Sparse shift-varying FIR preconditioners for fast volume denoising

Madison G. McGaffin and Jeffrey A. Fessler

***Abstract*—Splitting-based CT reconstruction algorithms decompose the reconstruction problem into a iterated sequence of "easier" subproblems. One relatively memory-efficient algorithm decomposes the reconstruction problem into a several subproblems, including a volume denoising problem. While easier to solve in isolation than jointly, these subproblems have highly shift-varying Hessians that are challenging to effectively precondition with circulant operators. In this work, we present an algorithm to design a positive-definite, Schatten p-norm optimal, finite impulse response (FIR) approximation to a given circulant matrix. With this algorithm, we generate efficient space-varying preconditioners for the volume denoising problem. We demonstrate that PCG with an efficient space-varying preconditioner can converge at least quickly as a split-Bregman-like algorithm while using considerably less memory.**

I. Introduction

Consider a statistical image reconstruction problem

$$\widehat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \left\{ J(\mathbf{x}) = \frac{1}{2}||\mathbf{A}\mathbf{x} - \mathbf{y}||_{\mathbf{W}}^2 + \mathsf{R}(\mathbf{C}\mathbf{x}) \right\}, \quad (1)$$

where $\mathbf{A} \in \mathbb{R}^{M \times N}$ is the system matrix, $\mathbf{W}$ is a diagonal matrix of statistical weights, and $\mathsf{R}(\mathbf{C}\mathbf{x})$ is a convex, smooth and edge-preserving regularizer:

$$\mathsf{R}(\mathbf{C}\mathbf{x}) = \sum_{d=1}^{N_d} \beta_d \sum_{j=1}^{N} \kappa_{d,j} \phi\Big([\mathbf{C}_d \mathbf{x}]_j \Big). \quad (2)$$

The $\{\mathbf{C}_d\}_{d=1}^{N_d}$ are circulant first-order difference matrices, *e.g.*, $N_d = 13$ for 26-neighbor differences in 3D CT, and the object-dependent but constant $\{\kappa_{d,j}\}_{d=1,j=1}^{N_d,N}$ control local regularizer strength [4]. The potential function ϕ is convex, smooth, nonnegative and even.

This minimization problem is challenging to solve directly due to the large dimension of $\mathbf{A}$, the nonlinearity of the regularizer, and the high spatial variance of the data-fit and regularizer Hessians, $\mathbf{A}'\mathbf{W}\mathbf{A}$ and $\nabla^2 \mathsf{R}(\mathbf{C}\mathbf{x})$, respectively.

Variable splitting may be used to introduce auxiliary variables to separate the terms in (1). Enforcing equality constraints between the new variables and linear functions of $\mathbf{x}$ then converts (1) into a new, equivalent, constrained minimization problem. The alternating directions methods of multipliers (ADMM) [2] may then be used to solve the new constrained optimization problem via an iterated sequence of optimization problems in each variable. This approach has the effect of splitting jointly difficult terms from one another, *e.g.*, the data-fit and regularization terms in (1). This technique has proved quite fruitful, and can handle both non-smooth regularizers and additional constraints like nonnegativity [10].

A relatively memory-efficient splitting introduces two auxiliary variables $\mathbf{u} = \mathbf{A}\mathbf{x}$ and $\mathbf{v} = \mathbf{x}$ to separate the data-fit and regularizer terms [8]. Applying ADMM to the resulting constrained optimization problem leads to an algorithm involving the following nontrivial inner optimization problems:

$$\mathbf{x}^{(j+1)} = \underset{\mathbf{x}}{\operatorname{argmin}} \frac{\mu_u}{2} \left|\left| \mathbf{A}\mathbf{x} - \left(\mathbf{u}^{(j)} - \boldsymbol{\eta_u}^{(j)} \right) \right|\right|^2 + \frac{\mu_v}{2} \left|\left| \mathbf{x} - \left(\mathbf{v}^{(j)} - \boldsymbol{\eta_v}^{(j)} \right) \right|\right|^2, \quad (3)$$

$$\mathbf{v}^{(j+1)} = \underset{\mathbf{v}}{\operatorname{argmin}} \frac{\mu_v}{2} \left|\left| \mathbf{v} - \left(\mathbf{x}^{(j+1)} + \boldsymbol{\eta_v}^{(j)} \right) \right|\right|^2 + \mathsf{R}(\mathbf{C}\mathbf{v}), \quad (4)$$

with the scalar parameters μ_u and μ_v and the dual variables $\boldsymbol{\eta_u}$ and $\boldsymbol{\eta_v}$ introduced by the ADMM algorithm.

