A Splitting-Based Iterative Algorithm for Accelerated Statistical X-Ray CT Reconstruction

Sathish Ramani*, *Member, IEEE*, and Jeffrey A. Fessler, *Fellow, IEEE*

Abstract—**Statistical image reconstruction using penalized weighted least-squares (PWLS) criteria can improve image-quality in X-ray computed tomography (CT). However, the huge dynamic range of the statistical weights leads to a highly shift-variant inverse problem making it difficult to precondition and accelerate existing iterative algorithms that attack the statistical model directly. We propose to alleviate the problem by using a variable-splitting scheme that separates the shift-variant and ("nearly") invariant components of the statistical data model and also decouples the regularization term. This leads to an equivalent constrained problem that we tackle using the classical method-of-multipliers framework with alternating minimization. The specific form of our splitting yields an alternating direction method of multipliers (ADMM) algorithm with an inner-step involving a "nearly" shift-invariant linear system that is suitable for FFT-based preconditioning using cone-type filters. The proposed method can efficiently handle a variety of convex regularization criteria including smooth edge-preserving regularizers and non-smooth sparsity-promoting ones based on the ℓ_1-norm and total variation. Numerical experiments with synthetic and real *in vivo* human data illustrate that cone-filter preconditioners accelerate the proposed ADMM resulting in fast convergence of ADMM compared to conventional (nonlinear conjugate gradient, ordered subsets) and state-of-the-art (MFISTA, split-Bregman) algorithms that are applicable for CT.**

Index Terms—**Alternating minimization, iterative algorithm, method of multipliers, regularization, statistical image reconstruction.**

I. INTRODUCTION

STATISTICAL image reconstruction methods in X-ray computed tomography (CT) minimize a cost function consisting of a data-fidelity term that accommodates the measurement statistics and the geometry of the data-acquisition process, and a regularization term that reduces noise. For example, penalized weighted least-squares (PWLS) cost functions for X-ray CT use a (statistically) weighted quadratic data-fidelity term [1], [2] and can provide improved image-quality compared to filtered back-projection (FBP) [1], [2]. However, computation-intensive iterative methods are needed to

minimize such cost functions. This paper describes a new minimization algorithm that uses variable splitting to provide accelerated convergence.

Several types of iterative algorithms have been proposed for statistical image reconstruction in X-ray CT, including iterative coordinate descent (ICD) methods [1], block-based coordinate descent [3], ordered subsets (OS) algorithms based on separable quadratic surrogates (SQS) [4], [5], and (preconditioned) nonlinear conjugate gradient (NCG) methods [6]. For fast computation on multiprocessor computers, (P)NCG-type methods appear to be particularly amenable to efficient parallelization because they update all voxels simultaneously using all measurements.

Developing suitable preconditioners for (P)NCG is challenging for X-ray CT because the enormous dynamic range of the transmission data causes the Hessian of the statistical data-fidelity term to be highly shift-variant [6]. Clinthorne *et al.* [7] showed that for unweighted least-squares reconstruction, one can precondition the problem effectively using FFTs with a kind of cone filter. This cone filter amplifies high spatial frequencies, helping to accelerate convergence. But that cone filter is ineffective for (P)NCG in the PWLS case [6]. Delaney *et al.* [8] considered a very special type of shift-invariant weighting and also demonstrated accelerated convergence, but for low-dose X-ray CT the appropriate statistical weighting does not satisfy the assumptions in [8]. Shift-variant preconditioners based on multiple FFTs were proposed in [6] for 2-D transmission tomography, but never became popular due to their complexity and were never investigated for 3-D problems. Another way to introduce a cone filter is the iterative FBP approach [9], [10]. Initially these algorithms "converge" rapidly compared to (P)NCG methods, but typically they do not have any theoretical convergence properties and "too many" iterations lead to undesirably noisy images. Furthermore, it is unclear how to include regularization while ensuring convergence.

The challenges described above apply regardless of the form of the regularizer. Additional difficulties arise when one uses nonsmooth regularizers such as total variation (TV) [11] and sparsity-promoting ones based on the ℓ_1-norm [12]. These regularizers are not differentiable everywhere precluding optimization by conventional gradient-descent methods (e.g., NCG). Differentiable approximations (e.g., using "corner-rounding" [12, Sec. VI-A], [13, App. A]) can be employed, but even with such modifications the Hessian of the regularizer can have very high curvature leading to slow convergence of conventional gradient-descent methods [14]. While some state-of-the-art algorithms such as (M)FISTA [15], [16] and split-Bregman-type schemes (that split only the regularization term) [17], [18] are able to handle nonsmooth regularizers exactly (i.e., without corner rounding), when

*S. Ramani is with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109 USA (e-mail: sramani@umich.edu).

J. A. Fessler is with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109 USA (e-mail: fessler@umich.edu).

applied to X-ray CT, they must minimize a cost function that involves the original statistical data-fidelity term and are in turn hindered by the shift-variance of its Hessian (see Sections II-B and IV).

In this work, we propose to use a variable-splitting technique that not only decouples the regularization term in the spirit of [17], but also dissociates the statistical and geometrical components in the data-fidelity term. This forms the key feature of our approach that enables us to "isolate" the shift-variant element in the statistical data-fidelity term thereby alleviating the problem. Our splitting procedure uses auxiliary constraint variables to transform the original PWLS problem into an equivalent constrained optimization task that we solve using the classical method-of-multipliers [19], [20] and alternating direction optimization [21]–[23] frameworks. This leads to an alternating direction method of multipliers (ADMM) algorithm for solving the original PWLS problem that, apart from requiring simple operations (such as inverting a diagonal matrix, solving 1-D denoising problems), involves the solving of a "nearly" shift-invariant linear system, which is amenable to fast Fourier transform (FFT)-based preconditioning using cone-type filters [7]. Experimental results with synthetic and real *in vivo* human data indicate that the proposed ADMM converges faster than conventional (NCG and ordered subsets) and state-of-the-art (MFISTA and split-Bregman) methods, illustrating the efficacy of our splitting scheme and the potential of cone-filter preconditioners for accelerating the proposed ADMM. The proposed ADMM can also be used with a variety of convex regularization criteria (see Section VI-A) including smooth edge-preserving regularizers and nonsmooth ones such as TV and ℓ_1-regularization.

The paper is organized as follows. In Section II, we mathematically formulate X-ray CT reconstruction as a PWLS problem and briefly discuss drawbacks of some existing algorithms for X-ray CT. Section III discusses the proposed splitting strategy and the development of the ADMM algorithm in detail. In Section IV, we compare our ADMM algorithm with the split-Bregman technique applied for CT, schematically. Section V is dedicated to numerical experiments and results, while Section VI discusses possible extensions of this work to 3-D CT and other statistical models. Finally, we draw our conclusions in Section VII.

II. STATISTICAL X-RAY CT RECONSTRUCTION

A. Problem Formulation

For CT, an accurate statistical model for the data is quite complicated [24], [25] and is often replaced by a Gaussian approximation [1], [2] with a suitable diagonal weighting term $\mathbf{W}$ whose components $\{w_i\}$ are inversely proportional to the measurement variances [1], [2]. We consider a penalized weighted least-squares (PWLS) formulation of statistical CT reconstruction [1]

$$\mathbf{P0}: \quad \arg\min_{\mathbf{x}} \left\{ J(\mathbf{x}) \triangleq J_{\text{data}}(\mathbf{y}, \mathbf{Ax}) + \Psi(\mathbf{Rx}) \right\} \quad (1)$$

$$J_{\text{data}}(\mathbf{y}, \mathbf{Ax}) \triangleq \frac{1}{2} \|\mathbf{y} - \mathbf{Ax}\|_{\mathbf{W}}^2 \quad (2)$$

where $\mathbf{y}$ is the $M \times 1$ data vector (log of transmission data), $\mathbf{A}$ is the $M \times N$ system matrix, $\mathbf{Ax}$ represents the forward projection operation (e.g., line integrals), $\mathbf{W} = \text{diag}\{w_i\}$ is a $M \times M$ diagonal matrix consisting of statistical weights[1], and $\|\mathbf{u}\|_{\mathbf{W}}^2 \triangleq \mathbf{u}^\top \mathbf{Wu}$. We use a general family of regularizers of the form [12]

$$\Psi(\mathbf{Rx}) = \lambda \sum_r \kappa_r \Phi_r \left(\sum_{p=1}^{P} |[\mathbf{R}_p \mathbf{x}]_r|^m \right) \quad (3)$$

where $\lambda > 0$ is the regularization parameter, $\kappa_r > 0 \; \forall r$ are user-provided weights that govern the spatial resolution in the reconstructed output [26], Φ_r are potential functions, the $R \times N$ matrix $\mathbf{R} \triangleq [\mathbf{R}_1^\top \cdots \mathbf{R}_P^\top]^\top$ constitutes regularization operators $\mathbf{R}_p$ (e.g., finite differences, frames, etc.) of size $L \times N$, where $R = PL$. We concentrate on values of m and instances of Φ that result in a convex regularizer Ψ in (3).

The above general regularizer is in the "analysis" form [27], i.e., Ψ is specified as a function of the reconstructed image $\mathbf{x}$. The method proposed in this paper can also be easily extended to handle "synthesis" forms [27], e.g., by writing $\mathbf{x} = \mathbf{S}\boldsymbol{\theta}$ and considering $J(\boldsymbol{\theta}) = J_{\text{data}}(\mathbf{y}, \mathbf{AS}\boldsymbol{\theta}) + \Phi(\boldsymbol{\theta})$ in $\mathbf{P0}$, for some potential function Φ and synthesis operator $\mathbf{S}$. We focus on the analysis form (3) as it includes popular nonsmooth criteria such as TV (for $\Phi_r(x) = \sqrt{x}$ and $m = 2$), analysis ℓ_1-wavelets (for $\Phi_r(x) = x$, $m = 1$) and a variety of smooth convex edge-preserving regularizers (e.g., Huber [28], [29], Fair [6], [30], etc.).

B. Previous Approaches

Conventional gradient-descent methods, e.g., NCG, for $\mathbf{P0}$ depend on the Hessian of J_{data}: $\mathbf{H}_{\text{data}} = \mathbf{A}^\top \mathbf{WA}$, which is highly shift-variant in CT particularly due to the large dynamic range of $\mathbf{W}$. As a result, it becomes difficult to precondition and accelerate such methods [6]. Fessler *et al.* [6] directly attacked $\mathbf{P0}$ using NCG and proposed a shift-variant preconditioner to tackle $\mathbf{H}_{\text{data}}$. But their preconditioner is data-dependent and requires at least one pair of FFT-iFFT operations per NCG-iteration.

