Convolutional analysis operator learning: Application to sparse-view CT

(Invited Paper)

Il Yong Chun and Jeffrey A. Fessler
Department of Electrical Engineering and Computer Science, The University of Michigan
Ann Arbor, MI 48019-2122 USA
Email: iychun@umich.edu, fessler@umich.edu

***Abstract*—Convolutional analysis operator learning (CAOL) methods train an autoencoding convolutional neural network (CNN) in an unsupervised learning manner, to more accurately solve inverse problems. *Block Proximal Gradient method using a Majorizer* (BPG-M) achieved fast and convergent CAOL, by using sharp majorizers and the momentum terms. This paper proposes a model-based image reconstruction (MBIR) method using autoencoding CNNs trained via CAOL, for sparse-view computational tomography (CT). We apply BPG-M to rapidly and stably solve the corresponding block multi-nonconvex optimization problem. Numerical experiments show that, for sparse-view CT, *1)* the proposed MBIR method outperforms the standard MBIR method using edge-preserving regularization; *2)* larger parameter dimensions of autoencoding CNNs improve reconstruction accuracy of the proposed MBIR method; and *3)* when using BPG-M, sharper majorization is more critical for accelerating its convergence than giving more weights on extrapolation.**

I. Introduction

Convolutional analysis operator learning (CAOL) trains an autoencoding CNN using unsupervised learning to more accurately solve inverse problems [1]. CAOL has several theoretical and/or practical benefits for both training and testing. First, CAOL can benefit from "big data", i.e., training data consisting of many (high-dimensional) signals [2]. Second, the autoencoding CNNs trained via CAOL have signal recovery guarantees when applied to compressed sensing [2]. Third, CAOL is useful for unsupervised training of deep layered CNNs [1].

Block Proximal Gradient method using a Majorizer (BPG-M) [1], [3] uses block-wise extrapolation (i.e., a sequence update using *momentum* terms), and is the state-of-the-art optimization framework for solving block multi-(non)convex problems, when used with a sharp majorizer. By using a more general Lipschitz continuity assumption for block-wise gradients, BPG-M is particularly useful for rapidly calculating majorizers involved with large-scale problems. In particular, BPG-M is effective for quickly learning convolutional operators from big data, e.g., CAOL [1] and convolutional dictionary learning [3].

This paper proposes a model-based image reconstruction (MBIR) method using autoencoding CNNs trained via CAOL, for sparse-view computational tomography (CT). This model additionally incorporates a weighting technique that promotes uniform resolution or noise properties in the reconstructed image. We apply BPG-M to rapidly and stably solve the corresponding block multi-nonconvex optimization problem. Numerical experiments show that, for sparse-view CT, *1)* the proposed MBIR method outperforms the standard MBIR method using edge-preserving (EP) regularization; *2)* larger parameter dimensions of autoencoding CNNs improve performance of the proposed MBIR model; and *3)* when solving the proposed model via BPG-M, sharper majorization is more critical in accelerating its convergence than giving more weights on extrapolation.

This work is supported in part by the Keck Foundation, NIH U01 EB018753, and NIH R01 EB022075.

II. CAOL: Review

This section reviews the (single-layer) CAOL model in [2]. The goal of CAOL is to find a set of filters that "best" sparsify a set of training images. In particular, CAOL aims to learn diverse filters, considering that diverse filters can generate diverse sparse features. Consider the following CAOL model [2]:

$$\underset{D=[d_1,\dots,d_K]}{\operatorname{argmin}} \min_{\{z_{l,k}\}} F(D,\{z_{l,k}\}) + \beta g(D), \tag{1}$$

$$F(D,\{z_{l,k}\}) := \sum_{l=1}^{L}\sum_{k=1}^{K} \frac{1}{2}\left\|d_k \circledast x_l - z_{l,k}\right\|_2^2 + \alpha\|z_{l,k}\|_0,$$

where $\circledast$ denotes a convolution operator (see boundary condition details in [1]), $\{x_l \in \mathbb{C}^J : l = 1,\dots,L\}$ is a set of training images, $\{d_k \in \mathbb{C}^R : k = 1,\dots,K\}$ is a set of convolutional kernels, $\{z_{l,k} \in \mathbb{C}^J : l = 1,\dots,L, k = 1,\dots,K\}$ is a set of sparse codes, $g(D)$ is a regularizer or constraint that encourages filter diversity or incoherence, $\alpha > 0$ is a thresholding parameter controlling the sparsity of features $\{z_{l,k}\}$, and $\beta > 0$ is a regularization parameter for $g(D)$. We often group the K filters into a matrix $D = [d_1,\dots,d_K] \in \mathbb{C}^{R\times K}$. To learn diverse filters via CAOL (1), we proposed a (nonconvex) orthogonality constraint that enforces a tight-frame (TF) condition:

