An Adversarial Learning Based Approach for 2D Unknown View Tomography

Mona Zehni and Zhizhen Zhao

***Abstract*—The goal of 2D tomography is to recover an image given its projections from various views. It is often presumed that viewing angles associated with the projections are known in advance. Under certain situations, however, these angles are known only approximately or are completely unknown. It becomes more challenging to reconstruct the image from a collection of random projections with unknown viewing directions. We propose an adversarial learning based approach to recover the image and the viewing angle distribution by matching the empirical distribution of the measurements with the generated data. Fitting the distributions is achieved through solving a min-max game between a generator and a critic based on Wasserstein generative adversarial network structure. To accommodate the update of the viewing angle distribution through gradient back propagation, we approximate the loss using the Gumbel-Softmax reparameterization of samples from discrete distributions. Our theoretical analysis verifies the unique recovery of the image and the projection distribution up to a rotation and reflection upon convergence. Our extensive numerical experiments showcase the potential of our method to accurately recover the image and the viewing angle distribution under noise contamination.**

***Index Terms*—2D unknown view tomography, categorical distribution, generative adversarial learning, Gumbel-softmax, Hartley-Bessel expansion.**

I. Introduction

Multitude of imaging modalities rely on reconstructing an unknown signal either in 2D or 3D domain given a set of partial measurements. Examples of such are medical imaging and cryo-electron microscopy (cryo-EM) for imaging macro-molecules, to name a few. More specifically, in a tomography setup, the measurements i.e. projections, are the line or plane integrals of the underlying object along various angles. In imaging applications such as computed tomography (CT), the viewing angles are known a-priori through the acquisition process. However, this does not hold when reconstructing macromolecular structures in cryo-EM [1]. Thus, it is important to develop solutions for tomography with unknown projection directions. In this paper, we focus on 2D unknown view tomography (UVT) with the goal of jointly recovering the unknown 2D image and the viewing angle distribution given a large set of noisy projections.

Tomographic inversion with known viewing angles is typically a linear inverse problem and is solved by filtered backprojection (FBP), direct Fourier methods [2], or solving a regularized optimization problem [3], [4], [5], [6]. Moreover, deep learning solutions, training on rich datasets, exist that either learn the reconstruction from sinogram to image [7], [8], [9], [10], [11], denoise the FBP reconstructed images from a low-dose sinogram [12], [13], [14], [15], [16], [17] in a supervised manner or provide a prior i.e. regularizer, over the space of target images [18], [19].

However, the knowledge of the viewing angles is not always available or accurate. To avoid adverse effects on the quality of the reconstructed image, it is important to account for uncertainties in the viewing angles. Previous methods devoted to 2D UVT estimate the viewing angles either prior to [20], [21], [22], [23], [24], [25] or jointly with the image reconstruction [26]. In addition, in limited settings, [27], [28], [29], [30] bypass the estimation of the projection views via the use of invariant features.

In this paper, we present an unsupervised adversarial learning based approach for 2D tomography with unknown random viewing angles, namely *UVTomo-GAN*. Our approach does not require large paired training sets and reconstructs an image given merely its unordered tomographic measurements. By employing generative adversarial networks (GAN) [31], our approach recovers the image and viewing angle distribution through matching the distributions of the generated projections with the measurements. Our proposed method is inspired by CryoGAN [32] in which a 3D cryo-EM map is reconstructed given a large set of noisy projection images with unknown orientations by employing Wasserstein-GAN [33]. The main assumption in CryoGAN is that the distribution of the orientations of the particles is known beforehand. However, in cryo-EM experiments, the distribution of the orientations is hard to obtain a-priori. Therefore, under the 2D UVT set-up, we remove the assumption that the viewing angle distribution is given and develop a new approach to recover both the viewing angle distribution and the 2D image simultaneously.

To recover the viewing angle distribution in a GAN framework, the original generator's loss involves sampling from the viewing angle distribution which is non-differentiable. To enable the flow of gradients in the backward pass through this non-differentiable operator, we modify the loss function at the

Manuscript received 16 January 2022; revised 2 May 2022 and 18 July 2022; accepted 20 July 2022. Date of publication 10 August 2022; date of current version 26 August 2022. This work was supported in part by NSF under Grants DMS-1854791 and OAC-1934757, and in part by Alfred P. Sloan Foundation. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Tobias Lasser. *(Corresponding author: Mona Zehni.)*

The authors are with the Department of Electrical and Computer Engineering and Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: mzehni2@illinois.edu; zhizhenz@illinois.edu).

Digital Object Identifier 10.1109/TCI.2022.3197939

generator side using Gumbel-Softmax approximation of samples from a categorical distribution [34]. Our proposed idea is general and applicable to a vast range of similar inverse problems which involve latent variables with unknown probability distributions such as multi-segment reconstruction [35].

This manuscript is an extension of our previous work [36]. In this paper, we use the truncated Hartley-Bessel expansion of the image in the Hartley domain in our reconstruction pipeline. This truncated expansion regularizes the images and allows for the direct use of central slice theorem (CST) to generate the projections efficiently. As noted in [24], 2D tomography from noisy projections taken at unknown random directions with non-uniform distribution is more challenging than its 3D analogue, since we cannot directly use the geometric constraints given by CST in 3D. Our theoretical analysis and numerical results affirm the ability of our method in recovering the image and projection angle distribution accurately from both clean and noisy measurements.

The organization of this paper is as follows. Section II summarizes related work to UVT. We introduce the projection formation model and the reconstruction method in Sections III and IV. The analysis and experimental results are described in Sections V and VI. The discussions and future directions are presented in Section VII. We conclude the paper in Section VIII.

II. Related Work

In this section, we review related literature on 2D UVT and unsupervised solutions for 3D UVT task.

A. 2D UVT

One family of 2D UVT solutions determine the viewing angles first [20], [21], [22], [23], [24], [25] and reconstruct the image given the estimated views subsequently. Other approaches include iterative methods that solve for the 2D image and the viewing angles in alternating steps [26]. While proven effective, these methods are computationally expensive and sensitive to initialization. In another class of methods, to circumvent the estimation and refinement of the viewing angles, a set of rotation invariant features are estimated from the noisy projections. These features are later on used to reconstruct the unknown image [27], [28], [29], [30]. Note that these methods require only one pass through the projection dataset and are therefore computationally more efficient. However, they are mainly used, when the underlying object is sparse [27], [28], projections in the form of tilt series are available [30] or to recover a low-resolution ab-initio model [29].

B. Adversarial Learning for 3D UVT

Gupta et al. in CryoGAN [32] proposed an unsupervised learning approach through a distribution matching lens for cryo-EM single particle reconstruction. In CryoGAN, the goal is to estimate the underlying 3D density such that the distribution of the observed projection image dataset and the one generated from the estimated volume match. Due to its distribution matching criterion, CryoGAN bypasses the estimation of individual projection parameters. In CryoGAN, the distribution distance is chosen as Wasserstein-1 (W_1), i.e. Earth Mover's distance. Thus, the reconstruction problem is stated as:

$$v^* = \operatorname*{argmin}_{v} W_1(P_{\text{sim}}(v; p_{\text{latent}}), P_{\text{real}}) \tag{1}$$

where P_{real} is the distribution of the observed (i.e. real) projection image dataset. Also, $P_{\text{sim}}(v; p_{\text{latent}})$ is the distribution of the simulated projection image dataset generated from the volume v following an a-priori *known* distribution for the latent variables p_{latent}. In a cryo-EM setup, each projection image is obtained from the volume following a forward model. This forward model is parameterized by the projection view, in-plane translation and the contrast transfer function (CTF) parameters corresponding to the projection image. Given a projection image dataset, the collection of these parameters (projection view, in-plane translation and CTF parameters), is considered a *random latent variable* with p_{latent} probability distribution, which in CryoGAN is assumed to be known. Thus, to sample from P_{sim} given v and p_{latent}, one samples latent variables based on p_{latent} and then adopt the projection forward model to generate random simulated projections of v.

As computing W_1 between two high-dimensional distributions is highly intractable, W_1 minimization is often done in its dual form, following Kantrovich-Rubinstein duality [33]:

$$v^* = \operatorname*{argmin}_{v} \max_{f:\|f\|_L \leq 1} (\mathbb{E}_{y \sim P_{\text{real}}}[f(y)] - \mathbb{E}_{x \sim P_{\text{sim}}(v; p_{\text{latent}})}[f(x)]) \tag{2}$$

where f represents a 1-Lipschitz function, mapping its input (i.e. a projection image) to a single real-valued score.

Due to the close link between (2) and Wasserstein-GAN (WGAN) frameworks [33], CryoGAN specifically proposes the use of WGAN with gradient-penality (GP), WGAN-GP [37], to solve (2). In a WGAN-GP setup, the mapping f is modeled via a neural network named *critic* and its 1-Lipschitz continuity constraint is enforced via the GP term.

In this paper, we extend the CryoGAN framework for the 2D UVT problem defined in Section III. In a 2D UVT setting, the projection views form the underlying latent variable. Unlike CryoGAN, we assume the latent variable probability distribution (p_{latent} in CryoGAN context) is unknown and we develop a novel approach to handle its joint recovery with the image. In addition, we compare our method against the baselines formed by the adaptations of CryoGAN for 2D UVT in Section VI.

III. Projection Formation Model and Problem Formulation

We define the 1D projection formation model as,

$$\zeta_\ell = \mathcal{P}_{\theta_\ell} I + \varepsilon_\ell, \quad \ell \in \{1, 2, \ldots, L\} \tag{3}$$

where $I : \mathbb{B}_2 \to \mathbb{R}_1$ is an unknown 2D image compactly supported in the unit ball $\mathbb{B}_2$ we wish to estimate. We restrict I to the space of absolute and square integrable functions on $\mathbb{B}_2$, i.e. $I \in \mathcal{L}_1(\mathbb{B}_2) \cap \mathcal{L}_2(\mathbb{B}_2)$. $\mathcal{P}_\theta$ denotes the tomographic projection operator that takes the line integral along the parallel beams whose normal direction makes an angle $\theta \in [0, 2\pi)$ with the

x-axis,

$$(\mathcal{P}_\theta I)(x) = \int_{-\infty}^{\infty} I(R_\theta\,\mathbf{x})dy \tag{4}$$

where $\mathbf{x} = [x, y]^T$ represents the 2D Cartesian coordinates. R_θ is a 2×2 rotation matrix associated with θ. As I is compactly supported in $\mathbb{B}_2$, its projection along any direction would also be compactly supported in the unit ball, i.e. $\mathcal{P}_\theta I \in \mathcal{L}_1(\mathbb{B}_1) \cap \mathcal{L}_2(\mathbb{B}_1)$. We assume the viewing angles $\{\theta_\ell\}_{\ell=1}^L$ are unknown and randomly drawn from an *unknown* distribution p. Finally, the discretized projection lines of length m are corrupted by additive white Gaussian noise ε_ℓ with zero mean and variance σ^2. Here we consider σ to be known, although an unbiased estimator of σ is attainable from the variance of the boundary pixels of the projections that only contain noise [24].

In this paper, given a large set of noisy projections, i.e. $\{\zeta_\ell\}_{\ell=1}^L$, we aim to recover the image I and the unknown distribution of the viewing angles p.

IV. Method

A. Image Representation

To alleviate the computational cost of generating projections in practice, (3) is evaluated in Fourier domain using non-uniform fast Fourier transform [38] according to central slice theorem (CST). CST states that the Fourier transform of the projection corresponds to the central slice in the 2D Fourier domain,

$$\mathcal{F}(\mathcal{P}_\theta I)(\xi) = \mathcal{F}(I)(\xi, \theta). \tag{5}$$

with $\mathcal{F}$ denoting the Fourier transform and (ξ, θ) the polar coordinates. This motivates us to directly adopt CST to generate the projections. Therefore, in our pipeline we seek to recover the image in Fourier domain rather than pixel domain.

We use the Hartley transform of the images, which is a real representation closely related to Fourier transform and defined as:

$$\mathcal{H}(I) = \text{real}\{\mathcal{F}(I)\} - \text{imag}\{\mathcal{F}(I)\}, \tag{6}$$

where $\mathcal{H}$ denotes the Hartley transform. We assume the image I has essential bandlimit $0 \le s \le \frac{1}{2}$ and is concentrated in the spatial domain with radius $R \le \frac{m}{2}$. Therefore, $\mathcal{H}(I)$ can be expanded on an orthonormal basis on a disk of radius s. We choose real-valued steerable Hartley-Bessel (HB) expansion as a continuous representation which implicitly regularizes the image I and enables the use of CST for generating the projections. Based on the Fourier-Bessel basis introduced in [39], [40], we construct the real-valued HB basis $u_s^{k,q}(\xi, \theta) = J_s^{k,q}(\xi)\text{cas}(k\theta)$ with radial functions

$$J_s^{k,q}(\xi) = \begin{cases} N_{k,q} J_k\left(R_{k,q}\frac{\xi}{s}\right), & \xi \le s, \\ 0, & \xi > s, \end{cases} \tag{7}$$

where J_k is the Bessel function of the first kind and integer order k, $R_{k,q}$ denotes the q-th root of J_k, and $N_{k,q} = (s\sqrt{\pi}|J_{k+1}(R_{k,q})|)^{-1}$ is the normalization factor. The angular part of the HB basis is $\text{cas}(k\theta) = \cos(k\theta) + \sin(k\theta)$. We can expand $\mathcal{H}(I)$ on the HB basis,

$$\mathcal{H}(I)(\xi, \theta) = \sum_{k=-\infty}^{\infty} \sum_{q=1}^{\infty} c_{k,q} J_s^{k,q}(\xi)\,\text{cas}(k\theta). \tag{8}$$

Note that, q and k correspond to radial and angular frequencies. We can truncate the expansion in (8) for functions that are well concentrated in real and Fourier space using a sampling criterion $R_{k,q} \le 2\pi sR$ [39], [41]. The maximum angular frequency index is denoted by $K_{\max}$ and the maximum radial frequency for k-th angular frequency is denoted by p_k. The expansion coefficients $c = \{c_{k,q} \,|\, \forall(k, q) \text{ s.t. } |k| \le K_{\max}, 1 \le q \le p_k\}$ are the unknown parameters of I we aim to recover. For an image with $s < 0.5$ or $R < \frac{m}{2}$, c has less number of terms than the number of pixels I, i.e. the cardinality of $c < m^2$. Thus, c would constitute a compressed representation of the image.

