Bi-level Optimization and Implicit Differentiation as a Framework for Optimal Experimental Design in Tomography

Hamid Fathi[1], Alexander Skorikov[1], and Tristan van Leeuwen[1,2]

[1] Computational Imaging Group, Centrum Wiskunde & Informatica (CWI), Science Park 123, 1098 XG Amsterdam, The Netherlands
[2] Mathematical Institute, Utrecht University, Budapestlaan 6, 3584 CD Utrecht, The Netherlands

Abstract. Total Variation (TV) regularized reconstruction is one of the most relevant methods to improve the quality of limited-angle tomographic reconstructions. Nevertheless, the accuracy of computed tomography (CT) reconstructions with a limited number of measurements can be further improved by selecting the most informative acquisition angles. This optimal experimental design (OED) task can be formulated as a bi-level optimization problem, with selecting optimal angle combinations (experimental design parameter) on the upper-level and tomographic reconstruction on the lower-level. However, integrating TV regularized reconstruction into the bi-level optimization approach is non-trivial because of the large number of iterations required for the algorithm convergence, which impedes naive computation of gradients of the upper-level objective with respect to the experimental design parameter. In this work, we address this problem by employing implicit differentiation approach to calculate the upper-level objective gradient. Moreover, we utilize inexact methods to dynamically adjust the accuracy of the lower-level solver, refining the gradient calculation as needed. We demonstrate that this approach makes OED with TV regularized reconstruction applicable to realistic 3D data. Our numerical results demonstrate that the angles selected by our bi-level optimization framework significantly outperform the standard equidistant angular selection. The proposed approach is therefore effective in minimizing experimental time and radiation dose requirements for CT reconstruction of objects benefiting from TV regularization, and can be readily extended to other types of computationally demanding iterative reconstruction algorithms.

Keywords: Bi-level Optimization · Tomographic Imaging · Implicit Differentiation · Optimal Experimental Design.

1 Introduction

Tomographic imaging, a technique widely used in medical diagnostics and materials science, involves reconstructing the internal structure of an object from

a series of projections taken at different angles [1,2,3]. This imaging modality is crucial for providing detailed cross-sectional views of complex structures. However, the reconstruction process often faces challenges related to noise and incomplete data, which can degrade image quality. To mitigate these issues, variational methods considering prior information on the image can be employed [4]. For instance, the total variation (TV) prior promotes sparsity in the gradient domain, effectively preserving edges while reducing noise and artifacts in the reconstructed images [5]. By leveraging regularization, tomographic imaging achieves higher fidelity reconstructions, particularly in cases of limited data or under-sampled projections.

In limited-angle tomography, the quality of the image can be further enhanced if we collect the most informative measurements. To achieve this, optimal experimental design (OED), which has been widely applied in fields such as MRI and seismic imaging, provides a systematic approach to improving experimental efficiency [7,8]. Bayesian OED is a mathematical framework that facilitates the design of experimental parameters while minimizing experimental costs [6]. Using prior knowledge and Bayesian inference, it refines the selection of experimental settings, thereby improving data collection efficiency and decision-making effectiveness. One can use ground truth images as prior knowledge to approximate the design parameters by minimizing the expected error between reconstruction and actual ground truth. This is known as the A-optimality criterion in OED literature papers [9]. The OED optimization problem in tomographic imaging can be formulated as a stochastic bi-level optimization problem as follows,

$$\hat{\boldsymbol{p}} := \underset{\boldsymbol{p}\in\Omega}{\arg\min}\ \mathbb{E}\ \tfrac{1}{2}\|\hat{\boldsymbol{x}}(\boldsymbol{p},\boldsymbol{y}^{\mathrm{gt}}) - \boldsymbol{x}^{\mathrm{gt}}\|^2, \tag{1}$$

$$s.t.\ \hat{\boldsymbol{x}}(\boldsymbol{p}) := \underset{\boldsymbol{x}}{\arg\min}\ g(\boldsymbol{x},\boldsymbol{y}^{\mathrm{gt}},\boldsymbol{p}). \tag{2}$$

The expectation is taken over the training data $(\boldsymbol{x}^{\mathrm{gt}},\boldsymbol{y}^{\mathrm{gt}})$ and $\boldsymbol{p}$ denotes the design parameter. The upper-level objective function evaluates the quality of the lower-level solution $\hat{\boldsymbol{x}}$ with respect to ground truth $\boldsymbol{x}^{\mathrm{gt}}$ while the feasible set Ω captures constraints on the experimental setting. The lower-level problem (2) aims to retrieve $\hat{\boldsymbol{x}}$ from given data $\boldsymbol{y}^{\mathrm{gt}}$ for given experimental settings $\boldsymbol{p}$. The function g typically includes data-fidelity and regularization terms and it is assumed to be strongly convex in $\boldsymbol{x}$ to guarantee the uniqueness of $\hat{\boldsymbol{x}}$.

