ECE558 Project: CNN-Based Projected Gradient Descent for Consistent X-ray CT Scatter Correction

Berk Iskender
* Electrical and Computer Engineering, University of Illinois Urbana-Champaign

***Abstract*—In the reviewed paper, a method that utilizes a deep convolutional neural network (DCNN) as the projection operator in projected gradient descent (PGD) method is described. The method is important in the sense that it provides a feedback mechanism for the image reconstruction task which is usually non-existent for DCNN based regression approaches. The proposed approach uses a recursive approach to enforce consistency using gradient descent, and uses the DCNN to project the intermediate solution to the feasible space of solutions. Moreover, a specific method is proposed for training the projector DCNN. The performance of the method is tested on sparse-view CT reconstruction. In the second part of the report, X-ray CT scatter correction problem is described, and the proposed method is adapted to the respective problem and several experiments are conducted.**

I. Introduction

As we have seen throughout the ECE558 lectures, it is possible to improve the reconstructed signal quality by incorporating prior information about the signal itself. In many areas of the medical imaging, it is especially true and important since it also provides a way to reduce harmful effects (e.g. high exposure to X-rays etc.) of the imaging modality on patients.

A. Problem Statement

In many areas such as CT and MRI, the measurement operator is linear and the process can be formulated as

$$y = Hx + n \tag{1}$$

where $y \in \mathbb{R}^M$ is the vectorized measured signal, $x \in \mathbb{R}^N$ is the space-domain image that needs to be reconstructed, $H \in \mathbb{R}^{M \times N}$ is the linear measurement (imaging) operator and $n \in \mathbb{R}^M$ is the intrinsic noise or some degradation such as scatter in X-ray CT. One important point is that usually when we would like to reduce the harmful effects or reduce the scan time, n gets more significant and it is necessary to perform some kind of artifact reduction to address degradations due to it. Another problem is when the number of measurements is less than number of unknowns, $M << N$, there exists the issue of non-uniqueness of the solution to the inverse problem, with infinitely many feasible solutions y. To bring y on a feasible space, different algorithms have been proposed.

B. Background and motivation for the proposed method

- *Classical Algorithms:* These methods apply the pseudo-inverse of the measurement operator, $H^\dagger$, or an approximation of it to the measurement, which provides the least-squares solution. In CT, this amounts to filtered backprojection

$$x^* = H^T F y \tag{2}$$

where $F \in \mathbb{R}^{M \times M}$ is a linear filter, and H^T is the backprojection operator. We covered these topics during the lectures and homeworks of ECE558 to a great extent. As we verified in our homeworks, these methods provide really good results when the number measurements is sufficiently large (i.e. the signal is sampled sufficiently in the Fourier domain since Fourier transform of the measurements are central slices of the Fourier transform of the image). However, when the number of measurements are significantly small $M << N$, or there are significant artifacts due to n, they fail.
- *Iterative Algorithms:* These methods introduce a regularization term to the optimization problem and by this way try to constrain the solution space,

$$\mathbf{x}^* = \arg\min_{\mathbf{x} \in \mathbb{R}^N} (E(\mathbf{H}\mathbf{x}, \mathbf{y}) + \lambda R(\mathbf{x})), \tag{3}$$

where $E : \mathbb{R}^M \times \mathbb{R}^M \rightarrow \mathbb{R}^+$ is a data-fidelity term (usually the Frobenius norm operator) verifying the consistency of the solution, and $R : \mathbb{R}^N \rightarrow \mathbb{R}^+$ is the regularization operator to incorporate the prior knowledge to restrict the solution space, with a constant λ controlling the trade-off between these two sources of information. (3) can be reformulated as

$$\mathbf{x}^* = \arg\min_{\mathbf{x} \in \mathcal{S}_R} E(\mathbf{H}\mathbf{x}, \mathbf{y}) \tag{4}$$

where $\mathcal{S}_R = \left\{\mathbf{x} \in \mathbb{R}^N : R(\mathbf{x}) \leq \tau\right\}$ denotes the set of feasible solutions, constrained by the constant $\tau(\lambda)$. Common choices for R are total variation, $||\nabla x||$, ℓ_1-norm of the basis expansion coefficients c of the image in some transform domain, $c = Wx$, where W can be wavelets or DCT, as we have studied in the lectures. These methods restrict the type of prior knowledge that can be used.
- *Learning Algorithms:* These methods either attempt to learn a non-linear inverse operator directly working on measurements $x^* = N(y)$, or working on an initial reconstruction of the object $x^* = N(\tilde{x})$, where $\tilde{x} = Ay$. By adding nonlinearity to the operator, it becomes possible to address the non-linear degradations, n, which

sometimes can be a function of the object x as scatter. These methods require the training of N using the pairs $\{x_i, y_i\}$, $i \in \{1, \dots, K-1\}$ which include the ground-truth images x_i. The main-drawback of these methods are
  - having no worst-case performance guarantee,
  - lacking an iterative feedback-mechanism to preserve data-fidelity.
- *Iterative Learning Algorithms:* Several methods that can be included in this group try to unroll an iterative scheme k steps, and then learn it as an DCNN [1], [2]. Their drawback is the need to learn the parameters of the network dependent on the iteration scheme. Another group called plug-and-play (PnP) [3] variations utilizes the ADMM scheme, but replaces the proximal operator from soft-thresholding to different denoisers. These methods still lack the theoretical guarantees and interpretation. Also, generative models are incorporated for projecting the solution onto a feasible manifold [4]. Finally, as authors of this paper indicated there is a very similar method to the one that is proposed in this paper but it lacks theoretical justification and guarantees.