Given the image expanded on HB basis, following CST, the Hartley transform of the projection from angle θ_ℓ is simply obtained by setting $\theta = \theta_\ell$ in (8) and is written as:

$$\mathcal{H}(\mathcal{P}_{\theta_\ell} I)(\xi) = \sum_{k=-K_{\max}}^{K_{\max}} \sum_{q=1}^{p_k} c_{k,q} J_s^{k,q}(\xi)\text{cas}(k\theta_\ell) = H_{\theta_\ell}(\xi)c. \tag{9}$$

Therefore, we rewrite (3) in Hartley domain as:

$$\widetilde{\zeta}_\ell = H_{\theta_\ell} c + \widetilde{\varepsilon}_\ell,\ \theta_\ell \sim p, \quad \ell \in \{1, 2, \ldots, L\}, \tag{10}$$

with $\widetilde{\zeta} = \mathcal{H}(\zeta)$ and $\widetilde{\varepsilon} = \mathcal{H}(\varepsilon)$. The Hartley transform is unitary due to its self-adjoint and self-inverse properties. Therefore, the distribution of the Gaussian additive noise is preserved after taking the Hartley transform, i.e. $\widetilde{\varepsilon}_\ell \sim \mathcal{N}(\mathbf{0}_m, \sigma^2 I_m)$ where $\mathbf{0}_m$ is a vector of zeros of length m and I_m is an $m \times m$ identity matrix.

From the HB expansion coefficient c, we can reconstruct the image in the spatial domain,

$$I(r, \varphi) = \sum_{k=-K_{\max}}^{K_{\max}} \sum_{q=1}^{p_k} c_{k,q}\, \mathcal{H}\left(u_s^{k,q}\right)(r, \varphi) \tag{11}$$

where

$$\mathcal{H}\big(u_s^{k,q}\big)(r, \varphi) = \frac{2\sqrt{2\pi}s(-1)^{(q+l)} R_{k,q} J_k(2\pi sr)}{(2\pi sr)^2 - R_{k,q}^2} \times \cos\left(k\varphi + \frac{\pi}{4}\right), \tag{12}$$

and $l = \frac{k+1}{2}$ for odd k and $l = \frac{k}{2}$ for even k. Since we have the analytical form of the basis function, we can easily evaluate the function values on Cartesian coordinates $[x, y]$ with $x = r\cos\varphi$ and $y = r\sin\varphi$.

B. Adversarial Learning for 2D UVT

Similar to CryoGAN, our reconstruction criterion is matching the distribution of the real projection dataset and the projections generated by c and p following (10). As GANs have proven suitable for matching a target distribution, we employ an adversarial learning framework presented in Fig. 1.

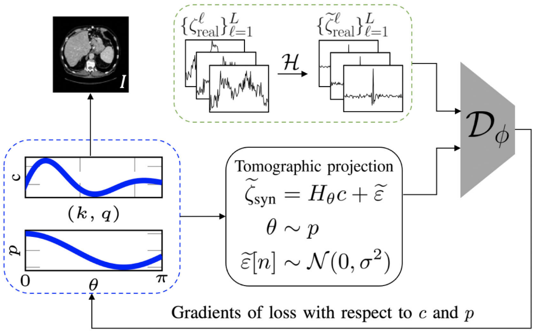
I
$\{\zeta^{\ell}_{\text{real}}\}_{\ell=1}^{L}$
$\mathcal{H}$
$\{\widetilde{\zeta}^{\ell}_{\text{real}}\}_{\ell=1}^{L}$
$\mathcal{D}_\phi$
c
(k, q)
p
0
θ
π
Tomographic projection
$\widetilde{\zeta}_{\text{syn}} = H_\theta c + \widetilde{\varepsilon}$
$\theta \sim p$
$\widetilde{\varepsilon}[n] \sim \mathcal{N}(0, \sigma^2)$
Gradients of loss with respect to c and p

Fig. 1. An illustration of our pipeline for adversarial learning based 2D UVT. Given the projections $\{\zeta^{\ell}_{\text{real}}\}_{\ell=1}^{L}$ (green dashed box), we recover the truncated Hartley-Bessel expansion coefficients c of the image and viewing angle distribution p (blue dashed box).

Our adversarial learning approach consists of a critic $\mathcal{D}_\phi$ and a generator $\mathcal{G}$. Unlike classic GAN models with generators parameterized by neural networks with learnable weights, we specify the generator $\mathcal{G}$ by the known projection model defined in (10), the parameters of the image and viewing angle distribution, i.e. c and p. The generator's goal is to output projections that are close to the real projection dataset $\{\widetilde{\zeta}^{\ell}_{\text{real}}\}_{\ell=1}^{L}$ in distribution and hence fool the critic. For our model, the unknowns we seek to estimate at the generator side are c and p. On the other hand, the critic $\mathcal{D}_\phi$, parameterized by ϕ, tries to distinguish between the observations and the generated projections. Our pipeline is depicted in Fig. 1.

We use WGAN [33] loss, express the loss function in terms of c, p and ϕ and state the min-max problem as,

$$\mathcal{L}(c, p, \phi) = \sum_{b=1}^{B} \mathcal{D}_\phi(\widetilde{\zeta}^{b}_{\text{real}}) - \mathcal{D}_\phi(\widetilde{\zeta}^{b}_{\text{syn}}) \tag{13}$$

$$\widehat{c}, \widehat{p} = \underset{c,p}{\text{argmin}} \max_{\phi} \mathcal{L}(c, p, \phi), \tag{14}$$

where $\mathcal{L}$ denotes the loss, B and b represent the batch size and the index of a sample in the mini-batch, respectively. Also, $\widetilde{\zeta}_{\text{real}}$ and $\widetilde{\zeta}_{\text{syn}}$ mark the real and synthesized projections in Hartley domain. $\widetilde{\zeta}_{\text{syn}}$ is generated from the estimated image $\widehat{c}$ and projection distribution $\widehat{p}$ following $\widetilde{\zeta}_{\text{syn}} = H_\theta \widehat{c} + \widetilde{\varepsilon}$, $\theta \sim \widehat{p}$. In our experiments, we used spectral normalization (SN) [42] to regularize the critic and found that SN is sufficient for stabilizing the training. Following common practice, we solve (14) by alternating updates between ϕ and the generator's variables, i.e. c and p, based on the associated gradients.

The loss at the generator side for a fixed $\mathcal{D}_\phi$ is,

$$\mathcal{L}_G(c, p) = -\sum_{b=1}^{B} \mathcal{D}_\phi(H_{\theta_b} c + \widetilde{\varepsilon}_b),\ \theta_b \sim p. \tag{15}$$

While (15) is differentiable with respect to c, its gradient of p is not defined, as it involves sampling θ_b from the distribution p. This hinders updating p through gradient back-propagation. To address this, we aim to design an alternative approximation of (15) which is differentiable with respect to p.

Algorithm 1: UVTomo-GAN.

```
Require: $\alpha_\phi$, $\alpha_c$, $\alpha_p$: learning rates for $\phi$, $c$ and $p$. $n_{\text{disc}}$: the number of updates of the critic per generator update.
Input: $\{\widetilde{\zeta}^{\text{real}}_{\ell}\}_{\ell=1}^{L}$. Random initialization of $c$. The distribution $p$ is initialized with $\text{Unif}(0, 2\pi)$.
Output: Estimates of $I$ and $p$.
while $\phi$ has not converged do
    for $t = 0, \ldots, n_{\text{disc}} - 1$ do
        Sample a batch from real data, $\{\widetilde{\zeta}^{b}_{\text{real}}\}_{b=1}^{B}$.
        Sample a batch of simulated projections using estimated $c$ and $p$, i.e. $\{\widetilde{\zeta}^{b}_{\text{syn}}\}_{b=1}^{B}$ following (10).
        Update the critic following gradient ascent steps using the gradient of (13) with respect to $\phi$.
    end for
    Sample a batch of $\{r_{i,b}\}_{b=1}^{B}$ using (19).
    Update $c$ and $p$ using stochastic gradient descent steps by taking the gradients of (20) with respect to $c$ and $p$.
end while
```

To accommodate this approximation, we first discretize the support of the viewing angles, i.e. $[0, 2\pi)$ into N_θ equal-sized bins. This makes p a probability mass function (PMF) of length N_θ with the following properties:

$$\sum_{i=0}^{N_\theta - 1} p_i = 1, \text{and } p_i \geq 0, \forall i \in \{0, \ldots, N_\theta - 1\}. \tag{16}$$

Now p corresponds to a discrete or categorical distribution over θ, which implies the sampled viewing angles from p can only belong to N_θ discrete categories. Therefore, we re-write the loss function (15) as:

$$\mathcal{L}_G(c, p) = -\sum_{b=1}^{B} \sum_{t=0}^{N_\theta - 1} \delta(\theta_t - \theta_b) \mathcal{D}_\phi(H_{\theta_t} c + \widetilde{\varepsilon}_b),\ \theta_b \sim p. \tag{17}$$

A closer look at (17) reveals that $\delta(\theta_t - \theta_b)$, $\theta_b \sim p$ is a sample from the discrete distribution p. This enables us to incorporate the notion of Gumbel-Softmax distribution and approximate (15) as:

$$\mathcal{L}_G(c, p) \approx -\sum_{b=1}^{B} \sum_{i=0}^{N_\theta - 1} r_{i,b}(p) \mathcal{D}_\phi(H_{\theta_i} c + \widetilde{\varepsilon}_b), \tag{18}$$

$$r_{i,b}(p) = \frac{\exp\left((g_{b,i} + \log(p_i))/\tau\right)}{\sum_{j=0}^{N_\theta - 1} \exp\left((g_{b,j} + \log(p_j))/\tau\right)},$$
$$g_{b,i} \sim \text{Gumbel}(0, 1), \tag{19}$$

where τ is the softmax temperature factor. As $\tau \to 0$, $r_{i,b}(p) \to \text{one-hot}(\text{argmax}_i[g_{b,i} + \log(p_i)])$. Moreover, to obtain samples from the $\text{Gumbel}(0, 1)$ distribution, it suffices to draw $u \sim \text{Unif}(0, 1)$, $g = -\log(-\log(u))$ [34]. Note that due to the reparametrization trick applied in (18), the approximated generator's loss has a tangible gradient with respect to p. We also add prior knowledge on the image and projection distribution in the form of regularization terms. Hence, the regularized loss function we optimize at the generator side is:

$$\mathcal{L}(c,p)=\mathcal{L}_G(c,p)+\gamma_1 g_{\text{TV}}(c)+\gamma_2\|c\|^2+\gamma_3\text{TV}(p)+\gamma_4\|p\|^2, \tag{20}$$

where we include total variation (TV) and ℓ_2 regularization terms for the image, with γ_1 and γ_2 weights. To construct the TV of the image in terms of c, we use (11) to render I on a Cartesian grid in spatial domain and then compute total variation of I. Furthermore, we assume that the unknown PMF is a piece-wise smooth function of viewing angles (which is a valid assumption especially in single particle analysis in cryo-EM [43]), therefore adding TV and ℓ_2 regularization terms for the PMF with γ_3 and γ_4 weights. We present the pseudo-code for UVTomo-GAN in Algorithm 1.

C. Maximum Marginalized Likelihood Estimation Via Expectation-Maximization

As a baseline for UVTomo-GAN, we consider maximum marginalized likelihood estimation (MMLE). We solve MMLE in Fourier domain via expectation-maximization (EM) and represent $\mathcal{F}(I)$ with its expansion coefficients a on Fourier-Bessel bases. Thus, MMLE is formulated as

$$\widehat{a},\widehat{p}=\underset{a,p}{\text{argmax}}\sum_{\ell=1}^{L}\log\left(\sum_{i=0}^{N_\theta-1}P(\mathcal{F}(\zeta_\ell)|a,\,\theta_i)p_i\right). \tag{21}$$

To solve (21), we take the gradients with respect to a and p and set them to zero. For p, we further impose $\sum_{i=0}^{N_\theta-1}p_i=1$. This yields the following alternating updates for a and p, in the form of:

$$(\text{E-step}):r_{i,j}^t=\frac{\exp\left(-\frac{\|\mathcal{F}(\zeta_i)-H_{\theta_j}a^{t-1}\|^2}{2\sigma^2}\right)}{\sum_{j=0}^{N_\theta-1}p_j^{t-1}\exp\left(-\frac{\|\mathcal{F}(\zeta_i)-H_{\theta_j}a^{t-1}\|^2}{2\sigma^2}\right)}, \tag{22}$$

$$(\text{M-step}):\begin{cases}\boldsymbol{A}^t a^t=\boldsymbol{b}^t,\\ p_j^t=\frac{\sum_{i=1}^{L}r_{i,j}^t}{\sum_{i=1}^{L}\sum_{j=0}^{N_\theta-1}r_{i,j}^t},\end{cases} \tag{23}$$

where

$$\boldsymbol{A}^t((k,q),(k,'\,q'))=\widehat{p^t}(k-k')\sum_{\xi=1}^{N_\xi}J_s^{k,q}(\xi)J_s^{k,'q'}(\xi) \tag{24}$$

$$\widehat{p^t}(k)=\sum_{j=0}^{N_\theta-1}p_j^t\exp\left(-\imath\frac{2\pi kj}{N_\theta}\right) \tag{25}$$

$$\boldsymbol{b}^{\boldsymbol{t}}(k,q)=\sum_{\xi=1}^{N_\xi}\sum_{j=0}^{N_\theta-1}J_s^{k,q}(\xi)\exp\left(-\imath\frac{2\pi kj}{N_\theta}\right)\sum_{i=1}^{L}r_{i,j}\mathcal{F}(\zeta_i) \tag{26}$$

where $r_{i,j}$ denotes the probability that the $i-$th projection is associated with θ_j angle and t is the iteration index. Also, $H_\theta a$ generates the projection at θ direction in Fourier domain given FB expansion coefficients a. In (23), $\boldsymbol{A}^t$ is indexed by (k,q) and $(k,'\,q')$ pairs and the discretization in ξ is identical to the projection dataset. The advantages of using truncated FB expansion is that: (1) similar to HB representation, it provides an implicit regularization on the image, and (2) building matrix $\boldsymbol{A}^t$ in (24) in each iteration only requires rescaling the entries of a pre-computed matrix $\boldsymbol{J}((k,q),(k,'\,q'))=\sum_{\xi=1}^{N_\xi}J_s^{k,q}(\xi)J_s^{k,'q'}(\xi)$ by $\widehat{p^t}(k-k')$.

