Bayesian Experimental Design for Computed Tomography with the Linearised Deep Image Prior

Riccardo Barbano[*]   RICCARDO.BARBANO.19@UCL.AC.UK
Department of Computer Science, University College London

Johannes Leuschner*   JLEUSCHN@UNI-BREMEN.DE
Center for Industrial Mathematics, University of Bremen

Javier Antorán*   JA666@CAM.AC.UK
Department of Engineering, University of Cambridge

Bangti Jin   B.JIN@UCL.AC.UK
Department of Computer Science, University College London

José Miguel Hernández-Lobato   JMH233@CAM.AC.UK
Department of Engineering, University of Cambridge

Abstract

We investigate adaptive design based on a single sparse pilot scan for generating effective scanning strategies for computed tomography reconstruction. We propose a novel approach using the linearised deep image prior. It allows incorporating information from the pilot measurements into the angle selection criteria, while maintaining the tractability of a conjugate Gaussian-linear model. On a synthetically generated dataset with preferential directions, linearised DIP design allows reducing the number of scans by up to 30% relative to an equidistant angle baseline.

1. Introduction and related work

Linear inverse problems in imaging aim to recover an unknown image $x \in \mathbb{R}^{d_x}$ from measurements $y \in \mathbb{R}^{d_y}$, which are often described by the application of a forward operator $A \in \mathbb{R}^{d_y \times d_x}$, and the addition of Gaussian noise $\epsilon \sim \mathcal{N}(0, \sigma_y^2 I_{d_y})$ as

$$y = Ax + \epsilon. \tag{1}$$

This acquisition model is ubiquitous in machine vision, computed tomography (CT), and magnetic resonance imaging among other applications. Due to the inherent ill-posedness of the task (e.g. $d_y \ll d_x$), suitable regularisation or prior assumptions are crucial for the stable and accurate recovery of x (Tikhonov and Arsenin, 1977; Ito and Jin, 2014). In this work, we focus on X-ray imaging, a setting with application to both medical and industrial settings (Buzug, 2011).

In CT, an emitter sends X-ray quanta through the object being scanned. The quanta are captured by d_p detector elements placed opposite the emitter. Each row of A tells us about which regions (pixels) the X-ray quanta will pass through before reaching a detector element (cf. fig. 1). The number of X-ray quanta measured by a detector pixel conveys information about the attenuation coefficient of the material present along the quanta's path. This procedure is repeated at $d_{\mathcal{B}}$ angles, yielding a measurement of dimension $d_y = d_p \cdot d_{\mathcal{B}}$.

---

*. Authors contributed equally. Our code is at github.com/educating-dip/bayesian_experimental_design
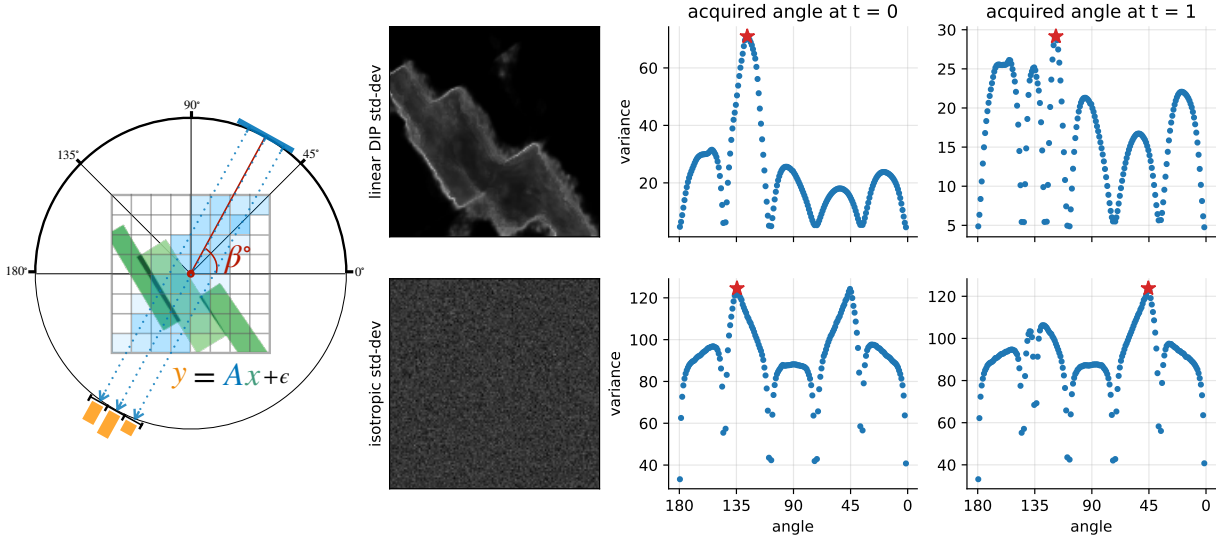
Figure 1: Left: A schematic diagram of 2D parallel beam CT geometry, used in the experiments. Top row: the linearised DIP assigns prior variance to pixels where edges are present, guiding angle selection so that X-ray quanta cover these pixels. Bottom row: the isotropic linear model's variance does not depend on the measurements. Angles $45, 135$ are chosen since they are oblique and maximise quanta path-length in the image.

In CT, Bayesian experimental design employs prior assumptions to select scanning angles which will yield the highest fidelity reconstruction. Adaptive design further incorporates information gained at previous angles to inform subsequent angle selections (Chaloner and Verdinelli, 1995). These methods are of great practical interest since they promise to reduce radiation dosages and scanning times. Alas, existing CT design methods often struggle to improve over equidistant angle choice (Shen et al., 2022). Furthermore, the requisite of additional computations before subsequent scans makes adaptive methods impractical for many applications.

Critically important to experimental design is the choice of prior (Feng, 2015; Foster, 2021). Linear models allow for tractable computation of quantities of interest for design, but their predictive uncertainty is independent of previously measured values, disallowing adaptive design (Burger et al., 2021). More complex model choices make inference difficult, necessitating approximations which can degrade performance (Helin et al., 2022; Shen et al., 2022).

This work aims to make adaptive design practical by considering a setting where the CT scan is performed in two phases. First, a sparse pilot scan is performed to provide data with which to fit adaptive methods. These are then used to select angles for a full scan. We demonstrate this procedure with a synthetic dataset where a different "preferential" angle is most informative for each image. Preferential directions appear commonly in industrial CT for material science and in medical CT for medical implant assessment. We use the linearised Deep Image Prior (DIP) (Barbano et al., 2022) as a data-dependent prior for adaptive design which preserves the tractability of conjugate Gaussian-linear models. Unlike simple linear models, the linearised DIP outperforms the equidistant angle baseline. Finally, we show that designs obtained with the linearised DIP perform well under traditional (non DIP-based) regularised-reconstruction.

2. Regularised reconstruction and deep image prior

Total Variation (TV) is the most popular regulariser for CT reconstruction (Rudin et al., 1992; Chambolle et al., 2010). The anisotropic TV semi-norm of an image vector $x \in \mathbb{R}^{d_x}$ is given by

$$\mathrm{TV}(x) = \sum_{i,j} |X_{i,j} - X_{i+1,j}| + \sum_{i,j} |X_{i,j} - X_{i,j+1}|, \tag{2}$$

where $X \in \mathbb{R}^{h \times w}$ denotes the vector $x \in \mathbb{R}^{d_x}$ reshaped into an image of height h by width w, and $d_x = h \cdot w$. The corresponding regularised reconstruction is obtained by

$$x^{\star} \in \underset{x \in \mathbb{R}^{d_x}}{\operatorname{argmin}} \|\mathrm{A}x - y\|^2 + \lambda \mathrm{TV}(x), \tag{3}$$

where the hyperparameter $\lambda > 0$ determines the strength of regularisation.

The DIP (Ulyanov et al., 2020; Baguer et al., 2020) reparametrises the reconstruction x as the output of a U-net $x(\theta)$ (Ronneberger et al., 2015) with a fixed input, which we omit for clarity, and parameters $\theta \in \mathbb{R}^{d_\theta}$. The resulting reconstruction problem reads

$$\theta^{\star} \in \underset{\theta \in \mathbb{R}^{d_\theta}}{\operatorname{argmin}} \|\mathrm{A}x(\theta) - y\|^2 + \lambda \mathrm{TV}(x(\theta)) \quad \text{and} \quad x^{\star} = x(\theta^{\star}). \tag{4}$$

We follow Barbano et al. (2021) in accelerating optimisation of eq. (4) using pre-trained U-nets.