$$\underset{D}{\operatorname{argmin}} \min_{\{z_{l,k}\}} F(D,\{z_{l,k}\}) \quad \text{subj. to } DD^H = \frac{1}{R}\cdot I, \tag{P1}$$

and a regularizer that promotes diversity between filters [1]. In particular, when one applies the "learned" CAO (P1) to

MBIR, the learned CAO (P1) works as an autoencoding CNN [4]–[7]; see Remark 5 below. For fast and convergent CAOL, we applied BPG-M as summarized in the next section.

III. BPG-M: Review

This section reviews *block multi-nonconvex* problems and summarizes BPG-M [1]. Consider the optimization problem

$$\operatorname{argmin} F(u_1,\dots,u_B) := f(u_1,\dots,u_B) + \sum_{b=1}^{B} g_b(u_b), \quad (2)$$

where variable u is decomposed into B blocks $u_1,\dots,u_B$ ($\{u_b \in \mathbb{R}^{n_b} : b = 1,\dots,B\}$), f is assumed to be continuously differentiable, but functions $\{g_b : b = 1,\dots,B\}$ are not necessarily differentiable. The function g_b can incorporate the constraint $u_b \in \mathcal{U}_b$, by allowing the g_b functions to be extended-valued, e.g., $g_b(u_b) = \infty$ if $u_b \notin \mathcal{U}_b$, for $b = 1,\dots,B$. We assume that both f and $\{g_b\}$ are closed and proper and the sets $\{\mathcal{U}_b\}$ are closed and nonempty. Importantly, we do *not* assume that f, $\{g_b\}$, or $\{\mathcal{U}_b\}$ are convex.

In BPG-M, we consider a more general concept than Lipschitz continuity of the gradient and define the following:

Definition 1 (M-Lipschitz continuity [1])**.** *A function $g : \mathbb{R}^n \to \mathbb{R}^n$ is M-*Lipschitz continuous *on $\mathbb{R}^n$ if there exist a (symmetric) positive definite matrix M such that*

$$\|g(u) - g(v)\|_{M^{-1}} \le \|u - v\|_M, \quad \forall u, v,$$

where $\|x\|_M^2 := x^T M x$.

If the gradient of a function is M-Lipschitz continuous, then we obtain the following quadratic majorizer at a given point y without assuming convexity:

Lemma 2 (Quadratic majorization via M-Lipschitz continuous gradients [1])**.** *Let $f(u) : \mathbb{R}^n \to \mathbb{R}$. If ∇f is M-Lipschitz continuous, then*

$$f(u) \le f(v) + \langle \nabla_u f(v), u - v\rangle + \frac{1}{2}\|u - v\|_M^2, \quad \forall u, v \in \mathbb{R}^n.$$

Using Definition 1 and Lemma 2, BPG-M is given as follows. To solve (2), we minimize majorizers of F cyclically over each block $u_1,\dots,u_B$, while fixing the remaining blocks at their previously updated variables. Let $u_b^{(i+1)}$ be the value of u_b after its ith update, and define

$$f_b^{(i+1)}(u_b) := f\Big(u_1^{(i+1)},\dots,u_{b-1}^{(i+1)}, u_b, u_{b+1}^{(i)},\dots,u_B^{(i)}\Big),$$

for all b, i. At the bth block of the ith iteration, we apply Lemma 2 to functional $f_b^{(i+1)}(u_b)$ with a M-Lipschitz continuous gradient at the extrapolated point $\acute{u}_b^{(i+1)}$, and minimize the majorized function. The corresponding proximal mapping problem is given by

$$\begin{aligned} u_b^{(i+1)} &= \operatorname{Prox}_{g_b}\Big(\acute{u}_b^{(i+1)} - \Big(\widetilde{M}_b^{(i+1)}\Big)^{-1}\nabla f_b^{(i+1)}(\acute{u}_b^{(i+1)}); \widetilde{M}_b^{(i+1)}\Big) \\ &:= \operatorname*{argmin}_{u_b} \langle \nabla f_b^{(i+1)}(\acute{u}_b^{(i+1)}), u_b - \acute{u}_b^{(i+1)}\rangle \\ &\qquad + \frac{1}{2}\Big\|u_b - \acute{u}_b^{(i+1)}\Big\|^2_{\widetilde{M}_b^{(i+1)}} + g_b(u_b) \end{aligned}$$