In (22)-(23), we update the probabilistic angular assignments for the projections in the E-step while updating a and p in the M-step. Note that, in the absence of noise, i.e. $\sigma=0$, the E-step reduces to template matching [44]. To solve a^t from the equation $\boldsymbol{A}^t a^t=\boldsymbol{b}^t$, we use preconditioned conjugate gradient descent [45].

D. Computational Complexity

We conclude this section by comparing the computational complexity per iteration of UVTomo-GAN and EM.

1) UVTomo-GAN Complexity: Based on Algorithm 1, we split the computational cost of UVTomo-GAN between: 1) the critic and 2) the generator (i.e. c and p) updates. Let $C_{\mathcal{D}}$ denote a fixed computational cost related to forward and backpropagation passes through the critic $\mathcal{D}_\phi$. As expected, $C_{\mathcal{D}}$ depends on the batch size, network architecture and the size of its input. Thus, the larger the critic network, the higher the $C_{\mathcal{D}}$. For our critic architecture, we use a cascade of $N\ll m$ fully connected (FC) layers with intermediate ReLU non-linearities. Therefore, $C_{\mathcal{D}}$ points to the cost of matrix multiplications and backward passes through these N layers. Furthermore, we keep the input and output sizes of these FC layers to be $O(m)$ (m is the image/projection size). Therefore, $C_{\mathcal{D}}=O(m^2N)=O(m^2)$. As these operations can be parallelized on GPU, forward and backward passes through $\mathcal{D}_\phi$ are time-efficient. For batch size $B=O(m)$, the cost of critic update is $O(B\,C_{\mathcal{D}})=O(m^3)$.

For updating the generator according to (18), first we generate $N_\theta=O(m)$ projections or templates. This is done in $O(m^3)$. A thorough discussion on the derivation of this computational complexity term is deferred to Appendix A.

In our implementation of (18), instead of using B different noise realizations $\{\widetilde{\varepsilon}_b\}_{b=1}^{B}$ for each of the clean templates, we consider N_θ noisy templates in total. This means the loss function we use at the generator side is:

$$\mathcal{L}_G(c,p)\approx-\sum_{b=1}^{B}\sum_{i=1}^{N_\theta}r_{i,b}(p)\mathcal{D}_\phi(H_{\theta_i}c+\widetilde{\varepsilon}_i). \tag{27}$$

Indeed in the absence of noise, (27) matches (18). However, in the noisy case, the benefits of (27) are two-fold: 1) having the same performance as (18) empirically, 2) reducing the number of passes through the critic.

Consequently, adding up the cost of passing N_θ projection templates through $\mathcal{D}_\phi$ leads to a total computational cost of $O(m^3+mC_{\mathcal{D}})$ per generator update step. We update c and p every n_{disc} iterations. Therefore, the average cost of UVTomo-GAN per iteration including the generator and critic's updates is $O\left(\frac{(n_{\text{disc}}-1)m^3+(m^3+mC_{\mathcal{D}})}{n_{\text{disc}}}\right)=O(m^3)$.

EM Complexity: For EM, we specify the computational cost of E-step and M-step. At each E-step, we generate N_θ projection templates. If these templates are generated following

CST and using the non-uniform Fourier transform of the image, they require $O(m^2 \log m)$ computations. Next, we update the angular assignments of L projections by comparing them against $O(m)$ templates, hence a cost of $O(m^2 L)$. Then, the total cost of E-step is $O(m^2 \log m + m^2 L) = O(m^2 L)$. For the M-step, computing $\boldsymbol{b}^t$ from the projections costs $O(m^2 L)$ (or $O(m \log mL)$ if using FFT) while updating FB coefficients a in (23) using conjugate gradient descent has $O(\sqrt{\kappa}\omega)$ computational cost [45] where ω is the number of non-zeros of $\boldsymbol{A}^t$ and κ is its condition number. Note that $\omega = O(\eta\, m^3)$ depends on the number of non-zero elements in $\widehat{p^t}$, i.e. η. If all entries in $\widehat{p^t}$ are non-zero ($\eta = O(m)$), then the M-step's computational cost is $O(\sqrt{\kappa}\, m^4)$. Finally, the overall computational complexity for EM is $O(\sqrt{\kappa}\, \eta\, m^3 + m^2 L)$.

In terms of convergence, we empirically observe that UVTomo-GAN requires more training iterations. We attribute this to the difference between the convergences of stochastic gradient descent used in UVTomo-GAN versus full batch processing in EM. On the other hand, we show that while UVTomo-GAN is robust to the choice of initialization, EM is likely to get stuck in a bad locally optimal solution with random initialization. This observation is also reported in cryo-EM settings in [43], [46].

V. Analysis

In this section, we first define our notations and then formally state the reconstruction guarantees of UVTomo-GAN.

A. Notations

We assume the image $f \in \mathcal{L}_1(\mathbb{B}_2) \cap \mathcal{L}_2(\mathbb{B}_2)$ has a bandlimit $0 < s \leq 0.5$ and compactly supported in the unit ball $\mathbb{B}_2$. In addition, $f \in \text{span}\{u^s_{k,q}\}_\Omega$, $\Omega = \{(k,q) \,|\, |k| \leq K_{\max}, 1 \leq q \leq p_k\}$ with $u^{k,q}_s = J^{k,q}_s(\xi)\text{cas}(k\theta)$. Thus, the Hartley transform of f is expanded on a HB basis set. A measurement ζ associated with the projection angle $\theta \sim p$ is $\zeta = \mathcal{P}_\theta f + \varepsilon$ with $\varepsilon[n] \sim q_\varepsilon$ denoting additive IID noise. We assume q_ε has full support in Fourier domain, i.e. $\{\mathcal{F}q_\varepsilon\}(\omega) \neq 0, \forall \omega$.

Let $\mathrm{O}(2)$ denote the group of all possible rotations and reflections, i.e. $\Gamma^T\Gamma = I$ and $\det(\Gamma) = \pm 1, \forall \Gamma \in \mathrm{O}(2)$. The action of the $\mathrm{O}(2)$ group on f is defined as,

$$(\Gamma f)(\mathbf{x}) = f(\Gamma^{-1}\mathbf{x}), \forall \Gamma \in \mathrm{O}(2) \tag{28}$$

where $\mathbf{x} = [x, y]$ denotes the Cartesian coordinate. On the other hand, the action of Γ on a probability distribution p defined over $[0, 2\pi)$ manifests as a combination of flip or circular shift. The group $\mathrm{O}(2)$ partitions the space of $\text{span}\{u^s_{k,q}\}_\Omega$ into a set of equivalence classes where $[f] = \{\Gamma f, \forall \Gamma \in \mathrm{O}(2)\}$. Let $P^{\text{clean}}_{f,p}$ and $P^{\text{noisy}}_{f,p}$ denote the probability distributions induced by clean and noisy projections, i.e. $\zeta_{\text{clean}} = \mathcal{P}_\theta f$ and $\zeta_{\text{noisy}} = \mathcal{P}_\theta f + \varepsilon$ with $\theta \sim p$, respectively.

B. Theoretical Results

Here we elaborate upon the theoretical reconstruction guarantees of our proposed method.

Theorem 1: Consider $f, g \in \mathcal{L}_1(\mathbb{B}_2) \cap \mathcal{L}_2(\mathbb{B}_2)$ and the associated bounded probability distributions p_f and p_g on the viewing angles distributed in $[0, 2\pi)$. Then,

$$P^{\text{clean}}_{f,p_f} = P^{\text{clean}}_{g,p_g} \Rightarrow [f] = [g],\, [p_f] = [p_g]. \tag{29}$$

Furthermore, if $f = \Gamma g$, $\Gamma \in \mathrm{O}(2)$, then $p_f = \Gamma p_g$.

The proof is provided in Appendix B. Intuitively, Theorem 1 states that if f and g have the same induced clean projection distribution, then the underlying objects and projection distributions are equivalent up to a rotation and reflection. We link the proof of this theorem to unique angular recovery in unknown view tomography [20], [21].

Theorem 2: Assume $f \in \mathcal{L}_1(\mathbb{B}_2) \cap \mathcal{L}_2(\mathbb{B}_2)$ denoting the ground truth (GT) image and p representing the bounded GT probability distribution over the viewing angles $\theta \in [0, 2\pi)$. Let $\widehat{f}$ and $\widehat{p}$ stand for the recovered image and the bounded probability distribution after the convergence of UVTomo-GAN. Consider the asymptotic case as $L \to \infty$. Then,

$$P^{\text{noisy}}_{f,p} = P^{\text{noisy}}_{\widehat{f},\widehat{p}} \Rightarrow \widehat{f} = \Gamma f,\ \widehat{p} = \Gamma p, \tag{30}$$

for a unique $\Gamma \in \mathrm{O}(2)$.

The proof is available in Appendix C. This theorem validates that upon the convergence of UVTomo-GAN in the presence of noise and infinite number of noisy projections, the GT image and viewing angle distribution is recovered up to a rotation-reflection transformation. We defer the study of sample complexity of UVTomo-GAN with finite size projection dataset to future work.

VI. Numerical Results

A. Experiment Setup

1) Datasets: To evaluate the generalization of our method to images of different properties, we conduct experiments on four different images (for additional results, refer to Appendix F). Two are biomedical images of lung and abdomen from low dose CT (LDCT) dataset [47]. We defined the third image as a set of randomly located and shaped ellipses with various intensities. For the last image, we generated the 3D map of 100S Ribosome [48] using its protein sequence in Chimera [49] and took a 2D projection of the generated map along a random view. All images are resized to 101×101 dimension. We refer to these images as Lung, Abdomen, Ellipses and a 2D projection image of 100S Ribosome (Rib-Proj). We synthesize the real projection dataset in Hartley domain following (10) where p is a smooth probability distribution over the viewing angles and is chosen randomly. To generate the real dataset, we discretize the projection angle domain $[0, \pi)$ with 240 equal sized bins and use non-uniform polar FFT [50] and CST to generate the projections. We also add the flipped projections to the dataset, such that θ covers $[0, 2\pi)$. This means $p(\theta) = p(\theta + \pi)$, for $\theta \in [0, \pi)$. Therefore, when estimating p, we only recover p in $[0, \pi)$ range. Throughout this draft, we visualize p on $[0, \pi)$. For the reconstruction, we consider a coarser grid for the viewing angles with $N_\theta = 240$ bins for the interval $[0, 2\pi)$. This way we are taking into account the approximated discretization of θ at the reconstruction time which might differ from how the real

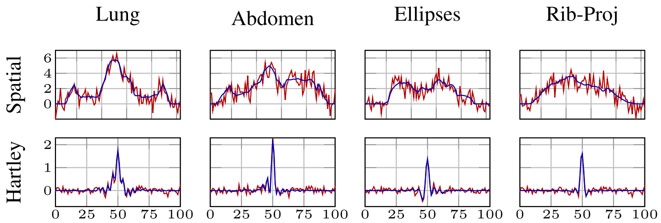

Fig. 2. Samples of clean (blue) and noisy (red) projections in spatial (first row) and Hartely (second row) domain. For noisy data SNR = 3.

viewing angles are obtained. We study two noise regimes: 1) no noise, and 2) noisy with SNR = 3, SNR denoting the ratio of signal-to-noise variance of the projections,

$$\mathrm{SNR} = \frac{\mathrm{Var}\{\zeta_{\text{clean}}\}}{\mathrm{Var}\{\zeta_{\text{noisy}} - \zeta_{\text{clean}}\}} \tag{31}$$

where ζ_{clean} and ζ_{noisy} stand for the clean and noisy projections in spatial domain, respectively. Examples of clean and noisy projections in both spatial and Hartley domains are illustrated in Fig. 2. In our experiments with clean data, the number of projections before adding the flipped versions is $L = 2 \times 10^3$, while for noisy experiments, $L = 2 \times 10^4$.

Training and Network Architecture: We set a batch-size of $B = 200$. We fix the regularization weights on the PMF as $\gamma_3 = 0.01$ and $\gamma_4 = 0.04$ unless otherwise stated. For the Lung and Abdomen images in the clean case, the default image regularization weights are $\gamma_1 = 10^{-5}$ and $\gamma_2 = 5 \times 10^{-5}$. We set $\gamma_1 = 0.001$ and $\gamma_2 = 0$ for Ellipses while having $\gamma_1 = \gamma_2 = 0$ for the Rib-Proj reconstruction. In the noisy case, to obtain the best results in various settings and take into account the difference in the projection datasets, we select γ_1 from $\{0.0005, 0.001, 0.002, 0.005\}$ and γ_2 from $\{0.0005, 0.005, 0.02, 0.04\}$.