3. Linear(ised) models for CT experimental design

Let $\mathcal{B}_a$ be the set of *all* possible angles at which we can scan. The task is to choose the subset of angles $\mathcal{B} \subset \mathcal{B}_a$ which produces the highest-fidelity reconstruction. We shall add angles sequentially over T steps. The set $\mathcal{B}^{(t)}$ denotes the chosen angles up to step $t < T$, and $\bar{\mathcal{B}}^{(t)} = \mathcal{B}_a \setminus \mathcal{B}^{(t)}$ the angles left to choose from. $\mathcal{B}^{(0)}$ denotes the set of angles used in the initial pilot scan, and $\mathcal{B} = \mathcal{B}^{(T)}$ the full design. We incorporate a decision to scan at angle $\beta \in \bar{\mathcal{B}}^{(t)}$ by concatenating the matrix $\mathrm{A}^{\beta} \in \mathbb{R}^{d_p \times d_x}$, which contains a row for each detector pixel at angle β, to the operator. After step t, the operator $\mathrm{A}^{(t)} \in \mathbb{R}^{d_p \cdot d_{\mathcal{B}^{(t)}} \times d_x}$ stacks $d_{\mathcal{B}^{(t)}}$ of these matrices, with $d_{\mathcal{B}^{(t)}} = |\mathcal{B}^{(t)}|$. $\bar{\mathrm{A}}^{(t)} \in \mathbb{R}^{d_p \cdot d_{\bar{\mathcal{B}}^{(t)}} \times d_x}$ denotes the forward operator for the angles left to choose from.

For design, we place a multivariate Gaussian prior on x with zero mean and covariance matrix $\Sigma_{xx} \in \mathbb{R}^{d_x \times d_x}$. Together with the Gaussian noise model in eq. (1), this gives a conjugate Gaussian-linear model. The vector $y^{(t)} \in \mathbb{R}^{d_p \cdot d_{\mathcal{B}^{(t)}}}$ of all measurements at step t is distributed as

$$y^{(t)}|x \sim \mathcal{N}(\mathrm{A}^{(t)}x, \, \sigma_y^2 \mathrm{I}_{d_y}) \quad \text{with} \quad x \sim \mathcal{N}(0, \Sigma_{xx}).$$

Thus, $\Sigma_{yy}^{(t)} = \mathrm{A}^{(t)} \Sigma_{xx} (\mathrm{A}^{(t)})^{\top} + \sigma_y^2 \mathrm{I}$ is the measurement covariance and the posterior over x is

$$x|y^{(t)} \sim \mathcal{N}(\mu_{x|y^{(t)}}, \Sigma_{x|y^{(t)}}), \quad \text{with}$$

$$\mu_{x|y^{(t)}} = \Sigma_{xx}(\mathrm{A}^{(t)})^{\top}(\Sigma_{yy}^{(t)})^{-1}y^{(t)}, \quad \text{and} \quad \Sigma_{x|y^{(t)}} = \Sigma_{xx} - \Sigma_{xx}(\mathrm{A}^{(t)})^{\top}(\Sigma_{yy}^{(t)})^{-1}\mathrm{A}^{(t)}\Sigma_{xx}. \tag{5}$$

The predictive covariance $\Sigma_{x|y^{(t)}}$ completely characterises the uncertainty of the reconstruction at step t and is the building block for the angle selection criteria in section 3.1. Note that natural images often exhibit heavy-tailed non-Gaussian statistics (Seeger and Nickisch, 2011). Additionally, by eq. (5), $\Sigma_{x|y^{(t)}}$ depends on the choice of angles through $\mathrm{A}^{(t)}$, but not on the measurements made at said angles $y^{(t)}$, precluding adaptive design. In section 3.2, we construct Σ_{xx} with correlations between nearby pixels, imitating the effects of the TV regulariser eq. (2), and with dependence on previous measurements, recovering adaptive design capability. In the experiments, we use linear models for angle selection and afterwards we discard the predictive mean $\mu_{x|y}$ and employ the regularised approaches from section 2 for reconstruction.

3.1 Experimental design with linear models

Acquisition objectives. Since the linear design task is submodular (Seeger, 2009), we greedily add one single angle per acquisition step [1]. We consider two popular acquisition objectives.

The first objective, *Expected Information Gain* (EIG) (Mackay, 1992a), is the expected reduction in the posterior entropy $H(x|y)$ from scanning at angle β. At step t, it is given by

$$\text{EIG} := \text{H}(x|y^{(t)}) - \mathbb{E}_{p(y^\beta|y^{(t)})}[\text{H}(x|y^{(t)}, y^\beta)] = \text{logdet}(\sigma_y^2 \text{I}_{d_{\mathcal{B}^{(t)}}} + \text{A}^\beta \Sigma_{x|y^{(t)}}(\text{A}^\beta)^\top) + C \quad (6)$$

where the constant $C = -\text{logdet}(\sigma_y^2 \text{I})$ is independent of the angle choice. We give a derivation in appendix A for completeness. Intuitively, the determinant of the matrix $\text{A}^\beta \Sigma_{x|y^{(t)}}(\text{A}^\beta)^\top \in \mathbb{R}^{d_p \times d_p}$ penalises angles for which different detector elements make correlated measurements and the log term encourages the measurements from all detector pixels to be similarly informative.

The second objective, which we find to perform better empirically, is to choose the angles for which our prediction has the largest *expected squared error* (ESE) in measurement space

$$\text{ESE} := \mathbb{E}_{p(y^\beta, x|y^{(t)})}[(y^\beta - \text{A}^\beta x)^\top (y^\beta - \text{A}^\beta x)] = \text{Tr}(\text{A}^\beta \Sigma_{x|y^{(t)}}(\text{A}^\beta)^\top) + C. \quad (7)$$

This objective is a bijection of EIG in the setting where our detector has a single pixel.

Efficient acquisition. Constructing the matrix $\text{A}^\beta \Sigma_{x|y^{(t)}}(\text{A}^\beta)^\top$ repeatedly for each candidate angle $\beta \in \bar{\mathcal{B}}^{(t)}$ requires $\mathcal{O}(d_p \cdot d_{\bar{\mathcal{B}}^{(t)}})$ matrix vector products, which is very costly even for moderate size scanners. Instead, we estimate the matrix for every angle simultaneously by drawing K samples from $\mathcal{N}(0, \bar{\text{A}}^{(t)} \Sigma_{x|y^{(t)}}(\bar{\text{A}}^{(t)})^\top)$ with $\mathcal{O}(K)$ matrix vector products. That is, we sample $\mathbb{R}^{d_p \cdot d_{\bar{\mathcal{B}}^{(t)}}}$ sized vectors built by concatenating the "pseudo measurements" for each unused angle $\beta \in \bar{\mathcal{B}}^{(t)}$. We use Matheron's rule (Hoffman and Ribak, 1991; Wilson et al., 2021)

$$\bigoplus_{\beta \in \bar{\mathcal{B}}^{(t)}} y_k^\beta = \bar{\text{A}}^{(t)} \left(x_k - \Sigma_{xx}(\text{A}^{(t)})^\top \Sigma_{yy}^{-1}(\eta_k + \text{A}^{(t)} x_k) \right) \quad \text{with}$$

$$x_k \sim \mathcal{N}(0, \Sigma_{xx}) \quad \text{and} \quad \eta_k \sim \mathcal{N}(0, \sigma_y^2 \text{I}), . \quad (8)$$

Here, $k \in \{1...K\}$ indexes different samples and $\bigoplus$ denotes vector concatenation. We compute

$$\text{A}^\beta \Sigma_{x|y^{(t)}}(\text{A}^\beta)^\top \approx K^{-1} \sum_{k=1}^{K} y_k^\beta (y_k^\beta)^\top,$$

which is then used to estimate the acquisition objective eq. (6) or eq. (7). The log term makes EIG estimates only asymptotically unbiased (i.e. as $K \to \infty$) but we find the bias to be insignificant. Once the angle β that maximises eq. (6) or eq. (7) is chosen, we update $\Sigma_{yy}^{(t+1)}$ as

$$\Sigma_{yy}^{(t+1)} = \begin{bmatrix} \Sigma_{yy}^{(t)} & \text{A}^{(t)} \Sigma_{xx}(\text{A}^{(t+1)})^\top \\ \text{A}^{(t+1)} \Sigma_{xx}(\text{A}^{(t)})^\top & \text{A}^{(t+1)} \Sigma_{xx}(\text{A}^{(t+1)})^\top \end{bmatrix}, \quad (9)$$

and repeat the procedure, i.e. return to eq. (8).