Iterative shrinkage-thresholding (IST) [31] and its variants ((M)FISTA [15], [16], and (M)TWIST [32]) that are applicable to $\mathbf{P0}$ depend on the Lipschitz constant L_{data} of $J_{\text{data}}(\mathbf{y}, \mathbf{Ax})$

$$L_{\text{data}} = \sigma_{\max}\{\mathbf{A}^\top \mathbf{WA}\} \quad (4)$$

where $\sigma_{\max}$ represents the maximum eigenvalue. The convergence speed of these algorithms is primarily determined by (4): A large value of L_{data} results in small gradient steps [15, Sec. 1.1] leading to slow convergence. Since $\mathbf{W}$ has a large dynamic range and due to the (approximately) $1/r$-type decay of the elements of $\mathbf{A}^\top \mathbf{A}$, L_{data} can be large for CT decreasing convergence speed of IST-type algorithms. Optimization transfer-based methods (e.g., [33, Sec. IV-B1]) face a similar issue in that the surrogate functions end up having high curvature [5] due to $\mathbf{W}$, which again leads to small update-steps and slow convergence.

---

[1]For simplicity we used $w_i = e^{-y_i}$ in our experiments.

In summary, the weighting term $\mathbf{W}$, although crucial for improving reconstruction quality, poses a challenge for optimization. Compared to $\mathbf{A}^\top \mathbf{W} \mathbf{A}$, the term $\mathbf{A}^\top \mathbf{A}$ is "more" shift-invariant and is appropriate for preconditioning using cone filters. This property has been used to accelerate unweighted least-squares reconstruction for tomographic image reconstruction [7]. Therefore, our idea to mitigate the shift-variance of $\mathbf{H}_{\text{data}}$ is to untangle $\mathbf{W}$ from $\mathbf{H}_{\text{data}}$ thereby making the resulting problem "more" shift-invariant and suitable to circulant preconditioning. To do so, we adopt a variable-splitting strategy.

Variable splitting (VS) refers to the process of introducing auxiliary constraint variables to separate coupled components in the cost function [12], [17], [18], [34]–[42]. This procedure transforms the original minimization problem into an equivalent constrained optimization problem that can be effectively solved using classical constrained optimization schemes [19], [20]. The VS approach is appealing as it renders the resulting constrained problem tractable to alternating minimization schemes that decouple it in terms of the auxiliary variables and simplify optimization [12], [17], [18], [34], [36], [37], [39]–[42].

The VS approach has become popular recently for solving reconstruction problems in image processing [17], [34]–[37], MRI [12], [39], [40] and CT [18], [41], [42]. Many authors have focussed on splitting the regularization term [17], [18], [34], [37], [39]–[41] as it is hard to tackle in inverse problems (especially nonsmooth ones such as TV and ℓ_1-regularization). Splitting the regularization term enables one to handle it exactly (i.e., without the need for "corner-rounding" [12, Sec. VI-A], [13, App. A] for nonsmooth criteria) via simple denoising problems [12], [17], [37], [42]. However, in PWLS problems for CT, the data-term adds to the complexity (as it leads to a shift-variant Hessian $\mathbf{H}_{\text{data}}$) and therefore demands attention. So in this work, besides splitting the regularization term, we also split the data-term.

III. PROPOSED METHOD

A. Equivalent Constrained Optimization Problem

We introduce auxiliary constraint variables $\mathbf{u} \in \mathbb{R}^M$ and $\mathbf{v} \in \mathbb{R}^R$ and write $\mathbf{P0}$ as the following equivalent constrained problem:

$$\arg \min_{\mathbf{x}, \mathbf{u}, \mathbf{v}} \left\{ f(\mathbf{u}, \mathbf{v}) \triangleq J_{\text{data}}(\mathbf{y}, \mathbf{u}) + \Psi(\mathbf{v}) \right\}$$
$$\text{s.t.} \quad \mathbf{u} = \mathbf{A}\mathbf{x}, \quad \mathbf{v} = \mathbf{R}\mathbf{x} \tag{5}$$

where $\mathbf{u}$ separates the effect of $\mathbf{W}$ on $\mathbf{A}\mathbf{x}$ and $\mathbf{v}$ splits the regularization term as in [17]. Afonso et al. [36] and Figueiredo et al. [38] have utilized data-term-splitting in the context of image restoration [36], [38] and reconstruction from partial Fourier observations [36]. However, our emphasis here is on CT reconstruction where $\mathbf{u}$ plays an important role: It leads to a subproblem that is "nearly" shift-invariant and suitable to preconditioning using cone filters [42] as explained in Section III-C.

In general, the proposed splitting strategy (5) can be applied to any PWLS problem of the form $\mathbf{P0}$ so as to exploit shift-invariant features in the data-model, e.g., deconvolution of blurred images corrupted with nonstationary noise.

Before proceeding, we rewrite (5) concisely as

$$\mathbf{P1}: \quad \arg \min_{\mathbf{x}, \mathbf{z}} f(\mathbf{z}) \quad \text{s.t.} \quad \mathbf{z} = \mathbf{C}\mathbf{x} \tag{6}$$

where

$$\mathbf{z} \triangleq \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix}, \mathbf{C} \triangleq \begin{bmatrix} \mathbf{A} \\ \mathbf{R} \end{bmatrix}. \tag{7}$$

Since $\mathbf{P1}$ is equivalent to $\mathbf{P0}$, solving $\mathbf{P1}$ for $\mathbf{x}$ yields the desired reconstruction in (1).

B. Method of Multipliers

To solve $\mathbf{P1}$, we use the classical framework of the method of multipliers [19], [20] and construct an augmented Lagrangian (AL) function [12], [19], [20], [42]

$$\mathcal{L}(\mathbf{x}, \mathbf{z}, \boldsymbol{\gamma}) \triangleq f(\mathbf{z}) + \boldsymbol{\gamma}^\top (\mathbf{z} - \mathbf{C}\mathbf{x}) + \frac{\mu}{2} \|\mathbf{z} - \mathbf{C}\mathbf{x}\|_{\boldsymbol{\Lambda}}^2 \tag{8}$$

that combines a multiplier term $\boldsymbol{\gamma}^\top (\mathbf{z} - \mathbf{C}\mathbf{x})$ with Lagrange multiplier $\boldsymbol{\gamma} \triangleq [\boldsymbol{\gamma}_{\mathbf{u}}^\top \boldsymbol{\gamma}_{\mathbf{v}}^\top]^\top \in \mathbb{R}^{M+R}$ and a quadratic penalty term $\mu/2 \|\mathbf{z} - \mathbf{C}\mathbf{x}\|_{\boldsymbol{\Lambda}}^2$, where $\mu > 0$ is the AL penalty parameter and $\boldsymbol{\Lambda} \succ 0$ is a symmetric weighting matrix. The multiplier term can be absorbed into the penalty term in (8) (by completing the square) for ease of manipulation leading to

$$\mathcal{L}(\mathbf{x}, \mathbf{z}, \boldsymbol{\eta}) \triangleq f(\mathbf{z}) + \frac{\mu}{2} \|\mathbf{z} - \mathbf{C}\mathbf{x} - \boldsymbol{\eta}\|_{\boldsymbol{\Lambda}}^2 + \mathfrak{c}_{\boldsymbol{\eta}} \tag{9}$$

where $\boldsymbol{\eta} \triangleq [\boldsymbol{\eta}_{\mathbf{u}}^\top \boldsymbol{\eta}_{\mathbf{v}}^\top]^\top = -1/\mu \boldsymbol{\Lambda}^{-1} \boldsymbol{\gamma}$ and $\mathfrak{c}_{\boldsymbol{\eta}} \triangleq -\mu/2 \|\boldsymbol{\eta}\|_{\boldsymbol{\Lambda}}^2$ is a constant independent of $\mathbf{x}$ and $\mathbf{z}$. Unlike standard approaches [36], [38] that set $\boldsymbol{\Lambda} = \mathbf{I}_{M+R}$, we propose to use

$$\boldsymbol{\Lambda} = \begin{bmatrix} \mathbf{I}_M & \mathbf{0} \\ \mathbf{0} & \nu \mathbf{I}_R \end{bmatrix} \tag{10}$$

where $\nu > 0$. This is crucial in CT because the elements of $\mathbf{A}$ and $\mathbf{R}$ can differ by several orders of magnitude and it is imperative to balance them to avoid numerical instabilities in the resulting algorithm and to achieve faster convergence [42].

The classical AL scheme for solving $\mathbf{P1}$ alternates between a joint-minimization step and an update step [12, Sec. III]

$$\left(\mathbf{x}^{(j+1)}, \mathbf{z}^{(j+1)} \right) = \arg \min_{\mathbf{x}, \mathbf{z}} \mathcal{L} \left(\mathbf{x}, \mathbf{z}, \boldsymbol{\eta}^{(j)} \right) \tag{11}$$

$$\boldsymbol{\eta}^{(j+1)} = \boldsymbol{\eta}^{(j)} - \left(\mathbf{z}^{(j+1)} - \mathbf{C}\mathbf{x}^{(j+1)} \right) \tag{12}$$

respectively. Unlike pure penalty methods, remarkably, the AL formalism does not require increasing $\mu \to \infty$ to ensure convergence of (11) and (12) to a solution of $\mathbf{P1}$ [19].

C. Alternating Direction Minimization

It is numerically appealing to replace the more difficult joint-minimization step (11) by alternating direction optimization that decouples (11) as [21]–[23]

$$\mathbf{x}^{(j+1)} = \arg \min_{\mathbf{x}} \mathcal{L} \left(\mathbf{x}, \mathbf{z}^{(j)}, \boldsymbol{\eta}^{(j)} \right) \tag{13}$$

$$\mathbf{z}^{(j+1)} = \arg\min_{\mathbf{z}} \mathcal{L}\left(\mathbf{x}^{(j+1)}, \mathbf{z}, \boldsymbol{\eta}^{(j)}\right). \qquad (14)$$

Thus, at the jth iteration, instead of (11) and (12), we perform (ignoring constant terms)

$$\mathbf{x}^{(j+1)} \stackrel{\epsilon_\mathbf{x}^{(j)}}{=} \arg\min_{\mathbf{x}} \left\| \mathbf{z}^{(j)} - \mathbf{C}\mathbf{x} - \boldsymbol{\eta}^{(j)} \right\|_{\boldsymbol{\Lambda}}^2 \qquad (15)$$

$$\mathbf{z}^{(j+1)} \stackrel{\epsilon_\mathbf{z}^{(j)}}{=} \arg\min_{\mathbf{z}} f(\mathbf{z}) + \left\| \mathbf{z} - \mathbf{C}\mathbf{x}^{(j+1)} - \boldsymbol{\eta}^{(j)} \right\|_{\boldsymbol{\Lambda}}^2 \qquad (16)$$

$$\boldsymbol{\eta}^{(j+1)} = \boldsymbol{\eta}^{(j)} - \left(\mathbf{z}^{(j+1)} - \mathbf{C}\mathbf{x}^{(j+1)}\right) \qquad (17)$$

where we write $\mathbf{x} \stackrel{\epsilon}{=} \mathbf{x}^\star$ to mean that $\|\mathbf{x} - \mathbf{x}^\star\|_2 \leq \epsilon$, i.e., we allow for inexact updates in (15) and (16) in the spirit of [43]. Although (15) and (16) is an approximation to (11), the following theorem adapted from [43, Th. 8] to **P1** guarantees convergence of (15)–(17) to a solution of (5) (and **P0**).