We have separate learning rates for $\mathcal{D}_\phi$, c and p denoted by α_ϕ, α_c and α_p, but often choose $\alpha_\phi = \alpha_c$. We select the initial values of α_ϕ, α_c and α_p from $[0.002, 0.01]$ with a step-decay schedule. We update $\mathcal{D}_\phi$, c and p using stochastic gradient descent (SGD) steps. We clip the gradients of $\mathcal{D}_\phi$ and c by 1 and 10 respectively and normalize the gradients of p to have norm 0.1. We train the critic $n_{\text{disc}} = 4$ times per updates of c and p. Although, after training for a while, we increase the frequency of updating c and p by setting $n_{\text{disc}} = 2$. Once converged, we use the reconstructed HB expansion coefficients to re-render the image in spatial domain according to (11).

Our critic consists of four fully connected (FC) layers with ℓ, $\ell/2$, $\ell/4$, and 1 output sizes with ReLU [51] in between. We choose $\ell = 512$ for no noise and $\ell = 256$ for noisy experiments. Our justification for adopting a smaller critic network in noisy case is to avoid overfitting to noisy projections and reduce the leak of noise in the final reconstruction.

To improve the stability of the GAN training, we use spectral normalization [42], applied to all critic layers. To enforce p to have non-negative values while summing up to one, we set it to be the output of a `Softmax` layer. We initialize each entry of c independently with a random variable drawn from $\mathcal{N}(0, 4 \times 10^{-4})$. We set p to be a uniform distribution initially. For the critic, we randomly initialize the weights of the FC layers with a zero-mean Gaussian distribution and standard deviation 0.05 and set the biases to zero. Our implementation is in PyTorch and runs on single GPU.

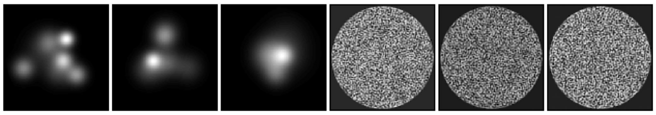

Fig. 3. Examples of the initialization images used in EM.

Evaluation Metrics: To assess the quality of the reconstructed image, we use peak signal to noise ratio (PSNR) and normalized cross correlation (CC). Higher value of these metrics signals better quality of the reconstruction. Also, to evaluate $\widehat{p}$ compared to the ground truth p, we use total variation distance (TV) defined as:

$$d_{\text{TV}} = \frac{1}{2}\|p - \widehat{p}\|_1. \tag{32}$$

B. Baselines

We benchmark UVTomo-GAN with unknown p against five baselines, including graph Laplacian tomography GLT [24], MADE [25] + GL, MMLE with EM, Adapted CryoGAN [32] and Adapted CryoGAN with unif. p. We defer the details of the first two baselines to Appendix E.

In our first baseline GLT, the projections with unknown views are sorted following [24] and the image is reconstructed accordingly. Note that compared to [23], [24] is more resilient to noise.

For our second baseline, we combine MADE [25] and graph Laplacian (GL) to obtain the angle corresponding to each projection. We name this baseline MADE+GL. Unlike GLT, MADE uses a moment-based approach to estimate the angular differences between any two projections. GL is applied to get robust estimation of individual projection angles from the estimates of angular differences.

As our third baseline, we compare against MMLE (21) solved by EM (22)-(23). We initialize EM with 10 random initializations. We test two different forms of initializations, 1) randomly located Gaussian blobs with random standard deviations, 2) initializing each pixel with Uniform distribution within a circular mask, i.e. $I[x, y] \sim \mathrm{Unif}(0, 1)$. In our experiments, we report the best results for EM out of these 10 random initializations, hence the name *EM best random init* for this baseline. Examples of initializations for EM are provided in Fig. 3.

To evaluate the effect of estimating the viewing angle distribution p, we consider two GAN-based benchmarks. In the first, we assume that p is given in advance. For our second GAN-based benchmark, we assume p to be a uniform distribution. In both GAN-based baselines, we follow Algorithm 1. However, we skip the SGD updates on p and instead sample directly from the GT p or the uniform distribution and use (15) as the generator loss. Note that, these two baselines are adapted from CryoGAN [32] (where the distribution of the latent variables is presumed to be known or uniform) to the 2D UVT problem. We refer to these

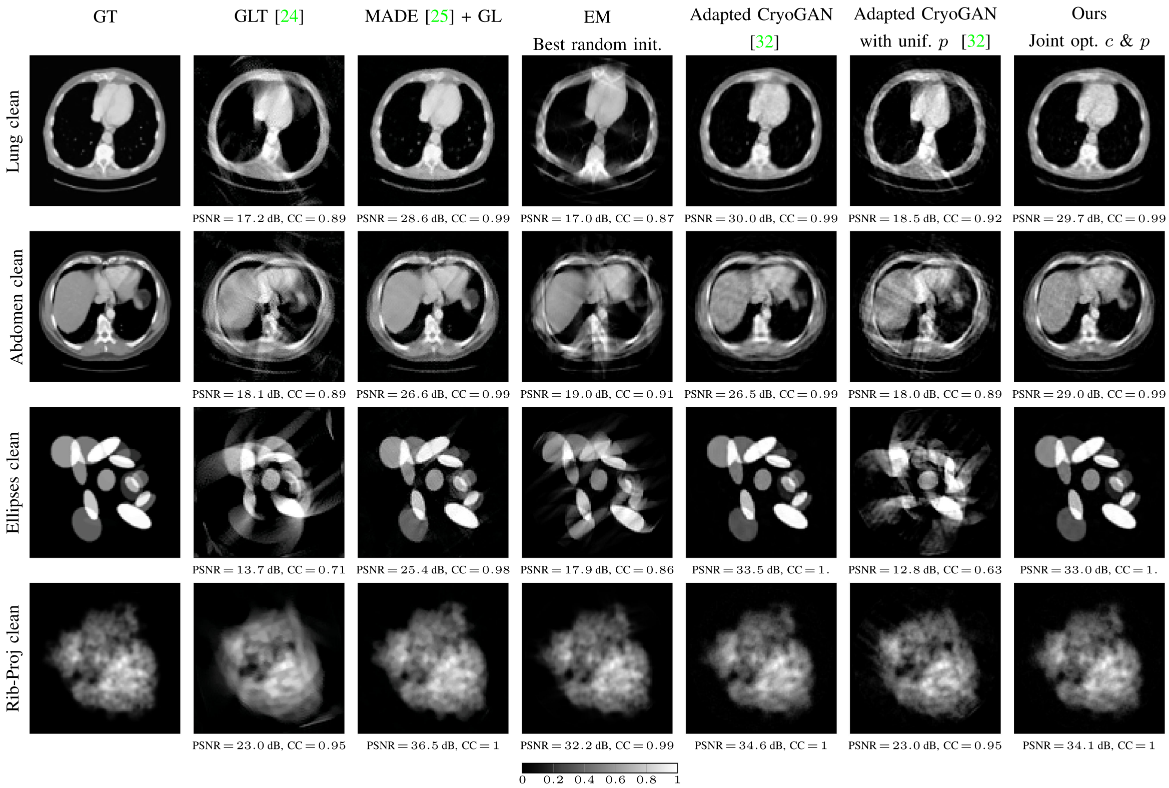
GT
GLT [24]
MADE [25] + GL
EM
Best random init.
Adapted CryoGAN
[32]
Adapted CryoGAN
with unif. p [32]
Ours
Joint opt. c & p
Lung clean
PSNR = 17.2 dB, CC = 0.89
PSNR = 28.6 dB, CC = 0.99
PSNR = 17.0 dB, CC = 0.87
PSNR = 30.0 dB, CC = 0.99
PSNR = 18.5 dB, CC = 0.92
PSNR = 29.7 dB, CC = 0.99
Abdomen clean
PSNR = 18.1 dB, CC = 0.89
PSNR = 26.6 dB, CC = 0.99
PSNR = 19.0 dB, CC = 0.91
PSNR = 26.5 dB, CC = 0.99
PSNR = 18.0 dB, CC = 0.89
PSNR = 29.0 dB, CC = 0.99
Ellipses clean
PSNR = 13.7 dB, CC = 0.71
PSNR = 25.4 dB, CC = 0.98
PSNR = 17.9 dB, CC = 0.86
PSNR = 33.5 dB, CC = 1.
PSNR = 12.8 dB, CC = 0.63
PSNR = 33.0 dB, CC = 1.
Rib-Proj clean
PSNR = 23.0 dB, CC = 0.95
PSNR = 36.5 dB, CC = 1
PSNR = 32.2 dB, CC = 0.99
PSNR = 34.6 dB, CC = 1
PSNR = 23.0 dB, CC = 0.95
PSNR = 34.1 dB, CC = 1
0 0.2 0.4 0.6 0.8 1

Fig. 4. Visual comparison of UVTomo-GAN with different baselines in no noise setting and $L = 2 \times 10^3$. The description of the columns: 1) ground truth image (GT), 2) graph Laplacian tomography (GLT) [24], 3) angular difference estimation [25] + graph Laplacian (MADE+GL), 4) EM with random initialization, 5) Adapted CryoGAN [32], 6) Adapted CryoGAN [32] with uniform p, 7) Ours, UVTomo-GAN with unknown p (jointly recovering c and p). The PSNR and CC between the reconstructed images and the GT are provided.

baselines as *Adapted CryoGAN* and *Adapted CryoGAN with unif. p*.

C. Experimental Results

1) Quality of Reconstructed Image: Figs. 4–5 (and Figs. 9–10 in Appendix F) compare the results of UVTomo-GAN jointly optimizing for c and p against the GT image and the aforementioned baselines for no noise and noisy scenarios. We also include the profiles of the middle vertical line of the reconstructed images against GT in Fig. 6. The results of UVTomo-GAN jointly optimizing for c and p closely resembles the Adapted CryoGAN baseline, both qualitatively and quantitatively. However, with unknown p, the reconstruction is more challenging. Note that, although by assuming p to be uniform (second to last column in Figs. 4–5, Adapted CryoGAN with unif. p baseline) the overall shape of the GT image emerges, the details are not successfully recovered. This highlights the importance of updating p to retrieve details accurately in the reconstruction. A similar observation, although in a different setting is reported in [32], [52].

Furthermore, in the clean case, GLT is able to recover the correct ordering of the viewing angles. However, as the viewing angle distribution is non-uniform, assigning equi-spaced angles to the sorted projections causes a distorted reconstructed image. On the other hand, MADE+GL is able to reconstruct the image accurately.

For SNR $= 3$, while GLT's performance on the Lung and Ellipses images is similar to the clean case, GLT's sorting of the projections for Abdomen and Rib-Proj images is erroneous despite tuning the hyperparameters (see Appendix E). Furthermore, we find the angle differences output by MADE for SNR $= 3$ extremely noisy. This led to an erroneous angular difference estimation and incorrect projection embedding. As MADE+GL failed in reconstructing all images at SNR $= 3$, we excluded the results of this baseline in Fig. 5.

In the presence of noise, we noticed that to obtain better results for EM starting from a random initialization, in the E-step (22), we need to inflate the noise standard deviation σ, otherwise EM can get stuck easily at poor local optima. In our EM experiments, we inflated σ by $\sqrt{2}$ for all datasets.

Fig. 5 (and Fig. 10 in Appendix F) display the effect of noise in the final reconstruction. We observe that the presence of noise makes the reconstruction task more challenging and degrades the reconstruction quality compared to the no noise case. This happens as the critic is having a harder time distinguishing signal from noise given the noisy projections. Overall, in the no noise setting, among baselines with unknown or assumed uniform projection angle distribution, MADE+GL and our method perform the best in terms of PSNR and CC. In the noisy case,