II. Proposed Method: RPGD

A. Major objectives

The major objectives of the proposed method can be summarized as in three main topics:

- learning a DCNN that acts a projection operator onto a set $\mathcal{S}$ which preferably represents the manifold of feasible (artifact-free) solutions,
- having a good initialization point,
- performing an iterative optimization where in each step data consistency is preserved by performing gradient descent on the data-fidelity term E and projecting onto set $\mathcal{S}$ to provide convergence guarantees under certain conditions.
- having the network (the projector) parameters independent of the iterative scheme.

As described in the previous section, the main objective is essentially to provide convergence guarantees and provide results that are consistent with the forward operator which state-of-the-art learning-based methods lack.

B. Theoretical Framework

In this paper, properties of a generalized PGD (which may not be orthogonal) onto a nonconvex set are shown. Important notation for the results can be listed as follows. $\mathcal{B}_\epsilon$ denotes the ℓ_2-ball centered at x with radius $\epsilon > 0$.

$$\mathcal{B}_\varepsilon(\mathbf{x}) = \left\{\mathbf{z} \in \mathbb{R}^N : \|\mathbf{z} - \mathbf{x}\|_2 \leq \varepsilon\right\} \tag{5}$$

The operator $T : \mathbb{R}^N \to \mathbb{R}^N$ is Lipschitz-continuous with constant L if

$$\|T(\mathbf{x}) - T(\mathbf{z})\|_2 \leq L\|\mathbf{x} - \mathbf{z}\|_2, \quad \forall \mathbf{x}, \mathbf{z} \in \mathbb{R}^N \tag{6}$$

If $L < 1$, then T is contractive, and non-expansive if $L = 1$.

Any projector onto set $\mathcal{S}$, $P_\mathcal{S}$ should be idempotent, $P_\mathcal{S} P_\mathcal{S} = P_\mathcal{S}$. The projector is orthogonal if $P_\mathcal{S}(\mathbf{x}) = \inf_{\mathbf{z}\in\mathcal{S}} \|\mathbf{x} - \mathbf{z}\|_2$, $\forall x \in \mathbb{R}^N$.

C. Constrained Least Squares

We can solve the following least-squares problem to obtain consistent reconstructions while imposing the prior knowledge that the solution must reside in the set $\mathcal{S}$,

$$\min_{\mathbf{x}\in\mathcal{S}} \frac{1}{2}\|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2. \tag{7}$$

The constraint acts as a regularizer in this formulation. If there is a one-to-one relationship between measurements y and images x, then this formulation picks a unique estimation $x^* \in \mathcal{S}$, solving the non-uniqueness issue and making the problem well-posed. When the measurements are perturbed, this formulation returns the LS solution in the set $\mathcal{S}$. Finally the estimate x^* can be called a local minimizer of (7) if it also resides on the ℓ_2-ball,

$$\exists \varepsilon > 0 : \|\mathbf{H}\mathbf{x}^* - \mathbf{y}\|_2 \leq \|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2, \quad \forall \mathbf{x} \in \mathcal{S} \cap \mathcal{B}_\varepsilon(\mathbf{x}^*). \tag{8}$$

D. Projected Gradient Descent

When $\mathcal{S}$ is a closed convex set, solution to (7) can be obtained by projected gradient descent,

$$\mathbf{x}_{k+1} = P_\mathcal{S}\left(\mathbf{x}_k - \gamma\mathbf{H}^\mathrm{T}\mathbf{H}\mathbf{x}_k + \gamma\mathbf{H}^\mathrm{T}\mathbf{y}\right), \tag{9}$$

where γ is the step size of the gradient descent and chosen such that $\gamma < 2/||H^T H||_2$. However, in the problems of interest, the set $\mathcal{S}$ consists of the objects that are usually medical phantoms and possibly not convex. The authors of the paper propose to use with some "possibly not orthogonal" projector $P_\mathcal{S}$, and provide conditions for when the method achieves local minimum of (7). Any local minimizer is a fixed point of the PGD and it must be reached as iterations proceed for the algorithm to be convergent:

$$\begin{aligned} \mathbf{x} = G_\gamma(\mathbf{x}) &= P_\mathcal{S}\left(\mathbf{x} - \gamma\mathbf{H}^\mathrm{T}\mathbf{H}\mathbf{x} + \gamma\mathbf{H}^\mathrm{T}\mathbf{y}\right) \\ \mathbf{x_{k+1}} &= G_\gamma(\mathbf{x_k}), \quad k \to \infty. \end{aligned} \tag{10}$$

A sufficient condition for each fixed point of G_γ to be a local minimizer of (7):

Proposition 1: Let $\gamma > 0$ and $P_\mathcal{S}$ be such that for all $\mathbf{x} \in \mathbb{R}^N$,

$$\langle \mathbf{z} - P_\mathcal{S}\mathbf{x}, \mathbf{x} - P_\mathcal{S}\mathbf{x}\rangle \leq 0, \quad \forall \mathbf{z} \in \mathcal{S} \cap \mathcal{B}_\varepsilon(P_\mathcal{S}\mathbf{x}) \tag{11}$$

for some $\varepsilon > 0$. Then, any fixed point of the operator G_γ in (10) is a local minimizer of (7). If (11) is satisfied globally,

$$\langle \mathbf{z} - P_\mathcal{S}\mathbf{x}, \mathbf{x} - P_\mathcal{S}\mathbf{x}\rangle \leq 0, \quad \forall \mathbf{x} \in \mathbb{R}^N, \mathbf{z} \in \mathcal{S} \tag{12}$$

then any fixed point of G_γ is a solution of (7).

It needs to be noted that (12) implies convexity of $\mathcal{S}$ and (11) is a relaxed and easier to achieve version of it.

Proposition 2: If $\mathcal{S}$ is a union of a finite number of closed convex sets in $\mathbb{R}^N$, the orthogonal projector $P_\mathcal{S}$ onto $\mathcal{S}$ satisfies (11).