---

1. Submodularity guarantees this procedure obtains a score within a $(1 - 1/e)$ factor of the optimal strategy.
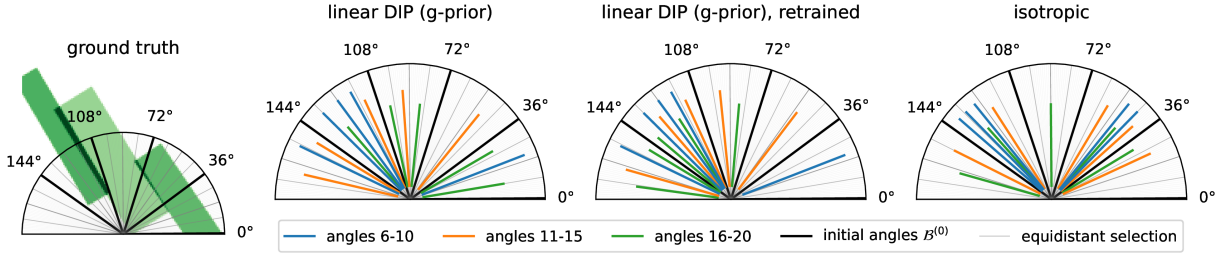
Figure 2: First 20 angles selected by each method under consideration for an example image.

3.2 Construction of the prior covariance Σ_{xx}

Now we describe the construction of the Gaussian prior covariance over reconstructions $\Sigma_{xx} \in \mathbb{R}^{d_x \times d_x}$. We consider a range of models, building from very simple models to flexible data-driven ones that allows for adaptive design.

Isotropic model. The simple choice $\Sigma_{xx} = \sigma_x^2 I_{d_x}$ assumes uncorrelated pixels, and it implies a ridge regulariser for the reconstruction, which is known to perform poorly in imaging.

Matern-$1/2$ Process. Antorán et al. (2022) employ the Matern-$1/2$ covariance $[\Sigma_{xx}]_{ij,i'j'} = \sigma_x^2 \exp(-\ell^{-1}\sqrt{(i-i')^2 + (j-j')^2})$, where i, j index the pixel locations in the image x, as a surrogate for TV. With the hyperparameters σ_x^2 and ℓ properly chosen, the prior samples and posterior inferences closely match those obtained with an intractable TV prior.

Linearised deep image prior (Barbano et al., 2022; Antorán et al., 2022). This data-driven prior is constructed by first fitting a DIP model on the measurements taken during the pilot scan with eq. (4), and then adopting a linear model on the basis expansion given by the Jacobian of the trained U-net $x(\cdot)$ with respect to θ evaluated at the optimal point $\theta^\star$, i.e. $\nabla_\theta x(\theta)_{|\theta=\theta^\star} =: \mathrm{J} \in \mathbb{R}^{d_x \times d_\theta}$ (Immer et al., 2021b). The resulting prior over x is given by

$$x = \mathrm{J}\theta, \quad \theta \sim \mathcal{N}(0, \Sigma_\theta) \quad \text{and thus} \quad x \sim \mathcal{N}(0, \mathrm{J}\Sigma_\theta \mathrm{J}^\top).$$

The covariance $\Sigma_{xx} = \mathrm{J}\Sigma_\theta \mathrm{J}^\top$ incorporates information about the pilot measurements through the features J. It assigns higher prior variance being near the edges in the reconstruction, cf. fig. 1, which are most sensitive to a change in U-net parameters. The covariance $\Sigma_\theta \in \mathbb{R}^{d_\theta \times d_\theta}$ weights different Jacobian entries. We consider two different structures for Σ_θ.

- The filter-wise block-diagonal matrix of Antorán et al. (2022) uses a separate prior for every block in the U-net (cf. appendix D.2). This choice uses a large number of hyperparameters. It risks overfitting to the pilot scan measurements resulting in uncertainty underestimation.

- The neural g-prior (Zellner, 1986; Antoran et al., 2022) is a maximally uninformative diagonal Gaussian prior with covariance matching the diagonal of U-net's inverse Fisher information matrix, denoted s^{-1}, scaled by a constant g (see appendix C for extended discussion). That is

$$\Sigma_\theta = g \cdot s^{-1}\mathrm{I}, \quad s = d_{y^{(t)}}^{-1} \sum_{i=1}^{d_y} ([\mathrm{A}^{(t)}\mathrm{J}]_i)^2 \in \mathbb{R}^{d_\theta}, \quad \text{and we choose} \quad g = (d_{y^{(0)}} d_\theta)^{-1} \sum_{i=1}^{d_y} ((y_i^{(0)})^2 - \sigma_y^2),$$

where $[\mathrm{AJ}]_i$ refers to the ith row of the matrix AJ. Computing s does not require measurement values and we update it every 5 acquired angles. We compute g once using the measurements from the pilot scan. Our choice of s ensures that the Jacobian entries corresponding to all U-net weights contribute equally to the marginal prior variance over measurements. Our choice of g ensures this marginal variance is equal to the empirical second moment of pilot measurements.

All models discussed have a number of free parameters $\sigma_y^2, \sigma_x^2, \ell, \Sigma_\theta$, which we choose to maximise the model evidence given the pilot scan measurements. See appendix B for details.
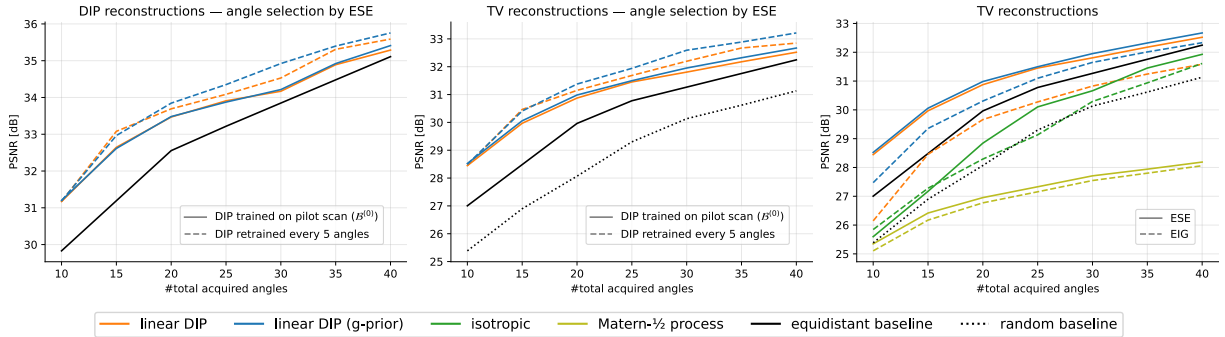
5

Figure 3: Reconstruction PSNR vs n. angles scanned, averaged across 30 images (5% noise).

4. Experiments and analysis

We simulate CT measurements y from 128×128 pixel images displaying rectangles of random proportions aligned along a randomly chosen "preferential" direction (see fig. 2 and fig. 4). The forward operator A is the discrete Radon transform, and either 5% or 10% white noise is added to the measurement y. We divide the range $[0°, 180°)$ into 200 selectable angles (i.e. $|\mathcal{B}_a| = 200$). The pilot scan measures at 5 equidistant angles, on which we fit all models' hyperparameters and the linearised DIP's U-net (see appendix B). Then, we apply the methods in section 3.1 to produce designs consisting of 35 additional angles. For every 5 acquired angles, we evaluate reconstruction quality using both the DIP (i.e. eq. (4)), and the traditional TV regularised approach (i.e. eq. (3)). We include equidistant and random angle selection as baselines. On an NVIDIA A100 GPU, a full linearised DIP acquisition step with $K = 3000$ samples takes 9 seconds and the full design takes 5 minutes. Appendix D contains full experimental details.

For the **linearised DIP**, we consider training our U-net and prior hyperparameters only on the pilot scan, and also retraining every 5 angles. Figure 2 shows both approaches can identify and prioritise the preferential direction, leading to reconstructions that *outperform the equidistant angle baseline by over 1.5 dB* in the range of $[10, 15]$ angles (see fig. 3). During this initial stage, the linearised DIP requires roughly *30% less scanned angles* to match the equidistant baseline's performance. The performance gap decreases as we select more angles, although linearised DIP remains more efficient even after 40 angles. Retraining the U-net provides most benefits in the large angle regime. It increases focus on preferential directions and consistently provides gains >0.5dB after 20 angles. All gains over the equidistant baseline are obtained with both DIP (i.e. eq. (4)) and traditional TV regularised reconstruction (i.e. eq. (3)). In the high 10% noise setting, gains from experimental design are smaller, but still significant (see appendix E).

