

# Quantifying Model Uncertainty in Inverse Problems via Bayesian Deep Gradient Descent

Riccardo Barbano Chen Zhang Simon Arridge Bangti Jin

University College London

# Outline

1 Aim

2 Unrolled Optimisation

3 Bayesian Deep Gradient Descent

4 Results & Discussion

5 Summary

# Introduction

- Deep learning (DL) image reconstruction techniques have remarkable results but lack estimates of uncertainty
- This is critical in sensitive domains such as medical imaging
- There are many *types* of uncertainty but the most common in medical imaging are
  - **Epistemic** - uncertainty in the parameters (i.e., model uncertainty)
  - **Aleatoric** - stochastic variability in data generation
- **Our Goal:** a DL reconstruction method that allows to account for *epistemic uncertainty* in medical image reconstruction

# Unrolled Optimisation

- **Unrolled optimisation** mimics iterative methods but
  - 1 Executes only a finite number of iterations
  - 2 Computes the updates using DNNs
- The refinements are computed as residual updates with a feasibility projection:

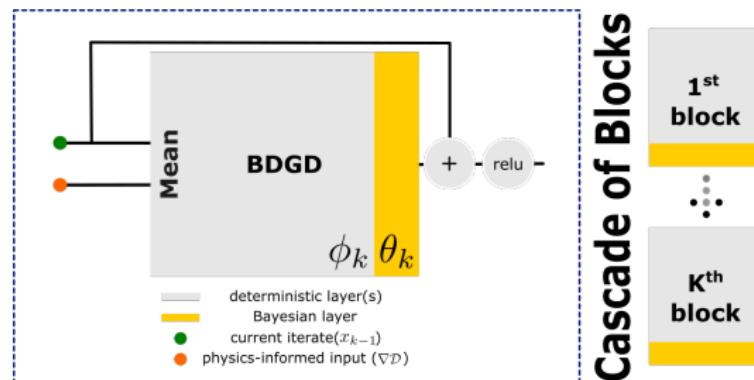
$$\mathbf{x}_k = \text{ReLU}(\mathbf{x}_{k-1} + \delta\mathbf{x}_{k-1})$$

- The increments  $\delta\mathbf{x}_{k-1}$  are computed as

$$\delta\mathbf{x}_{k-1} = f_{\varphi_k}(\nabla \mathcal{D}(\mathbf{y}, \mathbf{A}\mathbf{x}_{k-1}), \mathbf{x}_{k-1}) =: f_{\varphi_k}(\nabla \mathcal{D}, \mathbf{x}_{k-1})$$

# Bayesian Deep Gradient Descent

- Each block (network) of the cascade consists of two parts
  - Deterministic layers with weights  $\phi_k$
  - A final Bayesian layer with (random) weights  $\theta_k$



- To estimate the posterior  $p(\theta|X, Y)$  we use **variational inference** which uses an approximate, simple to compute, distribution  $q_{\psi}^*$
- The optimal distribution is computed by minimising the negative **ELBO**

$$q_{\psi_k}^* \in \underset{q_{\psi_k} \in \mathcal{Q}_k}{\operatorname{argmin}} \mathcal{L}_k(q_{\psi_k}; X, Y) := - \int q_{\psi_k}(\Theta_k) \log p(X|Y, \Theta_k) d\Theta_k + \text{KL}(q_{\psi_k}(\Theta_k) \| p(\Theta_k))$$

# Estimating Predictive Uncertainty

- We use Monte Carlo (MC) estimators to estimate the statistics of the approximate predictive posterior defined as:

$$q^*(\mathbf{x}|\mathbf{y}) = \int p(\mathbf{x}|\mathbf{y}, \Theta_K) q_{\psi_K}^*(\Theta_K) d\Theta_K$$