Bi-level problems, such as the one outlined above, are challenging to solve because of their non-convex nature and computational complexity. Moreover, in tomographic imaging we are typically faced with a high-dimensional design space. This precludes the use of model-free methods like grid and random search [10] and model-based methods such as Bayesian optimization [11]. Therefore, we focus on projected-gradient methods that find an optimal set of design parameters by utilizing the gradient of the upper-level objective. In order to compute this gradient, we need to compute the derivatives of $\hat{\boldsymbol{x}}$ w.r.t. $\boldsymbol{p}$. To accurately compute derivatives for an approximate lower-level solution, two common approaches are reverse-mode automatic differentiation (AD) [12] and implicit differentiation (ID) [13]. Although both schemes can be efficiently implemented

using automatic differentiation tools, reverse-mode AD usually has a (much) higher memory cost than ID [14]. This distinction becomes particularly important in the context of large-scale problems, where iterative solvers are often employed to solve the lower-level problem. These solvers generally require numerous iterations to approximate the solution accurately, making the gradient calculations with unrolling prohibitively expensive. While modified AD-based approaches exist to approximate derivatives without storing the full iteration history [15,16], they introduce additional approximation errors and computational overhead, making them less suitable for large-scale bi-level optimization problems.

To further speed up the computations, inexact methods have been considered [17]. Here, the lower-level problem is solved up to a dynamically adjusted tolerance in order to improve the computational efficiency of the upper-level objective gradient calculation.

1.1 Contributions and outline

In this paper, we optimize a set of acquisition angles tailored to a specific dataset, which can later be used to scan similar objects with the same orientation. We pose experimental design for 3D limited-angle tomography as a bi-level optimization problem. The lower-level problem involves 3D tomographic reconstruction using a Total Variation prior and for the upper-level problem we adopt the adaptive inexact method proposed by [17]. We use the implicit differentiation approach to efficiently calculate the gradient of the upper-level problem with low memory imprint, making the method feasible for large-scale problems.

The remainder of the paper is organized as follows. In section 2 we provide some background on the use of implicit differentiation for bi-level problems. In section 3, we pose the experimental design problem for tomographic imaging and given details on the implementation. Numerical results are presented in section 4, the discussion is provided in Section 5, and section 6 concludes the paper.

2 Background

2.1 Notation

Throughout, scalars and vector elements are denoted by lower-case italic letters, vectors by lower-case boldface letters, and matrices by upper-case boldface letters. The notation $\cdot^\top$ stands for the transpose of either a vector or matrix. By $\|\boldsymbol{x}\| = \sqrt{\sum_{i=1}^{n} x_i^2}$ we denote the ℓ_2-norm of a vector and $\#$ denotes the cardinality of a vector. (i.e., the number of non-zero elements).

2.2 Bi-level framework

For ease of notation, we consider it for a single data-point $(\boldsymbol{x}^{\mathrm{gt}}, \boldsymbol{y}^{\mathrm{gt}})$ and rewrite the bi-level optimization in (1) as

$$\hat{\boldsymbol{p}} := \underset{\boldsymbol{p}\in\Omega}{\arg\min}\ f(\boldsymbol{p}), \tag{3}$$

$$s.t.\ \hat{\boldsymbol{x}}(\boldsymbol{p}) = \mathcal{A}(\hat{\boldsymbol{x}}(\boldsymbol{p}), \boldsymbol{p}), \tag{4}$$

where $f(\boldsymbol{p}) = \ell(\hat{\boldsymbol{x}}(\boldsymbol{p}))$ and (4) represents the optimality condition for the lower-level problem. Specifically, the minimizer to the lower-level problem (2) corresponds to the fixed point of $\mathcal{A}(\cdot, \boldsymbol{p})$. Given that the lower-level problem is strongly convex and assuming small enough step size, the operator $\mathcal{A}(\cdot, \boldsymbol{p})$ is a contraction, guaranteeing convergence of the corresponding fixed-point iteration to a unique fixed point representing the minimizer of the problem [14]. Furthermore, we assume that both ℓ and $\mathcal{A}$ are continuously differentiable, allowing us to leverage smoothness properties in the analysis.

The gradient of the upper-level function f now reads

$$\nabla_{\boldsymbol{p}} f(\boldsymbol{p}) = \partial_{\boldsymbol{p}} \hat{\boldsymbol{x}}(\boldsymbol{p})^{\top} \nabla_{\boldsymbol{x}} \ell(\hat{\boldsymbol{x}}(\boldsymbol{p})), \tag{5}$$