The **isotropic and Matern-**$1/2$ models' uncertainty estimates are independent of the pilot measurements. These models prioritise clustered sets of oblique angles which maximise the length of quanta trajectories in the image. They perform similar to or worse than random. We explore this negative result in appendix E, finding it due to overfitting of hyperparameters.

ESE outperforms EIG across models. For the linearised DIP, this gap is smaller when using the g-prior. We hypothesise that model misspecification and hyperparameter overfitting may result in poor measurement covariance estimates, in turn degrading EIG estimates.

5. Conclusion and future work

Our results suggest that dependence on the measurement data, i.e. adaptivity, is key to out-performing equidistant angle selection in CT reconstruction, a notoriously difficult task (Shen

6

et al., 2022; Helin et al., 2022). Distinctly from previous work, our methods only necessitate a pilot scan instead of being fully online, increasing applicability. We observe the largest gains in the 10 to 20 angle regime, where our designs reduce the angle requirement by roughly 30% without loss of reconstruction quality. This is true for both traditional TV-regularised and DIP reconstructions. In future, we aim to apply linearised DIP designs to real measurements.

Acknowledgments

References

J. Antoran and A. Miguel. Disentangling and learning robust representations with natural clustering. In *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, pages 694–699, 2019.

Javier Antorán, James Urquhart Allingham, and José Miguel Hernández-Lobato. Depth uncertainty in neural networks. In Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin, editors, *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/781877bda0783aac5f1cf765c128b437-Abstract.html.

Javier Antorán, Umang Bhatt, Tameem Adel, Adrian Weller, and José Miguel Hernández-Lobato. Getting a CLUE: A method for explaining uncertainty estimates. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021. URL https://openreview.net/forum?id=XSLF1XFq5h.

Javier Antoran, James Urquhart Allingham, David Janz, Erik Daxberger, Eric Nalisnick, and José Miguel Hernández-Lobato. Linearised laplace inference in networks with normalisation layers and the neural g-prior. In *Fourth Symposium on Advances in Approximate Bayesian Inference*, 2022. URL https://openreview.net/forum?id=uUH8x-h9zdB.

Javier Antorán, Riccardo Barbano, Johannes Leuschner, José Miguel Hernández-Lobato, and Bangti Jin. A probabilistic deep image prior for computational tomography. Preprint, arXiv:2203.00479, 2022.

Javier Antorán, David Janz, James Urquhart Allingham, Erik A. Daxberger, Riccardo Barbano, Eric T. Nalisnick, and José Miguel Hernández-Lobato. Adapting the linearised laplace model evidence for modern deep learning. *CoRR*, abs/2206.08900, 2022. doi: 10.48550/arXiv.2206. 08900. URL https://doi.org/10.48550/arXiv.2206.08900.

Daniel Otero Baguer, Johannes Leuschner, and Maximilian Schmidt. Computed tomography reconstruction using deep image prior and learned reconstruction methods. *Inverse Problems*, 36(9):094004, 2020.

Riccardo Barbano, Johannes Leuschner, Maximilian Schmidt, Alexander Denker, Andreas Hauptmann, Peter Maaß, and Bangti Jin. Is deep image prior in need of a good education? *arXiv preprint arXiv:2111.11926*, 2021.

Riccardo Barbano, Javier Antorén, José Miguel Hernández-Lobato, and Bangti Jin. A probabilistic deep image prior over image space. In *Fourth Symposium on Advances in Approximate Bayesian Inference*, 2022. URL https://openreview.net/forum?id=qtFPfwJWowM.

Martin Burger, Andreas Hauptmann, Tapio Helin, Nutti Hyvönen, and Juha-Pekka Puska. Sequentially optimized projections in x-ray imaging. *Inverse Problems*, 37(7):075006, 2021. doi: 10.1088/1361-6420/ac01a4.

Thorsten M Buzug. Computed tomography. In *Springer handbook of medical technology*, pages 311–342. Springer, 2011.

Kathryn Chaloner and Isabella Verdinelli. Bayesian experimental design: A review. *Statistical Science*, 10(3):273 – 304, 1995. doi: 10.1214/ss/1177009939. URL https://doi.org/10.1214/ss/1177009939.

Antonin Chambolle, Vicent Caselles, Daniel Cremers, Matteo Novaga, and Thomas Pock. An introduction to total variation for image analysis. In *Theoretical foundations and numerical methods for sparse recovery*, pages 263–340. de Gruyter, 2010.

Erik A. Daxberger, Eric T. Nalisnick, James Urquhart Allingham, Javier Antorán, and José Miguel Hernández-Lobato. Bayesian deep learning via subnetwork inference. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event*, volume 139 of *Proceedings of Machine Learning Research*, pages 2510–2521. PMLR, 2021. URL http://proceedings.mlr.press/v139/daxberger21a.html.

V. V. Fedorov. *Theory of Optimal Experiments*. Academic Press New York, 1972. ISBN 0122507509.

Chi Feng. *Optimal Bayesian experimental design in the presence of model error*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, 2015.

Adam Evan Foster. *Variational, Monte Carlo and Policy-Based Approaches to Bayesian Experimental Design*. PhD thesis, University of Oxford, 2021.

Tapio Helin, Nuutti Hyvönen, and Juha-Pekka Puska. Edge-promoting adaptive Bayesian experimental design for X-ray imaging. *SIAM J. Sci. Comput.*, 44(3):B506–B530, 2022. ISSN 1064-8275. doi: 10.1137/21M1409330.

Yehuda Hoffman and Erez Ribak. Constrained realizations of Gaussian fields: a simple algorithm. *Astrophys. J. Lett.*, 380:L5–L8, 1991. doi: 10.1086/186160.

Alexander Immer, Matthias Bauer, Vincent Fortuin, Gunnar Rätsch, and Mohammad Emtiyaz Khan. Scalable marginal likelihood estimation for model selection in deep learning. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event*, volume 139 of *Proceedings of Machine Learning Research*, pages 4563–4573. PMLR, 2021a. URL http://proceedings.mlr.press/v139/immer21a.html.

Alexander Immer, Maciej Korzepa, and Matthias Bauer. Improving predictions of bayesian neural nets via local linearization. In Arindam Banerjee and Kenji Fukumizu, editors, *The 24th International Conference on Artificial Intelligence and Statistics, AISTATS 2021, April 13-15, 2021, Virtual Event*, volume 130 of *Proceedings of Machine Learning Research*, pages 703–711. PMLR, 2021b. URL http://proceedings.mlr.press/v130/immer21a.html.

Kazufumi Ito and Bangti Jin. *Inverse problems: Tikhonov theory and algorithms*, volume 22. World Scientific, 2014.

Jaehoon Lee, Samuel Schoenholz, Jeffrey Pennington, Ben Adlam, Lechao Xiao, Roman Novak, and Jascha Sohl-Dickstein. Finite versus infinite neural networks: an empirical study. *Advances in Neural Information Processing Systems*, 33:15156–15172, 2020.

David J. C. Mackay. Information-based objective functions for active data selection. *Neural Computation*, 4:590–604, 1992a.

David John Cameron Mackay. *Bayesian Methods for Adaptive Models*. PhD thesis, USA, 1992b. UMI Order No. GAX92-32200.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015.

Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268, 1992.

Matthias W. Seeger. On the submodularity of linear experimental design. Technical report, 2009. URL http://infoscience.epfl.ch/record/175483.

Matthias W. Seeger and Hannes Nickisch. Large scale bayesian inference and experimental design for sparse linear models. *SIAM J. Imaging Sci.*, 4(1):166–199, 2011. doi: 10.1137/090758775.

Ziju Shen, Yufei Wang, Dufan Wu, Xu Yang, and Bin Dong. Learning to scan: A deep reinforcement learning approach for personalized scanning in ct imaging. *Inverse Problems and Imaging*, 16(1):179–195, 2022.

Andrey N. Tikhonov and Vasiliy Y. Arsenin. *Solutions of ill-posed problems.* V. H. Winston & Sons, Washington, D.C.: John Wiley & Sons, New York, 1977. Translated from the Russian, Preface by translation editor Fritz John, Scripta Series in Mathematics.

Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. *Int. J. Comput. Vis.*, 128(7):1867–1888, 2020. ISSN 1573-1405. doi: 10.1007/s11263-020-01303-4.

James T. Wilson, Viacheslav Borovitskiy, Alexander Terenin, Peter Mostowsky, and Marc Peter Deisenroth. Pathwise conditioning of gaussian processes. *J. Mach. Learn. Res.*, 22:105:1–105:47, 2021. URL http://jmlr.org/papers/v22/20-1260.html.