- We summarise predictive uncertainty as the (entry-wise) predictive variance  $\text{Var}[\mathbf{x}]$  at the  $K^{\text{th}}$  step:

$$\begin{aligned} \text{Var}[\mathbf{x}] &= \text{Var}_{q_{\psi_K}(\Theta_K)}[\mathbb{E}(\mathbf{x}|\mathbf{y}, \Theta_K)] + \mathbb{E}_{q_{\psi_K}(\Theta_K)}[\text{Var}(\mathbf{x}|\mathbf{y}, \Theta_K)] \\ &\approx \sigma_K^2 + \underbrace{\frac{1}{T} \sum_{t=1}^T f_{\Theta_K^t}(\nabla \mathcal{D}, \mathbf{x}_0)^2 - \left( \frac{1}{T} \sum_{t=1}^T f_{\Theta_K^t}(\nabla \mathcal{D}, \mathbf{x}_0) \right)^2}_{\text{epistemic}} \end{aligned}$$

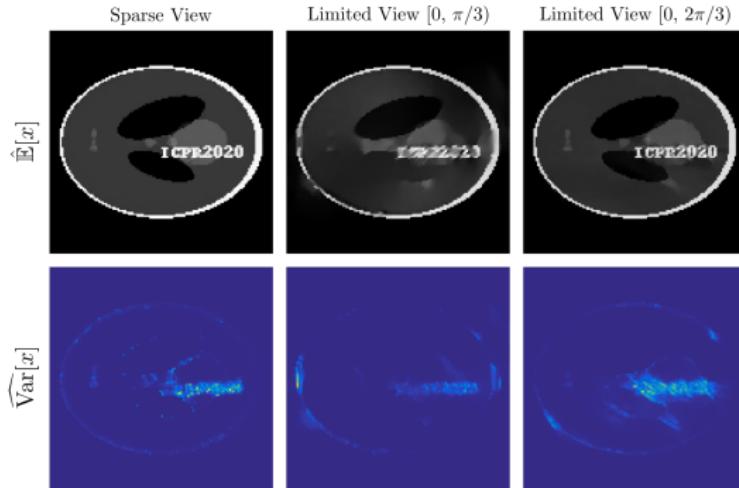
# Results & Discussion

**Table:** Sparse View CT (30 dirs)

Methods	Ellipses	Phantoms	SL Phantom
FBP	25.5264	18.4667	
TV	35.1587	37.2162	
LPD[1]	<b>44.5122 ± 0.4911</b>	44.0472 ± 0.4187	
DGD[2]	43.2577 ± 0.4183	44.6913 ± 0.6644	
BDGD - MFVI	<b>44.6642 ± 0.4637</b>	<b>47.2946 ± 0.5778</b>	
BDGD - MCDO	43.2126 ± 0.1285	45.1725 ± 0.4461	

**Table:** Limited View CT [0,  $2\pi/3$ )

Methods	Ellipses	Phantoms	SL Phantom
FBP	18.5958	17.1085	
TV	32.9134	29.2113	
LPD	40.7578 ± 0.3050	33.8427 ± 1.2380	
DGD	42.6994 ± 0.4243	42.8905 ± 0.5883	
BDGD - MFVI	<b>44.0297 ± 0.4698</b>	<b>45.5140 ± 0.8261</b>	
BDGD - MCDO	41.5367 ± 0.3884	41.4397 ± 0.6299	



**Figure:** Out-of-distribution reconstruction for different geometries by BDGD-MFVI: (*Left*) sparse view with 30 directions, (*Centre*) limited view  $[0, \pi/3]$ , (*Right*) limited view  $[0, 2\pi/3)$ .

# Summary

- We introduce a *tractable* and *statistically principled* framework that provides *epistemic uncertainty* by integrating a data-driven knowledge-aided framework with advances in BNNs and VI.
- We propose a *greedy* training scheme, which trains the framework *block-wise*, in a manner similar to deep gradient descent [2]. This allows for greatly reducing the training time.
- We further achieve computational efficiency for training a hybrid architecture where only a small portion of the whole network is Bayesian, together with the greedy training scheme.

# Bibliography



Jonas Adler and Ozan Öktem.

Learned primal-dual reconstruction.

*IEEE Trans. Med. Imag.*, 37(6):1322–1332, 2018.



Andreas Hauptmann, Felix Lucka, Marta Betcke, Nam Huynh, Jonas Adler, Ben Cox, Paul Beard, Sébastien Ourselin, and Simon Arridge.

Model-based learning for accelerated, limited-view 3-d photoacoustic tomography.

*IEEE Trans. Med. Imag.*, 37(6):1382–1393, 2018.