Combining Augmented Lagrangian Method with Ordered Subsets for X-Ray CT Reconstruction

Hung Nien and Jeffrey A. Fessler

Department of Electrical Engineering and Computer Science
University of Michigan, Ann Arbor, MI

***Abstract*—The augmented Lagrangian (AL) method (and its closely related cousin, the alternating direction method of multipliers, or in short, ADMM) is a powerful technique for solving ill-posed inverse problems using variable splitting. In this paper, inspired by the convergence analysis of a simplified CT problem with Tikhonov regularization, we focused on the diagonal preconditioned AL method, where the step size of each entry of the split variable is proportional to the statistical weighting in the penalized weighted least squares (PWLS) formulation. To solve the inner minimization problem efficiently, we used the ordered-subsets (OS) algorithm due to its fast convergence rate in early iterations. By combining AL method with OS, experimental results show that the standard OS algorithm can be accelerated remarkably.**

I. Introduction

The augmented Lagrangian (AL) method [1] has drawn more attention recently due to its scalability, simplicity, and fast convergence property. In the field of total-variation (TV) denoising and compressed sensing, the AL method is used to split a nonsmooth term, such as the TV-norm and ℓ_1-norm, in the variational formulation, yielding a subproblem that has a closed-form solution or can be solved almost exactly [2]. In the field of statistical X-ray computed tomography (CT) image reconstruction, the AL method is also used to separate the statistical weighting matrix (which has huge dynamic range) to make the inner least squares problem much easier to precondition [3]. Aside from the standard AL method, many extensions and variations have been proposed to further accelerate convergence. A survey can be found in [4].

One variation of the AL method is to precondition the ℓ_2 penalty term in the augmented Lagrangian by some positive definite matrix $\mathbf{G}$. For example, when $\mathbf{G}$ is a diagonal matrix with positive diagonal entries, we penalize each entry in the split variable differently, which means we can have larger steps for those entries that are still far from the solution by increasing the penalty. However, such a diagonal matrix is seldom used in practice because the diagonal preconditioning matrix sometimes can ruin the opportunity to exploit fast computation such as FFT and PCG for the inner problem in the AL method.

In statistical X-ray CT image reconstruction, the image reconstruction is usually formulated as a PWLS problem, and the ordered-subsets (OS) algorithm [5] can be used to accelerate its convergence in early iterations by a factor of M, the number of subsets. This M-time acceleration comes from the "subset balance condition" by grouping the projections into M ordered subsets and updating the image incrementally using the M subset gradients. Although the standard OS algorithm approaches some limit cycle eventually because of its incremental gradient descent structure, the M-time acceleration of solving a PWLS problem is still very promising for the AL method with *inexact* updates. In this paper, we first study the convergence of a simple quadratic PWLS problem using a general AL method to get intuition about how to choose the diagonal preconditioning matrix. Then, we relax the choice of preconditioned matrix by a scaling factor, apply it to the statistical X-ray CT image reconstruction problem, and solve the inner constrained PWLS problem by using the standard OS algorithm.

II. Method

To describe our proposed algorithm more clearly, we first define the statistically weighted CT reconstruction problem as follows:

$$\hat{\mathbf{x}} \in \operatorname*{argmin}_{\mathbf{x}\in\Omega} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_{\mathbf{W}}^2 + \mathsf{R}(\mathbf{x}) \right\}, \tag{1}$$

where $\mathbf{y}$ is the noisy post-logarithm sinogram, $\mathbf{A}$ is the system matrix of a CT scan, $\mathbf{W}$ is a diagonal weighting matrix accounting for measurement variance, R is an edge-preserving regularizer, and Ω is some convex set such as a box constraint on the solution. Instead of solving it directly using, for example, projected gradient descent method, we will focus on solving an equivalent constrained problem. That is, we are going to solve:

$$(\hat{\mathbf{x}}, \hat{\mathbf{u}}) \in \operatorname*{argmin}_{\mathbf{x}\in\Omega,\mathbf{u}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{u}\|_{\mathbf{W}}^2 + \mathsf{R}(\mathbf{x}) \right\} \text{ s.t. } \mathbf{u} = \mathbf{A}\mathbf{x}, \tag{2}$$