Arnold Zellner. *On assessing prior distributions and Bayesian regression analysis with g prior distributions*, volume 6 of *Studies in Bayesian Econometrics and Statistics*, pages 233–243. Elsevier, 1986.

Appendix A. Discussion on acquisition objectives

The descriptions that follow are well known within the experimental design community but may be of interest to readers with a background in CT. Thus we describe these facts for the convenience of readers, and refer readers to Fedorov (1972); Chaloner and Verdinelli (1995) for a comprehensive introduction to experimental design and to Mackay (1992a) for a Bayesian perspective on experimental design.

EIG quantifies the information (in nats) we expect to gain by observing the detector elements' measurements for an angle or set of angles (Mackay, 1992a). Since our experiments employ greedy angle selection, we derive EIG for measurements at a single angle β. The generalisation to the multi-angle setting is straightforward. EIG is the expected decrease in posterior entropy from observing the detector elements' measurements at β:

$$\text{EIG} = \text{H}(x|y^{(t-1)}) - \mathbb{E}_{p(y^\beta|y^{(t-1)})}[\text{H}(x|y^{(t-1)}, y^\beta)],$$

where we take an expectation over the new measurement y^β, since EIG is computed before the measurement y^β is made. For this, we use the posterior predictive distribution $p(y^\beta|y^{(t-1)})$ given our previous measurements $y^{(t-1)}$

$$p(y^\beta|y^{(t-1)}) = \int p(y^\beta|x)p(x|y^{(t-1)}) \, \mathrm{d}x.$$

For the linear-Gaussian case, this integral can be evaluated in closed form, although this will not be necessary for our purposes.

EIG is also equal to the mutual information $MI(x, y^\beta|y^{(t-1)})$ between the reconstruction x and the new measurement y^β conditional on the previous measurements $y^{(t-1)}$, giving an interpretation as aiming to select the angle β most informative towards the reconstruction. For fixed model hyperparameters, EIG is always greater or equal than 0 since making additional measurements cannot increase the uncertainty in the reconstruction.

The entropy of a multivariate Gaussian $\mathcal{N}(\mu, \Sigma)$ is $H = \frac{1}{2}\text{logdet}(\Sigma) + \frac{d}{2}(\log(2\pi) + 1)$. For a fixed dimensionality d, the second term is constant across design steps and thus we only need to focus on the log determinant. The entropy does not depend on the distribution mean but only its covariance. Thus, taking $y^{(t)} = [y^{(t-1)}, y^\beta]$, we can write

$$\text{EIG} = \text{logdet}(\Sigma_{x|y^{(t-1)}}) - \text{logdet}(\Sigma_{x|y^{(t)}}).$$

Since the covariance $\Sigma_{x|y^{(t-1)}}$ does not depend on the new angle choice β, maximising EIG is equivalent to choosing the angle which minimises the updated covariance log-determinant $\text{logdet}(\Sigma_{x|y^{(t)}})$. Hence, the EIG objective for linear models is also known as the D(eterminant)-optimal criterion.

We can obtain a more convenient expression for EIG by noting the sequential nature of Bayesian learning; when data is observed, the prior is updated to a posterior. This posterior represents the updated beliefs and, as such, acts as a prior distribution for further inferences

$$p(x|y^{(t)}) = \frac{p(y^\beta|x)p(x|y^{(t-1)})}{p(y^\beta|y^{(t-1)})}.$$

11

For conjugate Gaussian-linear models, we can apply this principle to obtain the posterior covariance at time t from the covariance at time $t-1$ using the matrix determinant lemma

$$\text{logdet}(\Sigma_{x|y^{(t)}}) = -\text{logdet}(\Sigma_{x|y^{(t-1)}}^{-1}) - \text{logdet}(\sigma_y^{-2}\text{I}) - \text{logdet}(\sigma_y^2\text{I} + \text{A}_0^{(t)}\Sigma_{x|y^{(t-1)}}\text{A}_0^{\top,(t)}).$$

Thus, we have

$$
\begin{aligned}
\text{EIG} &= \text{logdet}(\Sigma_{x|y^{(t-1)}}) - \text{logdet}(\Sigma_{x|y^{(t)}}) \\
&= \text{logdet}(\Sigma_{x|y^{(t-1)}}) - [-\text{logdet}(\Sigma_{x|y^{(t-1)}}^{-1}) - \text{logdet}(\sigma_y^{-2}\text{I}) - \text{logdet}(\sigma_y^2\text{I} + \text{A}^\beta\Sigma_{x|y^{(t-1)}}(\text{A}^\beta)^\top)] \\
&= -\text{logdet}(\sigma_y^2\text{I}) + \text{logdet}(\sigma_y^2\text{I} + \text{A}^\beta\Sigma_{x|y^{(t-1)}}(\text{A}^\beta)^\top) \\
&= \text{logdet}(\sigma_y^2\text{I} + \text{A}^\beta\Sigma_{x|y^{(t-1)}}(\text{A}^\beta)^\top) + C
\end{aligned}
$$

where the constant $C = -\text{logdet}(\sigma_y^2\text{I})$ is independent of angle choice, yielding the objective we use for angle selection in practise.

The ESE objective in eq. (7) aims to minimise the squared prediction error in measurement space. Objectives of this kind are commonly known as (A)verage-optimal. However, ESE is A-optimal over measurement space y, not over image space x. ESE is crucially different from minimising the arguably more relevant expected squared reconstruction error, a more computationally expensive criterion. ESE can be understood as a naive simplification of EIG, by discarding correlations between detector pixels, making $\text{logdet}(\text{A}^\beta\Sigma_{x|y^{(t-1)}}(\text{A}^\beta)^\top)$ match $\sum_{i<d_p} \log[\text{A}^\beta\Sigma_{x|y^{(t-1)}}(\text{A}^\beta)^\top]_{ii}$. Then, the order of log and sum are switched, something that will only be true if every element under the sum is the same. Having reached this point, since the log function is monotonic, it does not affect angle selection and the criterion matches the trace of $\text{A}^\beta\Sigma_{x|y^{(t-1)}}(\text{A}^\beta)^\top$.

Appendix B. Hyperparameter selection via model evidence maximisation

For the conjugate linear-Gaussian model, the model evidence can be computed in closed form

$$
\log p(y) = \log \mathcal{N}(y; 0, \Sigma_{yy}) = -\frac{1}{2}\left(y^\top \Sigma_{yy}^{-1} y + \text{logdet}(\Sigma_{yy})\right) + C
$$
$$
\text{with} \quad \Sigma_{yy} = \text{A}\Sigma_{xx}\text{A}^\top + \sigma_y^2 I_{d_y}
$$

and $C = -d_y/2 \log 2\pi$. This expression is straightforward to compute for the isotropic and Matern-$1/2$ models. The linear solve against Σ_{yy} and log-determiant operations, while costly, are tractable to perform when the dimensionality of y is low. This is the case in our experimental setup, where we use the measurements from our 5 angle pilot scan $y^{(0)}$, specifically, $d_y = d_p \cdot d_\mathcal{B} = 5 \cdot 183 = 915$. We refer to Antorán et al. (2022) for discussion of efficient computation of the model evidence for the linearised DIP. For additional discussion on the motivation for the model evidence objective, its applications and pitfalls, we refer to Mackay (1992b); Immer et al. (2021a); Antorán et al. (2022).

Selecting prior hyperparameters with the model evidence is often claimed to be immune from overfitting due to the flexibility of the prior model being relatively low. However, when the number of measurements is small, e.g. after performing the pilot scan, overfitting is still possible. Indeed we observe the Matern-$1/2$ model suffers due to this issue in our experiments. We further discuss this in appendix E.

The risk of overfitting is also high for the linearised DIP model of Barbano et al. (2022); Antorán et al. (2022). Here, the basis expansion is selected by training a U-net on the pilot measurements and the number of hyperparameters is twice the number of U-net blocks, making this prior class very flexible. This has motivated the use of the neural g-prior (Antoran et al., 2022), discussed in the following section.

Appendix C. Discussion on the neural g-prior

The neural g-prior $\Sigma_\theta = g \cdot s^{-1} I$ was introduced by Antoran et al. (2022) as an approach to "normalise" the second moment of the Jacobian feature expansion analogously to standard data normalisation. This normalisation ensures that the Jacobian entries corresponding to all network weights contribute equally to the predictions at the train points, or in our case, to the predictions at the already measured angles. We refer to Antoran et al. (2022) for a full derivation.