with $\partial_{\boldsymbol{p}} \hat{\boldsymbol{x}}$ denoting the Jacobian of $\hat{\boldsymbol{x}}$ w.r.t. $\boldsymbol{p}$. Note that in practice we only have access to an approximate solution $\tilde{\boldsymbol{x}}(\boldsymbol{p}) \approx \hat{\boldsymbol{x}}(\boldsymbol{p})$. One can proceed with unrolling method which can be efficiently implemented by reverse-mode AD. In *Reverse-mode automatic differentiation (AD)*, we apply automatic differentiation to the fixed-point iteration $\boldsymbol{x}_{k+1} = \mathcal{A}(\boldsymbol{x}_k, \boldsymbol{p})$. The advantage is that we can efficiently compute the exact derivatives of $\boldsymbol{x}_k$ w.r.t. $\boldsymbol{p}$ for any k and hence we can compute the exact derivative of the approximate objective $f(\boldsymbol{p}) \approx \ell(\tilde{\boldsymbol{x}}(\boldsymbol{p}))$. However, if the solver has not fully converged, this may still be an inexact approximation of the true gradient of $f(\boldsymbol{p})$. The main drawback of reverse-mode AD is its high memory cost, as it requires storing all intermediate iterates. In *Implicit differentiation (ID)*, we differentiate the fixed-point relation directly and employ the Implicit Function Theorem to compute the required Jacobian as

$$\partial_{\boldsymbol{p}} \hat{\boldsymbol{x}}(\boldsymbol{p}) = \left(\boldsymbol{I} - \partial_{\boldsymbol{x}} \mathcal{A}(\boldsymbol{x}, \boldsymbol{p})\right)^{-1} \partial_{\boldsymbol{p}} \mathcal{A}(\boldsymbol{x}, \boldsymbol{p})\big|_{\boldsymbol{x}=\hat{\boldsymbol{x}}(\boldsymbol{p})}. \tag{6}$$

The advantage is that it has a much smaller memory-imprint than AD. The disadvantage is that we need to solve a linear system to calculate the derivative.

3 OED in computed tomography

We consider the optimal experimental design (OED) problem for sparse-angle computed tomography (CT). Here, the goal is to select a subset of $K \ll l$ angles on which to base the tomographic reconstruction. Angle-selection will be encoded by the vector $\boldsymbol{w} \in \mathbb{R}^l$. We begin by introducing the image reconstruction problem for a particular design parameter followed by an upper-level optimization over the design parameter.

3.1 The lower-level problem

CT can be expressed as a linear inverse problem of retrieving a 3D image $\boldsymbol{x} \in \mathbb{R}^n$ from projection data $\boldsymbol{y} \in \mathbb{R}^m$, related via

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}, \tag{7}$$

where $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ denotes the forward operator. The forward operator exhibits a block-structure consisting of l blocks of rows corresponding to projections taken from their corresponding angles. Angle-selection is then modeled by weighting the i^{th} block of rows with its corresponding weight w_i, or

$$\boldsymbol{T}_{\boldsymbol{w}}(\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{T}_{\boldsymbol{w}}(\boldsymbol{y}), \tag{8}$$

where $\boldsymbol{T}_{\boldsymbol{w}} : \mathbb{R}^{m \times m} = \text{diag}(\boldsymbol{w}) \otimes \boldsymbol{I}$ is an $m \times m$ diagonal matrix that applies the weight w_i to the i^{th} block of rows.

Variational methods reconstruct $\boldsymbol{x}$ by minimizing an objective function comprising a fidelity term, which ensures consistency with the measurements, and a regularization term that incorporates prior knowledge about the image domain, addressing the ill-posedness of the problem. The corresponding lower-level problem then reads

$$\hat{\boldsymbol{x}}(\boldsymbol{p}) = \arg\min_{\boldsymbol{x}} g(\boldsymbol{x}, \boldsymbol{p}) := \tfrac{1}{2}\|\boldsymbol{T}_{\boldsymbol{w}}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y})\|^2 + r(\boldsymbol{x}, \boldsymbol{\lambda}), \tag{9}$$

where r is the regularization term controlled by $\boldsymbol{\lambda}$ that the dimension can change depending on variables, and $\boldsymbol{p} = (\boldsymbol{w}, \boldsymbol{\lambda})$. This means that we jointly optimize over the regularization parameter(s) and the weights in the bi-level problem. A natural choice for $\mathcal{A}$ is

$$\mathcal{A}(\boldsymbol{x}, \boldsymbol{p}) = \boldsymbol{x} - \alpha \nabla_{\boldsymbol{x}} g(\boldsymbol{x}, \boldsymbol{p}), \tag{10}$$

with $\alpha \in \mathbb{R}$ chosen to ensure $\mathcal{A}(\cdot, \boldsymbol{p})$ is a contraction for all $\boldsymbol{p} \in \Omega$. Note that we may choose to use a more efficient iterative method to solve (9) while still using this choice for $\mathcal{A}$ to compute the Jacobian. Below, we present a few examples of commonly-used regularizers.

Example 1: One can employ $r(\boldsymbol{x}, \lambda) = \frac{1}{2}\exp(\lambda)\|\boldsymbol{x}\|^2$ as the regularization term in case $\boldsymbol{A}$ is ill-conditioned or rank-deficient to stabilize the solution of the optimization. The objective function in (9) becomes $\exp(\lambda)$-strongly convex and guarantees a unique solution.