The ADMM does not require these subproblems to be solved exactly but only with summable absolute error taken over all iterations [3]. In practice, the ADMM algorithm will almost certainly not be run to convergence, and experience indicates that more accurate solutions to the iterated subproblems improve convergence of the algorithm as a whole. Consequently fast, even if not exact, solvers to problems (3) and (4) are desirable.

Solving (3) and (4) in isolation is "easier" than solving them jointly, but challenges in each problem remain. The tomography problem Hessian, $\mu_v \mathbf{I} + \mu_u \mathbf{A}'\mathbf{A}$, while free of shift variance induced by the statistical weights, is still far more shift-varying in cone-beam CT than in 2D, and evaluating the gradient of (3) remains very computationally expensive. While regularizer gradient evaluations are less expensive, the regularizer Hessian in (4) is highly shift-varying.

The preconditioned conjugate gradients (PCG) algorithm is an attractive candidate for both the tomography and denoising subproblems.[1] If an effective preconditioning operator $\mathbf{P} \approx \left(\nabla^2 J\right)^{-1}$ can be found, PCG converges quickly, has modest memory constraints, and updates all coordinates of the iterate simultaneously (which is attractive for high-dimensional problems and modern parallel hardware). However, designing such preconditioners can be challenging.

Department of Electrical Engineering and Computer Science, University of Michigan, 1301 Beal Ave., Ann Arbor, MI 48109-2122, U.S.A. Email: {mcgaffin, fessler}@umich.edu. Supported in part by NIH grant R01 HL 098686 and CPU donations by Intel. CT sinograms provided by GE Healthcare.

[1] Other rapidly converging algorithms exist for the denoising problem in particular, *e.g.*, split-Bregman-like algorithms [6]. However, these can require a prohibitive amount of memory for large reconstruction problems, *e.g.*, typical 3D helical CT problems.

Previously, the authors have preconditioned both the denoising problem and the tomography problem with circulant matrices [8]. Circulant preconditioners are attractive in part because they allow the algorithm designer to derive a shift invariant approximation of the Hessian, and immediately receive an efficient implementation of the approximation's inverse using FFTs. These features make them a good "default" preconditioner choice, but leave room for improvement by preconditioners which consider the spatial variance of the Hessian.

Because PCG with an appropriate preconditioner converges to an acceptably accurate solution using fewer gradient evaluations, we must measure the computational cost of a preconditioner relative to the computational cost of a gradient evaluation. If a preconditioner takes too much time to apply relative to a gradient evaluation, it may be more efficient to use a less computationally expensive preconditioner. For the denoising problem (4), gradient evaluations are relatively inexpensive, even compared to applying an FFT. A more computationally efficient preconditioner is desirable; sparse FIR filters may be sufficient to replace the FFT operations used to implement the conventional preconditioner.

In this work, we propose shift-varying preconditioners to tackle the denoising problem (4). In Section II we present an algorithm to design positive-definite FIR approximations to approximate a given circulant matrix. In Section III, we use these FIR filters to generate new preconditioners for the denoising problem. Results of the filter design algorithm and comparisons with other preconditioners and denoising algorithms are given in Section IV.

A. Notation

If $\mathbf{F} \in \mathbb{R}^{N \times N}$ is a circulant matrix, we use the lowercase $\mathbf{f} \in \mathbb{R}^N$ to indicate the first column or kernel of $\mathbf{F}$. The DFT of a vector $\mathbf{f}$ will be written using a hat, *e.g.*, $\text{DFT}\{\mathbf{f}\} = \widehat{\mathbf{f}}$. The vectors and matrices of all ones and zeros are written $\mathbf{1}$ and $\mathbf{0}$ respectively, whose dimension should be clear from context.