Antoran et al. (2022) learn the variance scale g with the model evidence objective. However, it is well known that this procedure can overfit in the small-data regime. To prevent overfitting, in this work we choose g using the heuristic

$$g = (d_y d_\theta)^{-1} \sum_{i=1}^{d_y} ((y_i)^2 - \sigma_y^2).$$

This choice is made so that the marginal predictive variance averaged across measurement locations matches the empirical second moment of the observed targets, which we will denote $\mathbb{E}[y^2] = d_y^{-1} \sum_{i=1}^{d_y} y_i^2$. In other words, when using this prior over weights, our prior over measurements will have roughly the "right" variance. To see this, first recall

$$s = d_y^{-1} \sum_{i=1}^{d_y} ([\mathrm{AJ}]_i)^2$$

where $[\mathrm{AJ}]_i$ refers to the ith row of the matrix AJ and we will use $[\mathrm{AJ}]_{ij}$ to index each scalar entry of this matrix. We now expand the average marginal variance across measurements when using the neural g-prior

$$
\begin{aligned}
d_y^{-1} \sum_{i=1}^{d_y} [\Sigma_{yy}]_{ii} &= d_y^{-1} \sum_{i=1}^{d_y} [\mathrm{AJ}(gs^{-1}I)\mathrm{A}^\top \mathrm{J}^\top]_{ii} + \sigma_y^2 \\
&= g d_y^{-1} \sum_{i=1}^{d_y} s_i^{-1} \sum_{j}^{d_\theta} [\mathrm{AJ}]_{ij}^2 + \sigma_y^2 \\
&= (\mathbb{E}[y^2] - \sigma_y^2) d_\theta^{-1} \sum_{i=1}^{d_y} \sum_{j=1}^{d_\theta} \frac{[\mathrm{AJ}]_{ij}^2}{\sum_{k=1}^{d_y}[\mathrm{AJ}]_{kj}^2} + \sigma_y^2 \\
&= (\mathbb{E}[y^2] - \sigma_y^2) d_\theta^{-1} \sum_{j=1}^{d_\theta} \frac{\sum_{i=1}^{d_y}[\mathrm{AJ}]_{ij}^2}{\sum_{k=1}^{d_y}[\mathrm{AJ}]_{kj}^2} + \sigma_y^2 \\
&= (\mathbb{E}[y^2] - \sigma_y^2) \left(d_\theta^{-1} \sum_{j=1}^{d_\theta} 1 \right) + \sigma_y^2 = \mathbb{E}[y^2],
\end{aligned}
$$

Figure 4: Examples of synthetic images.

showing the property.

For (Antorán et al., 2022), model evidence optimisation is the most computationally costly step of inference. Avoiding model evidence optimisation speeds up inference and thus angle selection, making our proposed procedure more attractive for a real deployment.

Additionally, in fig. 3, we observe that the EIG objective performs best when combined with the neural g-prior. Arguably, EIG is a better motivated selection criterion than ESE but performs worse than equidistant selection when combined with all models except the g-prior linearised DIP. We hypothesise that model misspecification introduces error in our estimates of relative marginal variances and covariances across detector pixel measurements, in turn degrading the performance of EIG. ESE is less sensitive to these, as discussed in appendix A. Since the neural g-prior is a maximally uninformative prior, it somewhat mitigates model misspecification, improving the performance of EIG acquisition.

Appendix D. Full experimental setup

D.1 Dataset generation

We use a synthetic dataset comprising images of rectangles with randomised shape, orientation and intensity values, and simulate CT measurements by applying the forward operator $A \in \mathbb{R}^{d_y \times d_x}$ and adding Gaussian noise with standard deviation of 5 % or 10 % of the average absolute value of the noiseless measurements Ax. Each image has resolution 128×128 px^2 and shows 3 superimposed rectangles, whose orientation is sampled from a single normal distribution with zero mean and standard deviation 2.86°. Thus, images in this class contain edges in roughly two perpendicular directions. Figure 4 shows example images from the dataset.

D.2 Implementation details for the linearised DIP

The key step of efficiently implementing the linearised DIP is the computation and Cholesky decomposition of the measurement covariance matrix Σ_{yy}. We describe this step in the following paragraphs and refer to Antorán et al. (2022), which we have followed in our implementation, for a complete set of details.

Computing the measurement covariance matrix $\Sigma_{yy}^{(t)}$ To assemble or multiply with $\Sigma_{yy}^{(t)}$, we employ matrix-free methods. Our workhorses are the matrix vector products $v_x^\top \Sigma_{xx}$ and $v_y^\top \Sigma_{yy}^{(t)}$ for $v_x \in \mathbb{R}^{d_x}$ and $v_y \in \mathbb{R}^{d_y}$. We efficiently compute these products through successive

	5 % noise				10 % noise			
#angles:	5	10-15	20-30	35-40	5	10-15	20-30	35-40
TV strength λ	$1e{-}2$	$3e{-}3$	$3e{-}3$	$3e{-}3$	$1e{-}2$	$1e{-}2$	$1e{-}2$	$3e{-}3$
iterations	60 000	30 000	10 000	10 000	60 000	30 000	10 000	10 000

Table 1: Hyperparameters for TV reconstruction. The values for λ are found by grid search on 10 validation images using 5, 10, 20 and 40 angles, and the numbers of iterations are chosen such that convergence is observed.

matrix vector product with the components of either Σ_{xx}, or $\Sigma_{yy}^{(t)}$, respectively. For instance,

$$v_y^\top \Sigma_{yy}^{(t)} = v_y^\top \left(A^{(t)} J \Sigma_\theta J^\top \left(A^{(t)} \right)^\top + \sigma_y^2 I \right).$$

For any vector v_θ of appropriate size, we compute Jacobian vector products $v_\theta^\top J^\top$ using forward mode automatic differentiation (AD) and $v_\theta^\top J$ using backward mode AD. For the non g-prior model, we efficiently compute products with Σ_θ by exploiting its block diagonal structure. Since the g-prior covariance matrix is diagonal, computing products with it is straightforward.

Numerically stable sample generation with Matheron's rule eq. (8) Numerical instabilities can arise during the sample generation with the Matheron's rule due to the inversion of Σ_{yy}, updated via eq. (9). We resort to a simple regularisation strategy, which consists in adding to $\Sigma_{yy}^{(t)}$ a small diagonal element ϵI, where ϵ is chosen from 1% to 10% of the diagonal mean, similarly to Lee et al. (2020).

D.3 Hyperparameters for TV and DIP reconstruction

The TV strength (*i.e.* λ) used in the DIP optimisation and the TV regularised objective, reported in table 2 and table 1, are found by grid search on 10 validation images. The DIP reconstruction quality from some images degrades when using many iterations Baguer et al. (2020), so an early stopping would be beneficial. For the PSNR evaluations of DIP reconstructions, we iterate for 30 000 steps and select the maximum PSNR for each image; this resembles the ideal early stopping by using the (in practice unknown) ground truth image, and is done in order to exclude the complexity of the stopping mechanism from our evaluations. For the DIP optimisations used for angle selection (i.e. the initial DIP on $\mathcal{B}^{(0)}$ and the DIPs retrained every 5 angles), the numbers of iterations in table 2 are used, which were found by grid search on 10 validation images.

	5 % noise				10 % noise			
#angles:	5	10-15	20-30	35-40	5	10-15	20-30	35-40
TV strength λ	$3e{-}3$	$3e{-}3$	$3e{-}3$	$1e{-}3$	$1e{-}2$	$1e{-}2$	$3e{-}3$	$3e{-}3$
iterations	19 000	9400	12 000	13 000	11 000	7500	12 000	7100

Table 2: Hyperparameters for DIP reconstruction (including TV regularisation). The values are found by grid search on 10 validation images using 5, 10, 20 and 40 angles.
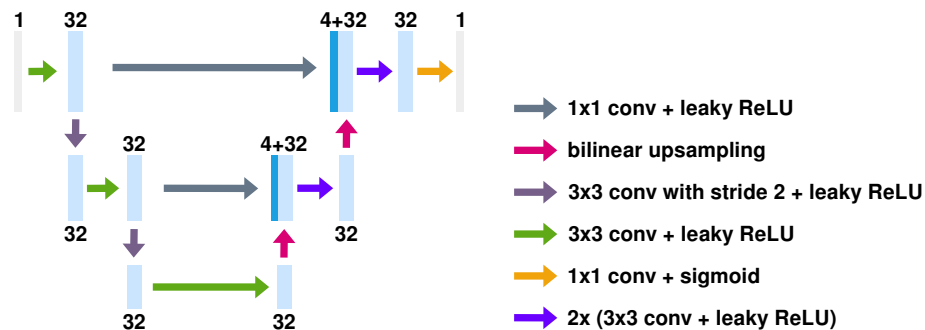
Figure 5: U-net architecture. Each light-blue box corresponds to a multi-channel feature map. The number of channels is set to 32 at every scale. The arrows denote the different operations.