Theorem 1: Consider **P1** in (6) where f is closed, proper, convex[2] and **C** has full column-rank. Let $\boldsymbol{\eta}^{(0)} \in \mathbb{R}^{M+R}, \mu > 0$

$$\sum_j \epsilon_\mathbf{x}^{(j)} < \infty, \quad \text{and} \quad \sum_j \epsilon_\mathbf{z}^{(j)} < \infty. \qquad (18)$$

If **P1** has a solution $(\mathbf{x}^\star, \mathbf{z}^\star)$, then the sequence of updates $\{(\mathbf{x}^{(j)}, \mathbf{z}^{(j)})\}_j$ generated by (15)–(17) converges to $(\mathbf{x}^\star, \mathbf{z}^\star)$. If **P1** has no solution, then at least one of the sequences $\{(\mathbf{x}^{(j)}, \mathbf{z}^{(j)})\}_j$ or $\{\boldsymbol{\eta}^{(j)}\}_j$ diverges. ∎

The result of Eckstein *et al.* [43, Th. 8] uses an AL function with $\boldsymbol{\Lambda} = \mathbf{I}$, so we apply [43, Th. 8] to (9) with $\boldsymbol{\Lambda}$ in (10) through a simple change of variables.[3] For CT, it can be readily ensured that **C** has full column-rank for a variety of regularization operators **R**. In the sequel, we explain how to perform the minimizations in (15) and (16).

Firstly, we see that due to the structure of $f(\mathbf{z})$ and **C**, (16) further dissociates into the following:

$$\mathbf{u}^{(j+1)}$$
$$\stackrel{\epsilon_\mathbf{u}^{(j)}}{=} \arg\min_{\mathbf{u}} \left\{ J_{\text{data}}(\mathbf{y}, \mathbf{u}) + \frac{\mu}{2} \left\| \mathbf{u} - \mathbf{A}\mathbf{x}^{(j+1)} - \boldsymbol{\eta}_\mathbf{u}^{(j)} \right\|_2^2 \right\} \qquad (19)$$

$$\mathbf{v}^{(j+1)}$$
$$\stackrel{\epsilon_\mathbf{v}^{(j)}}{=} \arg\min_{\mathbf{v}} \left\{ \Psi(\mathbf{v}) + \frac{\mu\nu}{2} \left\| \mathbf{v} - \mathbf{R}\mathbf{x}^{(j+1)} - \boldsymbol{\eta}_\mathbf{v}^{(j)} \right\|_2^2 \right\}. \qquad (20)$$

These subproblems are independent of each other and can therefore be solved simultaneously, where $[\epsilon_\mathbf{z}^{(j)}]^2 = [\epsilon_\mathbf{u}^{(j)}]^2 + [\epsilon_\mathbf{v}^{(j)}]^2 \forall j$. Subproblem (19) is quadratic and has a closed form solution

$$\mathbf{u}^{(j+1)} = \mathbf{D}_\mu^{-1} \left(\mathbf{W}\mathbf{y} + \mu \left(\mathbf{A}\mathbf{x}^{(j+1)} + \boldsymbol{\eta}_\mathbf{u}^{(j)} \right) \right) \qquad (21)$$

where $\mathbf{D}_\mu \triangleq (\mathbf{W} + \mu\mathbf{I}_M)$. Since $\mathbf{W}$ is diagonal, $\mathbf{D}_\mu$ can be inverted exactly, so that $\epsilon_\mathbf{u}^{(j)} = 0$ in (19) $\forall j$.

Minimization with respect to $\mathbf{v}$ (20) corresponds to a denoising problem that can be solved efficiently and/or exactly for a variety of instances of (3) including TV: This has been elucidated by many authors [12], [17], [36]–[40], [45], e.g., the techniques developed in [12, Sec. IV-A2–IV-A6] can be directly applied to (20). For brevity, we concentrate on two particular instances of (3) and solve (20) exactly so that $\epsilon_\mathbf{v}^{(j)} = 0$ in (20) $\forall j$.

• Analysis ℓ_1-regularization [$m = 1, \Phi_r(x) = x \forall r$ in (3)]

$$\Psi(\mathbf{R}\mathbf{x}) = \lambda \sum_{r=1}^R \kappa_r \left| [\mathbf{R}\mathbf{x}]_r \right| \qquad (22)$$

with the shift-invariant Haar wavelet transform (excluding the approximation level) for **R**, which is a sparsity-promoting criterion [12], [27], [36], [38].

• Smooth edge-preserving regularization [$P = 1\, m = 1$, $\Phi_r = \Phi_{\text{FP}} \forall r$ in (3)]

$$\Psi(\mathbf{R}\mathbf{x}) = \lambda \sum_{r=1}^R \kappa_r \Phi_{\text{FP}} \left(|[\mathbf{R}\mathbf{x}]_r| \right) \qquad (23)$$

using the Fair potential $\Phi_{\text{FP}}(x) = x/\delta - \log(1+x/\delta)$ with $\delta > 0$ [30] (also the smoothed Laplace function in [45, Eq. 4.11]) and finite-differences for **R**. This regularizer ensures a unique solution to **P0** as Φ_{FP} is strictly convex. It has also been successfully applied to PWLS problems in tomography [6].

For these regularizers, (20) separates into R 1-D minimization problems in terms of the components $\{v_r\}_{r=1}^R$ of $\mathbf{v}$

$$v_r^{(j+1)} = \arg\min_{v_r} \Psi(v_r) + \frac{\mu\nu}{2} \left(v_r - \varrho_r^{(j)} \right)^2 \qquad (24)$$

where $\varrho_r^{(j)}$ is the rth component of $\boldsymbol{\varrho}^{(j)} \triangleq \mathbf{R}\mathbf{x}^{(j+1)} + \boldsymbol{\eta}_\mathbf{v}^{(j)}$. For (22), the solution of (24) is given by the shrinkage rule[4] [46]

$$v_r^{(j+1)} = \text{sign}\left\{ \varrho^{(j)} \right\} \max\left(\left| \varrho^{(j)} \right| - \frac{\lambda\kappa_r}{\mu\nu}, 0 \right). \qquad (25)$$

For (23), (24) leads to a quadratic equation in v_r [45, Eq. 4.13] that yields

$$v_r^{(j+1)} = \text{sign}\left\{ \varrho_r^{(j)} \right\} \frac{\zeta_r^{(j)} + \sqrt{\left(\zeta_r^{(j)}\right)^2 + 4\delta \left| \varrho_r^{(j)} \right|}}{2} \qquad (26)$$

where $\zeta_r^{(j)} \triangleq |\varrho_r^{(j)}| - \delta - \lambda\kappa_r/(\delta\mu\nu)$.

Having addressed (16), we now consider (15) which can be easily solved analytically

$$\mathbf{x}^{(j+1)\star} = \mathbf{G}_\nu^{-1} \left(\mathbf{A}^\top \left(\mathbf{u}^{(j)} - \boldsymbol{\eta}_\mathbf{u}^{(j)} \right) + \nu\mathbf{R}^\top \left(\mathbf{v}^{(j)} - \boldsymbol{\eta}_\mathbf{v}^{(j)} \right) \right) \qquad (27)$$

where $\mathbf{x}^{(j+1)\star}$ represents the exact solution to (15) and

$$\mathbf{G}_\nu \triangleq \mathbf{C}^\top \boldsymbol{\Lambda}\mathbf{C} = (\mathbf{A}^\top\mathbf{A} + \nu\mathbf{R}^\top\mathbf{R}) \qquad (28)$$

---

[2]A convex function h is closed if and only if it is lower semi-continuous (LSC) [44, pp. 51, 52] and is proper if $h(\mathbf{x}) < +\infty$ for at least one $\mathbf{x}$ and $h(\mathbf{x}) > -\infty \forall \mathbf{x}$ [44, p. 24]. It can be shown that the convex functions J_{data}, Ψ (for a variety of regularizers such as TV and ℓ_1-regularization), and their sum, f (5), are LSC and proper [38].

[3]Writing $\mathbf{z}_0 = \boldsymbol{\Lambda}^{1/2}\mathbf{z}, \boldsymbol{\eta}_0 = \boldsymbol{\Lambda}^{1/2}\boldsymbol{\eta}$, and $\mathbf{M} = \boldsymbol{\Lambda}^{1/2}\mathbf{C}$, it is easy to see that (15)–(17) solve the constrained problem $\arg\min_{\mathbf{x},\mathbf{z}_0} f(\boldsymbol{\Lambda}^{-1/2}\mathbf{z}_0)\text{s.t.}\mathbf{z}_0 = \mathbf{M}\mathbf{x}$ that is equivalent to **P0** using the AL function $\mathcal{L}_0(\mathbf{x}, \mathbf{z}_0, \boldsymbol{\eta}_0) = f(\boldsymbol{\Lambda}^{-1/2}\mathbf{z}_0) + \mu/2\|\mathbf{z}_0 - \mathbf{M}\mathbf{x} - \boldsymbol{\eta}_0\|_2^2$ with an unweighted penalty term.

[4]An analytical update formula similar to (25) is available for the TV regularizer that is based on a vector shrinkage-rule, see e.g., [12, Sec. IV-A6].

is nonsingular because $\mathbf{\Lambda} \succ \mathbf{0}$ and $\mathbf{R}$ is chosen so that $\mathbf{C}$ has full column-rank. Although (27) is an exact analytical solution, the enormous size of $\mathbf{G}_\nu$ for CT makes it impossible to store and "invert" $\mathbf{G}_\nu$ exactly. So we propose to use the conjugate gradient (CG) method for (27) and obtain an approximate update $\mathbf{x}^{(j+1)} \stackrel{\epsilon_\mathbf{x}^{(j)}}{=} \mathbf{x}^{(j+1)\star}$. Since $\mathbf{G}_\nu$ is nonsingular, we have that

$$\epsilon_\mathbf{x}^{(j)} \leq \frac{\rho_\mathbf{x}^{(j)}}{\sigma_{\min}\{\mathbf{G}_\nu\}} \tag{29}$$

where $\rho_\mathbf{x}^{(j)}$ is the corresponding residue and $\sigma_{\min}\{\mathbf{G}_\nu\} > 0$ is the minimum eigenvalue of $\mathbf{G}_\nu$ that depends only on $\mathbf{A}$ and $\mathbf{R}$ and can be precomputed e.g., using stochastic techniques [47] or the Power method.[5] Therefore, using (29), one can monitor $\epsilon_\mathbf{x}^{(j)}$ in the CG-loop and design a suitable stopping rule to satisfy (18).