or equivalently, we must find a saddle point of the corresponding augmented Lagrangian of (2):

$$\mathcal{L}_A(\mathbf{x}, \mathbf{u}, \mathbf{d}) \triangleq \frac{1}{2} \|\mathbf{y} - \mathbf{u}\|_{\mathbf{W}}^2 + \mathsf{R}(\mathbf{x}) + \iota_\Omega(\mathbf{x}) + \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{u} - \mathbf{d}\|_{\mathbf{G}}^2, \tag{3}$$

where ι_Ω is the characteristic function of set Ω, $\mathbf{d}$ is the scaled dual variable of $\mathbf{u}$, and $\mathbf{G}$ is some positive definite matrix, e.g., $\mathbf{G} = \eta\mathbf{I}$ with $\eta > 0$. This problem can be solved by using the alternating direction method. In other words, we will minimize $\mathcal{L}_A$ with respect to $\mathbf{x}$ and $\mathbf{u}$ alternatively followed by a gradient ascent of $\mathbf{d}$, and the iterates will be:

$$\begin{cases} \mathbf{x}^{(j+1)} \in \underset{\mathbf{x}\in\Omega}{\operatorname{argmin}} \left\{ \frac{1}{2} \left\| \left(\mathbf{u}^{(j)} + \mathbf{d}^{(j)}\right) - \mathbf{A}\mathbf{x} \right\|_{\mathbf{G}}^2 + \mathsf{R}(\mathbf{x}) \right\} \\ \mathbf{u}^{(j+1)} = (\mathbf{W} + \mathbf{G})^{-1} \left(\mathbf{W}\mathbf{y} + \mathbf{G} \left(\mathbf{A}\mathbf{x}^{(j+1)} - \mathbf{d}^{(j)} \right) \right) \\ \mathbf{d}^{(j+1)} = \mathbf{d}^{(j)} - \mathbf{A}\mathbf{x}^{(j+1)} + \mathbf{u}^{(j+1)} . \end{cases} \tag{4}$$

A. Analysis of CT problem with Tikhonov regularization

To simplify the convergence rate analysis of the proposed algorithm, we first assume that $\mathsf{R}(\mathbf{x}) = \frac{\alpha}{2} \|\mathbf{C}\mathbf{x}\|_2^2$, and Ω is the entire space, where $\mathbf{C}$ is the finite difference matrix. Then, the iterates in (4) have closed-from expressions, and by doing some simple calculations, we can show that $\mathbf{u}^{(j)}$ converges to $\mathbf{A}\left(\mathbf{A}'\mathbf{W}\mathbf{A} + \alpha\mathbf{C}'\mathbf{C}\right)^{-1}\mathbf{A}'\mathbf{W}\mathbf{y} = \mathbf{A}\hat{\mathbf{x}}$ unconditionally and linearly with rate

$$\rho\Big((\mathbf{W} + \mathbf{G})^{-1} (\mathbf{G}\mathbf{A}\mathbf{F} + \mathbf{W}) \Big) , \tag{5}$$

where $\rho(\mathbf{K})$ denotes the spectral radius of matrix $\mathbf{K}$, and

$$\mathbf{F} \triangleq (\mathbf{A}'\mathbf{G}\mathbf{A} + \alpha\mathbf{C}'\mathbf{C})^{-1} \mathbf{A}' (\mathbf{G} - \mathbf{W}) . \tag{6}$$

Although there is no simple way to express the convergence rate in (5) using $\mathbf{G}$, one fairly good choice of $\mathbf{G}$ is $\mathbf{G} = \mathbf{W}$, thus leading to spectral radius of $1/2$, which is quite fast. However, if we set $\mathbf{G}$ to be $\mathbf{W}$, then the $\mathbf{x}$ subproblem is the original weighted CT problem with a different sinogram. In other words, the inner problem is as hard as the original problem itself, and we would gain nothing from the AL method.

