{k-1}-1}{\theta_k}(X_k - X_{k-1});$
$(3.2 : M_k)$ $\quad M_k = \widetilde{L}_k + \widetilde{S}_k - tE^*\big(E(\widetilde{L}_k + \widetilde{S}_k) - d\big).$

As in [3], we use the constant stepsize $t = \frac{1}{2\|E\|_2^2} = 0.5$, because the Lipschitz constant $l(g) \le 2$.

3.3. POGM update

In the smooth unconstrained setting, [6] introduces the optimized gradient method (OGM) that achieves a worst-case convergence bound twice as small as that of Nesterov's FGM. The POGM in [4] extends OGM to nonsmooth composite problems, with a numerical worst-case performance about twice better than that of FISTA. POGM uses an additional sequence $\overline{X}_k = \begin{bmatrix} \overline{L}_k \\ \overline{S}_k \end{bmatrix}$, while the computational simplicity

is preserved. POGM has a similar formulation as FISTA, requiring $O(Nd)$ arithmetic operations and $O(d)$ memory, for N iterations and d-dimensional variable. To obtain the expressions of the additional coefficients, [6] analytically optimizes the step size parameters of a fixed-step algorithm, by minimizing the upper bound on its worst-case. As in the FISTA update, we set $t = 0.5$, with the following initialization and update:

$(3.3 : I)$ $\quad \widetilde{X}_0 = \overline{X}_0 = X_0, \theta_0 = \zeta_0 = 1;$

$(3.3 : X_k)$ $\quad \widetilde{L}_k = M_{k-1} - S_{k-1},$

$\qquad \widetilde{S}_k = M_{k-1} - L_{k-1},$

$$\theta_k = \begin{cases} \frac{1+\sqrt{1+4\theta_{k-1}^2}}{2}, & k < N \\ \frac{1+\sqrt{1+8\theta_{k-1}^2}}{2}, & k = N \end{cases},$$

$\qquad \overline{X}_k = \widetilde{X}_k + \frac{\theta_{k-1}-1}{\theta_k}(\widetilde{X}_k - \widetilde{X}_{k-1})$

$\qquad + \frac{\theta_{k-1}}{\theta_k}(\widetilde{X}_k - X_{k-1}) + \frac{\theta_{k-1}-1}{\zeta_{k-1}\theta_k}t(\overline{X}_{k-1} - X_{k-1}),$

$\qquad \zeta_k = t\big(1 + \frac{\theta_{k-1}-1}{\theta_k} + \frac{\theta_{k-1}}{\theta_k}\big),$

$\qquad L_k = \text{SVT}_{\lambda_L}(\widetilde{L}_k),$

$\qquad S_k = T^{-1}\Big(\Lambda_{\lambda_S}\big[T(\overline{S}_k)\big]\Big);$

$(3.3 : M_k)$ $\quad M_k = L_k + S_k - tE^*\big(E(L_k + S_k) - d\big).$

3.4. General framework

For the L+S problem, the three algorithms share the same inputs, hence they can be viewed as different updates in a unified framework. While no further initialization is needed for $(3.1 : I)$, FISTA and POGM require initialization of the additional sequences for acceleration, as summarized below.

L+S Reconstruction Algorithm
Inputs:
$\quad d$: multicoil undersampled k-t data
$\quad E$: space-time multicoil encoding operator
$\quad T$: sparsifying transform
$\quad \lambda_L$: singular value threshold
$\quad \lambda_S$: sparsity threshold
$\quad i = 1, 2,$ or 3: update scheme
Initialization: $\quad M_0 = L_0 = E^*d, S_0 = 0, (3.i : I)$
for $k = 1, 2, \ldots, N$ **do**
$\quad X_k \leftarrow (3.i : X_k)$
$\quad M_k \leftarrow (3.i : M_k)$
end for
output: X_N

4. RESULTS

Based on the framework in Section 3.4, we implemented the L+S reconstruction methods in MATLAB, using two datasets considered in [1], with the same operators E, T and the same regularization parameters λ_L, λ_S.