Example 2: Smoothed Total Variation regularization incorporated with a ℓ_2 norm (strong convexity term), $r(\boldsymbol{x}, \boldsymbol{\lambda}) = \exp(\lambda_1)\text{TV}_{\exp(\lambda_2)}(\boldsymbol{x}) + \frac{1}{2}\exp(\gamma)\|\boldsymbol{x}\|^2$ can be used as an alternative to TV to ensure that g is smooth by design. Here, $\boldsymbol{\lambda} \in \mathbb{R}^2$ denotes the regularization and smoothing parameters that can be optimized through the learning process. In addition, it is less strict in enforcing piecewise constant solutions and the smoothing parameter can be optimized to fit a particular dataset.

Example 3: The non-smooth Total Variation regularization incorporated with a ℓ_2 term (strong convexity term), $r(\boldsymbol{x}, \lambda) = \exp(\lambda)\mathrm{TV}(\boldsymbol{x}) + \frac{1}{2}\exp(\gamma)\|\boldsymbol{x}\|^2$ can be used in CT reconstruction to address ill-posedness and improving image quality, under the assumption of piecewise constant images. As the objective g in (9) is non-smooth, we replace it by $g_\eta(\boldsymbol{x}, \boldsymbol{p}) = \frac{1}{2}\|\boldsymbol{T}_{\boldsymbol{w}}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y})\|^2 + \exp(\lambda)\mathrm{TV}_\eta(\boldsymbol{x}) + \frac{1}{2}\exp(\gamma)\|\boldsymbol{x}\|^2$, with TV_η a smooth approximation of TV, for the purpose of defining $\mathcal{A}$. Note that this case is different from Example 2, since the lower level objective is only smoothed while defining $\mathcal{A}$, and the smoothing parameter $\eta \in \mathbb{R}$ is fixed.

3.2 The upper-level problem

We aim to optimize the design parameter $\boldsymbol{w}$ while simultaneously determining the regularization parameter $\boldsymbol{\lambda}$. Let $\hat{\boldsymbol{x}}(\boldsymbol{p}, \boldsymbol{y})$ represent the reconstructed image corresponding to $\boldsymbol{p} = (\boldsymbol{w}, \boldsymbol{\lambda})$ and data $\mathbf{y}$. Given a data set $\{(\boldsymbol{x}_i^{\mathrm{gt}}, \boldsymbol{y}_i^{\mathrm{gt}})\}_{i=1}^N$ the objective is to identify the optimal $\hat{\boldsymbol{p}}$ that maximizes the quality of the reconstructed CT images under certain constraints. The resulting upper-level optimization problem reads

$$\min_{\boldsymbol{p}\in\Omega} \ \frac{1}{2N}\sum_{i=1}^{N} \|\hat{\boldsymbol{x}}_i(\boldsymbol{p}) - \boldsymbol{x}_i^{\mathrm{gt}}\|^2, \tag{11}$$

where $\hat{\boldsymbol{x}}_i(\boldsymbol{p}) \equiv \hat{\boldsymbol{x}}(\boldsymbol{p}, \boldsymbol{y}_i^{\mathrm{gt}})$ and

$$\Omega = \left\{ (\boldsymbol{w}, \boldsymbol{\lambda}) \mid \sum_i w_i = K, 0 \le w_i \le 1 \right\},$$

is constructed to promote solutions with K active sources and positive $\boldsymbol{\lambda}$. The simplex constraint on $\boldsymbol{w}$ is in fact a convex relaxation of the binary constraint $w_i \in \{0, 1\}, \sum_i w_i = K$ and has been proven useful in OED for tomographic imaging [18]. Note that we cannot guarantee that this relaxation will yield binary solutions. In such cases, we convert $\boldsymbol{w}$ to a binary one by setting the K largest elements to one and assigning zeros to the remaining elements.

3.3 Algorithm

The overall algorithm now needs three ingredients: *i)* an inner solver to solve the lower-level problem (9) up to a given tolerance, *ii)* a linear solver to solve the linear system in (6) up to a given tolerance, and *iii)* a procedure to update the parameters $\boldsymbol{p}$ based in inexact gradient information and taking the constraints in to account. In essence the resulting algorithm is based on the following update step

$$\boldsymbol{p}^{(k+1)} = \mathcal{P}_\Omega\left(\boldsymbol{p}^{(k)} - \boldsymbol{D}_k \boldsymbol{g}^{(k)}\right), \tag{12}$$

with

$$\boldsymbol{g}^{(k)} = \sum_{i=1}^{N} \left(\partial_{\boldsymbol{p}} \mathcal{A}(\tilde{\boldsymbol{x}}_i^{(k)}, \boldsymbol{p}^{(k)})\right)^\top \tilde{\boldsymbol{r}}_i^{(k)}, \tag{13}$$

where $\tilde{\boldsymbol{r}}_i^{(k)}$ is solved from

$$\left(\boldsymbol{I} - \partial_{\boldsymbol{x}}\mathcal{A}(\tilde{\boldsymbol{x}}_i^{(k)}, \boldsymbol{p})\right)\boldsymbol{r}_i = \left(\tilde{\boldsymbol{x}}_i^{(k)} - \boldsymbol{x}_i^{\mathrm{gt}}\right), \tag{14}$$

using the Conjugate Gradient method with relative tolerance δ_k. The solution to the inner problem $\tilde{\boldsymbol{x}}_i^{(k)} = \tilde{\boldsymbol{x}}_i(\boldsymbol{p}^{(k)})$ is obtained by solving (9) up to a relative tolerance of ϵ_k, i.e.