B. Diagonal preconditioned AL method for CT problem

To gain something from the AL method, we must add one more degree of freedom. In this paper, we consider the preconditioning matrix $\mathbf{G} = \eta\mathbf{W}$ with $\eta > 0$, and the resulting iterates become:

$$\begin{cases} \mathbf{x}^{(j+1)} \in \underset{\mathbf{x}\in\Omega}{\operatorname{argmin}} \left\{ \frac{1}{2} \left\| \left(\mathbf{u}^{(j)} + \mathbf{d}^{(j)}\right) - \mathbf{A}\mathbf{x} \right\|_{\mathbf{W}}^2 + \eta^{-1}\mathsf{R}(\mathbf{x}) \right\} \\ \mathbf{u}^{(j+1)} = \frac{1}{1+\eta} \left(\mathbf{y} + \eta \left(\mathbf{A}\mathbf{x}^{(j+1)} - \mathbf{d}^{(j)} \right) \right) \\ \mathbf{d}^{(j+1)} = \mathbf{d}^{(j)} - \mathbf{A}\mathbf{x}^{(j+1)} + \mathbf{u}^{(j+1)} . \end{cases} \tag{7}$$

Intuitively, this approach penalizes the *more important* line integrals more, thus leading to *larger* step sizes for those rays. By solving the last two equations in (7), we can get the identity

$$\eta\mathbf{d}^{(j+1)} = \mathbf{y} - \mathbf{u}^{(j+1)} . \tag{8}$$

Substituting (8) into (7), the final iterates are:

$$\begin{cases} \mathbf{x}^{(j+1)} \in \underset{\mathbf{x}\in\Omega}{\operatorname{argmin}} \left\{ \frac{1}{2} \left\| \mathbf{z}^{(j)} - \mathbf{A}\mathbf{x} \right\|_{\mathbf{W}}^2 + \eta^{-1}\mathsf{R}(\mathbf{x}) \right\} \\ \mathbf{u}^{(j+1)} = \frac{1}{1+\eta} \left(\mathbf{u}^{(j)} + \eta\mathbf{A}\mathbf{x}^{(j+1)} \right) , \end{cases} \tag{9}$$

where $\mathbf{z}^{(j)} \triangleq \eta^{-1}\mathbf{y} + \left(1 - \eta^{-1}\right)\mathbf{u}^{(j)}$. As can be seen from (9), the $\mathbf{x}$ subproblem is a weighted CT problem with an *updated* sinogram and a *scaled* regularizer.

To implement the proposed diagonal preconditioned AL method, we need a method to solve the inner weighted CT problem in (9). The OS algorithm is a good candidate here because it is usually fast in early iterations, and it is very easy to impose box constraints on the inner problem. Note that, when η is equal to one, the $\mathbf{x}$ subproblem is exactly the same as the original problem, and the iterates reduce to the standard OS algorithm. Intuitively, if we use a noisy FBP reconstruction as the initial guess and if we expect that the converged image should be less noisy, then we would like to choose a small η so the $\mathbf{x}$ iterate is more regularized. In general, we choose η to be between 0.3 and 1 so that we will not regularize $\mathbf{x}$ too much $(\eta < 0.3)$ or too little $(\eta > 1)$.

C. Practical implementation and discussion

Although (9) outlines the proposed algorithm, we usually do not implement the algorithm exactly in that way. According to the convergence theorem of ADMM methods [6, Theorem 8], it suffices for the errors of the inner minimization problems to be absolutely summable. Therefore, to try to improve the convergence behavior of our AL method, we run multiple OS iterations to refine $\mathbf{x}$ before updating the split variable $\mathbf{u}$. The practical algorithm should be as follows:

$$\begin{cases} \mathbf{x}^{(j+1)} = \mathcal{OS}_M^1\Big(\mathbf{x}^{(j)}; \frac{1}{2} \left\| \mathbf{z}^{(j)} - \mathbf{A}\mathbf{x} \right\|_{\mathbf{W}}^2 + \eta^{-1}\mathsf{R}(\mathbf{x}), \Omega \Big) \\ \mathbf{u}^{(j+1)} = \begin{cases} \mathbf{u}^{(j)} & \text{, if } \operatorname{mod}(j+1, P) \neq 0 \\ \frac{1}{1+\eta} \left(\mathbf{u}^{(j)} + \eta\mathbf{A}\mathbf{x}^{(j+1)} \right) & \text{, otherwise} , \end{cases} \end{cases} \tag{10}$$