In both experiments, since our only information is the multicoil undersampled data, we compared the convergence of the three updates by plotting $f^k - f^\infty$, where f^k is the
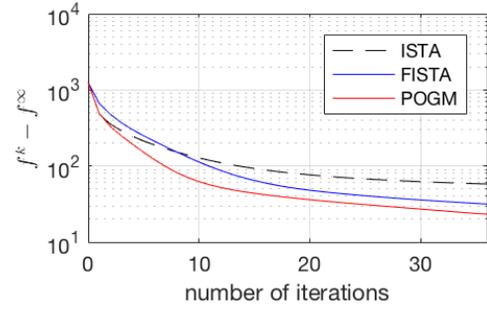

Fig. 1. Convergence of three updates in 36 iterations, on the cardiac perfusion dataset. POGM achieves faster convergence than ISTA and FISTA. Here the rank of L is 5 with ISTA, and 4 with FISTA and POGM, at 36 iterations.
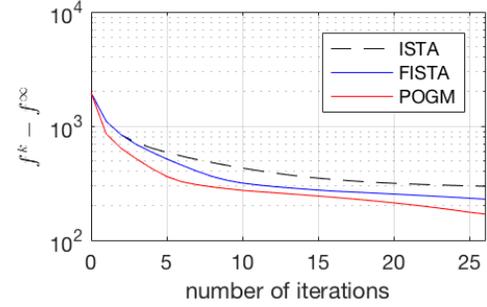

Fig. 2. Convergence of three updates in 26 iterations, on the cardiac cine dataset. FISTA and POGM achieve faster convergence than ISTA. Here the rank of L is 3 with ISTA, and 1 with FISTA and POGM, at 26 iterations.

cost function value of the kth iterate, and f^∞ is the average of the function values at the 100th iteration with FISTA and with POGM updates. Since the ISTA algorithm in [1] uses a stopping criterion, in each case we first performed ISTA until the criterion is met, and recorded the number of iterations and the function value. We then compared the results by FISTA and POGM with the same number of iterations, as well as the number of iterations needed for FISTA and POGM to obtain the same cost function value.

We first tested the framework on the undersampled cardiac perfusion data, where the image matrix size is 128×128, with 40 temporal frames and 12 coils [1]. The ISTA took 36 iterations to reach the stopping criterion that depends on the difference between successive function values. Fig. 1 shows that FISTA and POGM reach lower cost function values in 36 iterations. Reaching the same cost function value as ISTA required 17 iterations of FISTA, and only 11 iterations of POGM. FISTA in this case is slower than ISTA in the first few iterations, due to the smaller stepsize in the FISTA update. POGM, however, outperforms both ISTA and FISTA.

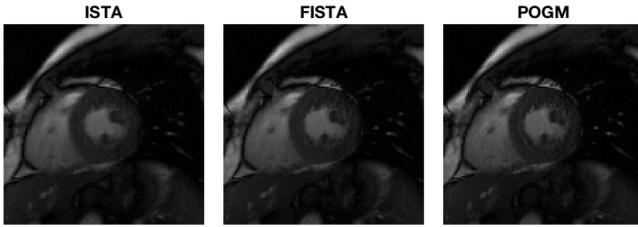Our second experiment compared the convergence of the three updates using the undersampled cardiac cine data,

Fig. 3. Comparison of the L+S reconstructed images on one temporal frame of the cardiac cine dataset, after 26 iterations of the three updates. Compared with ISTA, the results by FISTA and POGM provide better visualization of the structure.