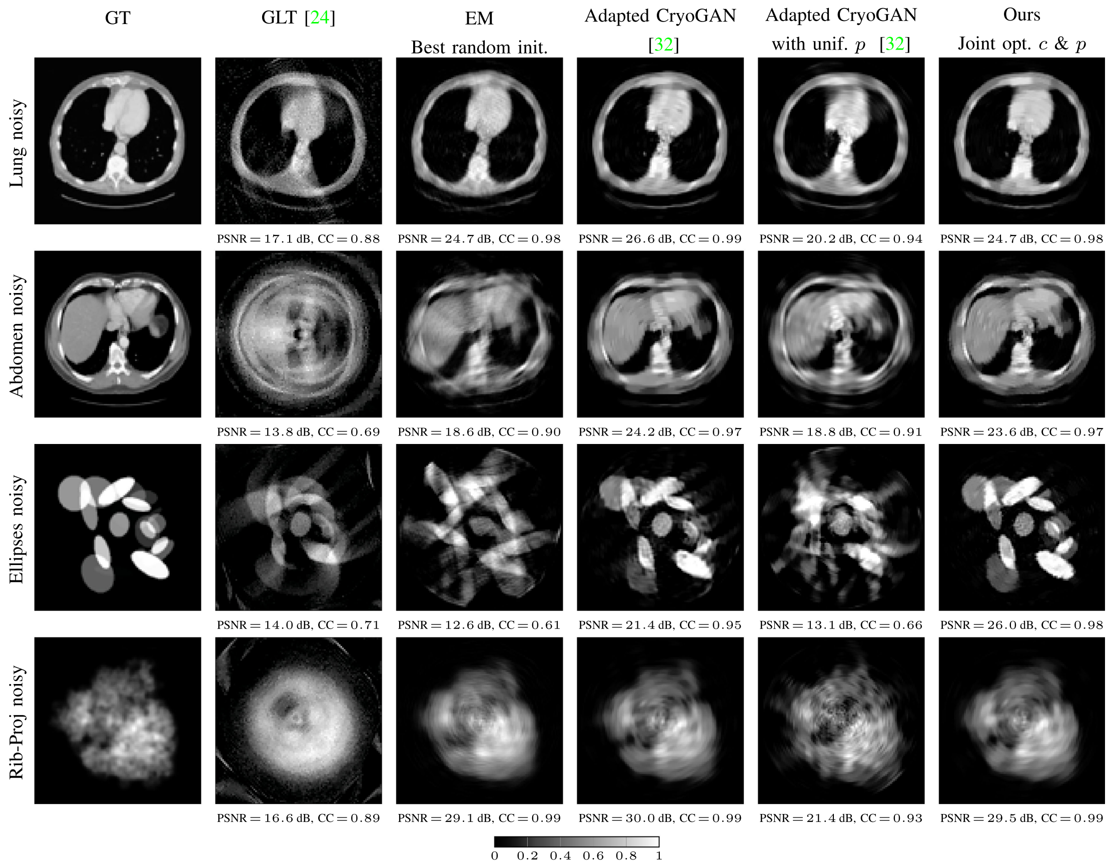
GT
GLT [24]
EM
Best random init.
Adapted CryoGAN
[32]
Adapted CryoGAN
with unif. p [32]
Ours
Joint opt. c & p
Lung noisy
PSNR = 17.1 dB, CC = 0.88
PSNR = 24.7 dB, CC = 0.98
PSNR = 26.6 dB, CC = 0.99
PSNR = 20.2 dB, CC = 0.94
PSNR = 24.7 dB, CC = 0.98
Abdomen noisy
PSNR = 13.8 dB, CC = 0.69
PSNR = 18.6 dB, CC = 0.90
PSNR = 24.2 dB, CC = 0.97
PSNR = 18.8 dB, CC = 0.91
PSNR = 23.6 dB, CC = 0.97
Ellipses noisy
PSNR = 14.0 dB, CC = 0.71
PSNR = 12.6 dB, CC = 0.61
PSNR = 21.4 dB, CC = 0.95
PSNR = 13.1 dB, CC = 0.66
PSNR = 26.0 dB, CC = 0.98
Rib-Proj noisy
PSNR = 16.6 dB, CC = 0.89
PSNR = 29.1 dB, CC = 0.99
PSNR = 30.0 dB, CC = 0.99
PSNR = 21.4 dB, CC = 0.93
PSNR = 29.5 dB, CC = 0.99
0 0.2 0.4 0.6 0.8 1

Fig. 5. Visual comparison of UVTomo-GAN with different baselines in noisy setting, i.e. SNR = 3 and $L = 2 \times 10^4$. The description of the columns: 1) ground truth image (GT), 2) graph Laplacian tomography (GLT) [24], 3) EM with random initialization, 4) Adapted CryoGAN [32], 5) Adapted CryoGAN [32] with uniform p, 6) Ours, UVTomo-GAN with unknown p (jointly recovering c and p). The PSNR and CC between the reconstructed images and the GT are provided.

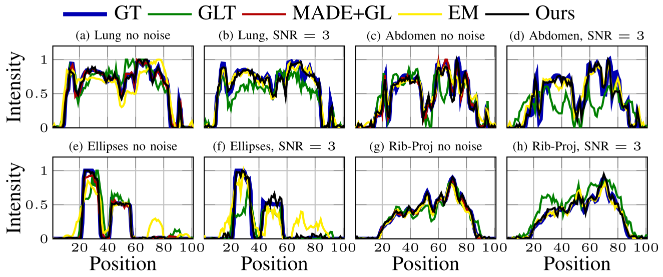

Fig. 6. Comparison between the line profile (middle vertical slice) of GT (blue) versus 1) GLT [24] (green), 2) MADE [25] + GL (red), 3) EM (yellow), 4) UVtomo-GAN jointly optimizing c and p (black).

our approach alongside Adapted CryoGAN with unif. p are the top-performing methods.

Quality of Reconstructed p: Comparison between the GT distribution of the viewing angles and the one recovered by UVTomo-GAN with unknown p is provided in Fig. 7 (and Fig. 11 in Appendix F). Note that the recovered p matches the GT distribution both visually and quantitatively in terms of TV distance. Although, the quality of the recovered PMF in the noisy cases (Fig. 7-(b), (d), (f), (h)) is not as good as the no noise case, it still closely resembles the GT projection distribution. This shows the ability of our approach to recover p accurately under different distributions and noise regimes.

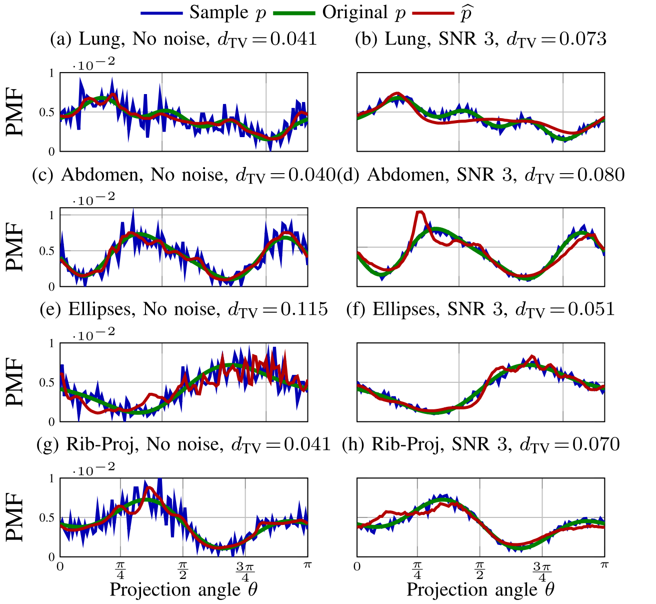

Fig. 7. Comparison between the original GT p (green) used to sample the viewing angles from, the empirical sample distribution of the viewing angles (blue) and the one estimated by our method $\widehat{p}$ (red). In each row, the subplots share the same vertical axis. For no noise settings, d_{TV} is computed between $\widehat{p}$ (red) and original p (green), while for the noisy case, d_{TV} is computed between $\widehat{p}$ (red) and sample estimation of p (blue).

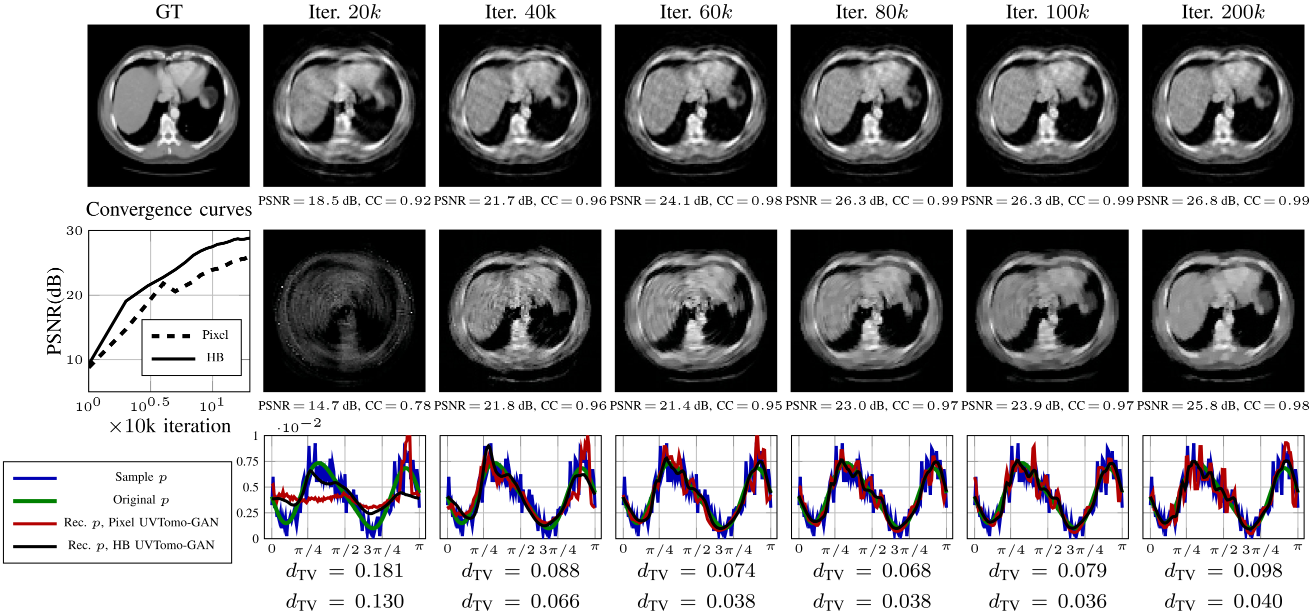

Fig. 8. Convergence of the reconstructed image and the viewing angle distribution throughout different iterations in a no noise setting for 1) UVTomo-GAN with HB image representation (first row), 2) UVTomo-GAN with pixel domain representation (second row). For each iteration, we report PSNR and CC of the reconstructed image and TV distance between the recovered p (red) and GT original p (green) from which the viewing angles are sampled, for both pixel and HB image representation (in second to last and last rows respectively). The sample estimation of p is plotted in blue. The evolution of PSNR throughout training iterations is plotted in second row, first subplot.

Convergence: To evaluate the effect of using HB representation on the convergence, we compare against an experiment with pixel domain representation of the image. We call this baseline *pixel UVTomo-GAN* versus our method *HB UVTomo-GAN*. In this comparison, we use the same dataset, initialization, batch-size, learning rate decay and schedules for both pixel and HB UVTomo-GANs. For HB UVTomo-GAN, to only examine the effect of the representation, we use no TV regularization on the image, i.e. $\gamma_1 = 0$. However, for pixel UVTomo-GAN, to further help with the convergence, we set a small TV regularization weight as 5×10^{-5} and enforce the image to be non-negative by defining it to be the output of a ReLU. For HB UVTomo-GAN, we choose $\alpha_\phi = \alpha_c = 0.008, \alpha_p = 0.0008$ while for pixel UVTomo-GAN, we fine-tuned these parameters as $\alpha_\phi = \alpha_I = 0.01$, $\alpha_p = 0.001$, α_I denoting the learning rate of the image. To implement the projection operator in pixel domain, we use Astra toolbox [53].

In Fig. 8, we show the results of this comparison. While both representations lead to accurate image and p recovery, their convergence behaviours are different. For HB UVTomo-GAN, as we are operating in Hartley domain and the images tend to have larger low-frequency components compared to the high-frequency details, initially the gradients corresponding to lower frequency components are larger, leading to faster updates of $c_{k,q}$s for smaller (k, q)s. This helps in more stable convergence of HB versus pixel UVTomo-GAN.

Note that, for HB UVTomo-GAN, we obtain a reasonable image and PMF at early stages of training, i.e., after 20k-40k iterations (which takes roughly 6-12 minutes). As expected, the image is further refined with more training iterations. In addition, we compare the convergence of our method versus Adapted CryoGAN in Fig. 12 in Appendix G.

VII. Discussion and Future Work

As noted in [24], the angular estimation problem in 2D UVT is more challenging than the 3D problem (such as the cryo-EM single particle reconstruction) when the distribution of the viewing angles is non-uniform. In the 3D problem, the central slice theorem implies that any two central slices share a common line of intersection that can be used to find the unknown imaging directions even when they are not uniformly distributed. However, in the 2D problem, since central slices all intersect at the origin, we can't directly use the corresponding geometric relation found in 3D. In this paper, we show that even though the distribution of the viewing angles is non-uniform, our adversarial learning based approach can simultaneously recover both the underlying object and the distribution of the viewing angles and provide theoretical justification for this.

Our framework can handle the uncertainty of various latent variables, such as rotations and translations, i.e. elements in $SE(2)$ group. Such latent variables also encode the rigid motion of an object in 2D. Furthermore, our approach can be extended to various imaging inverse problems, including 3D UVT. This can be explained through Fig. 1, where for a general imaging inverse problem, H_θ and θ represent the forward operator and the underlying random (multi-dimensional) latent variables with unknown distribution. These extensions are beyond the scope of this paper and we defer their study to future work.

VIII. Conclusion

In this paper, we present an adversarial learning approach for the 2D unknown view tomography problem. Since the viewing angles and their distribution are not known a-priori, we simultaneously recover the unknown image and the distribution of the viewing angles via a distribution matching formulation solved through a min-max game between a critic and a generator. To improve the computational efficiency and regularize the image, we employ a Fourier related representation of the image with truncated Hartley-Bessel expansion. For the GAN training, we show that the original loss function at the generator side is non-differentiable with respect to the viewing angle distribution. Thus, we use the Gumbel-Softmax approximation of samples from discrete distributions to allow the gradient update of the viewing angle distribution. Our analysis demonstrates that asymptotically unique recovery of the image and viewing angle distribution is achieved. Moreover, our simulation results show that our method outperforms the state-of-the-art methods in recovering the images from noisy projections.

Appendix

A. Computational Cost of UVTomo-GAN

Cost of Projection Generation: To generate $N_\theta = O(m)$ projection templates following (9), we first compute the inner summation over q, i.e.

$$f_k(\xi_j) = \sum_{q=1}^{p_k} c_{k,q} J_{k,q}(\xi_j). \tag{33}$$

On the radial line, we have $O(m)$ equally spaced points ξ_j. Given that $K_{\max} = O(m)$ and $p_k = O(m)$, computing $f_k(\xi_j)$, $\forall k, j$ requires $O(m^3)$ computations.