B. Schatten p-norms

If $\mathbf{F}$ is a matrix, the Schatten p-norm of $\mathbf{F}$ is the corresponding vector p-norm applied to the singular values of $\mathbf{F}$. The Schatten p-norms are unitarily invariant, so if $\mathbf{F}$ is circulant, $||\mathbf{F}||_p^p = \left|\left|\widehat{\mathbf{f}}\right|\right|_p^p$.

II. Preconditioner design

In this section, we describe an algorithm for designing a sparse, positive-definite, computationally efficient FIR filter to approximate a given circulant filter. In our experiments, the designed filters were restricted to symmetric $n \times n \times n$ blocks. One could instead use the following algorithm as an inner step of *e.g.*, the successive thinning algorithm [1] to algorithmically determine the filter footprint.

Let $\mathbf{G} \in \mathbb{R}^{N \times N}$ be a positive-definite circulant filter, and let $I \subset \{1, 2, \ldots, N\}$ indicate the desired filter footprint.

Let Ω be the set of circulant filters with the desired footprint. That is,

$$\Omega = \{\mathbf{X} : \mathbf{X} \text{ circulant and } [\mathbf{x}]_i \neq 0 \text{ only if } i \in I\}. \quad (5)$$

Our goal is to find the closest, in a Schatten p-norm sense, positive-definite filter in Ω to the given circulant matrix $\mathbf{G}$.

Satisfying both the positive-definiteness and the footprint constraints simultaneously is challenging. In fact, the positive-definiteness requirement (and the choice of any Schatten p-norm instead of the Schatten ∞-norm) distinguishes this problem from the one solved by the classical Parks-McClellan algorithm [9]. The Schatten p-norms are convex, and the set of positive-definite filters in Ω is convex, so we can use variable splitting to separate the constraints and use ADMM to solve the original problem.

Let $\mathbf{H} \in \mathbb{R}^{N \times N}$ be a circulant matrix and $\mathbf{\Gamma} \in \mathbb{R}^{N \times N}$ be the augmented Lagrangian dual variable. With the equality constraint $\mathbf{H} = \mathbf{F}$, we have the following saddle point problem involving the augmented Lagrangian-like function $\mathcal{L}$:

$$\min_{\mathbf{F} \in \Omega, \mathbf{H} \succ \mathbf{0}} \max_{\mathbf{\Gamma}} \mathcal{L} = \frac{1}{2}||\mathbf{H} - \mathbf{G}||_p^p + \frac{\mu}{2}||\mathbf{H} - (\mathbf{F} + \mathbf{\Gamma})||_F^2 \quad (6)$$

$$= \frac{1}{2}\left|\left|\widehat{\mathbf{h}} - \widehat{\mathbf{g}}\right|\right|_p^p + \frac{\mu N}{2}\left|\left|\widehat{\mathbf{h}} - \left(\widehat{\mathbf{f}} + \widehat{\boldsymbol{\gamma}}\right)\right|\right|_2^2, \quad (7)$$

with both $\mathbf{H}$ and $\mathbf{F}$ restricted to be circulant matrices. Equation (7) follows from the unitary invariance of the Schatten p-norms and the fact that the argument of the Frobenius norm is always a circulant matrix.[2]