Algorithm 1 BPG-M [1]

Require: $\{u_b^{(0)} = u_b^{(-1)} : \forall b\}$, $\{e_b^{(i)} \in [0,1], \forall b, i\}$, $i = 0$
  while a stopping criterion is not satisfied **do**
    for $b = 1,\dots,B$ **do**
      Calculate $M_b^{(i+1)}$, $\widetilde{M}_b^{(i+1)}$ by (4), and $E_b^{(i+1)}$ by (5)
      $\acute{u}_b^{(i+1)} = u_b^{(i)} + E_b^{(i+1)}\Big(u_b^{(i)} - u_b^{(i-1)}\Big)$
      $u_b^{(i+1)}$
      $= \operatorname{Prox}_{g_b}\Big(\acute{u}_b^{(i+1)} - \Big(\widetilde{M}_b^{(i+1)}\Big)^{-1}\nabla f_b^{(i+1)}(\acute{u}_b^{(i+1)}); \widetilde{M}_b^{(i+1)}\Big)$
    end for
    $i = i + 1$
  end while

where

$$\acute{u}_b^{(i+1)} = u_b^{(i)} + E_b^{(i+1)}\left(u_b^{(i)} - u_b^{(i-1)}\right), \quad (3)$$

$\nabla f_b^{(i+1)}(\acute{u}_b^{(i+1)})$ is the block-partial gradient of f at $\acute{u}_b^{(i+1)}$, a *scaled majorization matrix* is updated by

$$\widetilde{M}_b^{(i+1)} = \lambda_b \cdot M_b^{(i+1)} \succ 0, \qquad \lambda_b > 1, \quad (4)$$

and $M_b^{(i+1)} \in \mathbb{R}^{n_b \times n_b}$ is a symmetric positive definite *majorization matrix* of $\nabla f_b^{(i+1)}(u_b)$. In (3), the $\mathbb{R}^{n_b \times n_b}$ matrix $E_b^{(i+1)} \succeq 0$ is an *extrapolation matrix* that accelerates convergence in solving block multi-convex problems [3]. We design it in the following form:

$$E_b^{(i+1)} = \delta e_b^{(i)} \cdot \frac{\lambda_b - 1}{2(\lambda_b + 1)} \cdot \Big(M_b^{(i+1)}\Big)^{-1/2}\Big(M_b^{(i)}\Big)^{1/2}, \quad (5)$$

for some $\delta < 1$ and $\{0 \le e_b^{(i)} \le 1 : \forall b, i\}$. We apply the following increasing momentum coefficient formula to (5):

$$e_b^{(i+1)} = \frac{\theta^{(i)} - 1}{\theta^{(i+1)}}, \quad \theta^{(i+1)} = \frac{1 + \sqrt{1 + 4(\theta^{(i)})^2}}{2}, \quad (6)$$

which accelerated BPG-M in solving block multi-convex problems in some applications [3], [8]. Algorithm 1 summarizes these updates. Algorithm 1 has the following convergence result.

Theorem 3 (A limit point is a critical point [1])**.** *Under some mild conditions (e.g., lower-boundedness of F, lower semicontinuity of $\{g_b\}$, and existence of critical point of F, etc. [1, Assumptions i–iii]), any limit point of $x^{(i+1)}$ generated by Algorithm 1 is a critical point $\bar{x}$, i.e., $0 \in \partial F(\bar{x})$.*

Remark 4 In Algorithm 1, there is a tradeoff between majorization sharpness via (4) and extrapolation effect via (3) and (5). For CAOL with sufficiently sharp majorizers, we observed that emphasizing sharp majorization provides faster convergence than emphasizing extrapolation; for example, $\lambda_b = 1 + \epsilon$ gives faster convergence than $\lambda_b = 2$. See CAOL experiments with different λ_b values in [1]. When solving block multi-*convex* problems, BPG-M does not have this tradeoff [3].