These propositions imply that when S is not convex, we can only hope to find a fixed point as a local minimizer. The following theorem states the sufficient condition for PGD to converge to a unique fixed point of G_γ,.

Theorem 1: Let λ_{max} and λ_{min} be the largest and smallest eigenvalues of $H^T H$, respectively. If $P_\mathcal{S}$ satisfies (11) and is Lipschitz-continuous with $L < (\lambda_{max} + \lambda_{min})/(\lambda_{max} - \lambda_{min})$, then, for $\gamma = 2/(\lambda_{max} + \lambda_{min})$, the sequence $\{x_k\}$ generated by PGD converges to a local minimizer of (7), independent of the initialization, x_0.

Since for any two points $s_1,\ s_2 \in \mathcal{S}$ we have $||s_1 - s_2|| = ||P_\mathcal{S}(s_1) - P_\mathcal{S}(s_2)||$, the projector is not contractive and the iterations do not converge to zero. This is also verified by $L < 1$ for $\lambda_{min} = 0$ (H has a non-trivial nullspace) being not feasible. Still, the combined operator G_γ may not have a fixed point. It needs to be assumed that G_γ has a set of fixed points. With this assumption, the authors show that one of those fixed-points need to be achieved by iterating the averaged PGD (APGD)

$$\alpha \mathrm{Id} + (1-\alpha) G_\gamma \tag{13}$$

where $\alpha \in (0,1)$ and Id is the identity operator.

Theorem 2: Let λ_{max} be the largest eigenvalue of $H^T H$. If $P_\mathcal{S}$ satisfies (11) and is a non-expansive operator such that G_γ in (10) has a fixed point for some $\gamma < 2/\lambda_{max}$, then the sequence $\{x_k\}$ generated by APGD, with

$$\mathbf{x}_{k+1} = (1-\alpha)\mathbf{x}_k + \alpha G_\gamma(\mathbf{x}_k) \tag{14}$$

for $\alpha \in (0,1)$, converges to a local minimizer of (7), regardless of the initialization.

E. Relaxation with Guaranteed Convergence

Theorems 1 and 2 are not helpful when we employ a DCNN as a $P_\mathcal{S}$ because Lipschitz continuity cannot be enforced. In this case, the authors propose a third version of the PGD as the relaxed PGD (RPGD) where $P_\mathcal{S}$ is replaced by a general operator F. Also, two sequences $\{c_k\}$ to control the convergence rate and α_k as the relaxation parameters are introduced. The algorithm can be seen in Fig. 1.

Despite the relaxation step, the algorithm still converges under certain assumptions as stated in Theorem 3.

Theorem 3: Let the input sequence $\{c_k\}$ of RPGD algorithm be asymptotically upper-bounded by $C < 1$. Then, the following holds for reconstructed images $\{x_k\}$;

1) $x_k \to x^*$ as $k \to \infty$, for all choices of F;
2) if F is continuous and the relaxation parameters $\{\alpha_k\}$ are lower-bounded by $\varepsilon > 0$, then x^* is a fixed point of
$$G_\gamma(\mathbf{x}) = F\left(\mathbf{x} - \gamma \mathbf{H}^{\mathrm{T}}\mathbf{H}\mathbf{x} + \gamma \mathbf{H}^{\mathrm{T}}\mathbf{y}\right); \tag{15}$$
3) if, in addition to (ii), F is indeed a projector onto S that satisfies (11), then x^* is a local minimizer of (7).

While finding a local minima of (7) like PGD or APGD given that (ii) and (iii) are satisfied in Theorem 3, RPGD does not require F to be strictly a projection operator to guarantee convergence. The detailed proof of convergence for the proposed algorithm can be found in [5]. However, an open question is how meaningful is the fixed point that RPGD converges when (iii) in Theorem 3 is not satisfied since convergence does not necessarily means that the result is desirable.

Algorithm 1 Relaxed Projected Gradient Descent (RPGD)

Input: $\mathbf{H}$, $\mathbf{y}$, $\mathbf{A}$, nonlinear operator F, step size $\gamma > 0$, positive sequence $\{c_n\}_{n\geq 1}$, $\mathbf{x}_0 = \mathbf{A}\mathbf{y} \in \mathbb{R}^N$, $\alpha_0 \in (0,1]$.
Output: reconstructions $\{\mathbf{x}_k\}$, relaxation parameters $\{\alpha_k\}$.

$k \leftarrow 0$
while not converged **do**
  $\mathbf{z}_k = F(\mathbf{x}_k - \gamma\,\mathbf{H}^{\mathrm{T}}\mathbf{H}\mathbf{x}_k + \gamma\,\mathbf{H}^{\mathrm{T}}\mathbf{y})$
  if $k \geq 1$ **then**
    if $\|\mathbf{z}_k - \mathbf{x}_k\|_2 > c_k\,\|\mathbf{z}_{k-1} - \mathbf{x}_{k-1}\|_2$ **then**
      $\alpha_k = c_k\|\mathbf{z}_{k-1} - \mathbf{x}_{k-1}\|_2 / \|\mathbf{z}_k - \mathbf{x}_k\|_2\,\alpha_{k-1}$
    else
      $\alpha_k = \alpha_{k-1}$
    end if
  end if
  $\mathbf{x}_{k+1} = (1-\alpha_k)\mathbf{x}_k + \alpha_k \mathbf{z}_k$
  $k \leftarrow k+1$
end while

Fig. 1: Relaxed Projected Gradient Descent (RPGD) algorithm.