Next, using $f_k(\xi_j)$ we compute the outer sum in (9) with respect to k for N_θ viewing angles. A naive matrix multiplication implementation for this step leads to $O(m^3)$ cost (multiplying two matrices of size $O(m) \times O(m)$). This can be further reduced using FFT to $O(m^2 \log m)$. Finally, the total cost of generating N_θ projections using (9) is $O(m^3)$.

B. Proof of Theorem 1

First we prove:

$$P^{\text{clean}}_{f,p_f} = P^{\text{clean}}_{g,p_g} \Rightarrow [f] = [g]. \tag{34}$$

From $P^{\text{clean}}_{f,p_f} = P^{\text{clean}}_{g,p_g}$, it is implied that the support of the two distributions are the same. This means that f and g have the same projection set. In other words, $\{\mathcal{P}_{\theta_i} f\}_{i=1}^{N_\theta} = \{\mathcal{P}_{\widehat{\theta}_j} g\}_{j=1}^{N_\theta}$ where $\widehat{\boldsymbol{\theta}} = \{\widehat{\theta}_j\}_{j=1}^{N_\theta}$ can be a shuffled version of $\boldsymbol{\theta} = \{\theta_i\}_{i=1}^{N_\theta}$. Intuitively, one can imagine two objects f and g which have the same projections, however the order of the viewing angles of f can be a shuffled version of the viewing angles for g. Now the question that arises is: Given the class of functions f and g belong to, is it possible to have two distinct objects that produce identical projection sets?

This question is related to the feasibility of unique angle recovery in unknown view tomography, comprehensively studied in [20], [21]. Based on our discussions so far, we seek to prove the following:

$$\{\mathcal{P}_{\theta_i} f\}_{i=1}^{N_\theta} = \{\mathcal{P}_{\widehat{\theta}_j} g\}_{j=1}^{N_\theta} \Rightarrow [f] = [g]. \tag{35}$$

In (35), the LHS implies that f and g have the same set of projections, in other words we have: $\forall \gamma \in \{\mathcal{P}_{\theta_i} f\}_{i=1}^{N_\theta}, \gamma \in \{\mathcal{P}_{\widehat{\theta}_j} g\}_{j=1}^{N_\theta}$ and $\forall \gamma' \in \{\mathcal{P}_{\widehat{\theta}_j} g\}_{j=1}^{N_\theta}, \gamma' \in \{\mathcal{P}_{\theta_i} f\}_{i=1}^{N_\theta}$. To prove the above, we borrow the definitions and various theoretical results in [20]. Helgasson–Ludwig (HL) consistency conditions [54] link the geometric moments of a 2D object to its projections. Let v and μ define the geometric moment of the image f and its projection as:

$$v_{i,k}(f) = \int_{-1}^{1}\int_{-1}^{1} x^i y^k f(x,y)\,dx\,dy \tag{36}$$

$$\mu_d(\theta; f) = \int_{-1}^{1} x^d \{\mathcal{P}_\theta f\}(x)\,dx. \tag{37}$$

Object moments of order d are the ones that satisfy $i + k = d$. Let $\mathbf{v}(f)$, denote the set of geometric moments of order $d \in D$ for object f. Given the object moments $\mathbf{v}(f)$, we construct a family of trigonometric polynomials as:

$$\mathcal{Q}_d(\theta; \mathbf{v}(f)) = \sum_{r=0}^{d} \binom{d}{r} v_{r,d-r}(f)\,(\cos\theta)^r (\sin\theta)^{d-r}. \tag{38}$$

Given the definition (38), we state the HL conditions as:

$$\mathcal{Q}_d(\theta; \mathbf{v}(f)) = \mu_d(\theta; f). \tag{39}$$

We have defined equivalence for 2D images before. If two images are equivalent, then they are related through a rotation and reflection. Similarly, we can define equivalence on the viewing angles. Assume two vectors of viewing angles of length N_θ, $\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}} \in [-\pi, \pi]^{N_\theta}$. $\boldsymbol{\theta}$ is said to be equivalent to $\widehat{\boldsymbol{\theta}}$, i.e. $\boldsymbol{\theta} \sim \widehat{\boldsymbol{\theta}}$, if $\exists \eta \in \{-1, 1\}$ and $\alpha \in [-\pi, \pi]$ such that $\widehat{\theta}_i = \eta\theta_i + \alpha + 2\pi n_i$, for $n_i \in \mathbb{Z}$.

As the projection set for f and g objects are the same (based on (35)), we conclude $\forall \theta$, $\exists \widehat{\theta}$ such that:

$$\mu_d(\theta; f) = \mu_d(\widehat{\theta}; g). \tag{40}$$

After invoking HL conditions (39) for object f on the RHS of (40) we get:

$$\mathcal{Q}_d(\theta; \mathbf{v}(f)) = \mu_d(\widehat{\theta}; g), \quad \forall d \geq 0. \tag{41}$$

Note that, we have narrowed down the identical projection sets for f and g to (41). Now we restate our question as: what is the relationship between θ and $\widehat{\theta}$?

To find the answer to this question, we first limit the set of moment orders to $d \in D = \{1, 2\}$ (as (41) holds for $\forall d \geq 0$, we can simply do this). Note that for $\theta \in [0, 2\pi)$, the projections corresponding to $\theta \in [\pi, 2\pi)$ are a flipped version of projections associated to $\theta \in [0, \pi)$ and do not constitute new information [20]. Thus, in [20], the authors limit their analysis to the projections that are π-distinct, i.e., there are no two angles that are different by a factor of π. Following the same lines, given the projection sets corresponding to $\theta, \widehat{\theta} \in [0, 2\pi)$, we select a π-distinct projection subset by choosing a set of projections that have positive (or negative) 1st order geometric moment. We now invoke Corollary 5 of Theorem 9 in [20]. We restate this corollary in the following.

Corollary 1 (Corollary 5 of Theorem 9 [20])*:* Suppose θ is a set of π-distinct view angles and $N_\theta > 8$. Suppose $\mathbf{v}$ satisfies the following condition: $\nexists \beta, \gamma \in \mathbb{R}$ such that,

$$\mathcal{Q}_2(\theta; \mathbf{v}) = \beta \left(\mathcal{Q}_1(\theta; \mathbf{v})\right)^2 + \gamma, \, \forall \theta \in [0, \pi] \tag{42}$$

or equivalently,

$$\det \begin{bmatrix} v_{1,0}^2 & v_{2,0} & 1 \\ 2 v_{1,0} v_{0,1} & v_{1,1} & 0 \\ v_{0,1}^2 & v_{0,2} & 1 \end{bmatrix} \neq 0. \tag{43}$$

If $\theta \notin \mathrm{UAS}(\mathbf{v})$ with UAS (unidentifiable angle set) defined as:

$$\mathrm{UAS}(\mathbf{v}) = \left\{ \arg\left(\sqrt{\frac{-c_1^*}{c_1}}\right), \arg\left(-\sqrt{\frac{-c_1^*}{c_1}}\right) \right\} \tag{44}$$

where,

$$c_1 = \frac{1}{2}(v_{1,0} - i \, v_{0,1}) \tag{45}$$

then, the only view angles $\widehat{\theta}$ that produce the same projection moments of order $D = \{1, 2\}$ are equivalent to θ. This implies that $\theta \sim \widehat{\theta}$. ■

Adhering to Corollary 1, if $\mathbf{v}(f)$ satisfies the conditions in (42) or (43), then for $d \in \{1, 2\}$, the only viewing angles $\widehat{\theta}$ for which (41) holds are equivalent to θ and thus $\theta \sim \widehat{\theta}$. On the other hand, based on Corollary 1, the viewing angles recovered for f, i.e. θ, are equivalent to the GT viewing angles $\check{\theta}$ used for generating the projections of f, i.e $\theta \sim \check{\theta}$. Based on the transitivity property of equivalence relation, this leads to $\widehat{\theta} \sim \check{\theta}$

Given $\theta \sim \widehat{\theta} \sim \check{\theta}$ and the fact that the projection sets corresponding to the objects f and g are identical, the objects $\widehat{f}$ and $\widehat{g}$ reconstructed from the projection sets and viewing angles would also be the same (up to a rotation and reflection), i.e. $[\widehat{f}] = [\widehat{g}]$. We now link the reconstructed objects and their ground truths.

If we have sufficiently large N_θ, we can directly recover HB expansion coefficients c by solving a set of linear equations linking the projections to the HB expansion coefficients. Given the HB expansion coefficients, we have a continuous representation of the image as defined in (11). This leads to $\widehat{f} = f$ and $\widehat{g} = g$ and finally concludes $[f] = [g]$.

As $[f] = [g]$, $\exists \, \Gamma \in \mathrm{O}(2)$ such that $g = \Gamma f$. $P^{\text{clean}}_{f, p_f} = P^{\text{clean}}_{g, p_g}$ implies the TV distance between the two probability distributions is zero, i.e.

$$TV(P^{\text{clean}}_{f, p_f}, P^{\text{clean}}_{\Gamma f, p_g}) = 0. \tag{46}$$

Invoking Lemma 1 (stated in Appendix D), we know $P^{\text{clean}}_{\Gamma f, p_g} = P^{\text{clean}}_{f, \Gamma^{-1} p_g}$, therefore (46) becomes,

$$\begin{aligned} TV(P^{\text{clean}}_{f, p_f}, P^{\text{clean}}_{f, \Gamma^{-1} p_g}) &= TV(p_f, \Gamma^{-1} p_g) \\ &= \frac{1}{2} \| p_f - \Gamma^{-1} p_g \|_1. \end{aligned} \tag{47}$$

Following (46), the LHS of (47) is 0. Thus, based on the non-negativity property of $\|.\|_1$ norm, we have,

$$p_f = \Gamma^{-1} p_g \Rightarrow p_g = \Gamma p_f \tag{48}$$

implying $[p_f] = [p_g]$. ■

C. Proof of Theorem 2

Our proof follows closely the proof of Theorem 1 in [32]. We first show that,

$$P^{\text{noisy}}_{f, p} = P^{\text{noisy}}_{\widetilde{f}, \widetilde{p}} \Rightarrow P^{\text{clean}}_{f, p} = P^{\text{clean}}_{\widetilde{f}, \widetilde{p}}. \tag{49}$$

According to the forward model (3), we have $\zeta = \mathcal{P}_\theta f + \varepsilon$ where $\varepsilon[n] \sim q_\epsilon$ an IID additive noise which is independent of f and $\theta \sim p$. Note that we are considering a general model for the noise and not confining it to be a Gaussian. As ε is independent of the image and viewing angles, we have:

$$P^{\text{noisy}}_{f, p} = P^{\text{clean}}_{f, p} * q_\varepsilon \tag{50}$$

In Fourier domain, (50) becomes:

$$\mathcal{F}\{P^{\text{noisy}}_{f, p}\} = \mathcal{F}\{P^{\text{clean}}_{f, p}\} \mathcal{F}\{q_\varepsilon\}. \tag{51}$$

We have assumed ε to have full support in Fourier domain, therefore we can divide both sides of (51) by $\mathcal{F}\{p_\varepsilon\}$. Therefore given $\mathcal{F}\{P^{\text{noisy}}_{f, p}\}$, we have $\mathcal{F}\{P^{\text{clean}}_{f, p}\}$ and (49) is proved. Now, we show:

$$P^{\text{clean}}_{f, p} = P^{\text{clean}}_{\widetilde{f}, \widetilde{p}} \Rightarrow \widetilde{f} = \Gamma f \text{ and } \widetilde{p} = \Gamma p \tag{52}$$

for a unique $\Gamma \in \mathrm{O}(2)$. To prove (52), we invoke Theorem 1. Theorem 1 states that if the two images f and $\widetilde{f}$ have the same distribution of the clean projections, then the objects and their associated projection angle distributions are equivalent up to a rotation and reflection. This confirms $[f] = [\widetilde{f}]$, and $[p] = [\widetilde{p}]$, i.e. $\widetilde{f} = \Gamma f$ and $\widetilde{p} = \Gamma p$, for a $\Gamma \in \mathrm{O}(2)$. ■

D. Lemma 1

Assume $f \in \mathcal{L}_1(\mathbb{B}_2) \cap \mathcal{L}_2(\mathbb{B}_2)$, viewing angles θ are distributed following p, i.e. $\theta \sim p$ and $\Gamma \in \mathrm{O}(2)$. Then,

$$P^{\text{clean}}_{f, \Gamma^{-1} p} = P^{\text{clean}}_{\Gamma f, p} \tag{53}$$

Proof: For a given (f, p_f), if $\gamma \in \mathrm{O}(2)$ is applied to both f and p, then the induced probability distribution of the projections would be the same, i.e. $P^{\text{clean}}_{f, p} = P^{\text{clean}}_{\Gamma f, \Gamma p}$. After changing $p' = \Gamma p$, we have $P^{\text{clean}}_{f, \Gamma^{-1} p'} = P^{\text{clean}}_{\Gamma f, p'}$, thus concluding the proof.