IV. Sparse-View CT MBIR with "Learned" Regularizer via CAOL, and BPG-M

This section applies pre-trained filters from a representative dataset to recover images in *extreme* imaging that collects highly undersampled or noisy measurements. We choose a sparse-view CT application since it has interesting challenges in reconstructing images, including Poisson noise in measurements, nonuniform noise or resolution properties in reconstructed images, and complicated or unstructured system matrices. For CT, undersampling schemes can significantly reduce the radiation dose and cancer risk from CT scanning. The proposed approach can be applied to other image processing or reconstruction applications, e.g., (parallel) magnetic resonance imaging, deblurring, etc. (by replacing the data fidelity and spatial strength regularization terms in (P2) below).

We pre-learn TF filters $\{d_k^\star \in \mathbb{R}^K : k = 1, \dots, K\}$ via CAOL (P1) with a set of high-quality (e.g., normal-dose) CT images $\{x_l : l = 1, \dots, L\}$. To reconstruct a linear attenuation coefficient image $x \in \mathbb{R}^N$ from post-log measurement $y \in \mathbb{R}^m$ [9], [10], we apply the learned CAO (P1) as a regularizer, and solve the following block multi-nonconvex problem:

$$\underset{x \geq 0}{\operatorname{argmin}} \min_{\{z_k \in \mathbb{R}^N\}} \frac{1}{2} \|y - Ax\|_W^2 + \gamma \left(\sum_{k=1}^K \frac{1}{2} \|d_k^\star \circledast x - z_k\|_2^2 + \alpha' \sum_{j=1}^N \psi_j \phi((z_k)_j) \right). \quad \text{(P2)}$$

Here, $A \in \mathbb{R}^{m \times N}$ is a CT system matrix, $W \in \mathbb{R}^{m \times m}$ is a (diagonal) weighting matrix with elements $\{W_{l,l} = \rho_l^2/(\rho_l + \sigma^2) : l = 1, \dots, m\}$ based on a Poisson-Gaussian model for the pre-log measurements $\rho \in \mathbb{R}^m$ with electronic readout noise variance σ^2 [9]–[12], $\psi \in \mathbb{R}^J$ is a pre-tuned spatial strength regularization vector [13] with non-negative elements[1] $\{\psi_j = (\sum_{l=1}^m A_{l,j}^2 W_{l,l})^{1/2} / (\sum_{l=1}^m A_{l,j}^2)^{1/2} : j = 1, \dots, N\}$ that promotes uniform resolution or noise properties in the reconstructed image [10, Appx.], a function $\phi(a)$ is equal to 0 if $a = 0$, and is 1 otherwise, $z_k \in \mathbb{R}^N$ is unknown sparse code for the kth filter, and $\gamma > 0$ is a regularization parameter. We now solve (P2) via BPG-M in Section III with a two-block scheme.

A. Image Update

Given the current estimates of the sparse coefficients $\{z_k : k = 1, \dots, K\}$, the proximal mapping of the image update in (P2) is given by the following convex quadratic problem:

$$x^{(i+1)} = \underset{x \geq 0}{\operatorname{argmin}} \frac{1}{2} \left\| x - \eta^{(i+1)} \right\|_{\widetilde{M}_A}^2 + \frac{\gamma}{2} \sum_{k=1}^K \|d_k^\star \circledast x - z_k\|_2^2, \quad (7)$$

[1]The 2D CT geometry used here computes the area of overlap between each pixel and the strip between the X-ray source and the finite-width detector element, using the "`Gtomo2_dscmex.m`" routine in the Michigan Image Reconstruction Toolbox (MIRT) [14]. This particular on-the-fly projector also can compute the squared elements $\{A_{l,j}^2 : \forall l, j\}$ on the fly. For general CT geometries, $A_{l,j}^2$ may be impractical to compute, so one can replace the design in [13] with the newer design $\{\psi_j = \sum_{l=1}^m A_{l,j} W_{l,l} / \sum_{l=1}^m A_{l,j} : j = 1, \dots, N\}$ in [15].