D. Preconditioning Using Cone Filter

We see that $\mathbf{G}_\nu$ contains $\mathbf{A}^\top \mathbf{A}$, which is "nearly" shift-invariant, so for shift-invariant[6] $\mathbf{R}^\top \mathbf{R}$, $\mathbf{G}_\nu$ is amenable to preconditioning using suitable cone filters [6], [7]. We constructed a circulant matrix $\tilde{\mathbf{G}}_\nu$ from the central column of $\mathbf{G}_\nu$

$$\tilde{\mathbf{G}}_\nu = \text{circ}\{\mathbf{G}_\nu \mathbf{e}_c\} \tag{30}$$

and used its inverse, $\tilde{\mathbf{G}}_\nu^{-1}$, as the preconditioner, where $\mathbf{e}_c$ is a standard basis vector of $\mathbb{R}^N$ corresponding to the center pixel of the image and $\text{circ}\{\boldsymbol{\alpha}\}$ represents the construction of a circulant matrix from a vector $\boldsymbol{\alpha}$. The proposed preconditioner $\tilde{\mathbf{G}}_\nu^{-1}$ corresponds to a cone-type filter that amplifies high spatial frequencies and accelerates convergence of both the CG-loop for (27) and the overall ADMM scheme as demonstrated in Section V. Implementing $\tilde{\mathbf{G}}_\nu^{-1}$ requires only one FFT-iFFT per CG iteration and its construction[7] uses a product with $\mathbf{R}^\top \mathbf{R}$ and only one forward–backward projection that can be performed offline as $\tilde{\mathbf{G}}_\nu$ is independent of $\mathbf{W}$. In our experiments, we applied at most two preconditioned CG (PCG) iterations with warm starting [12] and found that $\rho_\mathbf{x}^{(j)}$ decreased sufficiently rapidly. Based on (15)–(29), we present our algorithm in Fig. 1 for solving **P1** (and thus **P0**). In principle, Steps 4 and 5 of ADMM may be executed in parallel as they are independent of each other, but in our implementation, we chose to execute all the steps sequentially for simplicity.

E. Selection of μ and ν

The parameters μ and ν do not affect the solution of **P1**, but only regulate the convergence speed of the proposed ADMM [12], [35, Sec. 4.4]. In general, choosing appropriate values for AL penalty parameters (such as μ and ν) is a nontrivial and application-dependent task. Several empirical rules have been

---

1. Select $\mathbf{x}^{(0)}$, μ, $\nu > 0$ and set $j = 0$
2. Set $\mathbf{u}^{(0)} = \mathbf{A}\mathbf{x}^{(0)}$, $\mathbf{v}^{(0)} = \mathbf{R}\mathbf{x}^{(0)}$, and $\boldsymbol{\eta}_\mathbf{u}^{(0)} = \boldsymbol{\eta}_\mathbf{v}^{(0)} = \mathbf{0}$
Repeat:
3. Obtain $\mathbf{x}^{(j+1)}$ by applying (P)CG iterations to (27)
4. Compute $\mathbf{u}^{(j+1)}$ using (21)
5. Compute $\mathbf{v}^{(j+1)}$ using (24)-(26)
6. $\boldsymbol{\eta}_\mathbf{u}^{(j+1)} = \boldsymbol{\eta}_\mathbf{u}^{(j)} - (\mathbf{u}^{(j+1)} - \mathbf{A}\mathbf{x}^{(j+1)})$
7. $\boldsymbol{\eta}_\mathbf{v}^{(j+1)} = \boldsymbol{\eta}_\mathbf{v}^{(j)} - (\mathbf{v}^{(j+1)} - \mathbf{R}\mathbf{x}^{(j+1)})$
8. Set $j = j + 1$
Until stop criterion is met

Fig. 1. ADMM for statistical X-ray CT reconstruction.

put forth by many authors for setting AL penalty parameters (to obtain good convergence speeds for AL-based iterative reconstruction schemes) in many applications, see e.g., [37], [38] for image restoration, [17] for denoising and compressed-sensing MRI, and [12], [39] for parallel MRI reconstruction.

In this paper, Step 3 is the only inexact step of the proposed ADMM. So the computational speed of ADMM is primarily determined by how efficiently (27) is solved, which in turn is governed by ν. We use an empirical rule for selecting ν that is based on [17]: Since ν balances $\mathbf{A}^\top \mathbf{A}$ and $\mathbf{R}^\top \mathbf{R}$ in $\mathbf{G}_\nu$ that have disjoint nontrivial null-spaces, the condition number $\kappa(\mathbf{G}_\nu)$ of $\mathbf{G}_\nu$ exhibits a minimum for some $\nu_{\min} > 0 : \nu_{\min} = \arg\min_\nu \kappa(\mathbf{G}_\nu)$. It was suggested in [17] to use this property to choose AL penalty parameters to ensure quick convergence of the CG-algorithm for solving a linear system such as (27). For $\mathbf{A}$ implemented using the distance-driven (DD) projector [49] and $\mathbf{R}$ in (22), (23), $\nu_{\min} \approx 10^5$, which yielded a very small $(\lambda/\mu\nu)$ in (20) and subsequently resulted in slow convergence of ADMM in our experiments. On the other extreme, setting $\nu = 1$ (corresponding to the standard case of $\mathbf{\Lambda} = \mathbf{I}_{M+R}$) yields a poorly conditioned $\mathbf{G}_\nu$ that was not favorable either.

Based on our experience with 2-D CT experiments, we found the empirical rule[8] $\nu_{\text{emp}} = 1/100 \arg\min_{\nu_0} \kappa(\tilde{\mathbf{G}}_{\nu_0})$ to yield good overall convergence speeds for ADMM, where $\tilde{\mathbf{G}}_\nu$ is the circulant matrix in (30). We also observed that ADMM was slightly more robust to the choice of μ than ν. We selected $\mu = \text{median}\{w_i\}$ to avoid outliers in $\mathbf{W}$; this yielded a well-conditioned $\mathbf{D}_\mu$ (with $\kappa\{\mathbf{D}_\mu\} \in [10, 40]$) that improved the numerical stability of ADMM.

IV. COMPARISON WITH THE SPLIT-BREGMAN APPROACH

The split-Bregman (SB) method [17] uses constraint variables to split the regularization term alone. For (1), this corresponds to using only $\mathbf{v} = \mathbf{R}\mathbf{x}$ which leads to following equivalent constrained problem:

$$\arg\min_{\mathbf{x},\mathbf{v}} \left\{ f_{\text{SB}}(\mathbf{x},\mathbf{v}) \triangleq J_{\text{data}}(\mathbf{y}, \mathbf{A}\mathbf{x}) + \Psi(\mathbf{v}) \right\} \tag{31}$$

$$\text{s.t.} \quad \mathbf{v} = \mathbf{R}\mathbf{x}. \tag{32}$$

This type of splitting has been investigated for CT reconstruction in [18], [41]. Applying the Bregman iterations [17, Eq.

---

[5]Since the Power method (PM) iteratively estimates the maximum eigenvalue (in absolute magnitude) of a matrix [48, p. 488], an estimate $\hat{\sigma}_{\max}\{\mathbf{G}_\nu\}$ of $\sigma_{\max}\{\mathbf{G}_\nu\}$ is first computed by applying PM on $\mathbf{G}_\nu$. Next, applying PM on $\mathbf{K} \triangleq \mathbf{G}_\nu - \hat{\sigma}_{\max}\{\mathbf{G}_\nu\}\mathbf{I}_N$ yields $|\hat{\sigma}_{\min}\{\mathbf{G}_\nu\} - \hat{\sigma}_{\max}\{\mathbf{G}_\nu\}|$ (as $\sigma_{\min}\{\mathbf{G}_\nu\} - \hat{\sigma}_{\max}\{\mathbf{G}_\nu\}$ is the largest eigenvalue of $\mathbf{K}$ in absolute magnitude) from which $\hat{\sigma}_{\min}\{\mathbf{G}_\nu\}$ can be easily obtained.

[6]The matrix $\mathbf{R}^\top \mathbf{R}$ is circulant when periodic boundary conditions are used for $\mathbf{R}$ in (22), (23).

[7]We only store the frequency response corresponding to $\tilde{\mathbf{G}}_\nu^{-1}$ to save memory.

[8]It would be ideal to consider $\kappa(\mathbf{G}_\nu)$ instead of $\kappa(\tilde{\mathbf{G}}_\nu)$ for selecting ν_{emp}, but estimating $\kappa(\mathbf{G}_\nu)$ (e.g., using the Power method) for a given ν for CT is computationally expensive (ignoring the fact that it is independent of $\mathbf{W}$ and could be computed offline). But as $\mathbf{G}_\nu$ is approximately shift-invariant, $\kappa(\tilde{\mathbf{G}}_\nu) \approx \kappa(\mathbf{G}_\nu)$, which leads to ν_{emp}.

3.7–3.8] with alternating minimization [17, Sec. 3.1] to (31) yields the following SB scheme:[9]

$$\mathbf{x}^{(j+1)}$$
$$= \arg\min_{\mathbf{x}} \left\{ \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_{\mathbf{W}}^2 + \frac{\mu}{2} \left\| \mathbf{v}^{(j)} - \mathbf{R}\mathbf{x} - \boldsymbol{\eta}_{\mathbf{v}}^{(j)} \right\|_2^2 \right\} \quad (33)$$

$$\mathbf{v}^{(j+1)} = \arg\min_{\mathbf{v}} \left\{ \Psi(\mathbf{v}) + \frac{\mu}{2} \left\| \mathbf{v} - \mathbf{R}\mathbf{x}^{(j+1)} - \boldsymbol{\eta}_{\mathbf{v}}^{(j)} \right\|_2^2 \right\} \quad (34)$$

$$\boldsymbol{\eta}_{\mathbf{v}}^{(j+1)} = \boldsymbol{\eta}_{\mathbf{v}}^{(j)} - \left(\mathbf{v}^{(j+1)} - \mathbf{R}\mathbf{x}^{(j+1)} \right). \quad (35)$$

The minimization in (34) is same as that in (20), so the techniques described for (20) apply to (34) as well. The main difference between the proposed method (15)–(17) and the SB scheme (33)–(35) is in the way $\mathbf{x}$ is updated. The minimization in (33) leads to

$$\mathbf{x}^{(j+1)\star} = \mathbf{B}_{\mu}^{-1} \left[\mathbf{A}^{\top}\mathbf{W}\mathbf{y} + \mu\mathbf{R}^{\top} \left(\mathbf{v}^{(j)} - \boldsymbol{\eta}_{\mathbf{v}}^{(j)} \right) \right] \quad (36)$$

where $\mathbf{x}^{(j+1)\star}$ represents the exact solution to (36) and

$$\mathbf{B}_{\mu} \triangleq (\mathbf{A}^{\top}\mathbf{W}\mathbf{A} + \mu\mathbf{R}^{\top}\mathbf{R}). \quad (37)$$

The matrix $\mathbf{B}_{\mu}$ contains the shift-variant component $\mathbf{W}$ that makes standard preconditioners (including cone filters) less effective for CG-based solving of (36). Nevertheless, we used PCG for (36) with a circulant preconditioner $\tilde{\mathbf{G}}_{\mu}^{-1}$ [obtained by setting $\mu \equiv \nu$ in (30)] in our implementation of the SB scheme and found that it improved upon the standard CG method for (36). We selected[10] $\mu = L_{\text{data}}/(100\sigma_{\max}\{\mathbf{R}^{\top}\mathbf{R}\})$ for SB [11] (see (4) for definitions of L_{data} and $\sigma_{\max}$). This choice is motivated by the discussion pertaining to ν in Section III-E.