$$\|\nabla_{\boldsymbol{x}} g(\tilde{\boldsymbol{x}}_i^{(k)}, \boldsymbol{p}^{(k)})\| \leq \mu\epsilon_k$$

for a fixed problem dependent parameter μ the lower bounds the strong convexity constant of $g(\cdot, \boldsymbol{p})$. The matrix $\boldsymbol{D}_k$ is a diagonal weighting to account for differences in scaling between $\boldsymbol{w}$ and $\boldsymbol{\lambda}$ and the two stepsizes are updated using a backtracking linesearch. The projection operator involves a projection on the simplex constraint, which can be done efficiently as outlined by [18]. To control the tolerances (ϵ_k, δ_k) for the inner solver and linear solver, we adopt the adaptive MAID algorithm for bi-level learning presented in [17]. As inner solvers we use either Nesterov's accelerated gradient method [19] (in case the objective is smooth) or the Chambolle-Pock method (for the non-smooth TV regularizer) [20].

4 Numerical Experiments

In this section, we present the optimal acquisition angles in CT obtained by our proposed bi-level framework. We first outline the experimental settings and then present the results.

4.1 Experimental settings

Implementation with PyTorch To implement the forward operator we use Tomosipo [21], a PyTorch-compatible interface for the ASTRA toolbox [22]. We utilize the available implementation of Chambolle-Pock algorithm for 3D image reconstruction[3] as described in [5]. Furthermore, Nesterov's accelerated gradient method is implemented and accessible within the Tomosipo algorithms package [4].

As the CG only requires matrix-vector products to solve the linear system in (14), we can efficiently compute the gradient in (13) using automatic differentiation (AD) to perform Jacobian-vector products (JVPs).

For the MAID algorithm, the parameter μ as the lower bounds strong convexity of g is set to $\mu = \exp(\gamma)$ while $\gamma = -3$ for OED with TV-regularized reconstruction and $\mu = \exp(\lambda)$ for OED with ℓ_2 regularized reconstruction while λ is the regularization parameter. In OED with TV-regularized reconstruction, we consider $\eta = \exp(-6)$ as the smoothing parameter and the initial accuracies of the lower-level solver and the CG solver are set to $\epsilon_0 = \exp(6)$ and $\delta_0 = \exp(3)$ while in OED with ℓ_2 regularized reconstruction $\epsilon_0 = 10^3$ and $\delta_0 = 10$ with a maximum of 500 iterations for each.

[3] `https://github.com/schoonhovenrichard/AutodiffCTWorkflows`

[4] `https://github.com/ahendriksen/ts_algorithms`

Data set As ground truth we use a CT scan of a cooling device shown in figure 1. The phantom is represented as a binary 3D volumetric image with dimensions $58 \times 224 \times 224$. We consider a CT setup with a cone beam geometry and a detector with 400×400 pixels. We generate a set of projections collected from 100 angles, evenly distributed between $0°$ and $360°$. The objective is to acquire $K < 100$ most informative angles to reconstruct the 3D image using regularized reconstruction either ℓ_2 or TV regularization.

(a) (b) (c) (d)

Fig. 1: Ground truth of the 3D phantom and central slices from different views.

Evaluation For all experiments we first jointly optimize over the parameters $\boldsymbol{p} = (\boldsymbol{w}, \lambda)$ using either ℓ_2 or TV-regularization. We then binarize the resulting solution $\boldsymbol{w}$ by setting the largest K elements to one and the remaining elements to zero. As a reference we will also use an equidistant design with the same number of angles. To ensure a fair (best-case) comparison of the different designs (ℓ_2, TV, or equidistant), we perform a TV-regularized reconstruction where the regularization parameter λ is optimized again (for fixed $\boldsymbol{w}$). The resulting reconstructed volumes are then compared qualitatively and quantitatively by computing the structural similarity index measure (SSIM) and mean square error (MSE) between the reconstruction and the ground-truth.

Finally, to investigate the robustness of the designs $\boldsymbol{w}$ and its corresponding regularization parameter $\boldsymbol{\lambda}$ obtained as outlined above, we evaluate their performance on noisy data.

4.2 Numerical results

In figure 2, we present the convergence of the method using MAID algorithm in case of OED with TV regularization. The optimized designs for varying K are shown in figure 3. The corresponding TV-regularized reconstructions for noiseless data are shown in figure 4. The results in terms of the reconstruction error are summarized in table 1.