where the image matrix size is 256×256, with 24 temporal frames and 12 coils [1]. Here ISTA took 26 iterations to reach the stopping criterion, while FISTA and POGM achieved the same cost function value with respectively 12 and 8 iterations. In this case, FISTA converges faster than ISTA, and POGM is faster still; see Fig. 2. Fig. 3 suggests that FISTA and POGM provide shaper image quality at 26 iterations. Additionally, ISTA gives rank$(L) = 3$, while FISTA and POGM give rank$(L) = 1$ at this number of iterations. Since problem (1) is not necessarily strictly convex, the algorithms may converge to different global minimizers. To compare the results, we performed more iterations of ISTA, and observed that it took 105 iterations to attain the same cost function value as with 36 iterations of FISTA, generating comparable image quality, and giving rank$(L) = 1$. This suggests that ISTA, in this case, converges to the same global minimizer as the other two updates. Hence the results with FISTA or POGM predict the convergence behavior of the low-rank component with the selected regularization parameter λ_L, as one performs more iterations of ISTA.

5. DISCUSSION AND CONCLUSION

We have presented an algorithm framework for the L+S dynamic image reconstruction problem. In place of using ISTA to estimate the low-rank and the sparse components, our framework considers accelerated methods with FISTA and POGM updates, while preserving the computational simplicity of ISTA. The faster convergence obtained by the new updates provides about three-fold acceleration in solving the L+S decomposition problems, with simple algorithm modifications. The new updates could also be used to accelerate the sparse coding step of dictionary-based algorithms for dynamic MRI reconstruction, such as the method in [7].

A variable splitting scheme with augmented Lagrangian has also been proposed to solve problem (1) [2]. There, ADMM is used to minimize the augmented Lagrangian function of four variables, leading to sub-problems with closed-form updates. Two of the sub-problems in [2] involve

quadratic cost functions that require computing $(E^*E + \delta I)^{-1}$ for some penalty parameter δ, but direct implementation of this inverse is impractical. However, one could circumvent this issue by further variable splitting. Specifically, one can represent $E = FC$, where F is a Fourier encoding matrix, and C an estimate of the sensitivity maps. Then the sub-problems are quadratic, but now with inverses of the forms $(F^*F + \delta I)^{-1}$ and $(C^*C + \delta I)^{-1}$, which can be computed efficiently due to the circulant behavior of F^*F and the diagonal structure of C^*C. This idea has been investigated in [8], and is an interesting future direction for further acceleration of the L+S reconstruction.

6. REFERENCES

[1] R. Otazo, E. Candès, and D. K. Sodickson, "Low-rank plus sparse matrix decomposition for accelerated dynamic MRI with separation of background and dynamic components," *Mag. Res. Med.*, vol. 73, no. 3, pp. 1125–1136, 2015.

[2] B. Trémoulhéac, N. Dikaios, D. Atkinson, and S. R. Arridge, "Dynamic MR Image Reconstruction–Separation From Undersampled (**k**,*t*)-Space via Low-Rank Plus Sparse Prior," *IEEE Trans. Med. Imag.*, vol. 33, no. 8, pp. 1689–1701, 2014.

[3] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[4] A. B. Taylor, J. M. Hendrickx, and F. Glineur, "Exact worst-case performance of first-order methods for composite convex optimization," *SIAM J. Optim.*, vol. 27, no. 3, pp. 1283–1313, 2017.

[5] Y. Nesterov, "A method of solving a convex programming problem with convergence rate $O(1/k^2)$," *Soviet Mathematics Doklady*, vol. 27, no. 2, pp. 372–376, 1983.

[6] K. Donghwan and J. A. Fessler, "On the convergence analysis of the optimized gradient method," *J. Optim. Theory Appl.* , vol. 172, no. 1, pp. 187–205, 2017.

[7] S. Ravishankar, B. E. Moore, R. R. Nadakuditi, and J. A. Fessler, "Low- rank and Adaptive Sparse Signal (LASSI) Models for Highly Accelerated Dynamic Imaging," *IEEE Trans. Med. Imag.*, vol. 36, no. 5, pp. 1116–1128, 2017.

[8] S. Ramani and J. A. Fessler, "Parallel MR image reconstruction using augmented Lagrangian methods," *IEEE Trans. Med. Imag.*, vol. 30, no. 3, pp. 694–706, 2011.