where $\mathcal{OS}_M^n(\mathbf{x}_0; \Psi, \mathcal{C})$ denotes n iterations (M sub-iterations per iteration) of the OS algorithm with initial guess $\mathbf{x}_0$, cost function Ψ, and constraint set $\mathcal{C}$, and P is the period of the split variable update. Furthermore, to minimize the error of $\mathbf{x}$ subproblem (at least for early iterations), we have to take advantage of the M-time acceleration of the OS algorithm, so the number of subsets should be large enough. However, using more subsets leads to a "larger" limit cycle, which will accelerate the error accumulation.

One could also accelerate the standard OS algorithm by starting from a larger regularization parameter (assuming the initial guess is noisy) and decreasing it gradually to one as the algorithm proceeds. The benefit of our proposed algorithm is that, thanks to the AL method, we do not have to use such continuation of the regularization parameter for convergence. Note that since the OS algorithm itself does not converge, instead of decreasing the regularization parameter (or η), we would reduce the number of subsets and increase the period of split variable update to achieve convergence in practice.

III. Result

In this section, we evaluate our proposed algorithm using a patient helical CT scan. To investigate the effects of η (the AL penalty parameter) and P (the update period), we consider three different AL penalty parameters (0.3, 0.5, and 0.7) and three different update periods (1, 5, and 10) in our experiment. The number of subsets is set to be 41. The standard OS algorithm is the baseline method. Note that each split variable update requires one "extra" forward projection compared to the standard OS algorithm. To have a fair comparison, we

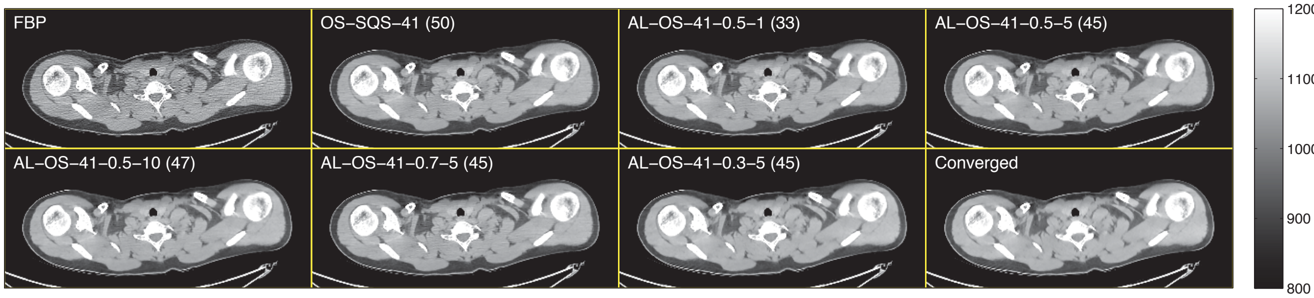


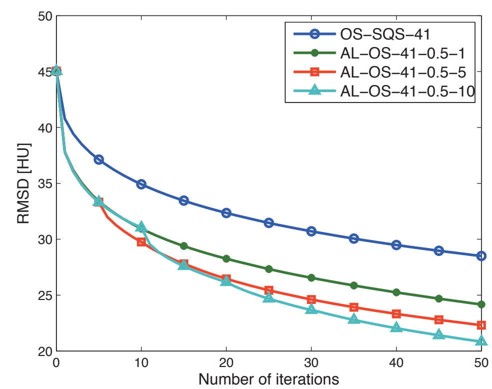
Fig. 1: Cropped images from the central slice of the reconstructed patient helical CT scan, where **FBP** denotes the FBP reconstruction, **OS-SQS-M** denotes the standard OS algorithm with M subsets, and **AL-OS-M-η-P** denotes the proposed algorithm with M subsets, the AL penalty parameter η, and the update period P. Numbers in parentheses show the number of iterations of each algorithm so that the total number of forward/back-projections is approximately 100. The **AL-OS-41-0.3-5** result after 45 iterations is very similar to the converged image, whereas the other images exhibit residual streak artifacts for the same computation time.