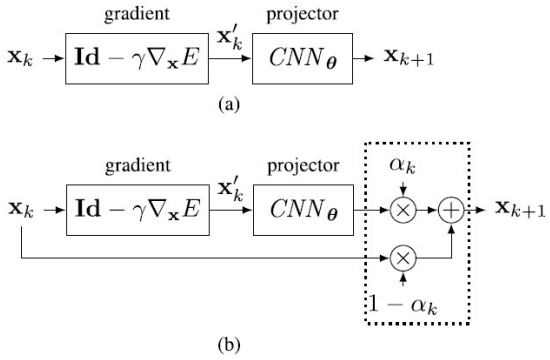


Fig. 2: Block diagram of (a) the PGD with CNN as the projector, and (b) Relaxed Projected Gradient Descent (RPGD).

F. Training a CNN as a Projector

Although it may not be possible to obtain a projector as a CNN, two desirable properties of a projector that we want the CNN to have can be listed as

1) idempotence $P_\mathcal{S}(\mathbf{x}) = \mathbf{x}$, and
2) $\mathbf{x} = P_\mathcal{S}(\tilde{\mathbf{x}})$ where $\tilde{\mathbf{x}}$ is a perturbed version of $\mathbf{x}$.

Thus, given the clean images, $N-1$ perturbed versions of each image is generated and the DCNN is trained by the following loss function for T epochs,

$$J(\boldsymbol{\theta}) = \sum_{n=1}^{N} \underbrace{\sum_{q=1}^{Q} \|\mathbf{x}^q - CNN_\theta\left(\tilde{\mathbf{x}}^{q,n}\right)\|_2^2}_{J_n(\boldsymbol{\theta})}. \tag{16}$$

An important point is that the clean images are also used as pairs ($n = 1$) to enforce idempotence of the model. To, prevent overfitting of a certain type, the type of perturbations are selected as

$$\begin{aligned} \tilde{\mathbf{x}}^{q,1} &= \mathbf{x}^q \\ \tilde{\mathbf{x}}^{q,2} &= \mathbf{AHx}^q \\ \tilde{\mathbf{x}}^{q,3} &= CNN_{\boldsymbol{\theta}_{t-1}}\left(\tilde{\mathbf{x}}^{q,2}\right), \end{aligned} \tag{17}$$

where A is the FBP operator, H is the radon transform. The third option stands as a dynamic linearly perturbed set where θ_t are the CNN parameters after the update of t epochs. This is useful since it provides a way to bound errors on CNN and provides a sense of Lipschitz continuity and greatly increases the training diversity.

III. Scatter in X-ray CT

In X-ray CT, measurements of the attenuation of rays as they pass through the object are used to find the linear attenuation coefficient map $f(x)$, $x \in \mathbb{R}^2$ of the object. The set of line integral projections is obtained from the object as

$$g(t,\theta) = (\mathcal{R}f)(t,\theta), \tag{18}$$

where $\mathcal{R}$ is the 2D Radon transform.

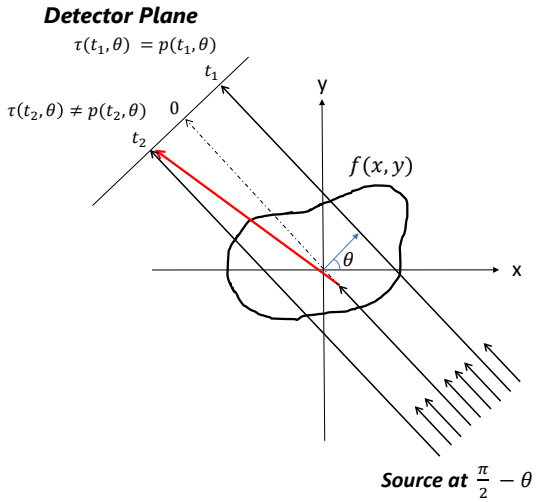


Fig. 3: 2D parallel-beam CT geometry, and x-ray scatter. An x-ray directed toward the detector position $(0,\theta)$ reaches the detector at (t_2,θ) after scattering in the object f, causing $\tau(t_2,\theta) \neq p(t_2,\theta)$.

The inverse problem is to compute the inverse Radon transform $f(x) = (\mathcal{R}^{-1}g)(x)$. Filtered-backprojection (FBP) $\widehat{R^{-1}}$ is one alternative solution for this problem.

Using an energy-integrating detector, the primary measurement at position t and angle θ is obtained as

$$p(t,\theta) = I_0 \int c(E) e^{-g_E(t,\theta)} dE \tag{19}$$

where I_0 is the vacuum (or bright field) fluence measurement, $c(E)$ denotes the fraction of source photon fluence scaled by the energy-dependent detector response at energy E, and where $\int c(E)dE = 1$, and $f_E(\mathbf{x})$ is the linear attenuation coefficient at energy E and at position $\mathbf{x}$.

With a monochromatic source with photon energy E_0, (19) reduces to

$$p(t,\theta) = I_0 e^{-g(t,\theta)}, \tag{20}$$

where the dependence on E_0 is omitted to simplify notation. In this case, the line integral projection is available from the measurement by a logarithm, $g(t,\theta) = -\ln[p(t,\theta)/I_0]$.

Considering the scattering events, the total measurement at angle θ that is obtained is

$$\tau(t,\theta) = p(t,\theta) + s(t,\theta) \tag{21}$$

where $s(t,\theta)$ is an additive scatter term, which is a nonlinear function of the object. This additive contribution leads to artifacts in the analytical FBP reconstruction if the total measurement is used to directly reconstruct the object,

$$\tilde{g} = -\ln[\tau(t,\theta)/I_0] \tag{22}$$

$$\tilde{f}(x) = (\widehat{R^{-1}}\tilde{g})(x). \tag{23}$$

IV. Problem statement for scatter correction

We assume that we are given a set total measurements $\boldsymbol{\tau}\{\boldsymbol{\tau}_\theta, \theta \in \Theta\}$ which, without the scatter, would suffice for accurate reconstruction of the object f by FBP. The main goal is to produce a reconstruction f^* that approximates the FBP reconstruction $\hat{f} = g$ that would be obtained from $g(t,\theta) = -\ln[p(t,\theta)/I_0]$, where p is the set of primary measurements.