Appendix E. Additional experimental results and analysis

In this section, we include additional experimental figures and discuss them.
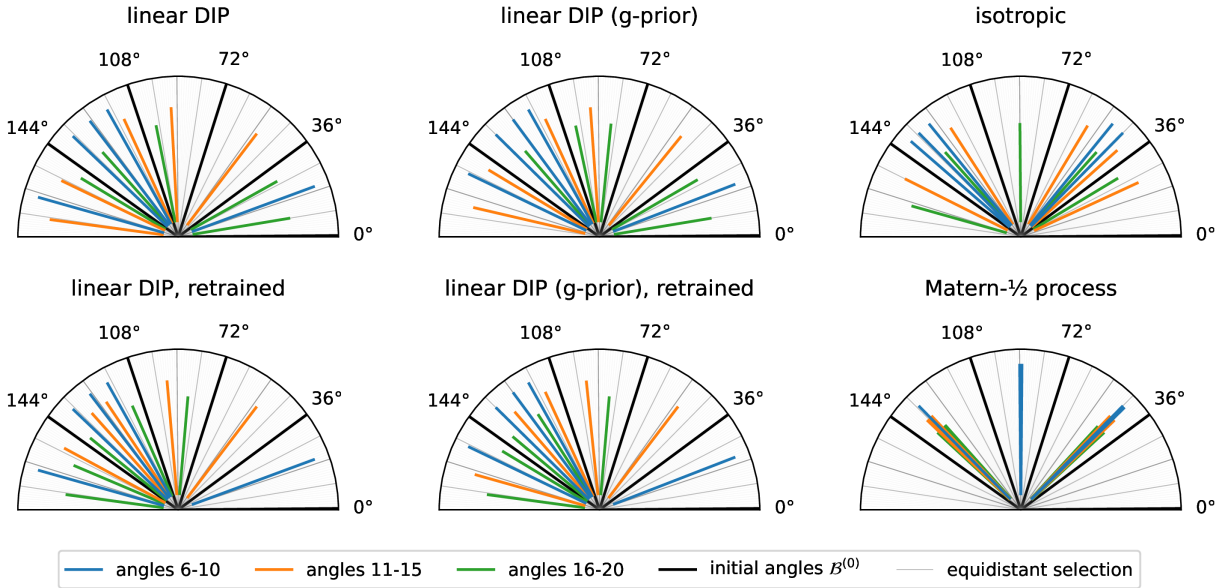
Figure 6: First 20 angles selected by each method under consideration for the example image shown in fig. 1

Figure 6 completes fig. 2 by showing the angles selected by all methods under consideration. Both linear DIP and linear DIP with g-prior choose very similar angles, with the g-prior resulting in a very slightly more diverse angle set. Retraining the linearised DIP every 5 angles to update the basis expansion results in a stronger focus on angles close to the preferential direction. As expected, the differences with the non-retrained DIP are more pronounced for later selected angles, i.e. angles 16-20.

The Matern-$1/2$ model concentrates its selection on oblique angles much more strongly than the isotropic model. This results in a very non-diverse angle set which achieves very poor performance. To understand why this happens we first remark that the Matern-$1/2$ model generalises the isotropic model and the two are equal when the lengthscale is set to $\ell = 0$. We investigate the hyperparameters chosen by the model evidence for the Matern-$1/2$ model and find that for all images the lengthscale is in the range [40-70]. This value is very large relative to the size of the image (128×128) and represents an assumption that the reconstructed image has only 2 or 3 regions with different pixel intensity values. Under this assumption, only taking measurements at 3 different angles is justified.

We verify this explanation by examining the ESE scores assigned by the isotropic and Matern-$1/2$ models to the first 8 angles chosen in fig. 7 and fig. 8 respectively. The isotropic model chooses oblique angles. After each new angle is included in the updated operator $A^{(t)}$, the predictive variance in a region spanning roughly $10°$ around the chosen angles decreases. This is the span of the detector elements. The uncertainty at other angles remains unchanged because the model assumes reconstruction pixels to be uncorrelated. By modelling correlations among detector
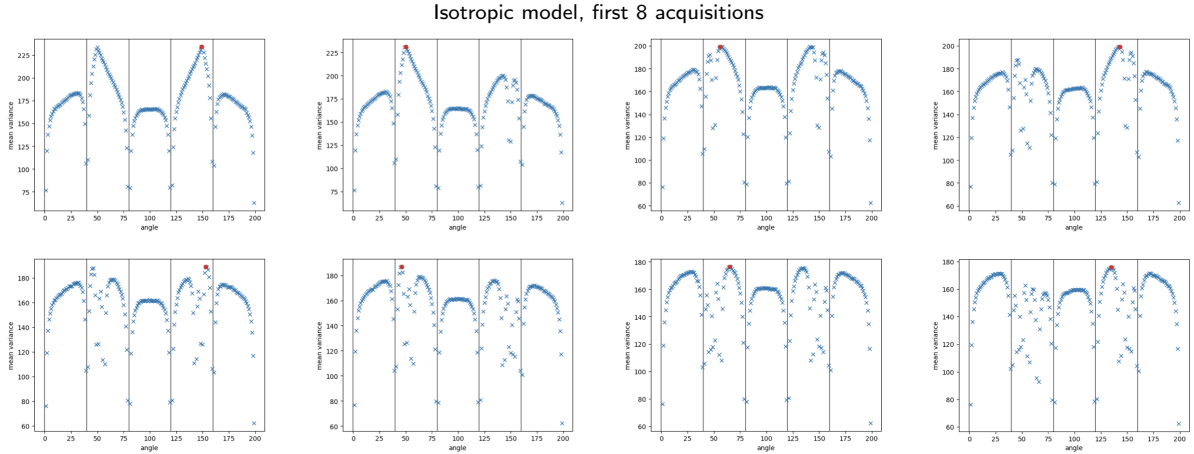
Isotropic model, first 8 acquisitions

Figure 7: Variance assigned to each candidate angle during the first 8 design steps by our Isotropic model.

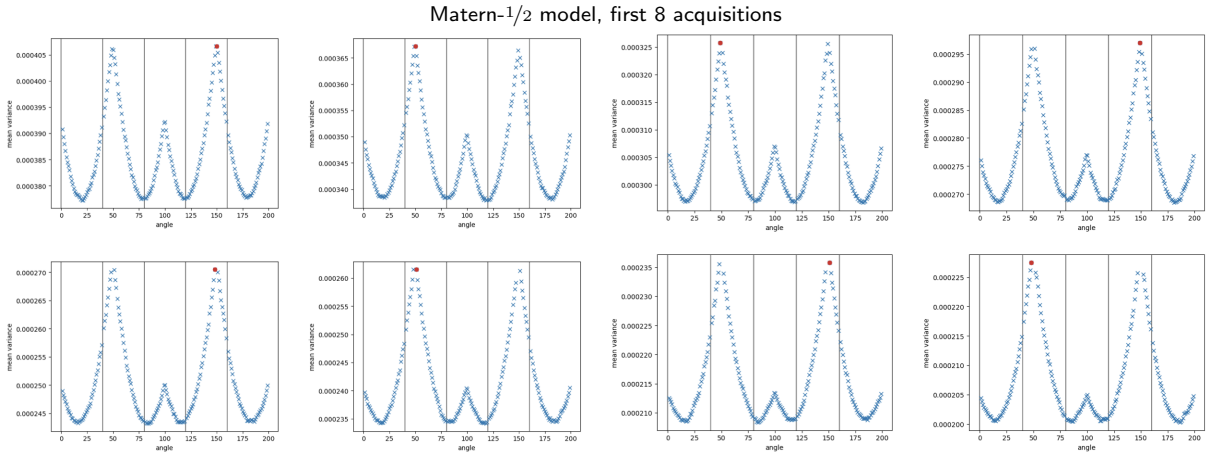Matern-$1/2$ model, first 8 acquisitions

Figure 8: Variance assigned to each candidate angle during the first 8 design steps by our Matern-$1/2$ model.

pixels, each additional angle should reduce the Matern-$1/2$ model's uncertainty in a larger angle range (set via the lengthscale), promoting exploration. However, because the lengthcale, which has overfit the pilot measurements, is very large, each new angle introduced into the operator reduces the predictive variance of every angle almost equally. As a result, the relative assignment of predictive variance in angle space remains roughly constant throughout design steps, and all of the chosen angles become very similar to each other.