In principle, it is possible to construct a shift-variant preconditioner for $\mathbf{B}_{\mu}$ in the spirit of [6], but such a preconditioner would invariably be data-dependent and may be computationally involved. Our approach (15)–(17) provides a simple and effective alternative using an extra constraint variable $\mathbf{u}$ in (5): Compared to the SB scheme (33)–(35), our method requires only an extra trivial operation of inverting a diagonal matrix $\mathbf{D}_{\mu}$ in (21).

V. EXPERIMENTAL RESULTS

We present numerical results for 2-D CT reconstruction from simulated NCAT phantom data and *in vivo* human head data. The proposed ADMM is also applicable, in principle, to 3-D CT reconstruction (see Section VI-A). We implemented the following algorithms in Matlab and conducted the experiments on a quad-core PC with 3.07-GHz Intel Xeon processors and 12-GB RAM.

- **NCG-** n: unpreconditioned nonlinear conjugate gradient algorithm with n line-search iterations that monotonically decrease the cost function J [6],

TABLE I
COMPUTATION TIME AND NUMBER OF PROJECTIONS REQUIRED PER
ITERATION OF ALGORITHMS COMPARED IN Section V

Algorithm	Time / Iteration (in seconds)		Projection operations / Iteration	
	Section V-A	Section V-B	Forward (**A**)	Backward (**A**$^{\top}$)
NCG-5	1.56	4.85	1	1
NCG-10	-	8.83	1	1
MFISTA-5	2.49	8.87	1	1
MFISTA-25	5.23	-	1	1
OS-4	-	10.19	1 (effective)	1 (effective)
OS-41	-	61.84	1 (effective)	1 (effective)
SB-CG-1	2.29	6.22	1	1
SB-CG-2	3.29	8.93	2	2
SB-PCG-1	2.29	6.25	1	1
SB-PCG-2	3.30	9.07	2	2
ADMM-CG-1	3.29	8.91	1	2
ADMM-CG-2	4.31	11.61	2	3
ADMM-PCG-1	3.32	8.94	1	2
ADMM-PCG-2	4.34	11.70	2	3

- **MFISTA-** n: Monotone Fast Iterative Shrinkage-Thresholding Algorithm [16] with n iterations for solving auxiliary denoising subproblems similar to [16, Eq. 3.13],
- **OS-** n: Ordered subsets algorithm [5] with n blocks,
- **SB-(P)CG-** n: Split-Bregman scheme from Section IV with n (P)CG iterations for solving (36),
- **ADMM-(P)CG-** n: Proposed ADMM with n (P)CG iterations for solving (27).

MFISTA is a state-of-the-art method developed by Beck *et al.* [16] for image restoration that is readily applicable to **P0** with the Lipschitz constant L_{data} in (4). Beck *et al.* [15] also proposed a backtracking strategy that does not require explicit computation of L_{data}, but we chose to estimate and use[11] L_{data} both for ease of implementation and because it is the smallest possible value [15, Ex. 2.2] that yields the fastest convergence for MFISTA. We applied the Chambolle-type method [50] for the inner-step (i.e., computing the proximal map [16, Eq. 3.13]) of MFISTA as that does not require smoothing of ("corners" of) ℓ_1-regularizers such as (22).

Since our task is to solve **P0**, we fixed the cost function J (that led to a visually appealing reconstruction) and focussed on the convergence speed of the algorithms. We quantified the convergence rate using the normalized ℓ_2-distance between $\mathbf{x}^{(j)}$ and $\mathbf{x}^{\star}$

$$\xi(j) = 20 \log_{10} \left(\frac{\left\| \mathbf{x}^{(j)} - \mathbf{x}^{\star} \right\|_2}{\|\mathbf{x}^{\star}\|_2} \right) \quad (38)$$

where $\mathbf{x}^{\star}$ is a solution to **P0** obtained numerically by running one of the above algorithms as described next. Since the algorithms have different computation load per (outer) iteration, we evaluated $\xi(j)$ as a function of algorithm runtime[12] t_j, i.e., the time elapsed from start until iteration j. We also plot $\xi(j)$ as a function of the iteration index j for completeness. We used the DD-projector [49] (with eight threads) for implementing matrix-vector products such as $\mathbf{A}\mathbf{x}$, $\mathbf{A}^{\top}\mathbf{u}$ and initialized all the algorithms with the image reconstructed using FBP (with the ramp filter) in all experiments.
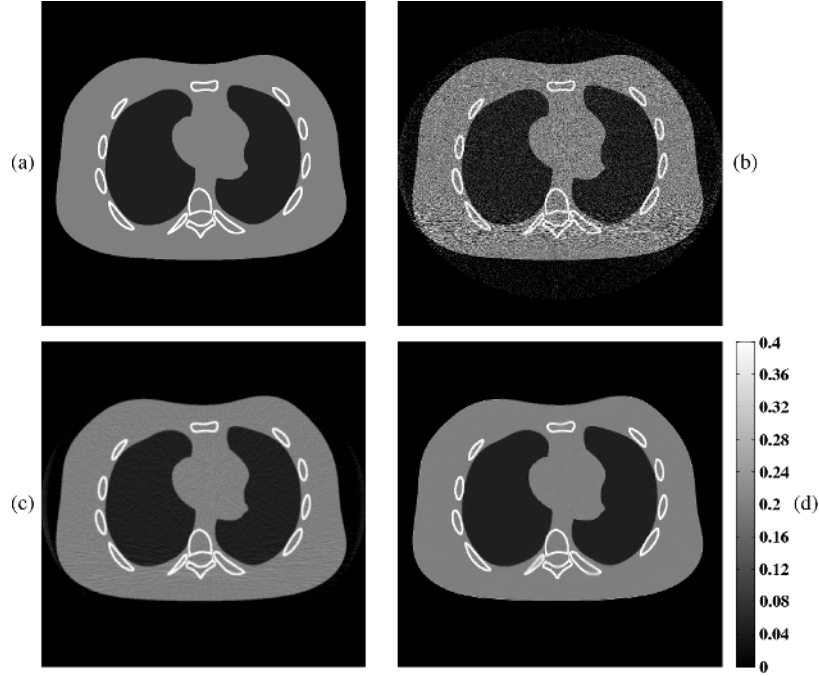
Fig. 2. Simulation with the NCAT phantom: (a) Noisefree NCAT phantom (in cm^{-1}), (b) FBP reconstruction with ramp filter, also the initial guess $\mathbf{x}^{(0)}$ for all iterative algorithms, (c) FBP reconstruction with Hanning filter, and (d) ℓ_1-regularized reconstruction, also the solution $\mathbf{x}^\star$ to $\mathbf{P0}$. Images in (a)–(d) have been normalized to the same color scale [as that of (a)] indicated beside (d). The ℓ_1-regularized reconstruction (d) is less noisy and has almost no streaky-artifacts compared to both FBP results.

Products with $\mathbf{A}$ and $\mathbf{A}^\top$ (corresponding to forward-and back-projections, respectively) are computation intensive in CT reconstruction problems and dominate the overall computation load of a reconstruction algorithm.[13] NCG and MFISTA both require only one product with $\mathbf{A}$ and $\mathbf{A}^\top$, respectively, per iteration. The OS method breaks products with $\mathbf{A}$ and $\mathbf{A}^\top$ in terms of block-rows of $\mathbf{A}$ and block-columns of $\mathbf{A}^\top$, respectively, and cycles through each block once per every iteration, so effectively, OS also requires only one product with $\mathbf{A}$ and $\mathbf{A}^\top$, respectively, per iteration. However, for each block, the OS method demands the evaluation of the gradient of the regularization term that increases computation time per iteration as indicated in Table I. For the SB scheme, we employ (P)CG for "inverting" $\mathbf{B}_\mu$ (that depends on $\mathbf{A}^\top \mathbf{W} \mathbf{A}$) in (36), so **SB-(P)CG-** n requires n products with $\mathbf{A}$ and $\mathbf{A}^\top$, respectively, per iteration of (33)–(35). In the case of ADMM, we apply (P)CG at Step 3 (see Fig. 1) for "inverting" $\mathbf{G}_\nu$ in (27), but that step also requires a product with $\mathbf{A}^\top$ in the RHS of (27), so overall **ADMM-(P)CG-** n uses n products with $\mathbf{A}$ and $n + 1$ products with $\mathbf{A}^\top$ per iteration of Steps 3–7 in Fig. 1. Table I summarizes this discussion and also shows the mean computation time per iteration (averaged over 10 iterations) of the above algorithms. Although the proposed ADMM(-PCG) requires more forward- and back-projections per iteration (and accordingly exhibits higher computation time per iteration) compared to other algorithms (with the exception of the OS method) in Table I, we demonstrate in the sequel that it converges faster in terms of algorithm run-time.

[13]NCG, MFISTA, OS, and SB require the evaluation of $\mathbf{A}^\top \mathbf{W} \mathbf{y}$ (e.g., see RHS of (36) for the SB scheme), but this quantity needs to be computed once, so we ignore this computation need for these schemes.

A. Simulation with the NCAT Phantom

We used a 1024×1024 2-D slice of the NCAT phantom [51] and numerically generated a 888×984-view noisy sinogram with GE LightSpeed fan-beam geometry [52] corresponding to a monoenergetic source with 2.5×10^4 incident photons per ray and no background events. We used the ℓ_1-regularization in (22) with $\kappa_r = \omega_{(r \bmod N)}$, where $\omega_n \triangleq \sqrt{[\mathbf{A}^\top \mathbf{W} \mathbf{1}]_n / [\mathbf{A}^\top \mathbf{1}]_n}$ is based on [26]. We reconstructed 512×512 images over a FOV of 65 cm; we obtained $\mathbf{x}^\star$ by running 5000 iterations of **MFISTA-25** as it does not require "corner-rounding" and is therefore guaranteed to converge to a solution of $\mathbf{P0}$. NCG cannot directly handle nonsmooth criteria such as (22) without smoothing it [13, App. A], so we used a smoothing value of $10^{-6}\ cm^{-1}$. The FBP reconstructions in Fig. 2(b) and (c) corresponding to the ramp and Hanning filters, respectively, are either noisy or blurred and streaked with artifacts. The ℓ_1-regularized reconstruction $\mathbf{x}^\star$ in Fig. 2(d) preserves image features and has lower RMSE than both FBP outputs.