V. Proposed method for CNN-based PGD for consistent scatter correction

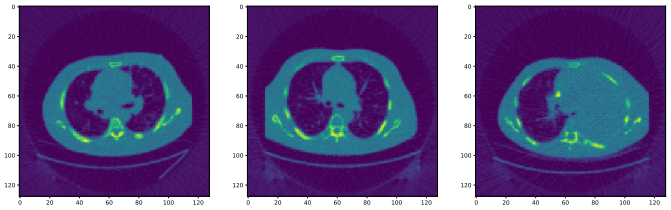
The main objective of the course project is to utilize the novel technique developed in [5] to perform a measurement consistent scatter correction in the reconstruction domain. For this purpose, a couple algorithms were tested. The formulations of sparse-view CT (the problem that is considered in [5]) and scatter correction in CT can be made almost equivalent as follows

$$\tilde{g} = Rf + n_s(f) \tag{24}$$

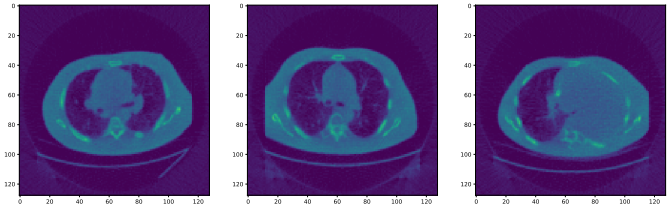
where $y \triangleq g$ is the measurement (scatter-corrupted line integral projections), $H \triangleq R$ is the measurement operator (Radon transform), and $n = n_s(f)$ is the process specific degradation term, related to the effect of scatter on the line integral projections. One important difference is that the degradation is dependent on the object for scatter correction. However, it is assumed to be independent for the scope of this project.

VI. Dataset

For the scatter correction experiments, 2D axial slices obtained from the phantoms of TCIA dataset [6] are used. Total of 30 phantoms were divided into 27 training and 3 test phantoms. These phantoms were mapped to a tissue-mapping pre-processing where the attenuation coefficient values were mapped to 5 different materials: (1) air, (2) lung, (3) adipose, (4) soft tissue and (5) bone. After this step, these phantoms were fed to a GPU-based Monte-Carlo X-ray simulator, namely MC-GPU [7], which simulates the propagation of rays for a given source, phantom, and detector setting for the cone beam CT geometry. Then, these total measurements were reconstructed using ASTRA toolbox using the parameters describing the simulation geometry. MC-GPU is capable of providing total and primary measurements separately, saving the need for computing theoretical projections using Beer's Law. Sample slices reconstructed from total and primary measurements are illustrated in Fig. 4.



(a) Primary measurement reconstructions



(b) Total measurement reconstructions

Fig. 4: Polychromatic CBCT reconstructions of central axial slices. Using (a) primary measurements $p(t,\theta)$ (assumed to be ground-truth reconstructions), (b) total (scatter-corrupted) measurements $\tau(t,\theta)$.

VII. Results

A. Total measurement initial reconstruction and RPGD

In this subsection, the proposed RPGD method was implemented as it is in the original paper. The loss function in (16) was minimized using

$$\begin{aligned}\tilde{\mathbf{x}}^{q,1} &= \mathbf{x}^q \\ \tilde{\mathbf{x}}^{q,2} &= \mathbf{A}\mathbf{H}\mathbf{x}^q \\ \tilde{\mathbf{x}}^{q,3} &= CNN_{\boldsymbol{\theta}_{t-1}}\left(\tilde{\mathbf{x}}^{q,2}\right)\end{aligned} \tag{25}$$

where H is the 2D radon transform, A is the FBP with ramp filtering, and CNN_{θ_t} is a DCNN with the UNET [8] structure with a skip connection between input and output.

As expected, enforcing measurement consistency directly without any preliminary scatter correction on the initial reconstruction $\mathbf{x}_0$ fails to perform scatter correction significantly. The convergence and reconstruction results for a test slice are provided in Figures 12, 13 and 14.

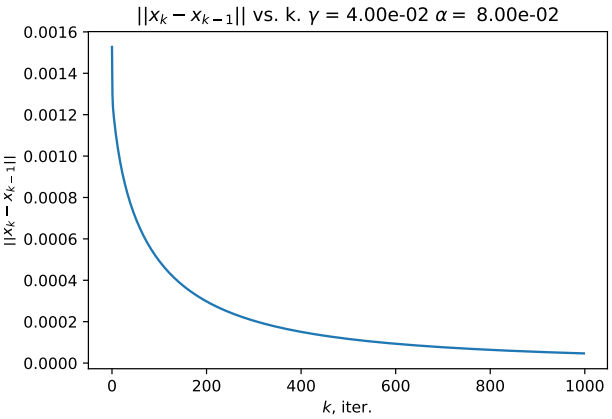


Fig. 5: ℓ_2 norm of the difference between consecutive iterations for total measurement initial reconstruction and RPGD setting for hyperparameters $\alpha_0 = 8.10^{-2}$, $\gamma = 4.10^{-2}$, and $c_n = (0.8)\mathbf{1}$.