Although, it is well known that experimental design is very sensitive to the choice of prior (Feng, 2015; Foster, 2021), the ease with which the relatively very simple Matern-$1/2$ model can

overfit the degree to which this degrades performance was unexpected to us. In future work we will investigate alternative methods for setting model hyperparameters.

Figure 9 and fig. 10 show the variance assigned to each angle in the first 8 acquidition steps on an example image (first image from fig. 4) for the linearised DIP and the linearised DIP with g-prior, respectively. Although the angles selected by the two models are different, both prioritise similar angle regions.
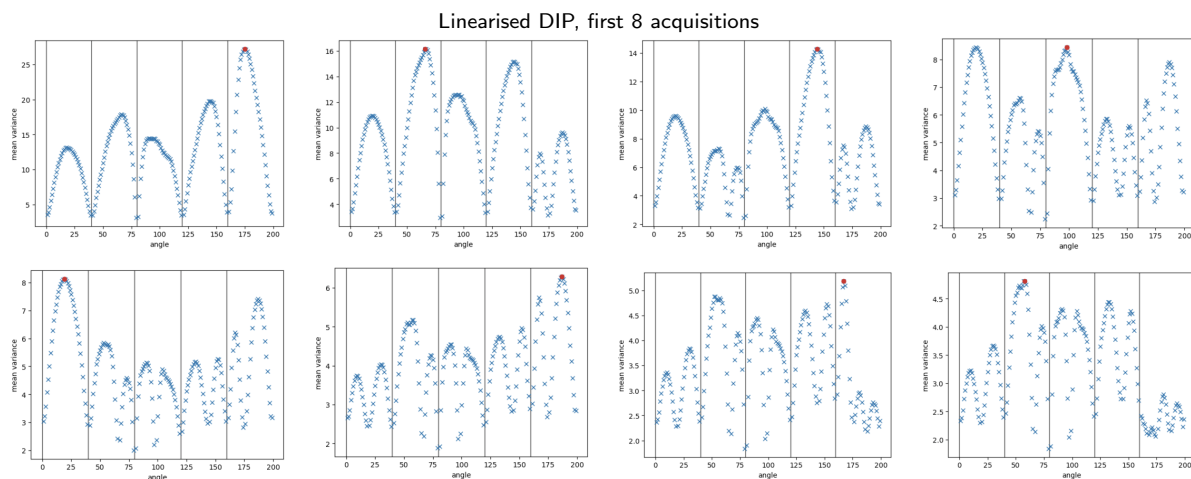

Figure 9: Variance assigned to each candidate angle during the first 8 design steps by our linearised DIP model.
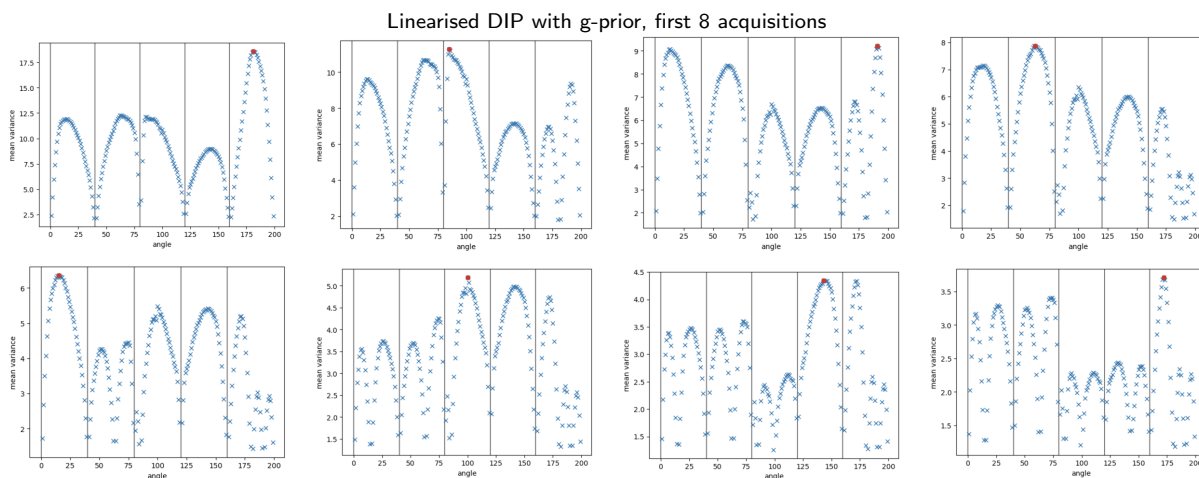

Figure 10: Variance assigned to each candidate angle during the first 8 design steps by our linearised DIP model with the g-prior.

Figure 11 is a more complete version of fig. 3, including the standard error. Given our 30 image runs, we can conclude that the linearised DIP provides a statistically significant improve-

ment over the equidistant baseline up to 20 selected angles. By retraining the DIP Jacobians every 5 angles, we can extend the significant improvements up to 35 scanned angles. In future we aim to make these statements stronger by running more experiments.

Figure 12 shows our findings on measurement data simulated adding 10% noise. The gains from experimental design are slightly reduced in the noisier setting, although the conclusions remain the same. From the EIG expression eq. (6), we can see that noisier measurements should push our score assignment to be more uniform across angles and thus closer to the equidistant baseline.
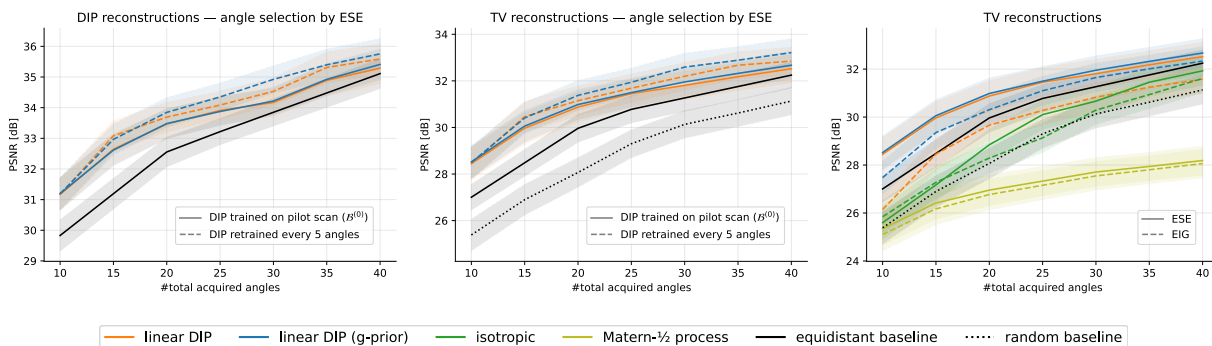
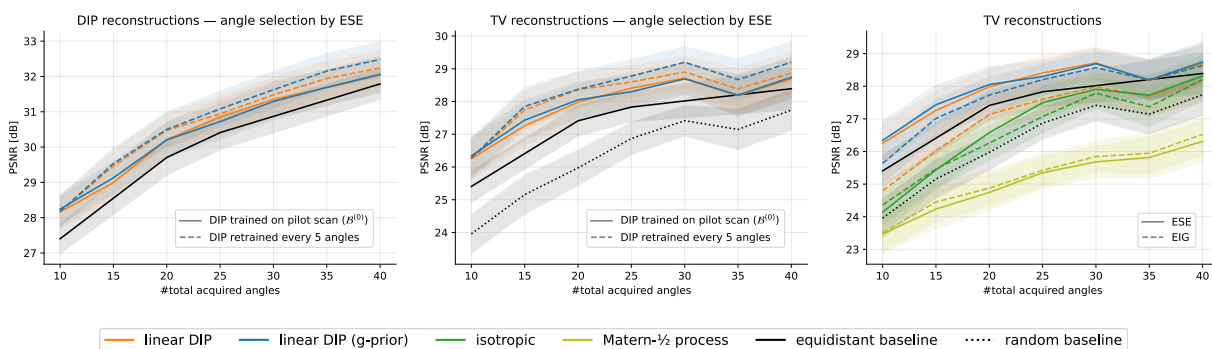Figure 11: Reconstruction PSNR vs n. angles scanned, averaged across 30 images (5% noise).

Figure 12: Reconstruction PSNR vs n. angles scanned, averaged across 30 images (10% noise).