TABLE I
RMSE (HU) of different CT reconstruction models (123 views and $\rho_0 = 10^5$)

FBP	EP	Proposed model (P2), $R = K = 25$	Proposed model (P2), $R = K = 49$
82.8	40.8	35.2	**34.7**

where

$$\eta^{(i+1)} = \acute{x}^{(i+1)} - \widetilde{M}_A^{-1} A^T W \left(A\acute{x}^{(i+1)} - y \right),$$
$$\acute{x}^{(i+1)} = x^{(i)} + E_A^{(i+1)} \left(x^{(i)} - x^{(i-1)} \right),$$

$\widetilde{M}_A = \lambda_A M_A$ by (4), and $M_A \succeq A^T W A$ is a diagonal majorization matrix in [1]. Because the trained filters $\{d_k^\star\}$ satisfy the TF condition (see [1]), we obtain an efficient closed form solution to (7) as follows:

$$x^{(i+1)} = \left[\left(\widetilde{M}_A + \gamma I_R \right)^{-1} \left(\widetilde{M}_A \eta^{(i+1)} + \gamma \sum_{k=1}^K (P d_k^\star) \circledast z_k \right) \right]_{\geq 0},$$

where $P \in \mathbb{C}^{R \times R}$ flips a column vector in the vertical direction (e.g., it rotates 2D filters by 180°).

B. Sparse Code Update

Given the current estimate of x, the sparse coding in (P2) is given by the following separable nonconvex problem:

$$\{z_k^{(i+1)}\} = \underset{\{z_k\}}{\operatorname{argmin}} \sum_{k=1}^K \frac{1}{2} \|d_k^\star \circledast x - z_k\|_2^2 + \alpha' \sum_{j=1}^N \psi_j \phi((z_k)_j). \quad (8)$$

Since we have a sharp majorizer, we solve (8) without extrapolation to reduce memory use (similar to CAOL [1]). We obtain its optimal solution by hard thresholding with a voxel-dependent threshold:

$$z_k^{(i+1)} = \mathcal{T}\left(d_k \circledast x, \sqrt{2\alpha'\psi} \right), \quad k = 1, \dots, K, \quad (9)$$

where the hard thresholding operator $\mathcal{T}(v, a) : \mathbb{R}^N \to \mathbb{R}^N$ is defined by

$$\mathcal{T}(v, a)_j := \begin{cases} 0, & |v_j| < a_j, \\ v_j, & |v_j| \geq a_j. \end{cases}$$

Remark 5 Note that when the learned CAOL (P1) is applied to MBIR, it works as an autoencoding CNN. To see this, substitute solution (9) to the proximal mapping problem in (7). Because the trained filters in (P2) satisfy the TF condition, i.e., $\sum_{k=1}^K \|d_k \circledast x\|_2^2 = \|x\|_2^2$ for some boundary conditions [1], we rewrite $\sum_{k=1}^K \|d_k^\star \circledast x - z_k\|_2^2$ as $\|x - \sum_{k=1}^K (P d_k^\star) \circledast z_k\|_2^2$. Substituting solution (9) into the proximal mapping (7) leads to the following non-zero mean vector for a Gaussian prior at the ith proximal mapping problem (7):

$$\sum_{k=1}^K (P d_k^\star) \circledast \mathcal{T}\left(d_k^\star \circledast x^{(i)}, \sqrt{2\alpha'\psi} \right).$$

This connection is an explicit mathematical motivation for constructing architectures of recurrent regression CNNs for MBIR, e.g., BCD-Net [4], [5] and Momentum-Net [6], [7].