Solving (7) with ADMM yields the following set of iterated updates:

$$\mathbf{f}^{(j+1)} = \text{proj}_\Omega\left(\text{IDFT}\left\{\widehat{\mathbf{h}}^{(j)} - \widehat{\boldsymbol{\gamma}}^{(j)}\right\}\right), \quad (8)$$

$$\widehat{\mathbf{h}}^{(j+1)} = \underset{\widehat{\mathbf{h}} \succ \mathbf{0}}{\text{argmin}} \ \frac{1}{2}\left|\left|\widehat{\mathbf{h}} - \widehat{\mathbf{g}}\right|\right|_p^p + \frac{\mu N}{2}\left|\left|\widehat{\mathbf{h}} - \left(\widehat{\mathbf{f}}^{(j+1)} + \widehat{\boldsymbol{\gamma}}^{(j)}\right)\right|\right|_2^2, \quad (9)$$

$$\widehat{\boldsymbol{\gamma}}^{(j+1)} = \widehat{\boldsymbol{\gamma}}^{(j)} + \widehat{\mathbf{f}}^{(j+1)} - \widehat{\mathbf{h}}^{(j+1)}. \quad (10)$$

The $\mathbf{f}$ update requires only an FFT and zeroing of unneeded coordinates, or an IFFT-like operation which efficiently computes a small number of coefficients.

The $\mathbf{h}$ update requires that $\mathbf{H}^{(j+1)}$ be a positive-definite matrix, *i.e.*, $\widehat{\mathbf{h}} > 0$ (in practice, $\widehat{\mathbf{h}} \geq \epsilon$). For all $p \in [1, \infty)$, the update (9) is separable. In these cases, the problem can be solved unconstrained, and, because each separable problem is convex, the solution in each coordinate can then be clamped to $[\epsilon, \infty)$. If $p = \infty$, the $\mathbf{h}$ update is still convex but is nonseparable. We suggest an inexact and relatively efficient solution to the $p = \infty$ problem in Figure 1.

III. Denoising preconditioner design

We could use the filter design algorithm in Section II to replace the denoising problem's conventional circulant preconditioner with a more efficient FIR filter. In this section, we

[2] Both $\mathbf{H}$ and $\mathbf{F}$ are always restricted to be circulant matrices. The dual variable update for $\mathbf{\Gamma}$ (10) with the standard initialization $\mathbf{\Gamma}^{(0)} = \mathbf{0}$ ensures that $\mathbf{\Gamma}$ is also always circulant.

1) Let $\mathbf{d} = \max\left\{\epsilon, \widehat{\mathbf{f}}^{(j+1)} + \widehat{\boldsymbol{\gamma}}^{(j+1)}\right\} - \widehat{\mathbf{g}}$, and $\widehat{\mathbf{h}}(\eta) = \widehat{\mathbf{g}} + \min\{|\mathbf{d}|, \eta\} \cdot \operatorname{sign}(\mathbf{d})$.
2) Compute $\eta_{\min}$ and $\eta_{\max}$ as the extrema of the coordinates of $|\mathbf{d}|$.
3) Perform a grid search over $[\eta_{\min}, \eta_{\max}]$ using a small number of points of $J(\eta) = \frac{1}{2}\eta + \frac{\mu N}{2}\left|\left|\widehat{\mathbf{h}}(\eta) - \left(\widehat{\mathbf{f}}^{(j+1)} + \widehat{\boldsymbol{\gamma}}^{(j+1)}\right)\right|\right|_2^2$ to find η_*.
4) Return $\widehat{\mathbf{h}}(\eta_*)$.

Fig. 1: Approximate algorithm for the $\ell_2 - \ell_\infty$ $\mathbf{h}$ update (9).

instead propose a collection of FIR filters to model the inverse of the Hessian in different regions of the volume.

The Hessian of the denoising problem (4) can be written

$$\nabla^2 J_v(\mathbf{v}) = \mu_v \mathbf{I} + \sum_{d=1}^{N_d} \beta_d \mathbf{C}_d{}' \mathbf{D}_{(\mathbf{v})} \mathbf{C}_d, \tag{11}$$

where $\mathbf{D}_{(\mathbf{v})}$ is a diagonal matrix related to the $\{\kappa_{d,j}\}_{d=1,j=1}^{N_d,N}$ and the second derivative of the potential function ϕ. At the jth voxel, we locally approximate the Hessian with a circulant filter parameterized by a voxel-dependent scalar. That is, for voxels k near j,

$$\mathbf{e}_k{}'\left[\nabla^2 J_v(\mathbf{v})\right]\mathbf{e}_j \approx \mathbf{e}_k{}'\left(\mu_v \mathbf{I} + \alpha_j \sum_{d=1}^{N_d} \beta_d \mathbf{C}_d{}' \mathbf{C}_d\right)\mathbf{e}_j. \tag{12}$$