5 Discussion

Here, we would like to highlight a few points related to our proposed bi-level framework: *i)* While we build upon an existing framework, ensuring robust convergence requires incorporating certain heuristics. For example, the optimization

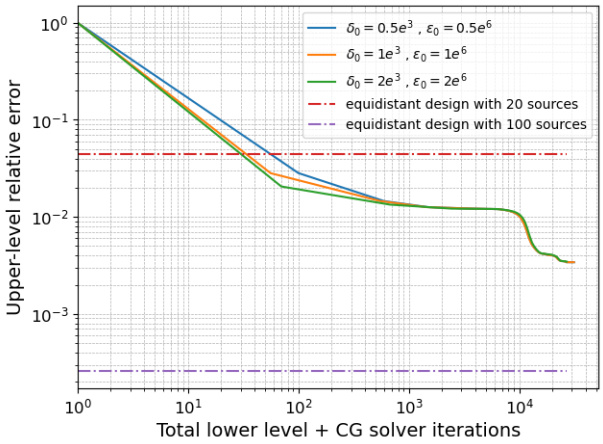


Fig. 2: Upper level relative error over the total number of iterations using TV-regularization with $K = 20$. While the MAID algorithm is initialized with three different sets of tolerance values (ϵ_0, δ_0), all configurations converge to the same level of accuracy. Two dashed lines present the relative errors of $K = 20, 100$ equidistant sources as upper and lower bounds of errors, respectively.

	$K = 10$	$K = 20$	$K = 30$
Equidistant	0.41	0.49	0.62
ℓ_2-design	0.55	0.59	0.62
TV-design	0.61	0.68	0.67

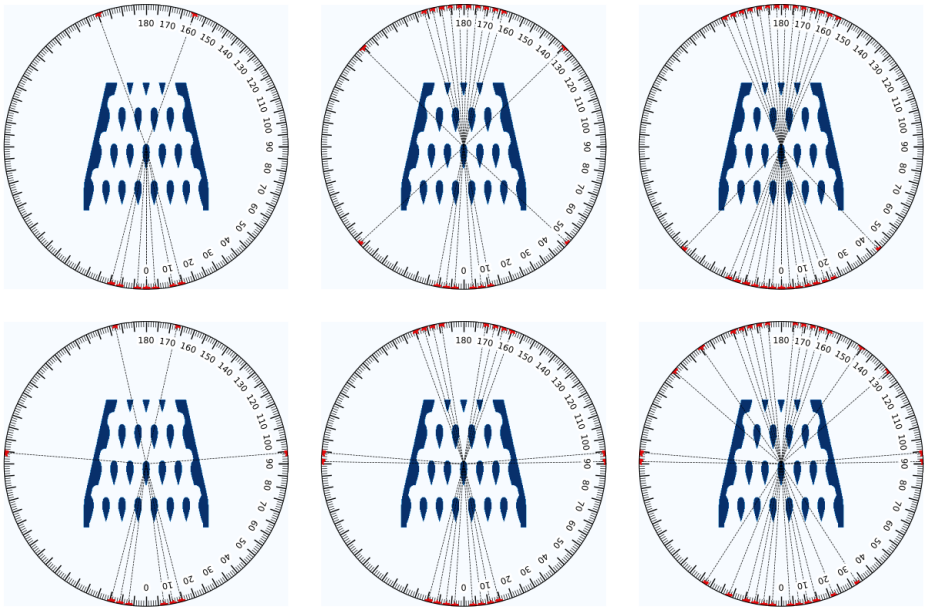
SSIM Results

	$K = 10$	$K = 20$	$K = 30$
Equidistant	5.7×10^{-2}	2.2×10^{-2}	0.3×10^{-2}
ℓ_2-design	1.7×10^{-2}	7.3×10^{-3}	4.4×10^{-3}
TV-design	4.7×10^{-3}	1.9×10^{-3}	0.8×10^{-3}

MSE Results

Table 1: Performance comparison of different angular selection methods in terms of SSIM and MSE for $K = 10$, $K = 20$, and $K = 30$.

process is initialized with a very small regularization value, and a small regularization parameter is added into the CG solver to aid convergence. Establishing formal convergence guarantees when using these heuristic techniques is a matter for a future research. *ii)* Since the upper-level objective is a non-convex function, there is no guarantee to achieve the global minimum. One potential approach to improve the robustness of our method is employing multiple random initialization of parameters. *iii)* The computational cost of implementation increases substantially when handling large-scale training datasets. To address this challenge, alternative strategies such as stochastic gradient descent (SGD) can be utilized. Rather than processing the entire dataset in each iteration, SGD performs parameter updates by randomly selecting a subset of data points, thereby reducing computational complexity. *iv)* Other regularization types can be incorporated in the lower-level objective of the proposed framework, provided they satisfy the assumptions outlined in this paper.