E. Details on Baselines

GLT [24]: For this baseline, a graph is constructed based on the pairwise distances of the compressed denoised projections. The tunable parameters in GLT are 1) number of nearest neighbors (NN) and, 2) Jaccard index threshold (β). The choice of NN affects the connectivity of the constructed graph (before denoising). On the other hand, Jaccard index thresholding reduces the shortcut edges in the graph. For the clean case, we choose $\mathrm{NN} = 111$ and $\beta = 0.41$ for all images except for the Shepp-Logan for which we set $\mathrm{NN} = 68$ and $\beta = 0.03$. In the noisy case, we set $\mathrm{NN} = 111$ and $\beta = 0.21$, $\beta = 0.31$

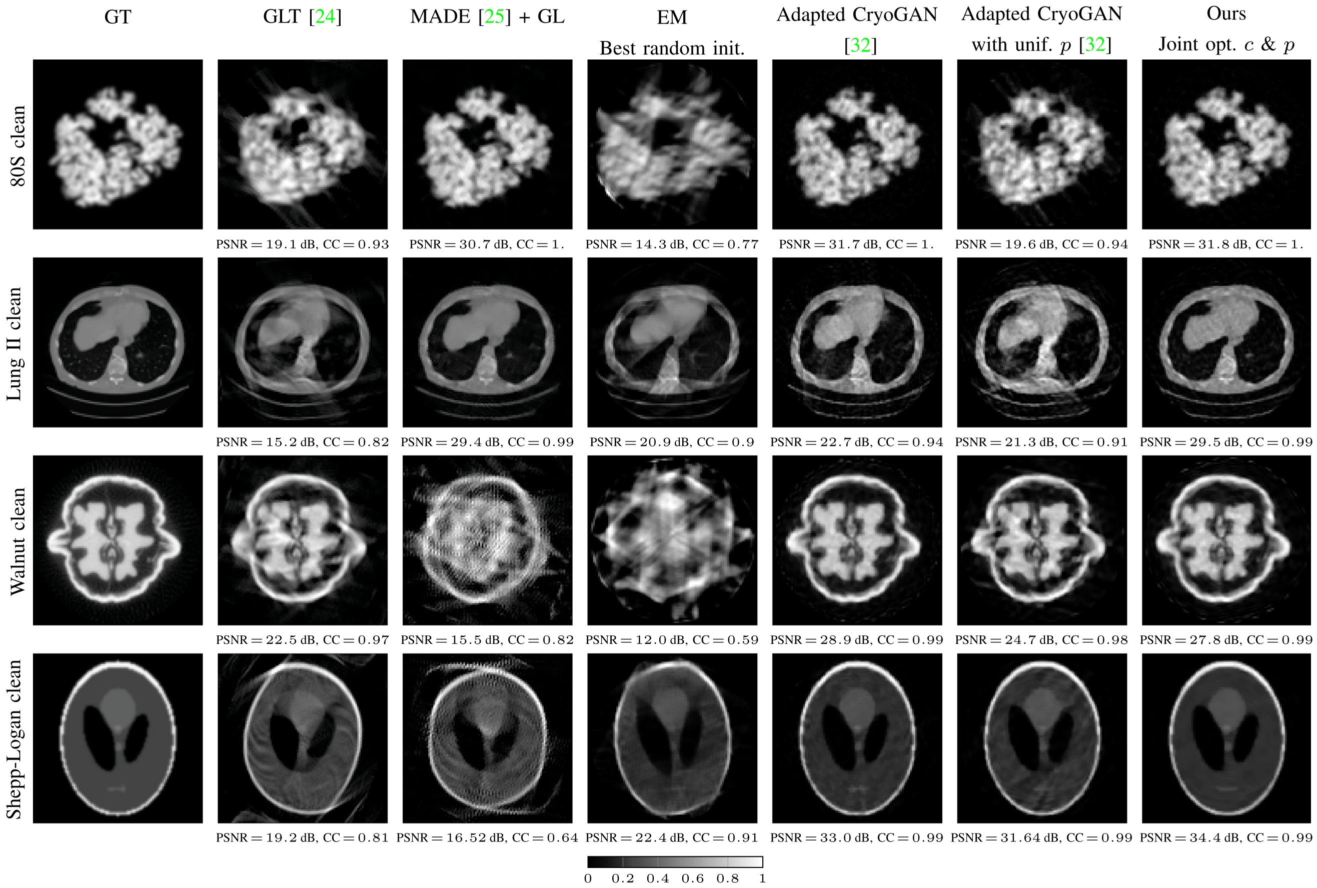
GT
GLT [24]
MADE [25] + GL
EM
Best random init.
Adapted CryoGAN
[32]
Adapted CryoGAN
with unif. p [32]
Ours
Joint opt. c & p
80S clean
Lung II clean
Walnut clean
Shepp-Logan clean
PSNR = 19.1 dB, CC = 0.93
PSNR = 30.7 dB, CC = 1.
PSNR = 14.3 dB, CC = 0.77
PSNR = 31.7 dB, CC = 1.
PSNR = 19.6 dB, CC = 0.94
PSNR = 31.8 dB, CC = 1.
PSNR = 15.2 dB, CC = 0.82
PSNR = 29.4 dB, CC = 0.99
PSNR = 20.9 dB, CC = 0.9
PSNR = 22.7 dB, CC = 0.94
PSNR = 21.3 dB, CC = 0.91
PSNR = 29.5 dB, CC = 0.99
PSNR = 22.5 dB, CC = 0.97
PSNR = 15.5 dB, CC = 0.82
PSNR = 12.0 dB, CC = 0.59
PSNR = 28.9 dB, CC = 0.99
PSNR = 24.7 dB, CC = 0.98
PSNR = 27.8 dB, CC = 0.99
PSNR = 19.2 dB, CC = 0.81
PSNR = 16.52 dB, CC = 0.64
PSNR = 22.4 dB, CC = 0.91
PSNR = 33.0 dB, CC = 0.99
PSNR = 31.64 dB, CC = 0.99
PSNR = 34.4 dB, CC = 0.99
0
0.2
0.4
0.6
0.8
1

Fig. 9. Visual comparison of UVTomo-GAN with different baselines in no noise setting and $L = 2 \times 10^3$. For description of each column and the evaluation metrics refer to Fig. 4 and Sections VI-A, VI-B.

and $\beta = 0.41$ for Lung, Abdomen and Rib-Proj images, respectively. For the rest of the images, we set the parameters as, Ellipse: $\mathrm{NN} = 50$, $\beta = 0.25$, 80S: $\mathrm{NN} = 1111$, $\beta = 0.3$, Lung II: $\mathrm{NN} = 1111$, $\beta = 0.3$, Walnut: $\mathrm{NN} = 50$, $\beta = 0.25$, Shepp-Logan: $\mathrm{NN} = 222$, $\beta = 0.11$.

MADE [25] + GL: To find the angular differences between any two projections we use MADE. The tunable parameters for MADE are similar to GLT. For the Lung, Abdomen and Walnut images, we set the number of nearest neighbors $\mathrm{NN} = 70$ while $\mathrm{NN} = 90$ for the rest. For all the images, we set $\beta = 0.1$. After obtaining the angular differences between the neighbor projections, through a shortest path algorithm, i.e. Djikstra, the absolute angle differences between any two projections are obtained. Next, we construct a weight matrix E based on the angle differences from MADE as:

$$E(i,j) = \begin{cases} e^{-\frac{|\theta_i - \theta_j|^2}{\epsilon}}, & |\theta_i - \theta_j| \leq 5^\circ \\ 0. & \text{o.w.} \end{cases} \tag{54}$$

where θ_i denotes the angle corresponding to the i-the projection. In our experiments, we set $\epsilon = 20$. Next, we normalize E similar to [24] and perform eigenvalue decomposition. In the clean case, the top two non-trivial eigenvectors of the normalized matrix form the embedding of the projections which is a circle. The angle of the i-th projection embedded on the circle is assigned as θ_i. Based on the assigned viewing angles, the image is reconstructed. For both GLT and MADE+GL baselines, after the estimation of the projection ordering and angles, we reconstruct the image via a TV regularized optimization solved by ADMM [55] using GlobalBioIm library [56].

F. Additional Numerical Results

In this section, we provide additional results on four other images described as: 1) *80S:* We generated the 3D map of 80S Ribosome [57] using its protein sequence in Chimera [49] and took a central slice of the 80S Ribosome molecule. 2) *Lung II:* A lung CT scan [12]. 3) *Walnut:* Tomographic X-ray reconstruction of a walnut. We used the projection data and code provided in [58][1] to generate this image. 4) *Shepp-Logan* phantom. We generated the projection datasets for both clean and noisy cases as described in section VI.

Figs. 9–10 compares the results of UVTomo-GAN jointly optimizing c and p versus several benchmarks, in clean and noisy ($\mathrm{SNR} = 3$) settings. Furthermore, in Fig. 11 we evaluate the performance of UVTomo-GAN in terms of the quality of the recovered projection angle distribution p. We notice that for the Walnut and Shepp-Logan images, in the noisy case, as shown in

[1] http://www.fips.fi/dataset.php

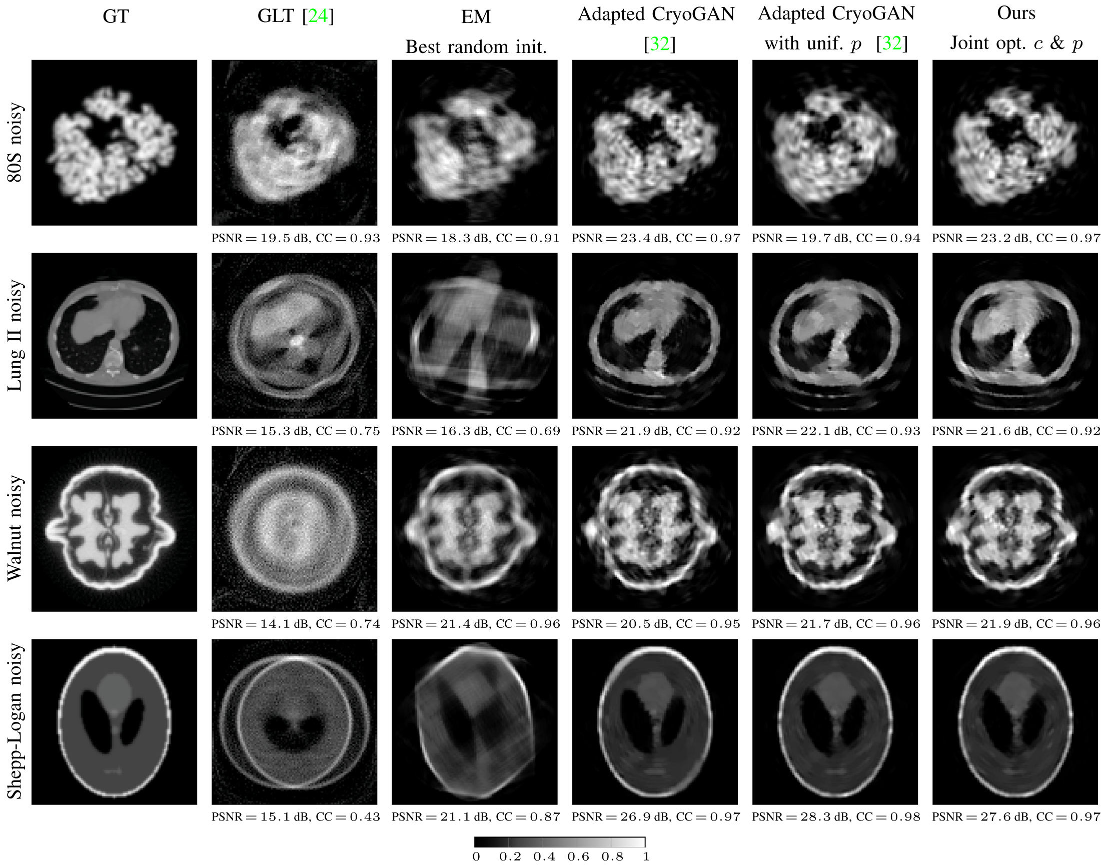
GT
GLT [24]
EM
Best random init.
Adapted CryoGAN
[32]
Adapted CryoGAN
with unif. p [32]
Ours
Joint opt. c & p
80S noisy
PSNR = 19.5 dB, CC = 0.93
PSNR = 18.3 dB, CC = 0.91
PSNR = 23.4 dB, CC = 0.97
PSNR = 19.7 dB, CC = 0.94
PSNR = 23.2 dB, CC = 0.97
Lung II noisy
PSNR = 15.3 dB, CC = 0.75
PSNR = 16.3 dB, CC = 0.69
PSNR = 21.9 dB, CC = 0.92
PSNR = 22.1 dB, CC = 0.93
PSNR = 21.6 dB, CC = 0.92
Walnut noisy
PSNR = 14.1 dB, CC = 0.74
PSNR = 21.4 dB, CC = 0.96
PSNR = 20.5 dB, CC = 0.95
PSNR = 21.7 dB, CC = 0.96
PSNR = 21.9 dB, CC = 0.96
Shepp-Logan noisy
PSNR = 15.1 dB, CC = 0.43
PSNR = 21.1 dB, CC = 0.87
PSNR = 26.9 dB, CC = 0.97
PSNR = 28.3 dB, CC = 0.98
PSNR = 27.6 dB, CC = 0.97
0 0.2 0.4 0.6 0.8 1

Fig. 10. Visual comparison of UVTomo-GAN with different baselines in no noise setting and $L = 2 \times 10^4$. For the description of each column and the evaluation metrics, please refer to Fig. 5, Sections VI-A, VI-B.

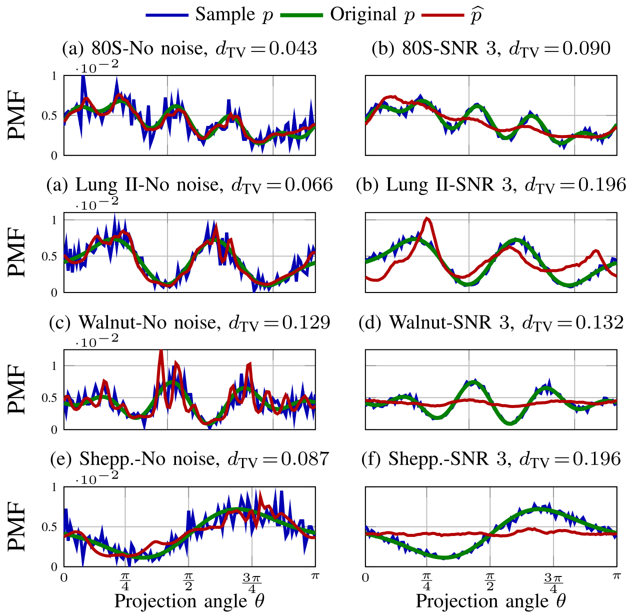

Fig. 11. Comparison between the original GT p (green) used to sample the viewing angles from, the empirical sample distribution of the viewing angles (blue) and the one estimated by our method $\widehat{p}$ (red). For more details on the computation of d_{TV}, refer to Section VI-A and Fig. 7.