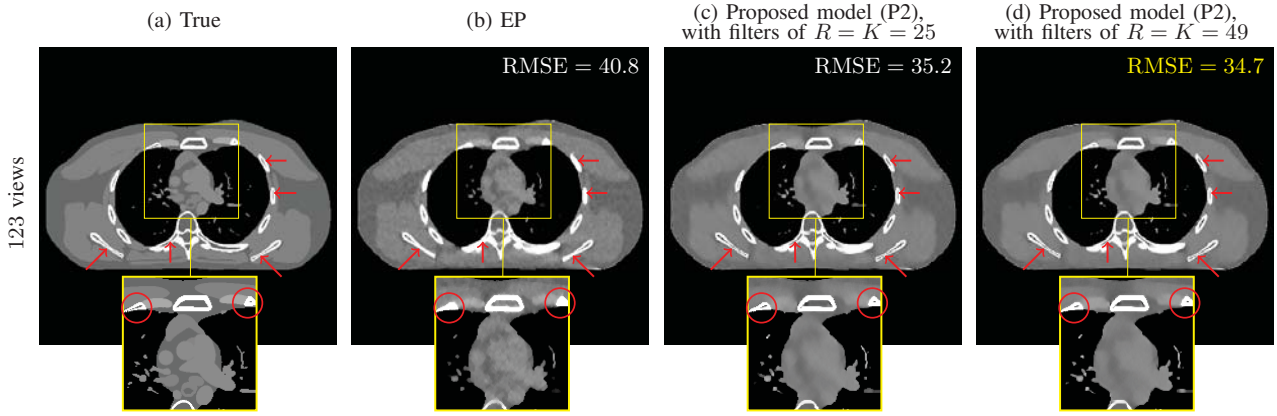


Fig. 1. Comparisons of reconstructed images from different MBIR methods in sparse-view CT (123 views (12.5% sampling) and $\rho_0 = 10^5$; for the proposed MBIR model (P2), TF filters were trained by CAOL (P1) – see Fig. 2; and display window is within $[800, 1200]$ HU). The proposed MBIR model (P2) using the autoencoder trained via CAOL (P1) shows higher image reconstruction accuracy compared to the EP reconstruction; see red arrows and circles. The larger-dimensional autoencoder improves reconstruction accuracy of the MBIR model (P2); compare the results in (c) and (d).

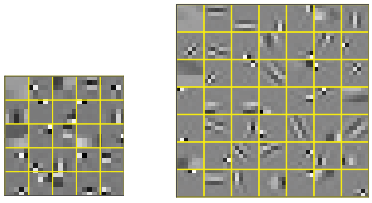
V. Results and Discussion

A. Experimental Setup

1) CAOL: We learned the CAOL model (P1) from the CT dataset of with $L = 10$ and $N = 512\times512$ from down-sampled 512×512 XCAT phantom slices [16]. We trained two sets of filters with parameter dimensions $\{R = 5\times5, K = 25\}$ and $\{R = 7\times7, K = 49\}$ to see their effectiveness in solving the proposed sparse-view CT MBIR model (P2). The parameters for the BPG-M algorithms were defined as follows.[2] We set the regularization parameter α as follows. We first trained filters of the parameter dimension $\{R = 5\times5, K = 25\}$, with some α's within $5\times[10^{-5}, 10^{-4}]$. We selected the filters learned with $\alpha = 2\times10^{-4}$, by considering their structural similarity to the transforms learned by the patch-based method in [12]. For training filters of the dimension $\{R = 7\times7, K = 49\}$, we set $\alpha = 10^{-4}$ by considering the $\frac{1}{R}$-scale of the filter constraint in (P1).

We applied the (scaled) exact Hessian majorizers for the D-update and (scaled) exact identity majorizer for the $\{z_{l,k}\}$-update in (P1); see the corresponding majorizaiton matrix designs in in [1]. We terminated the iterations if the relative error stopping criterion (e.g., [3, (44)]) is met before reaching the maximum number of iterations. We set the tolerance value as 10^{-13} and the maximum number of iterations to 2×10^4. See other training details in [3].

2) Sparse-View CT MBIR with Trained Filters via CAOL: To avoid an inverse crime, our imaging simulation used a 2D 840×840 slice (air cropped, $\Delta_x = \Delta_y = 0.4883$ mm) of the XCAT phantom that differs from the training slices. We simulated sparse-view sinograms of size 888×123 ('detectors or rays' $\times$ 'regularly spaced projection views or angles', where 984 is the number of full views) with GE LightSpeed fan-beam geometry corresponding to a monoenergetic source with $\rho_0 = 10^5$ incident photons per ray and no background events, and electronic noise variance $\sigma^2 = 5^2$. We reconstructed a 420×420 image with a coarser grid, where $\Delta_x = \Delta_y = 0.9766$ mm.



(a) $R = K = 25$ (b) $R = K = 29$

Fig. 2. Examples of trained filters of different parameter dimensions (filters in (a) and (b) were trained with $\alpha = 2\times10^{-4}$ and $\alpha = 10^{-4}$, respectively.)