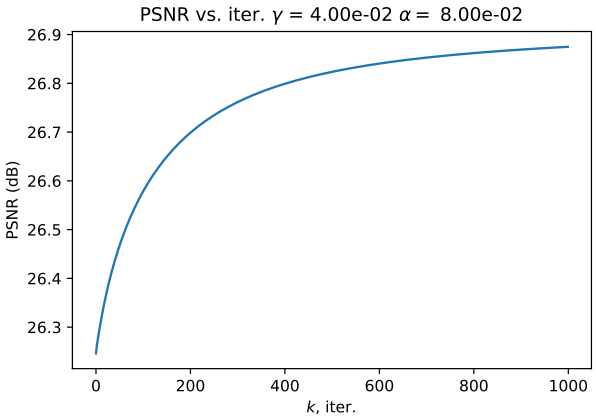


Fig. 6: PSNR values of consecutive iterations for total measurement initial reconstruction and RPGD setting for hyperparameters $\alpha_0 = 8.10^{-2}$, $\gamma = 4.10^{-2}$, and $c_n = (0.8)\mathbf{1}$.

B. Total measurement reconstruction and modified RPGD

In this setting, the proposed RPGD method was modified in the following form,

$$\begin{aligned}\tilde{\mathbf{x}}^{q,1} &= \mathbf{x}^q \\ \tilde{\mathbf{x}}^{q,2} &= \mathbf{A}\mathbf{H}\mathbf{x}^q \\ \tilde{\mathbf{x}}^{q,3} &= CNN_{\boldsymbol{\theta}_{t-1}}\left(\tilde{\mathbf{x}}^{q,2}\right) \\ \tilde{\mathbf{x}}^{q,4} &= \mathbf{A}\mathbf{y}^q\end{aligned} \tag{26}$$

where H is the 2D radon transform, A is the FBP with ramp filtering, and CNN_{θ_t} is a DCNN with the UNET [8] structure with a skip connection between input and output as before. As the fourth batch in the sequential training, scatter corrupted reconstructions were added to the RPGD training.

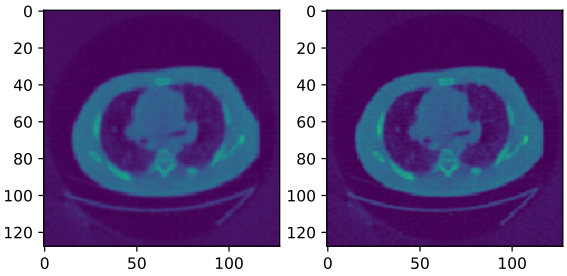


Fig. 7: Scatter-corrupted initial reconstruction (Left) and final output of RPGD iterations (Right) for hyperparameters $\alpha_0 = 8.10^{-2}$, $\gamma = 4.10^{-2}$, and $c_n = (0.8)\mathbf{1}$.

An important difference compared to other experiments, for this setting, whole training dataset was trained at once, rather than a sequential training as proposed in the original method. However, this scheme has significantly failed to provide convergence to an optimal point. Convergence results and reconstruction outputs are provided in Figures 8, 9, 10, and 11.

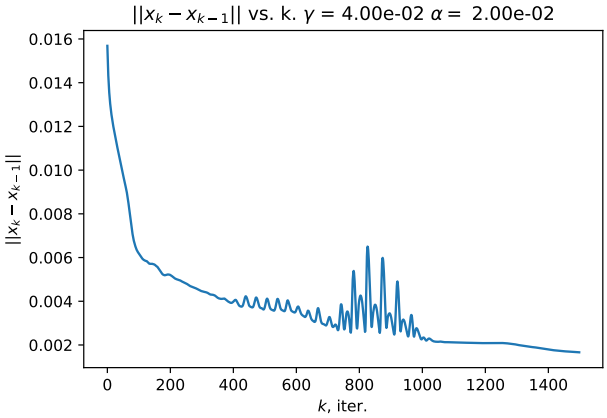


Fig. 8: ℓ_2 norm of the difference between consecutive iterations for total measurement reconstruction and modified RPGD setting for hyperparameters $\alpha_0 = 2.10^{-2}$, $\gamma = 4.10^{-2}$, and $c_n = (0.8)\mathbf{1}$.

C. Scatter-corrected initial reconstruction and RPGD

For these experiments, the proposed RPGD method was implemented as it is in the original paper, however, the initial reconstruction is pre-processed using another DCNN having the UNET architecture. Likewise, the loss function in (16) was minimized using

$$\begin{aligned} \tilde{\mathbf{x}}^{q,1} &= \mathbf{x}^q \\ \tilde{\mathbf{x}}^{q,2} &= CNN_{\xi}^{sc}(\mathbf{A}\mathbf{H}\mathbf{x}^q) \\ \tilde{\mathbf{x}}^{q,3} &= CNN_{\boldsymbol{\theta}_{t-1}}\left(\tilde{\mathbf{x}}^{q,2}\right) \end{aligned} \tag{27}$$

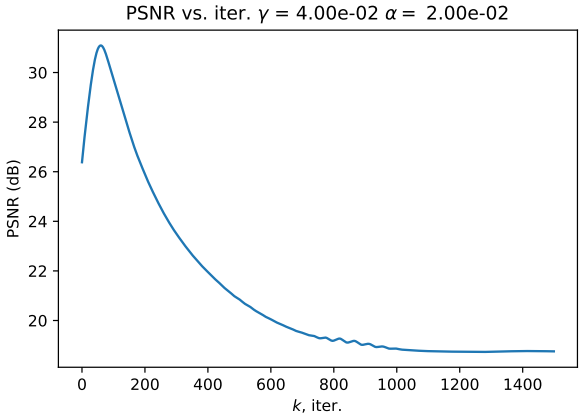


Fig. 9: PSNR values of consecutive iterations for total measurement reconstruction and modified RPGD setting for hyperparameters $\alpha_0 = 2.10^{-2}$, $\gamma = 4.10^{-2}$, and $c_n = (0.8)\mathbf{1}$.

