Differential optimisation for task- and constraint-aware design of particle detectors

Giles. C. Strong[*]
MODE Collaboration
INFN-Padova, Italy

Maxime Lagrange
MODE Collaboration
Université catholique de Louvain, Belgium

Aitor Orio
Muon Tomography Systems S.L., Spain
Universidad de Cantabria, Spain

Anna Bordignon
University of Padova, Italy

Florian Bury
University of Bristol, England

Tommaso Dorigo
MODE Collaboration
INFN-Padova, Italy

Andrea Giammanco
MODE Collaboration
Université catholique de Louvain, Belgium

Mariam Heikal
American University of Beirut, Lebanon

Jan Kieseler
MODE Collaboration
Karlsruhe Institute of Technology, Germany

Max Lamparth
MODE Collaboration
Technische Universität München, Germany

Pablo Martínez Ruíz del Árbol
MODE Collaboration
Instituto de Física de Cantabria, Spain

Federico Nardi
MODE Collaboration
University of Padova, Italy
Université Clermont-Auvergne, France

Pietro Vischia
MODE Collaboration
Universidad de Oviedo and ICTEA, Spain

Haitham Zaraket
MODE Collaboration
Lebanese University, Lebanon
Université Grenoble-Alpes, France

Abstract

We describe a software package, developed to optimise the geometrical layout and specifications of detectors designed for tomography by scattering of cosmic-ray muons. The software exploits differentiable programming for the modelling of muon interactions with detectors and scanned volumes, the inference of volume properties, and the optimisation cycle performing the loss minimisation. In doing so, we provide the first demonstration of end-to-end-differentiable and inference-aware optimisation of particle physics instruments. We study the performance of the software on a relevant benchmark scenario and discuss its potential applications.

[*]giles.c.strong@gmail.com

Machine Learning and the Physical Sciences Workshop, NeurIPS 2023.

1 Introduction

Particle detectors represent some of humankind's most complex and interconnected machines, with large experiments such as the CERN LHC [1] having $O(10^7)$ detection elements. Designing such detectors inevitably falls to optimising proxy objectives, rather than the final measurement the experiment seeks to make. In attempting to break this paradigm, we can look to differentiable programming to provide analytic gradients along which to shift design parameters to best improve end-goal performance in the context of a differentiable pipeline from detector, through inference, to measurement. Daunting though such a task may seem, we seek to demonstrate its potential in the simplified context of muon tomography.

Muon tomography is an imaging technique utilising cosmic muons and their scattering through material [2]. Many applications are being developed [3], which include: identification of nuclear material in waste drums [4] and at border controls [5], nuclear warhead verification in the context of non-proliferation treaties [6], and industrial monitoring and maintenance [7].

Each application considers different objectives and conditions, and so requires different detector configurations for optimal performance. Muon tomography represents a simplified testing ground for our ideas: unlike the complex collisions occurring at the LHC, we need only consider particles one by one; the inference processes are far easier to construct; and the detectors occupy a design space of much lower dimensionality than LHC experiments.

2 TomOpt

The simulation and optimisation routines used for this study are structured into a software package referred to as TOMOPT [8], which uses PYTORCH [9] as a backend.

Detector modelling is the starting point for differentiation: the path from detector parameters to loss value must be differentiable. Detectors are considered to be "panels" that record the position of muons as they pass through them, with a finite spatial resolution and probability (modelled as an efficiency weight). Still, the panel parameters (size and position) must be sensitive to muons that miss the panels, in order to move to more optimal states where they could have recorded the muons. To achieve this, in "training mode", the resolution and efficiency of the panels are allowed to extend beyond the physical panel with a smooth decay in hit reconstruction performance. In "validation mode", only muons that are within the physical panel are recorded.

Detector cost is modelled as being linear with panel area. The total cost is fixed, and an assignment per panel is made through a softmax-normalised array of learnable parameters. The assigned budget per panel is used to rescale panel areas such that their cost matches their budgets.

Muon generation and scattering is performed based on literature models [10, 11]. Volume inference begins by reconstructing the incoming and outgoing muon trajectories around the volume of interest. We perform maximum likelihood fits of straight lines to xyz hits, weighted by their associated efficiencies and resolutions; thus the trajectories are made differentiable with respect to the panel parameters. In the next step, the Point-of-Closest-Approach (PoCA) is found between the incoming and outgoing trajectory of each muon. This is assumed to be the location at which the muon scattered the most [12]. The PoCAs are also differentiable with respect to the detector parameters, and carry an uncertainty computed via automatic-differentiation-based propagation for the hit positions and resolutions. Information of the PoCAs can then be fed into dedicated task-specific inference stages. Similarly, a task-specific loss function can be used to guide updates of the detector parameters, allowing users to optimise directly for their end-goal measurement, rather than proxy objectives.

The update cycle is similar to that of a neural network: batches of volumes are scanned using many muons, and the loss for predictions on the volume batch is backpropagated to the detector parameters, where gradient descent can be used to update them in the direction that best improves performance.