We plot $\xi(j)$ for various algorithms as a function of time in Fig. 3(a). The SB-CG scheme appears to converge the slowest, while SB-PCG is faster indicating that the circulant preconditioner $\tilde{\mathbf{G}}_\mu^{-1}$ provides a moderate acceleration of CG for (36). MFISTA is slower than most of the algorithms for the reason explained in Section II. The CG-version of the proposed method, ADMM-CG, is slightly faster than MFISTA and SB-(P)CG but slower than NCG. The preconditioned version ADMM-PCG is the fastest among all algorithms illustrating that the cone-filter preconditioner $\tilde{\mathbf{G}}_\nu^{-1}$ is very effective in accelerating convergence of CG applied to (27) and ADMM-PCG. This is also corroborated by Fig. 3(c) where for a given number of iterations,
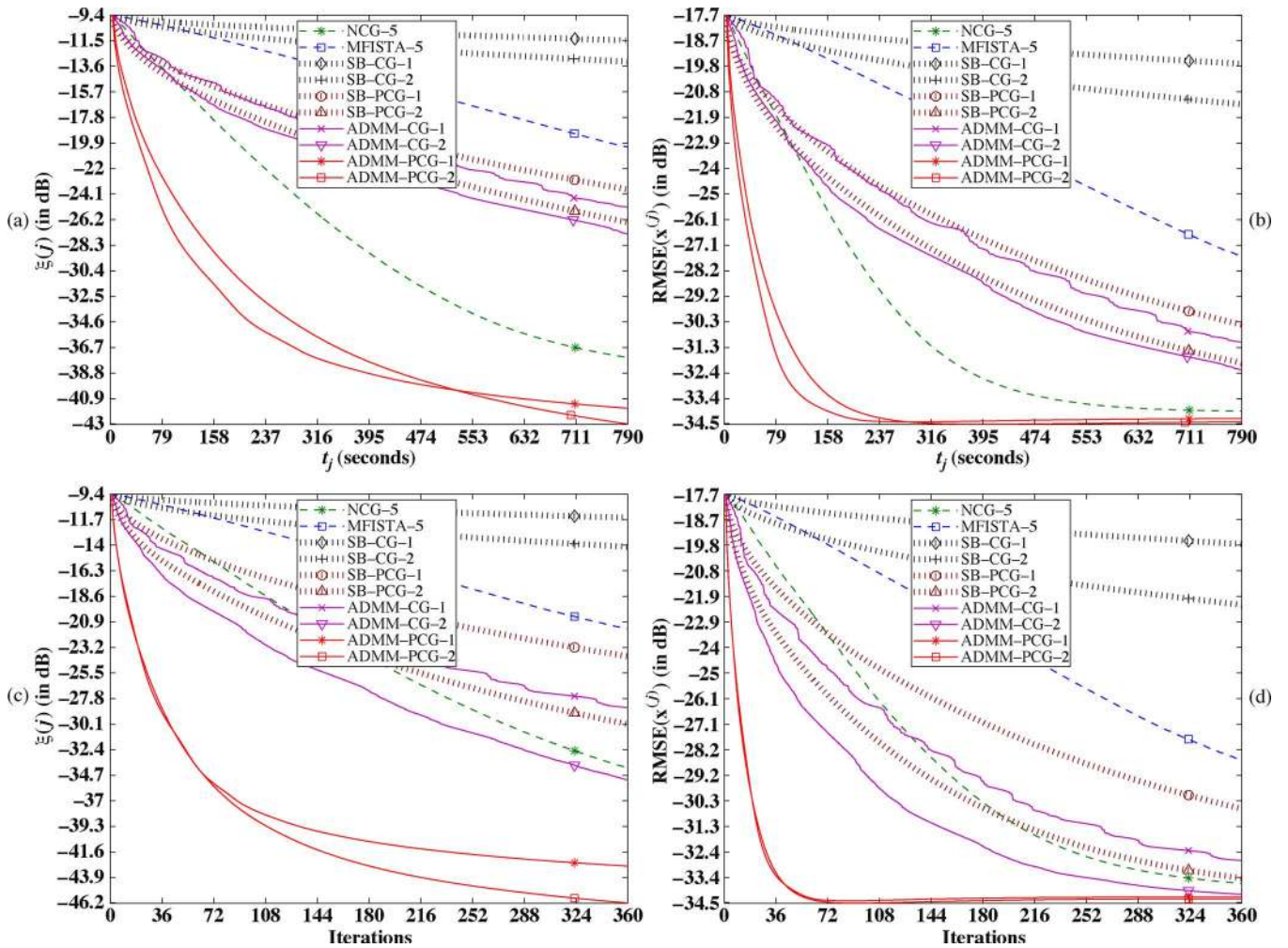
Fig. 3. Simulation with the NCAT phantom: (a), (b) Plot of $\xi(j)$ and $\mathrm{RMSE}(\mathbf{x}^{(j)})$, respectively, as a function of time t_j and (c), (d) Plot of $\xi(j)$ and $\mathrm{RMSE}(\mathbf{x}^{(j)})$, respectively, with respect to iterations, for various algorithms considered in this work. The unpreconditioned version of the proposed method, ADMM-CG, converges slightly faster than MFISTA and the split-Bregman scheme SB-(P)CG but is slower than NCG as seen in (a) and (b). But the preconditioned version, ADMM-PCG, converges rapidly both in terms of $\xi(j)$ and RMSE indicating that the cone-filter-preconditioner ($\tilde{\mathbf{G}}_\nu^{-1}$ in Section III-E) greatly accelerates convergence of the proposed ADMM.

ADMM-PCG produces a reconstruction that is closest to $\mathbf{x}^\star$ in terms of $\xi(j)$. Fig. 3(b) and (d) further substantiate the reconstruction speed-up of ADMM-PCG over other methods, where (both in terms of algorithm run-time and number of iterations) it rapidly leads to a RMSE-value close to $\mathrm{RMSE}(\mathbf{x}^\star)$.

B. Experiments with a In Vivo Human Head Data-Set

In this experiment, we used a *in vivo* human head data-set acquired with a GE scanner using 120 kVp source potential and 585 mA tube current with 0.6 s rotation. We reconstructed a 1024×1024 2-D slice with 50 cm FOV and 0.625 mm thickness from a 888×984-view sinogram. For Ψ in (1), we used the strictly convex regularizer (23) (with $\delta = 10$ HU) that guarantees a unique solution $\mathbf{x}^\star$ to **P0**. As NCG generally had faster convergence than MFISTA in our experiments, we obtained $\mathbf{x}^\star$ by running 5000 iterations of **NCG-10**. Fig. 4 shows the reconstruction results for this experiment. The regularized solution $\mathbf{x}^\star$ in Fig. 4(c) has reduced noise and better preserves

the anatomical features compared to the FBP reconstructions in Fig. 4(a) and (b) obtained using the ramp and Hanning filters, respectively.

Fig. 5(a)–(d) plots $\xi(j)$ as a function of t_j and iteration index j for all algorithms considered in this work. Here, we additionally compare the standard OS algorithm (that is not guaranteed to converge) in Fig. 5(b) and (d), where we used the implementation from [53] available currently for regularization criteria such as (23). The OS algorithm is faster than all algorithms (including ADMM-PCG) for the first few iterations but it does not converge to the minimizer as expected. In practice, it may be advantageous to run a few iterations of OS and use its output to initialize a more sophisticated iterative algorithm. Fig. 5(a) and (b) indicates that the convergence trends for MFISTA, NCG, SB-(P)CG and ADMM-CG are generally similar to those in Fig. 3(a) and (b). ADMM-PCG again provides notable reconstruction speed-up compared to all algorithms. This substantiates the potential of the cone-filter preconditioner
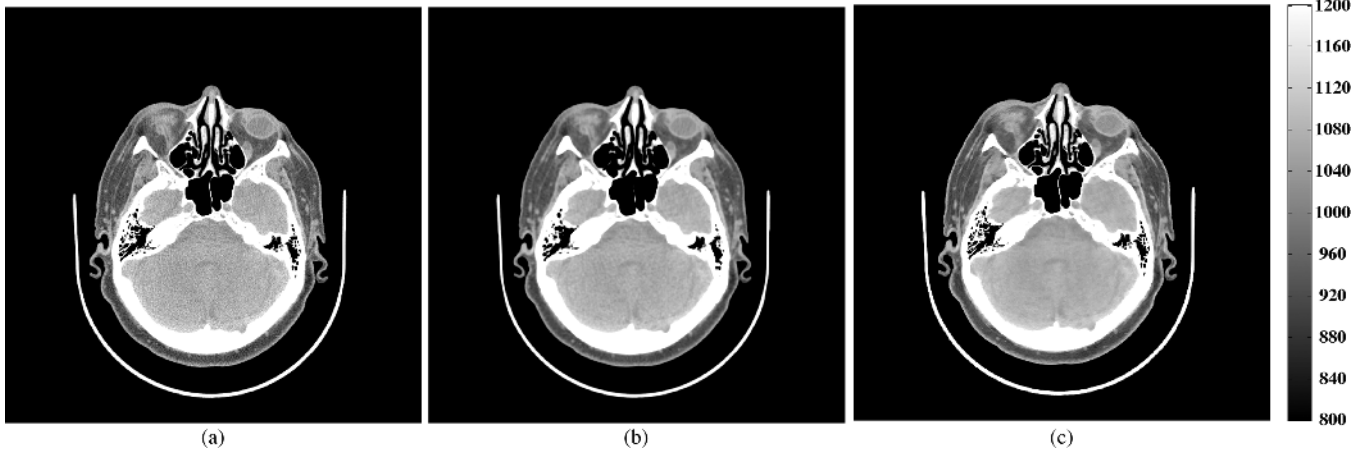
Fig. 4. Experiment with the *in vivo* human head data-set: (a) FBP reconstruction with the ramp filter, also the initial guess $\mathbf{x}^{(0)}$ for all iterative algorithms, (b) RBP reconstruction with Hanning filter, and (c) PWLS reconstruction with the strictly convex regularizer (23), also the unique solution $\mathbf{x}^\star$ to **P0**. Images in (a)–(c) are displayed in Hounsfield units indicated beside (c). The regularized reconstruction (c) is less noisy and preserves anatomical features compared to both FBP results.

(30) for the proposed ADMM and also demonstrates the benefit of our splitting scheme (5).

VI. DISCUSSION

A. Memory Requirements

Splitting-based algorithms simplify optimization at the expense of manipulating and storing auxiliary constraint variables (and corresponding Lagrange multipliers in the AL formalism) and therefore have additional memory requirements compared to conventional algorithms such as NCG. Although this does not pose much concern for 2-D reconstruction problems, it can represent a significant memory overhead for 3-D problems. Specifically, the SB (Section IV) and the proposed ADMM[14] (in Fig. 1) schemes use the constraint $\mathbf{v} = \mathbf{R}\mathbf{x}$ that requires the storage of $2P$ vectors ($\mathbf{v}$ and $\boldsymbol{\eta}_{\mathbf{v}}$) of size $L \times 1$. For instance, typically, the size of an image-volume in 3-D CT is $N = 512 \times 512 \times 512$ (≈ 1 GB of memory when stored in double-precision format in Matlab). Then, for finite-differences with $P = 13$ (there are 13 nearest-neighbors on one side of any voxel), this corresponds to storing at least 26 image-volumes (≈ 26 GB of memory) that might set a practical limitation on these methods from an implementation perspective.