$K = 10$ $K = 20$ $K = 30$

Fig. 3: Designs resulting from ℓ_2-regularization (top) and TV-regularization (bottom) for various K. We see that the optimized angles align well with the dominant edges of the phantom, as expected. For intermediate K, the ℓ_2 design seems biased towards angles around zero, whereas the TV-designs tend to include angles around ± 10 to align with the edges in the phantom.

.

6 Conclusion

In this research, we propose a bi-level framework for optimal experimental design in computed tomography, specifically optimal angular selection. Large-scale bi-level optimization problems often suffer from inaccuracies in the lower-level solution, which can propagate as inexact gradient computations at the upper-level, hindering convergence. To mitigate this issue, our framework incorporates implicit differentiation as an inexact approach to refine the lower-level accuracy without encountering memory limitations. By achieving a balance between computational efficiency and solution accuracy, the proposed framework is well-suited for large-scale and computationally demanding applications in tomography.

Acknowledgments. Our work was supported by funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 956172 (xCTing).

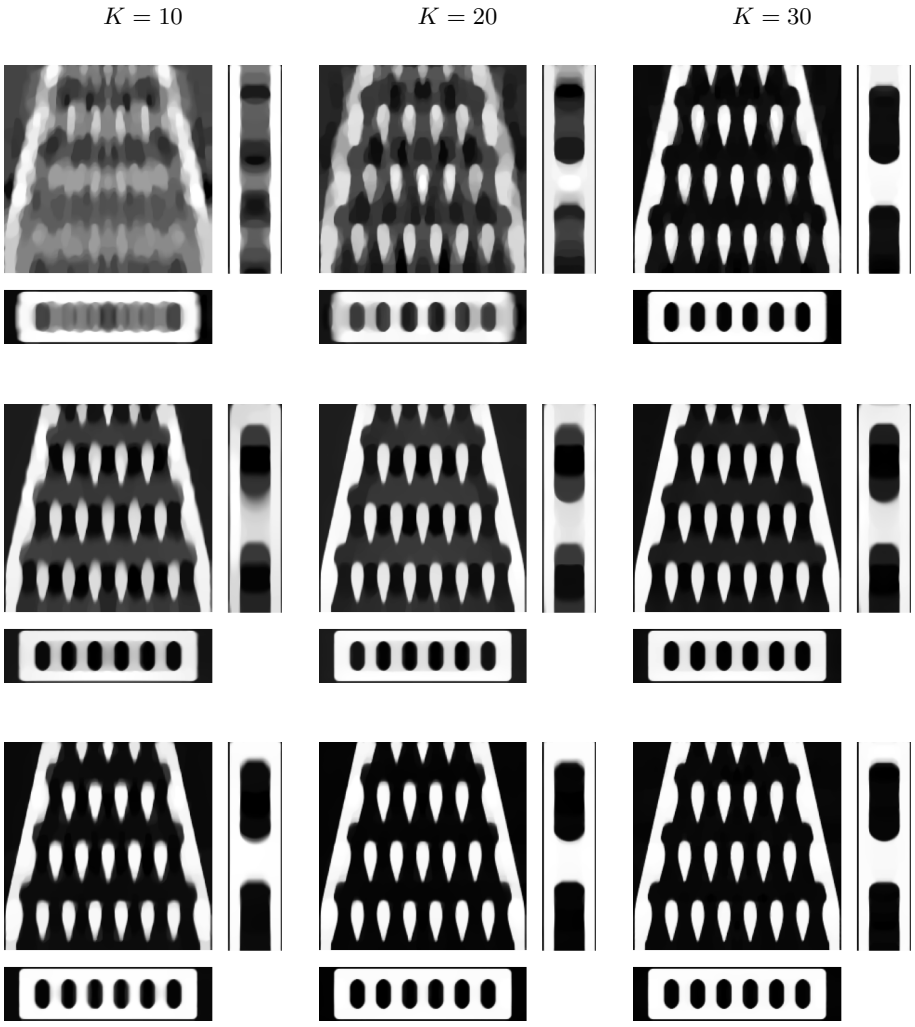


Fig. 4: At the top we show the central slices of the reconstruction with equidistant angles. In the middle and bottom rows, we display central slices of the reconstructed volumes corresponding to the designs shown in figure 3, respectively.

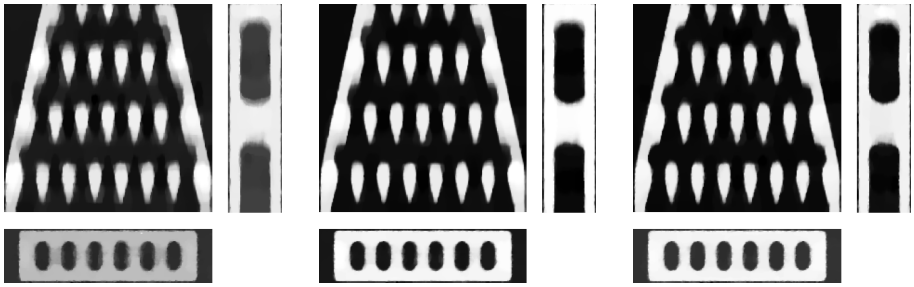


Fig. 5: Central slices of reconstructed volumes obtained from TV-regularized designs using noisy measurements with an SNR of 15 are shown. Comparing these results to the corresponding noise-free reconstructions in figure 4 (bottom row), we observe that the method demonstrates greater robustness for larger K values.