3 Benchmark: Fill-level estimation at metal refineries

Furnace ladles are structures used to transport molten metal produced in industrial plants before their contents are poured into moulds. Exact knowledge of the amount of molten metal inside the ladle can help prevent excess waste, however due to slag formation on top of the metal, optical measurements

cannot be made. In contrast, muon tomography can be used to predict the metal fill-height, by measuring incoming and outgoing muons and inferring the fill-level from their scattering. We will use the design of such a detector as a benchmark for TOMOPT.

3.1 Detector and volume

We model the ladle as a 1 m-sided, open-topped cube, with 10 cm thick steel walls and base. This will be filled with a varying amount of liquid steel, topped with a layer of slag (primarily CaO, Al_2O_3), as shown in Fig. 1.

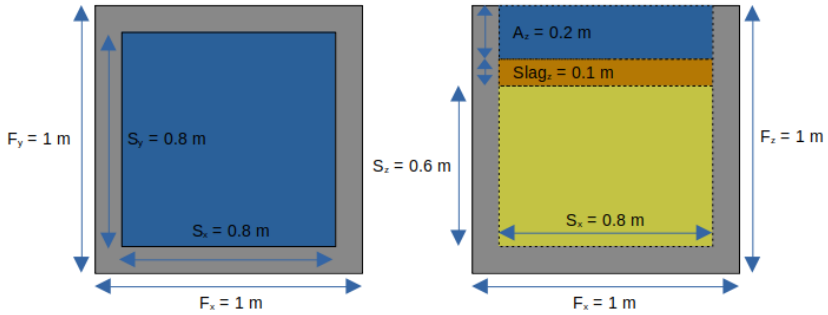


Figure 1: Top view (left), and front view (right) of the rectangular furnace ladle simulated. It is composed of walls (grey), liquid steel (yellow), slag (orange) and the air volume (blue) over the liquid steel and slag.

The initial detector consists of eight equally sized square panels placed in sets of four above, and four below, the passive volume. The available budget of six arbitrary units (a.u,) is initially distributed equally to all panels, which cost one a.u. per m^2, thus each panel measures 86.6 cm in both x and y length. Panels are defined to have nominal efficiencies of 90 %, and $x - y$ resolutions of 0.1 mm. The panels begin in a deliberate sub-optimal position (offset and closely separated, see Fig. 2a) that provides poor precision on the trajectory reconstruction and low muon exposure.

3.2 Inference method and loss function

In the spirit of an end-to-end optimisation pipeline, the inference must output the fill-height prediction and receive optimisation loss gradients based on this. Our inference is based on the observation that the mean z position of the PoCAs provides an indication of the centre of the fill, but must undergo a de-biasing procedure to accurately predict the fill height. Additionally, PoCAs are weighted to provide a greater emphasis on those with a low uncertainty on their z positions, are located near the centre of the furnace, and have a high trajectory efficiency. The weighted mean z position is then de-biased using a linear correction fitted using a range of volume scans at different fill-heights.

Even with the de-biasing, optimisation under e.g. mean squared-error risks being sensitive to any residual biases. Instead, a loss function that emphasises precise and separable predictions for fill-heights, whilst being invariant to their absolute value is desirable. The “spread over range" loss function captures the width of prediction distributions for each target fill-height, and the width of the distribution of their centres:

$$\mathcal{L} = \frac{1}{N_l} \frac{\sum_l \sigma(p_l)}{\sigma(\bar{p}_l)}, \tag{1}$$

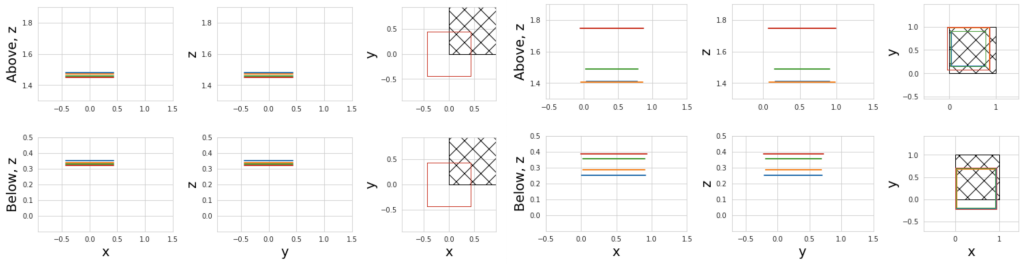
where $\sigma(p_l)$ is the standard deviation of predictions for fill-height index l, and $\sigma(\bar{p}_l)$ is the standard deviation of the mean prediction for each fill-height.