For filtered back-projection (FBP) methods, we used a conventional filtered back-projection method using a Hanning window. For EP reconstructions, we combined the penalized weighted-least squares term (PWLS) in (P2) and an EP regularizer $\sum_{j=1}^{J}\sum_{j'\in\mathcal{J}_j}\iota_j\iota_{j'}\varphi(x_j - x_{j'})$, where $\mathcal{J}_j$ is the set of indices of the neighborhood, ι_j and $\iota_{j'}$ are parameters that encourage uniform noise [15], and $\varphi(a) := \delta^2(|a/\delta| - \log(1 + |a/\delta|))$ ($\delta = 10$ in Hounsfield units[3], HU). We initialized the EP method with FBP, and ran the relaxed linearized augmented Lagrangian method with 12 ordered-subsets [17]. We finely tuned the regularization parameter (e.g., γ in (P2)) to achieve both good RMSE and SSIM values, and chose it as $2^{15.5}$.

For the proposed MBIR model (P2) using the filters trained by the experiment in Section V-A1, the parameters were defined as follows. For the trained filters of the dimension $R = K = 25$, we chose the reconstruction model parameters as $\{\gamma = 13\times10^6, \alpha' = 2\times10^{-10}\}$; the set shows a good tradeoff between the PWLS term and the learned regularize (P1). For the trained filters of the dimension $R = K = 49$, we chose the model parameters as $\{\gamma = 13\times10^6, \alpha' = 10^{-10}\}$, by considering the $\frac{1}{R}$-scale of the filter constraint in (P1). For the BPG-M algorithm, we set $\lambda_A = 1 + \epsilon$ as defaults, and terminated its iterations as similar in Section V-A1 – we set

[2]The remaining BPG-M parameters not described here are identical to those in [3, VII-A2].

[3]Modified Hounsfield units, where air is 0 HU and water is 1000 HU.

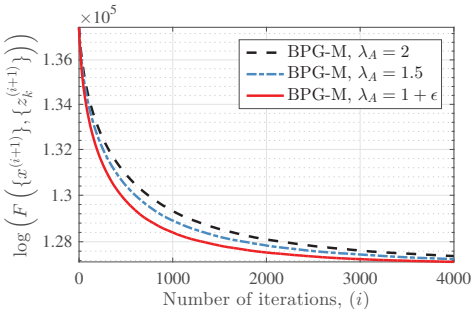


Fig. 3. Cost minimization comparisons in BPG-M-based MBIR (P2) with different λ_A values (the filters of $R = K = 25$ were learned via CAOL (P1) with the CT dataset). Under the sharp majorization regime, maintaining sharp majorization (i.e., $\lambda_A = 1 + \epsilon$) provides faster convergence than giving more weight on extrapolation (i.e., $\lambda_A = 2$).

the tolerance value and the maximum number of iterations to 10^{-6} and 4×10^3, respectively.

We evaluated the reconstruction quality by the root mean square error (RMSE; in HU) in a region of interest (ROI), where the ROI was a circular region (around center) including all the phantom tissues. The RMSE is defined by $\mathrm{RMSE}(x^\star, x^{\text{true}}) := (\sum_{j=1}^{N_{\text{ROI}}} (x_j^\star - x_j^{\text{true}})^2 / N_{\text{ROI}})^{1/2}$, where $x^\star$ is the reconstructed image, x^{true} is the ground truth image, and N_{ROI} is the number of pixels in a ROI.

B. Sparse-View CT MBIR with Trained Autoencoding CNN and BPG-M

In sparse-view CT using only 12.5% of full projections views, the proposed MBIR model (P2) using the trained autoencoder via CAOL (P1) outperforms EP reconstruction in terms of RMSE; it reduces RMSE by approximately 5.5HU. (The RMSE reduction is comparable to that of the patch-based method in [10] within a similar imaging setup.) See Table I and Fig. 1. The model (P2) can better recover high-contrast regions (e.g., bones) – see red arrows and circles in Fig. 1; however, as similarly observed in [10], [12], edges in low-contrast regions (e.g., soft tissues) become blurry while removing noise via (9). Fig. 2 shows that larger-dimensional filters capture more interesting or diverse features of training images; in particular, the trained filters of the dimension $R = K = 49$ capture some diagonal structures. This implies that diverse features in filters often captured in larger-dimensional filters can be useful in improving MBIR performances.

On the algorithmic side, the BPG-M framework guarantees the convergence of MBIR (P2). Under the sharp majorization regime in BPG-M, maintaining the majorization sharpness is more critical than having stronger extrapolation effects, as similarly shown in CAOL experiments (see Remark 4). Figs. 3 support these assertions.