The scalars $\{\alpha_j\}_{j=1}^N$, which are used to approximate the variation in $\mathbf{D}_{(\mathbf{v})}$, are computed for each voxel as

$$\alpha_j = \frac{\mathbf{e}_j{}'\nabla \mathsf{R}(\varepsilon \mathbf{e}_j)}{\mathbf{e}_j{}'\left(\sum_{d=1}^{N_d} \varepsilon \beta_d \mathbf{C}_d{}' \mathbf{C}_d\right)\mathbf{e}_j}, \tag{13}$$

with $\varepsilon > 0$ small enough (on the order of 10^{-2} for an image in HU) that $\mathbf{e}_j{}'\nabla \mathsf{R}(\varepsilon \mathbf{e}_j)/\epsilon$ approximates the diagonal of the regularizer Hessian.

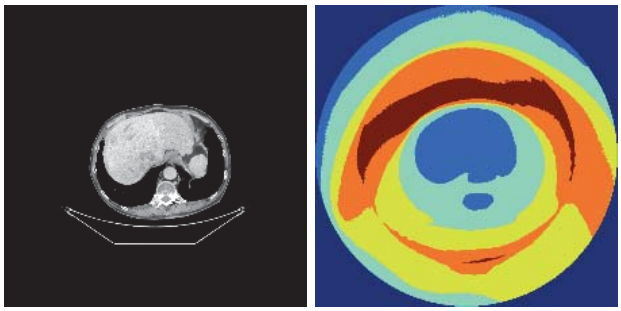
For a purely circulant filter, a single α_* is selected (*e.g.*, from the center of the volume) to form the preconditioner

$$\left(\mu_v \mathbf{I} + \alpha_* \sum_{d=1}^{N_d} \beta_d \mathbf{C}_d{}' \mathbf{C}_d\right)^{-1}. \tag{14}$$

In this paper, we instead now quantize the $\{\alpha_j\}_{j=1}^N$ into P classes, $\{b_p\}_{p=1}^P$, using *e.g.*, the k-means algorithm. Empirically, for reasonably small P, α_j slowly varies over the volume, yielding somewhat contiguous regions with similar Hessian behavior. Motivated by this property, we propose the following preconditioner,

$$\mathbf{P}_{\text{iir}} \triangleq \sum_{p=1}^{P} \mathbf{M}_p \left(\mu_v \mathbf{I} + b_p \sum_{d=1}^{N_d} \beta_d \mathbf{C}_d{}' \mathbf{C}_d\right)^{-1} \mathbf{M}_p, \tag{15}$$

where the $\{\mathbf{M}_p\}_{p=1}^P$ are diagonal matrices with 0 or 1 entries that partition the volume based on the voxel class assignments. Figure 2 illustrates one such partition.



(a) Image (b) Class assignments

Fig. 2: Example class assignments from the center slice of a volume with $P = 6$ classes. Note that the regions are somewhat contiguous and to a degree follow the anatomy in the volume, due to the object-dependence of the $\{\kappa_{d,j}\}_{d=1,j=1}^{N_d,N}$. Each colored region will receive a different preconditioner.

As written, (15) requires a pair of FFTs for each region in the image. This would be a significant cost for the volume denoising problem because gradient computations are relatively inexpensive. We suggest replacing the circulant inverses in (15) with a sparse FIR filter for each region designed using the algorithm described above:

$$\mathbf{P}_{\text{fir}} = \sum_{p=1}^{P} \mathbf{M}_p \mathbf{F}_p \mathbf{M}_p, \tag{16}$$

with $\mathbf{0} \prec \mathbf{F}_p \approx \left(\mu_v \mathbf{I} + b_p \sum_{d=1}^{N_d} \beta_d \mathbf{C}_d{}' \mathbf{C}_d\right)^{-1}$. This preconditioner attempts to handle the spatial variance of $\nabla^2 \mathsf{R}$, but requires no relatively expensive FFTs.