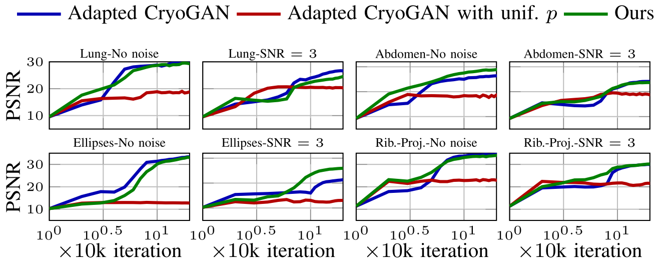

Fig. 12. Convergence results for the no noise and noisy (SNR = 3) experiments. The setting of the experiments are the same as the ones in Figs. 4–5. We compare UVTomo-GAN jointly optimizing for c and p (green) versus Adapted CryoGAN [32] (blue) and Adapted CryoGAN [32] with unif. p (red). Vertical axis shows the PSNR in dB and the horizontal axis is the training iteration number. The subplots in each row share the same vertical and horizontal axis.

Figs. 10–11, the reconstruction is less sensitive to the quality of the estimated p.

G. Convergence Results

We exhibit the convergence curves in terms of PSNR versus training iteration for no noise and noisy experiments in Fig. 12.

To obtain this curve, at each iteration, we align the reconstructions with the GT. We compare the convergence of UVTomo-GAN versus Adapted CryoGAN and Adapted CryoGAN with unif. p baselines.

For Adapted CryoGAN with unif. p baseline, after a certain number of iterations, we see no improvement in the reconstructed image. This is attributed to having an inaccurate PMF which hinders the correct distribution matching of synthetic and real measurements. Thus, the high frequency details in the final reconstructed image do not appear correctly (as also seen in Figs. 4–5, 9–10). This once again indicates the importance of recovering p to have high quality reconstructions.

Acknowledgment

The authors would like to thank Prof. Amit Singer and Prof. Hau-Tieng Wu for their helpful suggestions and sharing the code used as the GLT baseline.

References

[1] J. Frank, *Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State*. London, U.K.: Oxford Univ. Press, 2006.

[2] H. Stark, J. Woods, I. Paul, and R. Hingorani, "Direct fourier reconstruction in computer tomography," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 29, no. 2, pp. 237–245, Apr. 1981.

[3] 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. 4777–4807, Aug. 2008.

[4] S. Niu et al., "Sparse-view x-ray CT reconstruction via total generalized variation regularization," *Phys. Med. Biol.*, vol. 59, no. 12, pp. 2997–3017, May 2014.

[5] H. Zhang, J. Wang, D. Zeng, X. Tao, and J. Ma, "Regularization strategies in statistical image reconstruction of low-dose x-ray CT: A. review," *Med. Phys.*, vol. 45, no. 10, pp. e886–e907, 2018.

[6] C. Gong and L. Zeng, "Adaptive iterative reconstruction based on relative total variation for low-intensity computed tomography," *Signal Process.*, vol. 165, pp. 149–162, 2019.

[7] B. Zhu, J. Z. Liu, S. F. Cauley, B. R. Rosen, and M. S. Rosen, "Image reconstruction by domain-transform manifold learning," *Nature*, vol. 555, no. 7697, pp. 487–492, 2018.

[8] Y. Ge et al., "ADAPTIVE-NET: Deep computed tomography reconstruction network with analytical domain transformation knowledge," *Quantitative Imag. Med. Surg.*, vol. 10, no. 2, pp. 415–427, 2020.

[9] J. Adler and O. Ozan, "Solving ill-posed inverse problems using iterative deep neural networks," *Inverse Problems*, vol. 33, Nov. 2017, Art. no. 124007.

[10] J. Adler and O. Öktem, "Learned primal-dual reconstruction," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1322–1332, Jun. 2018.

[11] X. Yang et al., "Tomographic reconstruction with a generative adversarial network," *J. Synchrotron Radiat.*, vol. 27, no. 2, pp. 486–493, Mar. 2020.

[12] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4509–4522, Sep. 2017.

[13] H. Chen et al., "Low-dose CT with a residual encoder-decoder convolutional neural network," *IEEE Trans. Med. Imag.*, vol. 36, no. 12, pp. 2524–2535, Dec. 2017.

[14] H. Shan et al., "Competitive performance of a modularized deep neural network compared to commercial algorithms for low-dose CT image reconstruction," *Nature Mach. Intell.*, vol. 1, no. 6, pp. 269–276, 2019.

[15] Y. Han and J. C. Ye, "Framing U-Net via deep convolutional framelets: Application to sparse-view CT," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1418–1429, Jun. 2018.

[16] E. Kang, W. Chang, J. Yoo, and J. C. Ye, "Deep convolutional framelet denosing for low-dose CT via wavelet residual network," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1358–1369, Jun. 2018.

[17] Q. Yang et al., "Low-dose CT image denoising using a generative adversarial network with wasserstein distance and perceptual loss," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1348–1357, Jun. 2018.

[18] S. O. Lunz Öktem and C.-B. Schönlieb, "Adversarial Regularizers in Inverse Problems," in *Advances in Neural Information Processing Systems*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds., vol. 31. Red Hook, NY, USA: Curran Assoc., Inc., 2018.

[19] S. Mukherjee, M. O. Carioni Öktem, and C.-B. Schönlieb, "End-to-end reconstruction meets data-driven regularization for inverse problems," in *Advances in Neural Information Processing Systems*, M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, Eds., vol. 34. Red Hook, NY, USA: Curran Assoc., Inc., 2021, pp. 21413–21425.

[20] S. Basu and Y. Bresler, "Uniqueness of tomography with unknown view angles," *IEEE Trans. Image Process.*, vol. 9, no. 6, pp. 1094–1106, Jun. 2000.

[21] S. Basu and Y. Bresler, "Feasibility of tomography with unknown view angles," *IEEE Trans. Image Process.*, vol. 9, no. 6, pp. 1107–1122, Jun. 2000.

[22] Y. Fang, S. Murugappan, and K. Ramani, "Estimating view parameters from random projections for tomography using spherical MDS," *BMC Med. Imag.*, vol. 10, no. 1, 2010, Art. no. 12.

[23] R. R. Coifman, Y. Shkolnisky, F. J. Sigworth, and A. Singer, "Graph Laplacian tomography from unknown random projections," *IEEE Trans. Image Process.*, vol. 17, no. 10, pp. 1891–1899, Oct. 2008.

[24] A. Singer and H.-T. Wu, "Two-dimensional tomography from noisy projections taken at unknown random directions," *SIAM J. Imag. Sci.*, vol. 6, no. 1, pp. 136–175, Jan. 2013.

[25] M. S. Phan, É. Baudrier, L. Mazo, and M. Tajine, "Moment-based angular difference estimation between two tomographic projections in 2D and 3D," *J. Math. Imag. Vis.*, vol. 57, no. 2, pp. 164–182, 2017.

[26] B. B. Cheikh, E. Baudrier, and G. Frey, "A tomographical reconstruction method from unknown direction projections for 2D gray-level images," *Pattern Recognit. Lett.*, vol. 86, pp. 49–55, 2017.

[27] M. Zehni, S. Huang, I. Dokmanić, and Z. Zhao, "Geometric invariants for sparse unknown view tomography," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2019, pp. 5027–5031.

[28] M. Zehni, S. Huang, I. Dokmanić, and Z. Zhao, "3D unknown view tomography via rotation invariants," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2020, pp. 1449–1453.

[29] E. Levin, T. Bendory, N. Boumal, J. Kileel, and A. Singer, "3D Ab initio modeling in Cryo-EM by autocorrelation analysis," in *Proc. IEEE 15th Int. Symp. Biomed. Imag.*, 2018, pp. 1569–1573.

[30] L. Wang and Z. Zhao, "Two-dimensional tomography from noisy projection tilt series taken at unknown view angles with non-uniform distribution," in *Proc. IEEE Int. Conf. Image Process.*, 2019, pp. 1242–1246.

[31] I. Goodfellow et al., "Generative Adversarial Nets," in *Advances in Neural Information Processing Systems*, vol. 27. Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds. Red Hook, NY, USA: Curran Assoc., Inc., 2014, pp. 2672–2680.

[32] H. Gupta, M. T. McCann, L. Donati, and M. Unser, "CryoGAN: A. new reconstruction paradigm for single-particle Cryo-EM via deep adversarial learning," *IEEE Trans. Comput. Imag.*, vol. 7, pp. 759–774, 2021.

[33] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proc. 34th Int. Conf. Mach. Learn.*, 2017, vol. 70, pp. 214–223.

[34] E. Jang, S. Gu, and B. Poole, "Categorical reparameterization with Gumbel-softmax," in *Proc. Int. Conf. Learn. Representations*, 2017.

[35] M. Zehni and Z. Zhao, "MSR-GAN: Multi-segment reconstruction via adversarial learning," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2021, pp. 5115–5119.

[36] M. Zehni and Z. Zhao, "UVTOMO-GAN: An adversarial learning based approach for unknown view X-ray tomographic reconstruction," in *Proc. IEEE 18th Int. Symp. Biomed. Imag.*, 2021, pp. 1812–1816.

[37] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved training of Wasserstein GANs," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 5769–5779.

[38] L. Greengard and J.-Y. Lee, "Accelerating the nonuniform fast fourier transform," *SIAM Rev.*, vol. 46, no. 3, pp. 443–454, 2004.

[39] Z. Zhao, Y. Shkolnisky, and A. Singer, "Fast steerable principal component analysis," *IEEE Trans. Comput. Imag.*, vol. 2, no. 1, pp. 1–12, Mar. 2016.

[40] Z. Zhao and A. Singer, "Fourier bessel rotational invariant eigenimages," *J. Opt. Soc. Amer. A*, vol. 30, no. 5, pp. 871–877, May 2013.

[41] A. Klug and R. Crowther, "Three-dimensional image reconstruction from the viewpoint of information theory," *Nature*, vol. 238, no. 5365, pp. 435–440, 1972.

[42] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, “Spectral normalization for generative adversarial networks,” in *Proc. Int. Conf. Learn. Representations*, 2018.

[43] A. Punjani, M. A. Brubaker, and D. J. Fleet, “Building proteins in a day: Efficient 3D molecular structure estimation with electron cryomicroscopy,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 706–718, Apr. 2017.

[44] A. Barnett, L. Greengard, A. Pataki, and M. Spivak, “Rapid solution of the Cryo-EM reconstruction problem by frequency marching,” *SIAM J. Imag. Sci.*, vol. 10, pp. 1170–1195, Oct. 2016.

[45] J. Shewchuk, “An introduction to the conjugate gradient method without the agonizing pain,” Pittsburgh, PA, USA, Tech. Rep. 1994.

[46] S. H. Scheres, “RELION: Implementation of a bayesian approach to Cryo-EM structure determination,” *J. Struct. Biol.*, vol. 180, no. 3, pp. 519–530, 2012.

[47] T. R. Moen et al., “Low-dose CT image and projection dataset,” *Med. Phys.*, vol. 48, no. 2, pp. 902–911, 2021.

[48] B. Beckert et al., “Structure of a hibernating 100S ribosome reveals an inactive conformation of the ribosomal protein S1,” *Nature Microbiol.*, vol. 3, no. 10, pp. 1115–1121, 2018.

[49] E. F. Pettersen et al., “UCSF chimera—A visualization system for exploratory research and analysis,” *J. Comput. Chem.*, vol. 25, no. 13, pp. 1605–1612, Oct. 2004.

[50] A. Averbuch, R. Coifman, D. Donoho, M. Elad, and M. Israeli, “Fast and accurate polar fourier transform,” *Appl. Comput. Harmon. Anal.*, vol. 21, no. 2, pp. 145–167, 2006.

[51] B. Xu, N. Wang, T. Chen, and M. Li, “Empirical evaluation of rectified activations in convolutional network,” 2015, *arXiv:1505.00853*.

[52] A. Bora, E. Price, and A. G. Dimakis, “AmbientGAN: Generative models from lossy measurements,” in *Proc. Int. Conf. Learn. Representations*, 2018.

[53] W. V. Aarle et al., “Fast and flexible X-ray tomography using the ASTRA toolbox,” *Opt. Exp.*, vol. 24, no. 22, pp. 25129–25147, Oct. 2016.

[54] F. Natterer, *The Mathematics of Computerized Tomography*. Philadelphia, PA, USA: SIAM, 2001.

[55] S. Boyd, N. Parikh, and E. Chu, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[56] E. Soubies et al., “Pocket guide to solve inverse problems with GlobalBioIm,” *Inverse Problems*, vol. 35, no. 10, Sep. 2019, Art. no. 104006.

[57] H. Khatter, A. G. Myasnikov, S. K. Natchiar, and B. P. Klaholz, “Structure of the human 80S ribosome,” *Nature*, vol. 520, no. 7549, pp. 640–645, 2015.

[58] K. Hämäläinen, L. Harhanen, A. Kallonen, A. Kujanpää, E. Niemi, and S. Siltanen, “Tomographic X-ray data of a walnut,” Feb. 2015, *arXiv:1502.04064*.