VI. Conclusion

We reviewed the CAOL framework that trains autoencoding CNNs in an unsupervised fashion, and BPG-M that rapidly and stably solves block multi-(non)convex problems via sharp majorizers and the momentum. In addition, CAOL combined with BPG-M provides strong mathematical motivations in constructing recurrent regression CNNs, namely MomentumNet [6], [7], for fast and convergent MBIR.

For sparse-view CT, the proposed MBIR model using a (single-layered) autoencoder trained via CAOL, significantly improves reconstruction accuracy compared to EP-based MBIR. In particular, larger-dimensional autoencoders capture more diverse features in convolutional kernels, and further improve the performance of the proposed MBIR model. Finally, by using BPG-M with emphasis on sharp majorizers, we can rapidly and stably solve the proposed MBIR problem. Future work will explore the effectiveness of multi-layered CAOL in [1] to MBIR problems in extreme imaging.

References

[1] I. Y. Chun and J. A. Fessler, "Convolutional analysis operator learning: Acceleration and convergence," submitted, Jan. 2018.

[2] I. Y. Chun, D. Hong, B. Adcock, and J. A. Fessler, "Convolutional analysis operator learning: Dependence on training data and compressed sensing recovery guarantees," preprint, Aug. 2018.

[3] I. Y. Chun and J. A. Fessler, "Convolutional dictionary learning: Acceleration and convergence," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 1697–1712, Apr. 2018.

[4] ——, "Deep BCD-net using identical encoding-decoding CNN structures for iterative image recovery," in *Proc. IEEE IVMSP Workshop*, Zagori, Greece, Jun. 2018.

[5] Y. K. D. Hongki Lim, Jeffrey A. Fessler and I. Y. Chun, "Application of trained Deep BCD-Net to iterative low-count PET image reconstruction," in *Proc. IEEE NSS-MIC* (to appear), Sydney, Australia, Nov. 2018.

[6] I. Y. Chun, H. Lim, Z. Huang, and J. A. Fessler, "Fast and convergent iterative signal recovery using trained convolutional neural networkss," in *Proc. Allerton Conf. on Commun., Control, and Comput.* (to appear), Allerton, IL, Oct. 2018.

[7] ——, "Momentum-Net: Fast and convergent recurrent neural network for inverse problems," preprint, Nov. 2018.

[8] Y. Xu and W. Yin, "A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion," *SIAM J. Imaging Sci.*, vol. 6, no. 3, pp. 1758–1789, Sep. 2013.

[9] I. Y. Chun and T. Talavage, "Efficient compressed sensing statistical X-ray/CT reconstruction from fewer measurements," in *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, Lake Tahoe, CA, Jun. 2013, pp. 30–33.

[10] I. Y. Chun, X. Zheng, Y. Long, and J. A. Fessler, "Sparse-view X-ray CT reconstruction using ℓ_1 regularization with learned sparsifying transform," in *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, Xi'an, China, Jun. 2017, pp. 115–119.

[11] J. B. Thibault, C. A. Bouman, K. D. Sauer, and J. Hsieh, "A recursive filter for noise reduction in statistical iterative tomographic imaging," in *Proc. SPIE 6065, Computational Imaging IV*, vol. 6065, Feb. 2006, p. 60650X.

[12] X. Zheng, I. Y. Chun, Z. Li, Y. Long, and J. A. Fessler, "Sparse-view X-ray CT reconstruction using ℓ_1 prior with learned transform," submitted, Oct. 2017.

[13] 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. 1346–58, Sep. 1996.

[14] J. A. Fessler, "Michigan image reconstruction toolbox (MIRT) for Matlab," 2016, available from http://web.eecs.umich.edu/~fessler.

[15] J. H. Cho and J. A. Fessler, "Regularization designs for uniform spatial resolution and noise properties in statistical image reconstruction for 3-D X-ray CT," *IEEE Trans. Med. Imag.*, vol. 34, no. 2, pp. 678–689, Feb. 2015.

[16] W. P. Segars, M. Mahesh, T. J. Beck, E. C. Frey, and B. M. Tsui, "Realistic CT simulation using the 4D XCAT phantom," *Med. Phys.*, vol. 35, no. 8, pp. 3800–3808, Jul. 2008.

[17] H. Nien and J. A. Fessler, "Relaxed linearized algorithms for faster X-ray CT image reconstruction," *IEEE Trans. Med. Imag.*, vol. 35, no. 4, pp. 1090–1098, Apr. 2016.