IV. Experimental results

The following experiments were performed on a helical $600 \times 600 \times 101$-voxel dataset with 888 channels, 64 rows and 2080 views provided by GE. The regularizer used 26 voxel neighbors ($N_d = 13$) and the Fair potential function with $\delta = 10.0$ Hounsfeld units:

$$\phi(t) = \delta^2(|t/\delta| - \log(1 + |t/\delta|)). \tag{17}$$

All FFTs were computed using FFTW on a 2.8 GHz Intel Core i7 CPU with 8 threads. All other operations were performed on a NVIDIA GeForce GTX 480 with 1.5 GB of global memory using OpenCL through PyOpenCL. All data was kept on the GPU unless it was necessary to page a buffer off the GPU to RAM.

We generated a 1D version of the regularizer and fit several FIR filters to a circulant approximation of its Hessian using the filter design algorithm in Section II. The results are given in Figure 3, which illustrate the trade-offs, qualitatively speaking, between different choices of Schatten p-norm and filter size. We found that the choice of Schatten p-norm did not have a significant effect on the convergence rate of PCG. Empirically, the Schatten 2-norm filter design problem seemed converge more quickly than the other choices, so we designed the filters in our next experiment with $p = 2$.

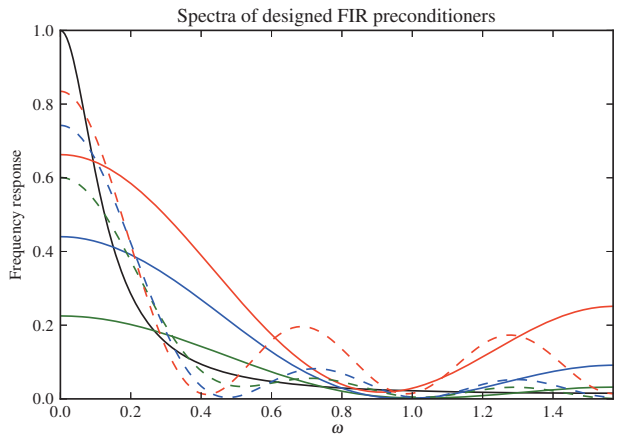


Fig. 3: Profiles of 1D filters generated by the algorithm in Section II. The target spectrum, a preconditioner for the denoising problem with $\beta = 2^4 \mu_v$, is in black. Green, blue and red correspond to the Schatten 1, 2 and ∞ norms, respectively. (These are the lower, middle and upper series on the LHS for each number of taps). The solid lines are 5-tap filters; the dashed lines are 11-tap filters.

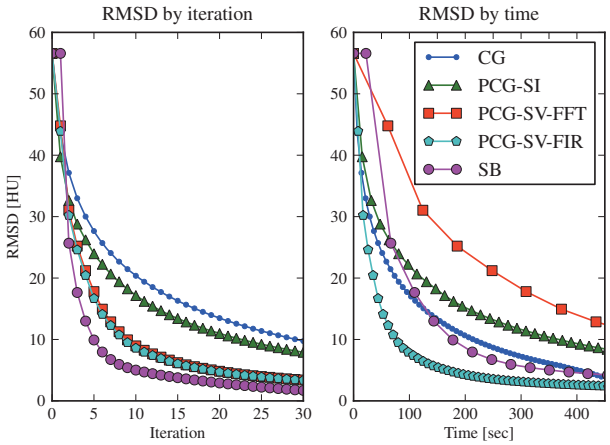


Fig. 4: Root mean-square differences (RMSD) to the converged solution of a denoising problem using (P)CG with shift-invariant (SI) and shift-varying (SV) preconditioners implemented with FFTs and $7 \times 7 \times 7$ FIR filters, and a split-Bregman (SB)-like algorithm. Six classes ($P = 6$ in (16)) were used for the shift-varying preconditioners.