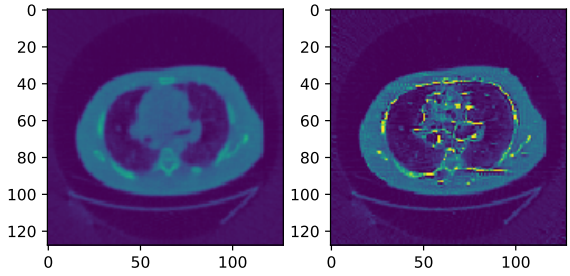


Fig. 10: Scatter-corrupted initial reconstruction (Left) and final output of RPGD iterations (Right) for hyperparameters $\alpha_0 = 2.10^{-2}$, $\gamma = 4.10^{-2}$, and $c_n = (0.8)\mathbf{1}$.

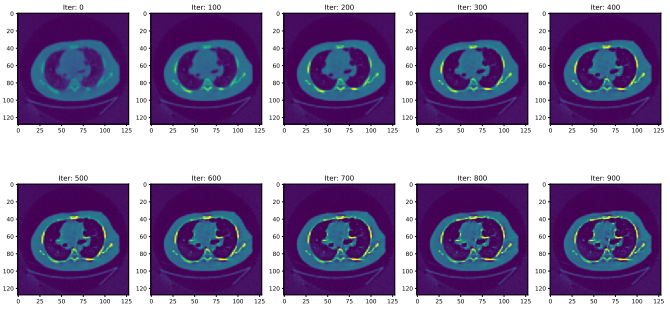


Fig. 11: Scatter-corrupted initial reconstruction (Left) and final output of RPGD iterations (Right) for hyperparameters $\alpha_0 = 2.10^{-2}$, $\gamma = 4.10^{-2}$, and $c_n = (0.8)\mathbf{1}$.

where H is the 2D radon transform, A is the FBP with ramp filtering, and CNN_{θ_t} is a DCNN with the UNET [8] structure with a skip connection between input and output as before. CNN_{ξ}^{sc} is the DCNN used to perform scatter correction on the initial reconstruction. This network was trained using the following loss function:

$$J_{sc}(\xi) = \sum_{q=1}^{Q} ||x^q - CNN_{\xi}^{sc}(Ay^q)||_2^2. \tag{28}$$

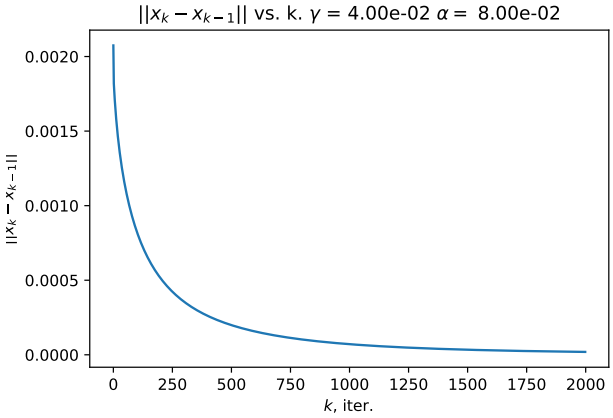


Fig. 12: ℓ_2 norm of the difference between consecutive iterations for scatter-corrected initial reconstruction and RPGD setting for hyperparameters $\alpha_0 = 8.10^{-2}$, $\gamma = 4.10^{-2}$, and $c_n = (0.8)\mathbf{1}$.

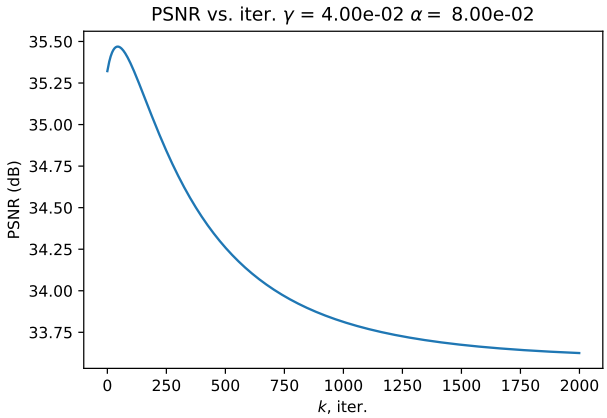


Fig. 13: PSNR values of consecutive iterations for scatter-corrected initial reconstruction and RPGD setting for hyperparameters $\alpha_0 = 8.10^{-2}$, $\gamma = 4.10^{-2}$, and $c_n = (0.8)\mathbf{1}$.

D. Comparison with a DCNN based projection-domain scatter correction method

A baseline method for data-driven projection-domain scatter correction, namely Deep Scatter Correction (DSE), is proposed in [9] where using a DCNN having a UNET [8] structure,

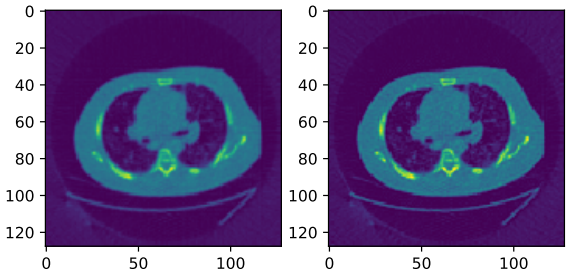


Fig. 14: Scatter-corrected initial reconstruction (Left) and final output of RPGD iterations (Right) for hyperparameters $\alpha_0 = 8.10^{-2}$, $\gamma = 4.10^{-2}$, and $c_n = (0.8)\mathbf{1}$.