plot the root mean square (RMS) difference between the reconstructed image and the converged reconstruction as a function of the number of *iterations* and the number of *forward/back-projections* (assuming that $\mathbf{Ax}$ and $\mathbf{A}'\mathbf{y}$ have the same computational complexity). Lastly, since the test helical scan contains gain fluctuations [7], we include blind gain correction [8] in all of our reconstruction algorithms. With this correction, the weighting matrix $\mathbf{W}$ and the preconditioning matrix $\mathbf{G}$ are diagonal plus a rank-1 matrix rather than pure diagonal, which is a simple extension of the proposed diagonal preconditioned AL method.

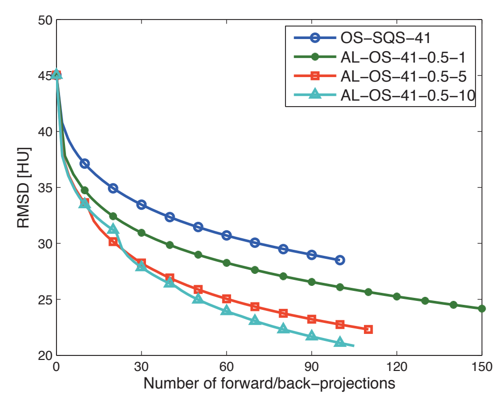
Figure 1 shows the initial noisy FBP image, the reconstructed images after about 100 forward/back-projections of the standard OS algorithm and the proposed algorithm using different values of η and P, and the converged image. As can be seen in Figure 1, the shading artifacts due to gain fluctuations are largely suppressed, and the proposed algorithm with all configurations outperforms the standard OS algorithm in image quality, especially for smaller η and larger P.

Figure 2 shows the convergence rate curves of the proposed algorithm with different values of P for the case $\eta = 0.5$, where **OS-SQS-M** denotes the standard OS algorithm with M subsets, and **AL-OS-M-η-P** denotes the proposed algorithm with M subsets, the AL penalty parameter η, and the update period P. As can be seen in Figure 2, the proposed algorithm with update period $P = 10$ converges much faster than the standard OS algorithm. There are sharp drops in the RMS difference when the split variable is updated, especially for larger P and in earlier iterations. This kind of acceleration diminishes as the algorithm proceeds because the speedup of OS algorithm saturates. To have more acceleration, we would need to either increase P or decrease M to solve the inner minimization problem in (9) more accurately.

Figure 3 shows the convergence rate curves of the proposed algorithm with different values of η for the case $P = 5$, where the naming convention is the same as in Figure 2. Note that the standard OS algorithm is just a special case of the proposed

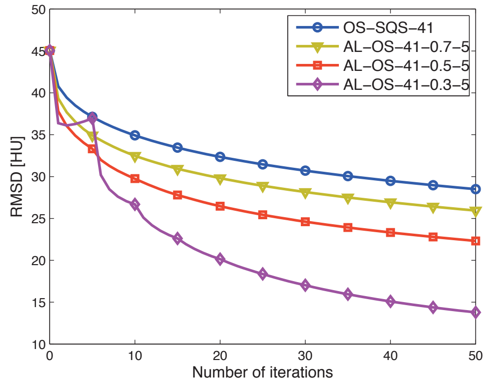


(a)

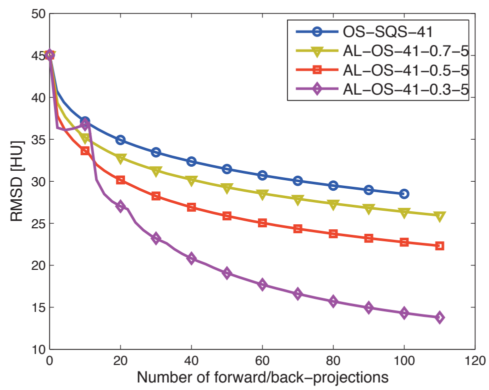


(b)

Fig. 2: RMS differences between the reconstructed image and the converged reconstruction as a function of (a) the number of iterations and (b) the number of forward/back-projections with different values of the update period P.