We solved the denoising problem using PCG with several preconditioners and with a split-Bregman (SB) like algorithm [6]. Figure 4 shows the RMSD of each algorithm to the converged solution as a function of time and iteration.

The shift-varying preconditioner significantly outperformed the conventional circulant preconditioner, and replacing the FFTs in (15) with FIR filters had nearly no effect on per-iteration convergence rate. Remarkably, PCG with the shift-varying preconditioner converged more quickly in time than the split-Bregman algorithm. This is due in part to implementing the split-Bregman algorithm's FFTs on the CPU, which incurred GPU-CPU data transfer costs. However, to some degree these costs are unavoidable for the split-Bregman algorithm, due to the GPU's limited memory and the split Bregman algorithm's significant memory requirement. Either way, the space-varying preconditioner is a dramatic improvement over the conventional circulant filter.

V. Conclusions and Future Work

We presented an algorithm to design a positive-definite sparse FIR filter that approximates a given circulant matrix. In our experiments, we heuristically chose filters with dense cubical support, but we have no guarantee that this choice is optimal. The successive thinning algorithm [1] provides a greedy way to select the footprint algorithmically. Another possible extension is to replace the Schatten p-norm-minimization with a minimum condition number criterion, as in [7].

Sparse FIR filters can be used as part of a space-varying preconditioner to significantly accelerate the convergence of PCG applied to the volume denoising problem. The resulting algorithm is memory efficient and performs comparably to the traditionally more rapidly converging split Bregman algorithm. However, efficient preconditioners for the 3D tomography problem are still needed. Such preconditioners will likely need to account for the spatial variance of $\mathbf{A}'\mathbf{A}$, and may benefit from the locality which FIR filters provide. Early work in this direction has already been done by Fu *et. al.* [5].

Finally, while the variable splitting framework used to separate the data-fit and regularization terms has been helpful, it may be useful to revisit frameworks which combine the two. In this case, preconditioners that simultaneously account for the spatial variance present in both the data-fit and regularizer Hessians will certainly be beneficial.

References

[1] T. Baran, D. Wei, and A. V. Oppenheim. Linear programming algorithms for sparse filter design. *IEEE Trans. Sig. Proc.*, 58(3):1605–17, March 2010.

[2] D. P. Bertsekas. Multiplier methods: A survey. *Automatica*, 12(2):133–45, March 1976.

[3] J. Eckstein and D. P. Bertsekas. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55(1-3):293–318, April 1992.

[4] J. A. Fessler and W. L. Rogers. Spatial resolution properties of penalized-likelihood image reconstruction methods: Space-invariant tomographs. *IEEE Trans. Im. Proc.*, 5(9):1346–58, September 1996.

[5] L. Fu, Z. Yu, J-B. Thibault, B. D. Man, M. G. McGaffin, and J. A. Fessler. Space-variant channelized preconditioner design for 3D iterative CT reconstruction. In *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, 2013. Submitted.

[6] T. Goldstein and S. Osher. The split Bregman method for L1-regularized problems. *SIAM J. Imaging Sci.*, 2(2):323–43, 2009.

[7] Z. Lu and T. K. Pong. Minimizing condition number via convex programming. *SIAM J. Matrix. Anal. Appl.*, 32(4):1193–211, 2011.

[8] M. G. McGaffin, S. Ramani, and J. A. Fessler. Reduced memory augmented Lagrangian algorithm for 3D iterative X-ray CT image reconstruction. In *Proc. SPIE 8313 Medical Imaging 2012: Phys. Med. Im.*, page 831327, 2012.

[9] T. W. Parks and J. H. McClellan. Chebyshev approximation for nonrecursive digital filters with linear phase. *IEEE Trans. Circ. Theory*, 19(2):189–99, March 1972.

[10] S. Ramani and J. A. Fessler. A splitting-based iterative algorithm for accelerated statistical X-ray CT reconstruction. *IEEE Trans. Med. Imag.*, 31(3):677–88, March 2012.