References

1. Withers, P. J., Bouman, C., Carmignato, S., Cnudde, V., Grimaldi, D., Hagen, C. K., Maire, E., Manley, M., Du Plessis, A., Stock, S. R.: X-ray computed tomography. *Nature Reviews Methods Primers*, **1**(1), 18 (2021).
2. Weyland, M., Midgley, P. A.: Electron tomography. *Materials Today*, **7**(12), 32–40 (2004).
3. Herman, G. T.: Fundamentals of computerized tomography: image reconstruction from projections. *Springer Science* & Business Media (2009).
4. Hansen, P. C., Jørgensen, J., Lionheart, W. R. B. (eds.): *Computed Tomography: Algorithms, Insight, and Just Enough Theory*. Society for Industrial and Applied Mathematics, Philadelphia, PA (2021).
5. Sidky, E. Y., Jørgensen, J. H., Pan, X.: Convex optimization problem prototyping for image reconstruction in computed tomography with the Chambolle–Pock algorithm. *Physics in Medicine & Biology*, **57**(10), 3065 (2012).
6. Chaloner, K., Verdinelli, I.: Bayesian Experimental Design: A Review. *Statistical Science*, **10**(3), 273–304 (1995).
7. Haldar, J. P., Kim, D.: OEDIPUS: An experiment design framework for sparsity-constrained MRI. *IEEE transactions on medical imaging*, **38**(7), 1545-1558 (2019).
8. Krampe, V., Edme, P., Maurer, H.: Optimized experimental design for seismic full waveform inversion: A computationally efficient method including a flexible implementation of acquisition costs. *Geophysical Prospecting*, **69**(1), 152-166 (2021).
9. Haber, E., Magnant, Z., Lucero, C., Tenorio, L.: Numerical methods for A-optimal designs with a sparsity constraint for ill-posed inverse problems. *Computational Optimization and Applications*, **52**, 293–314 (2012).
10. Bergstra, J., Bengio, Y.: Random search for hyper-parameter optimization. *J. Mach. Learn. Res.*, **13**, 281–305 (2012).
11. Snoek, J., Larochelle, H., Adams, R. P.: Practical Bayesian Optimization of Machine Learning Algorithms. In: Pereira, F., Burges, C. J., Bottou, L., Weinberger, K. Q. (eds.) *Advances in Neural Information Processing Systems*, vol. 25, Curran Associates, Inc., (2012).

12. Christianson, B.: Reverse accumulation and attractive fixed points. *Optimization Methods & Software*, **3**, 311–326 (1994).
13. Bengio, Y.: Gradient-Based Optimization of Hyperparameters. *Neural Comput.*, **12**(8), 1889–1900 (2000).
14. Grazzi, R., Franceschi, L., Pontil, M., Salzo, S.: On the Iteration Complexity of Hypergradient Computation. In: Daumé, H., Singh, A. (eds.) *Proceedings of the 37th International Conference on Machine Learning*, vol. 119, pp. 3748–3758, PMLR (2020).
15. Bolte, J., Pauwels, E., Vaiter, S.: One-step differentiation of iterative algorithms. *arXiv, abs/2305.13768* (2023).
16. Mehmood, S., Ochs, P.: Automatic Differentiation of Some First-Order Methods in Parametric Optimization. *International Conference on Artificial Intelligence and Statistics* (2019).
17. Salehi, M. S., Mukherjee, S., Roberts, L., Ehrhardt, M. J.: An adaptively inexact first-order method for bilevel optimization with application to hyperparameter learning. *arXiv:2308.10098* (2024).
18. Fathi, H., van Leeuwen, T.: Source Design Optimization for Depth Image Reconstruction in X-ray Imaging. *Mathematics*, **12**(10), 1524 (2024).
19. Nesterov, Y.: *Introductory Lectures on Convex Optimization: A Basic Course*. Springer Publishing Company, Incorporated, 1st edn. (2014).
20. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, **40**, 120–145 (2011).
21. Hendriksen, A., Schut, D., Palenstijn, W. J., Viganò, N., Kim, J., Pelt, D., van Leeuwen, T., Batenburg, K. J.: Tomosipo: Fast, Flexible, and Convenient 3D Tomography for Complex Scanning Geometries in Python. *Optics Express*, **29**(21), 32423–32441 (2021).
22. van Aarle, W., Palenstijn, W. J., Cant, J., Janssens, E., Bleichrodt, F., Dabravolski, A., De Beenhouwer, J., Batenburg, K. J., Sijbers, J.: Fast and flexible X-ray tomography using the ASTRA toolbox. *Optics Express*, **24**(22), 25129–25147 (2016).