A quick remedy is to consider the TV regularizer with finite-differences only along the three orthogonal directions ($P = 3$ corresponds to six image-volumes) which considerably reduces the memory load. Alternatively, one could also consider using an orthonormal transform (such as orthonormal wavelets[15]) for $\mathbf{R}$, so $P = 1$ and $L = N$. The SB and ADMM[14] schemes would then require storing only two image-volumes (corresponding to $\mathbf{v}$ and $\boldsymbol{\eta}_{\mathbf{v}}$). Moreover, an orthonormal $\mathbf{R}$ satisfies $\mathbf{R}^\top \mathbf{R} = \mathbf{I}_N$ that facilitates ADMM: $\mathbf{G}_\nu$ in (28)

[14]For ADMM, we have to additionally store two $M \times 1$ vectors, $\mathbf{u}$ and the associated multiplier $\boldsymbol{\eta}_{\mathbf{u}}$. This additional memory requirement is moderate for 2-D CT and can be high for 3-D CT depending on the size of the data.

[15]Quality-wise, shift-invariant wavelets are preferable to orthonormal ones [54], but due to their over-complete nature, they require significantly more memory (similar to finite differences) than orthonormal wavelets.

becomes $\mathbf{G}_\nu = (\mathbf{A}^\top \mathbf{A} + \nu \mathbf{I}_N)$ that is still "nearly" shift-invariant and can be effectively preconditioned using circulant preconditioners. With orthonormal wavelets, one also has the option of excluding the approximation coefficients from the regularization (as they are not sparse) by using scale dependent regularization parameters [50] and setting those parameters corresponding to the approximation level to zero.

B. Inclusion of Nonnegativity Constraint

In CT, a nonnegativity constraint is often imposed [1, Eq. 18], [11, Sec. 2.2] to model the positivity of the attenuation coefficient that is being reconstructed. Although we have not considered such a constraint in **P0**, it can be easily accommodated [38] as follows. We start with

$$\arg \min_{\mathbf{x}} \left\{ J(\mathbf{x}) \triangleq J_{\text{data}}(\mathbf{y}, \mathbf{A}\mathbf{x}) + \Psi(\mathbf{R}\mathbf{x}) + g(\mathbf{x}) \right\} \quad (39)$$

where g is an indicator function

$$g(\mathbf{x}) \triangleq \begin{cases} 0, & \text{if } \mathbf{x} \geq 0 \\ +\infty, & \text{otherwise} \end{cases} \quad (40)$$

that imposes the nonnegativity constraint, $\mathbf{x} \geq 0$, taken component-wise in (40). We then consider the following equivalent constrained version [38] that has an additional constraint compared to (5)

$$\arg \min_{\mathbf{x},\mathbf{u},\mathbf{v},\mathbf{w}} \left\{ f(\mathbf{u}, \mathbf{v}, \mathbf{w}) \triangleq J_{\text{data}}(\mathbf{y}, \mathbf{u}) + \Psi(\mathbf{v}) + g(\mathbf{w}) \right\}$$
$$\text{s.t.} \quad \mathbf{u} = \mathbf{A}\mathbf{x}, \ \mathbf{v} = \mathbf{R}\mathbf{x}, \ \mathbf{w} = \mathbf{x}. \quad (41)$$

Writing $\mathbf{z} \triangleq [\mathbf{u}^\top \mathbf{v}^\top \mathbf{w}^\top]^\top$, $\mathbf{C} \triangleq [\mathbf{A}^\top \mathbf{R}^\top \mathbf{I}_N]^\top$ and using $\boldsymbol{\Lambda} = \text{diag}\{\mathbf{I}_M, \nu_1 \mathbf{I}_R, \nu_2 \mathbf{I}_N\}$ in (9), we can design an ADMM-type algorithm similar to (15)–(17) for solving (41). It can be shown that the updates for $\mathbf{u}$ and $\mathbf{v}$ in this algorithm will be similar to (19) and (20), respectively, while the update of $\mathbf{x}$ would involve the "inversion" of $\mathbf{G}_{\nu_1 \nu_2} = (\mathbf{A}^\top \mathbf{A} + \nu_1 \mathbf{R}^\top \mathbf{R} + \nu_2 \mathbf{I}_N)$ and that of $\mathbf{w}$ would require a simple projection onto the positive orthant [38, Eq. 32]. Since $\mathbf{G}_{\nu_1 \nu_2}$ is also "nearly" shift-invariant, a cone-filter-type preconditioner similar to $\tilde{\mathbf{G}}_\nu^{-1}$ [see (30)] can be used for effective preconditioning of $\mathbf{G}_{\nu_1 \nu_2}^{-1}$. Moreover, the
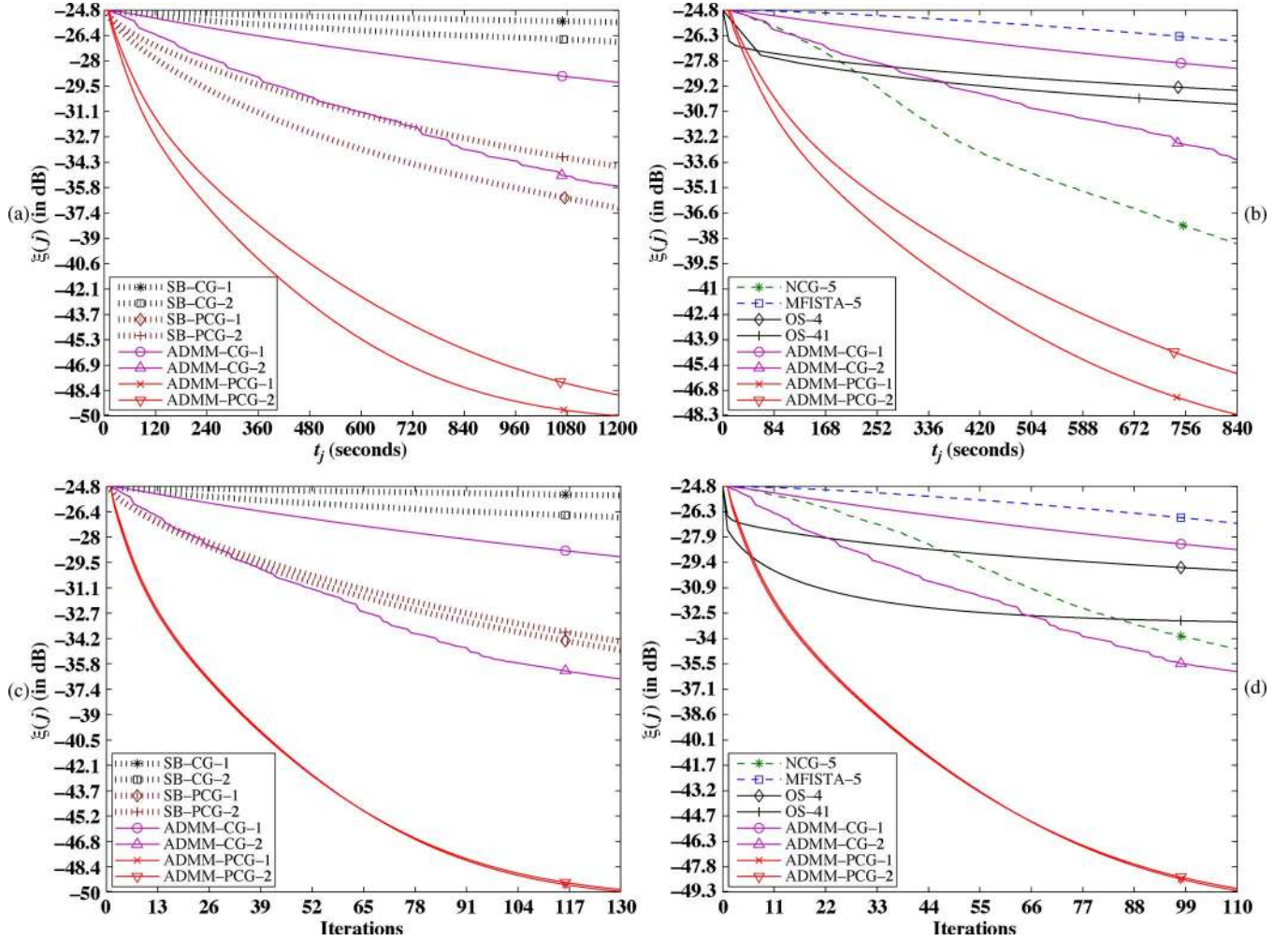
Fig. 5. Experiment with the *in vivo* human head data-set: (a), (b) Plot of $\xi(j)$ as a function of time t_j and (c), (d) Plot of $\xi(j)$ with respect to iterations, for various algorithms considered in this work. MFISTA and SB-CG appear to be the slowest. The proposed ADMM-(P)CG is generally faster than the split-Bregman scheme SB-(P)CG as seen in (a). Although ADMM-CG converges slower than NCG as seen in (b), the preconditioned version ADMM-PCG is the fastest among the considered algorithms, illustrating the benefit of the cone-filter-based preconditioner ($\bar{\mathbf{G}}_\nu$ in Section III-E) for the proposed ADMM.

above $\mathbf{C}$ has full column-rank, so this algorithm also satisfies Theorem 1 and is guaranteed to converge to a solution of (41) and (39).

C. Poisson-Likelihood Model for X-ray CT Reconstruction

The proposed strategy of splitting the data-term [i.e., the use of $\mathbf{u}$ in (5) and (41)] is also applicable for X-ray CT reconstruction using the Poisson-likelihood (PL) statistical model [5, Eq. 1] that may be more suitable for low-dose acquisitions. It can be shown that splitting the PL data-term yields separable 1-D problems in $\{u_i\}_{i=1}^{M}$ that can be solved simultaneously similar to [38, Eq. 30]. However, the PL model for X-ray CT may preclude exact updates like (21) for $\{u_i\}$. Moreover, the general PL model [5, Eq. 1] includes background events and can be ("mildly") nonconvex, so Theorem 1 cannot be directly applied to an ADMM-type algorithm developed for this problem. We plan to explore cost functions involving the PL model [5, Eq. 1] for transmission tomography reconstruction as part of future extensions to this work.

VII. Summary and Conclusion

Statistical X-ray CT reconstruction using penalized weighted least-squares (PWLS) criteria involve a diagonal weighting matrix $\mathbf{W}$ that poses a hindrance to several optimization methods due to its huge dynamic range and highly shift-variant nature. In this work, we employed a variable-splitting technique that, in addition to separating the regularization term like [17], also dissociates the statistical ($\mathbf{W}$) and the system ($\mathbf{A}$) components in the data term to decouple and mitigate the effect of $\mathbf{W}$. We applied the method of multipliers [19] with alternating minimization [21]–[23] for the resulting equivalent constrained problem and developed an alternating direction method of multipliers (ADMM) algorithm that chiefly involves three simple operations at each iteration: i) inverting a diagonal matrix that depends on $\mathbf{W}$, ii) minimizing a set of 1-D auxiliary denoising-cost-functions that can be performed efficiently and/or exactly for a variety of regularizers, and iii) solving a "nearly" shift-invariant linear system (involving $\mathbf{A}^\top \mathbf{A}$) using FFT-based preconditioning with cone-type filters [7].