(a)



(b)

Fig. 3: RMS differences between the reconstructed image and the converged reconstruction as a function of (a) the number of iterations and (b) the number of forward/back-projections with different values of the AL penalty parameter η.

algorithm when $\eta = 1$. In this case, the value of P does not matter because $\mathbf{z}^{(j)}$ in (10) is independent of $\mathbf{u}^{(j)}$. As can be seen in Figure 3, the convergence rate curve converges to the curve of the standard OS algorithm as η approaches to unity. Smaller η shows faster convergence rate because the converged image is smooth and edge-preserved; however, when η is too small, for example, when $\eta = 0.3$, we can see the problem (sharp increase in RMS difference) of over-regularization in early iterations since the inner minimization problem is too different from the original problem. When the inner minimization problem is solved properly, i.e., smaller error due to larger P or M, this "misdirection" can be corrected by split variable updates, for example, the purple curves in Figure 3. Furthermore, although we consider only the standard OS algorithm in this paper, any fast variation of the OS algorithm, e.g., [9] and [10], can be applied to the proposed diagonal preconditioned AL method.

IV. Conclusion

In this paper, we proposed to combine the AL method with OS. Inspired by the convergence analysis of the AL method for quadratic PWLS problems, we focused on a diagonal preconditioning matrix $\mathbf{G}$ that is proportional to the statistical weighting matrix $\mathbf{W}$. Experimental results show that the proposed algorithm accelerated the standard OS algorithm remarkably and provides a degree of freedom to fine tune the convergence rate. As possible future work, we will investigate different splits in the proposed diagonal preconditioned AL method. In addition, we are also interested in combining, for example, the frequency analysis of the AL method with tuning the AL penalty parameter η.

Acknowledgment

This work is supported in part by NIH grant R01-HL-098686 and by an equipment donation from Intel. The authors also want to thank GE for providing the patient helical CT sinogram data.

References

[1] M. R. Hestenes, "Multiplier and gradient methods," *J. Optim. Theory Appl.*, vol. 4, pp. 303–20, Nov. 1969.
[2] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Trans. Im. Proc.*, vol. 20, no. 3, pp. 681–695, 2011.
[3] S. Ramani and J. A. Fessler, "A splitting-based iterative algorithm for accelerated statistical X-ray CT reconstruction," *IEEE Trans. Med. Imag.*, vol. 31, pp. 677–88, Mar. 2012.
[4] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. & Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.
[5] H. Erdoğan and J. A. Fessler, "Ordered subsets algorithms for transmission tomography," *Phys. Med. Biol.*, vol. 44, pp. 2835–51, Nov. 1999.
[6] J. Eckstein and D. P. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Mathematical Programming*, vol. 55, pp. 293–318, Apr. 1992.
[7] J.-B. Thibault, Z. Yu, K. Sauer, C. Bouman, and J. Hsieh, "Correction of gain fluctuations in iterative tomographic image reconstruction," in *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, pp. 112–5, 2007.
[8] H. Nien and J. A. Fessler, "Splitting-based statistical X-ray CT image reconstruction with blind gain correction," in *Proc. SPIE 8668 Medical Imaging 2013: Phys. Med. Im.*, 2013. To appear as 8668-53.
[9] D. Kim and J. A. Fessler, "Parallelizable algorithms for X-ray CT image reconstruction with spatially non-uniform updates," in *Proc. 2nd Intl. Mtg. on image formation in X-ray CT*, pp. 33–6, 2012.
[10] J. H. Cho and J. A. Fessler, "Accelerating ordered-subsets image reconstruction for X-ray CT using double surrogates," in *Proc. SPIE Medical Imaging 2012: Phys. Med. Im.*, p. 83131X, 2012.