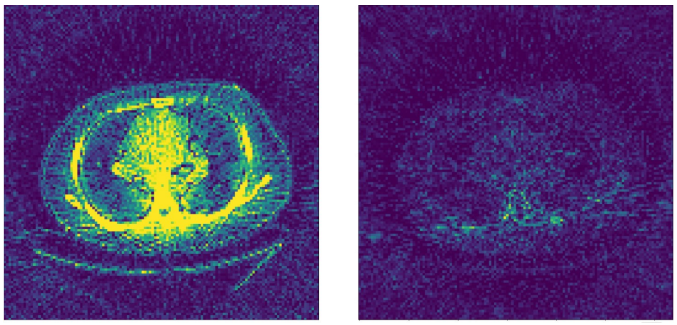


Fig. 15: Absolute error figures for scatter-corrupted reconstruction (Left) and Scatter-corrected initial reconstruction and RPGD output (Right). Color range is reduced to a tenth of the reconstruction figure range.

the primary measurements p are estimated from the total measurements τ as follows:

$$p(t,\theta)^* = CNN_{\beta}(\tau(t,\theta)) \tag{29}$$

Using the dataset in Section VI with the same 27 phantoms for training and the remaining 3 phantoms as the test set, CNN is trained using the following loss function

$$\beta \sum_{\theta} ||\tau_\theta - p_\theta||_2^2 \tag{30}$$

using the Adam optimizer. Results for the scatter corrected reconstruction for the same axial slice from the test set is shown in Figures 16 and 17. These results indicate that performing a measurement consistent correction on the reconstruction domain cannot perform as good as this baseline method that works over projection-domain data.

E. PSNR values of test reconstructions

In Table VII-E, the respective best PSNR values for each experiment in this section is described.

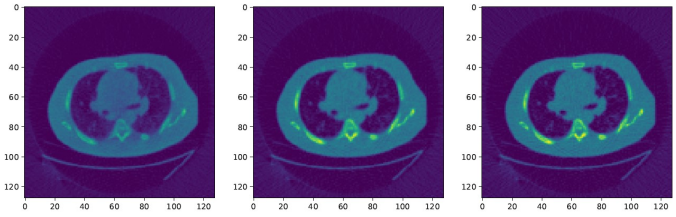


Fig. 16: Scatter-corrupted reconstruction (Left), scatter-corrected initial reconstruction (Middle) and output of DSE (Right).

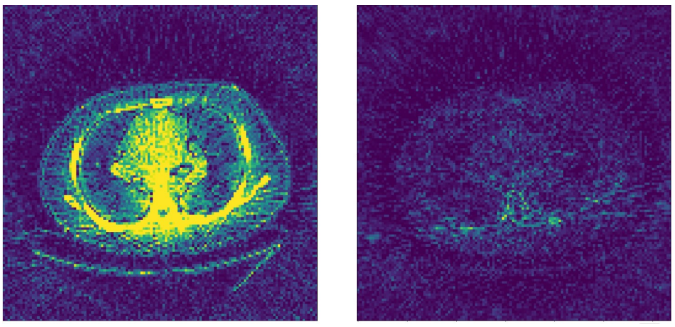


Fig. 17: Absolute error figures for scatter-corrupted reconstruction (Left) and DSE output (Right). Color range is reduced to a tenth of the reconstruction figure range.

Experiment/Method	PSNR (dB)
Total meas. init. recon. and RPGD	26.9
Total meas. recon. and modified RPGD	31.1
Scatter-corrected init. recon. and RPGD	35.5
DSE	37.7

TABLE I: Best PSNR values obtained by each experiment/method described in Section VII.

VIII. Conclusions

In this project, a convergent and iterative data-driven method RPGD investigated and its performance was tested on the scatter correction problem in X-ray CT. Although the original method does not provide sufficient scatter correction, it was possible to obtain satisfactory results after modifying the method on how it obtains the initial reconstruction. However, even after this modification, a baseline method which performs scatter correction on projection domain still performs significantly better than the modified RPGD method. An important reason for this observation can be the requirement of a larger training dataset for the projection operator.

References

[1] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *Proceedings of the 27th international conference on international conference on machine learning*, 2010, pp. 399–406. 2

[2] Y. Yang, J. Sun, H. Li, and Z. Xu, "Deep admm-net for compressive sensing mri," in *Proceedings of the 30th international conference on neural information processing systems*, 2016, pp. 10–18. 2

[3] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *2013 IEEE Global Conference on Signal and Information Processing*. IEEE, 2013, pp. 945–948. 2

[4] J. Rick Chang, C.-L. Li, B. Poczos, B. Vijaya Kumar, and A. C. Sankaranarayanan, "One network to solve them all–solving linear inverse problems using deep projection models," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 5888–5897. 2

[5] H. Gupta, K. H. Jin, H. Q. Nguyen, M. T. McCann, and M. Unser, "Cnn-based projected gradient descent for consistent ct image reconstruction," *IEEE transactions on medical imaging*, vol. 37, no. 6, pp. 1440–1453, 2018. 3, 4

[6] K. W. Clark, B. A. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. R. Phillips, D. Maffitt, M. Pringle, L. Tarbox, and F. Prior, "The cancer imaging archive (tcia): Maintaining and operating a public information repository," *Journal of Digital Imaging*, vol. 26, pp. 1045–1057, 2013. 5

[7] A. Badal and A. Badano, "Accelerating monte carlo simulations of photon transport in a voxelized geometry using a massively parallel graphics processing unit." *Medical physics*, vol. 36 11, pp. 4878–80, 2009. 5

[8] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," 2015. 5, 7

[9] J. Maier, S. Sawall, M. Knaup, and M. Kachelrieß, "Deep scatter estimation (dse): accurate real-time scatter estimation for x-ray ct using a deep convolutional neural network," *Journal of Nondestructive Evaluation*, vol. 37, no. 3, pp. 1–9, 2018. 7