The proposed ADMM algorithm is guaranteed to converge to a solution of the original PWLS problem under a mild condition

on the accuracy of operation iii) above. We demonstrated using simulations and experiments with real *in vivo* human data that cone-filter-type preconditioners are very effective for solving the linear system in iii) and that the preconditioned version of the proposed ADMM converges faster than conventional (NCG and ordered subsets) and state-of-the-art (MFISTA and split-Bregman) algorithms for CT. The proposed ADMM can handle a variety of regularization criteria for 2-D CT reconstruction and is also applicable to 3-D CT reconstruction, perhaps by using certain memory-conserving regularizers.

REFERENCES

[1] J.-B. Thibault, K. Sauer, C. Bouman, and J. Hsieh, "A three-dimensional statistical approach to improved image quality for multi-slice helical CT," *Med. Phys.*, vol. 34, no. 11, pp. 4526–44, Nov. 2007.

[2] K. Sauer and C. Bouman, "A local update strategy for iterative reconstruction from projections," *IEEE Trans. Signal Process.*, vol. 41, no. 2, pp. 534–48, Feb. 1993.

[3] T. M. Benson, B. K. B. D. Man, L. Fu, and J.-B. Thibault, "Block-based iterative coordinate descent," in *Proc. IEEE Nucl. Sci. Symp. Med. Im. Conf.*, 2010, pp. 2856–2859.

[4] C. Kamphuis and F. J. Beekman, "Accelerated iterative transmission CT reconstruction using an ordered subsets convex algorithm," *EEE Trans. Med. Imag.*, vol. 17, no. 6, pp. 1001–5, Dec. 1998.

[5] H. Erdogan and J. A. Fessler, "Ordered subsets algorithms for transmission tomography," *Phys. Med. Biol.*, vol. 44, no. 11, pp. 2835–2851, Nov. 1999.

[6] J. A. Fessler and S. D. Booth, "Conjugate-gradient preconditioning methods for shift-variant PET image reconstruction," *IEEE Trans. Image Process.*, vol. 8, no. 5, pp. 688–699, May 1999.

[7] N. H. Clinthorne, T. S. Pan, P. C. Chiao, W. L. Rogers, and J. A. Stamos, "Preconditioning methods for improved convergence rates in iterative reconstructions," *IEEE Trans. Med. Imag.*, vol. 12, no. 1, pp. 78–83, Mar. 1993.

[8] A. H. Delaney and Y. Bresler, "A fast and accurate Fourier algorithm for iterative parallel-beam tomography," *IEEE Trans. Image Process.*, vol. 5, no. 5, pp. 740–753, May 1996.

[9] J. W. Wallis and T. R. Miller, "Rapidly converging iterative reconstruction algorithms in single-photon emission computed tomography," *J. Nuc. Med.*, vol. 34, no. 10, pp. 1793–1800, Oct. 1993.

[10] J. Nuyts, B. De Man, P. Dupont, M. Defrise, P. Suetens, and L. Mortelmans, "Iterative reconstruction for helical CT: A simulation study," *Phys. Med. Biol.*, vol. 43, no. 4, pp. 729–737, Apr. 1998.

[11] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Phys. Med. Biol.*, vol. 53, no. 17, pp. 4777–4808, Sep. 2008.

[12] 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, Mar. 2011.

[13] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Mag. Res. Med.*, vol. 58, no. 6, pp. 1182–1195, Dec. 2007.

[14] S. Ramani and J. A. Fessler, "An accelerated iterative reweighted least squares algorithm for compressed sensing MRI," in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2010, pp. 257–260.

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

[16] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2419–2434, Nov. 2009.

[17] T. Goldstein and S. Osher, "The split Bregman method for L1-regularized problems," *SIAM J. Imag. Sci.*, vol. 2, no. 2, pp. 323–343, 2009.

[18] B. Vandeghinste, B. Goossens, J. D. Beenhouwer, A. Pizurica, W. Philips, S. Vandenberghe, and S. Staelens, "Split-Bregman-based sparse-view CT reconstruction," in *Proc. Int. Meeting Fully 3-D Image Recon. Rad. Nucl. Med*, 2011, pp. 431–434.

[19] D. P. Bertsekas, "Multiplier methods: A survey," *Automatica*, vol. 12, no. 2, pp. 133–145, Mar. 1976.

[20] J. Nocedal and S. J. Wright, *Numerical Optimization*. New York: Springer, 1999.

[21] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite element approximations," *Comput. Math. Appl.*, vol. 2, pp. 17–40, 1976.

[22] R. Glowinski and A. Marroco, "Sur l'approximation par element finis d'ordre un, et la resolution, par penalisation-dualité, d'une classe de problemes de Dirichlet nonlineares," *Revue Française d'Automatique, Informatique et Recherche Opérationelle 9*, vol. R-2, pp. 41–76, 1975.

[23] M. Fortin and R. Glowinski, "On decomposition-coordination methods using an augmented Lagrangian," in *Augmented Lagrangian Methods: Applications to the Solution of Boundary-Value Problems*, M. Fortin and R. Glowinski, Eds. Amsterdam, The Netherlands: Elsevier, 1983, vol. 15, Studies in Mathematics and Its Applications, pp. 97–146.

[24] I. A. Elbakri and J. A. Fessler, "Efficient and accurate likelihood for iterative image reconstruction in X-ray computed tomography," in *Proc. SPIE Med. Imag.*, 2003, pp. 1839–1850.

[25] B. R. Whiting, "Signal statistics in x-ray computed tomography," in *Proc. SPIE Med. Imag.*, 2002, pp. 53–60.

[26] J. A. Fessler and W. L. Rogers, "Spatial resolution properties of penalized-likelihood image reconstruction methods: Space-invariant tomographs," *IEEE Trans. Image. Process.*, vol. 5, no. 9, pp. 1346–1358, Sep. 1996.

[27] M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," *Inverse Prob.*, vol. 23, pp. 947–968, 2007.

[28] P. J. Huber *et al.*, *Robust Statistics*. New York: Wiley, 1981.

[29] M. Nikolova and M. K. Ng, "Analysis of half-quadratic minimization methods for signal and image recovery," *SIAM J. Sci. Comp.*, vol. 27, no. 3, pp. 937–966, 2005.

[30] R. C. Fair, "On the robust estimation of econometric models," *Ann. Econ. Social Measure.*, vol. 2, pp. 667–677, Oct. 1974.

[31] I. Daubechies, M. Defrise, and C. D. Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Comm. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, Nov. 2004.

[32] J. M. Bioucas-Dias and M. A. T. Figueiredo, "A new twIST: Two-step iterative shrinkage/Thresholding algorithms for image restoration," *IEEE Trans. Image Process.*, vol. 16, no. 12, pp. 2992–3004, Dec. 2007.

[33] S. Ramani, P. Thévenaz, and M. Unser, "Regularized interpolation for noisy images," *IEEE Trans. Image Process.*, vol. 29, no. 2, pp. 543–558, Feb. 2010.

[34] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *SIAM J. Imag, Sci.*, vol. 1, no. 3, pp. 248–272, 2008.

[35] M. K. Ng, P. Weiss, and X. Yuan, "Solving constrained total-variation image restoration and reconstruction problems via alternating direction methods," *SIAM J. Sci. Comp.*, vol. 32, no. 5, pp. 2710–2736, 2010.

[36] 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. Image Process.*, vol. 20, no. 3, pp. 681–695, Mar. 2011.

[37] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2345–2356, Sep. 2010.

[38] M. A. T. Figueiredo and J. M. Biouscas-Dias, "Restoration of Poissonian images using alternating direction optimization," *IEEE Trans. Image Process.*, vol. 19, no. 12, pp. 3133–3145, Dec. 2010.

[39] X. Ye, Y. Chen, and F. Huang, "Computational acceleration for MR image reconstruction in partially parallel imaging," *IEEE Trans. Med. Imag.*, vol. 30, no. 5, pp. 1055–1063, May 2011.

[40] X. Ye, Y. Chen, W. Lin, and F. Huang, "Fast MR image reconstruction for partially parallel imaging with arbitrary k-space trajectories," *IEEE Trans. Med. Imag.*, vol. 30, no. 3, pp. 575–585, Mar. 2011.

[41] Q. Xu, X. Mou, G. Wang, J. Sieren, E. A. Hoffman, and H. Yu, "Statistical interior tomography," *IEEE Trans. Med. Imag.*, vol. 30, no. 5, pp. 1116–1128, May 2011.

[42] S. Ramani and J. A. Fessler, "Convergent iterative CT reconstruction with sparsity-based regularization," in *Proc. Int. Meeting Fully 3-D Image Recon. Rad. Nucl. Med*, 2011, pp. 302–305.

[43] J. Eckstein and D. P. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Prog.*, vol. 55, no. 1–3, pp. 293–318, Apr. 1992.

[44] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ: Princeton Univ. Press, 1970.

[45] C. Chaux, P. L. Combettes, J.-C. Pesquet, and V. R. Wajs, "A variational formulation for frame-based inverse problems," *Inverse Prob.*, vol. 23, no. 4, pp. 1495–1518, Aug. 2007.

[46] A. Chambolle, R. A. DeVore, N. Y. Lee, and B. J. Lucier, "Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage," *IEEE Trans. Image Process.*, vol. 7, no. 3, pp. 319–335, Mar. 1998.

[47] J. D. Dixon, "Estimating extremal eigenvalues and condition numbers of matrices," *SIAM J. Num. Anal.*, vol. 20, no. 4, pp. 812–814, 1983.

[48] H. Anton and C. Rorres, *Elementary Linear Algebra—Applications Version*. Hoboken, NJ: Wiley, 2010.

[49] B. De Man and S. Basu, "Distance-driven projection and backprojection," in *Proc. IEEE Nucl. Sci. Symp. Med. Im. Conf.*, 2002, vol. 3, pp. 1477–1480.

[50] I. W. Selesnick and M. A. T. Figueiredo, "Signal restoration with overcomplete wavelet transforms: Comparison of analysis and synthesis priors," in *Proc. SPIE*, 2009, p. 74460D.

[51] W. P. Segars and B. M. W. Tsui, "Study of the efficacy of respiratory gating in myocardial SPECT using the new 4-D NCAT phantom," *IEEE Trans. Nucl. Sci.*, vol. 49, no. 3, pp. 675–679, Jun. 2002.

[52] J. Wang, T. Li, H. Lu, and Z. Liang, "Penalized weighted least-squares approach to sinogram noise reduction and image reconstruction for low-dose X-ray computed tomography," *IEEE Trans. Med. Imag.*, vol. 25, no. 10, pp. 1272–1283, Oct. 2006.

[53] J. A. Fessler, Image reconstruction toolbox Univ. Michigan [Online]. Available: http://www.eecs.umich.edu/fessler/code/index.html, 2011

[54] M. A. T. Figueiredo and R. D. Nowak, "Wavelet-based image estimation: An empirical Bayes approach using Jeffrey's noninformative prior," *IEEE Trans. Image Process.*, vol. 10, no. 9, pp. 1322–1331, Sep. 2001.