3.3 Optimisation process and results

Over the course of 10 batches of 40 volumes, the detector is updated, using 1-cycle learning rate schedule [13, 14] to scale the nominal learning rates over the course of the optimisation: initially they increase to allow large updates of the detector, and towards the end the decrease to much smaller values, allowing the panels to converge and stabilise. In addition, the detector model for resolution and

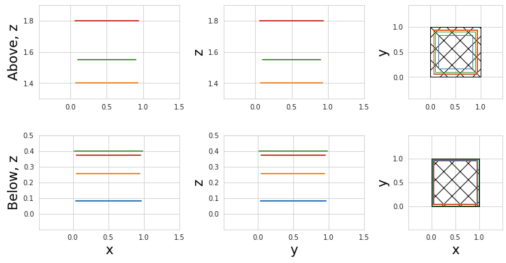
efficiency is annealed from an initially smooth, nonphysical configuration to a sharper, distribution of efficiency and resolution to better mimic a real detector. On a 2018 Macbook Pro with an Intel i7, this optimisation takes about 30 minutes to run (no GPU is required).

Figure 2b illustrates the detector configuration after this stage of the optimisation process. From this we can see that the panels have moved to be more central in $x - y$ to better capture muons that pass through the passive volume, and have expanded in z to improve trajectory precision.



(a) Initial detector configuration.

(b) Detector configuration after stage one optimisation process.



(c) Detector configuration after stage two optimisation process.

Figure 2: Comparison of detector configurations during the optimisation process. The coloured lines/squares indicate the positions and sizes of the panels, and the hatched area indicates the position and size of the passive volume. The top rows indicate panels above the passive volumes, and the bottom rows indicate panels below the passive volume. N.B. in the final detector configuration, two panels in the top section are close enough that they cannot be both viewed in the image; we model them as infinitely thin panels, but in reality their physical size would require them to have sufficient separation.

The initial optimisation process showed that TomOpt can take a poorly designed detector and quickly shift it into a configuration that provides much better performance. We can now refine the detector by fixing certain parameters and using lower learning rates. Having seen the panels move to become central, we use our knowledge of the problem symmetry to manually fix the panels at the $x - y$ centre, and attempt only to refine the panel z-positions and sizes.

Following a second optimisation stage we arrive at the final configuration, shown in Fig. 2c. Mainly we can see that the panels below the passive volume have further moved to cover a larger z range. While this could come at the expense of missing some high zenith-angle muons leaving the passive volume, the bottom panels have instead slightly increased in size by decreasing the intermediate panels above the passive volume.

Figure 3 indicates that the final design clearly improves over the initial design, providing a mean error on fill-level prediction of 2.5 cm.

4 Discussion

We believe that TomOpt provides the first demonstration of end-to-end-differentiable and inference-aware optimisation of particle physics instruments, and represents an important step in changing the way we design experimental apparatus. Given the number of free parameters considered here, it is

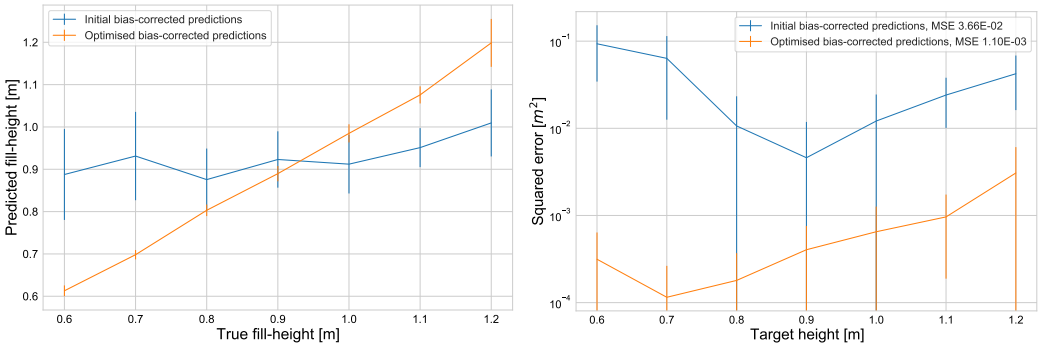


Figure 3: Performance of the optimised detector after stage 2 compared to the initial detector (both shown after the de-biasing process). Heights shown include the space below the passive volume.

worth mentioning that non-differentiable, black-box optimisers such as Bayesian optimisation (BO) could be well applicable. However given the parameter spaces of larger detectors, such as those at the CERN LHC, we believe that differentiable optimisation would be required. With that in mind, our primary aim of this paper was to demonstrate the viability of measurement-aware detector design in a differentiable pipeline.

Looking to the immediate future of this work, there are many possible extensions. Whilst TomOpt itself is still under development, we intend to make the framework open-source soon. So far, we have relied on “classical" inference algorithms, but inference via deep-neural networks are naturally compatible with TomOpt’s requirement on differentiable processes. We will be exploring their use in future publications, as these can help to ease the burden of designing task-specific differentiable inference algorithms. Additionally, as a quick simulator of the physics, TomOpt can be exploited for developing and testing advanced (non-differentiable) inference algorithms for muon-tomography in general. The available degrees of freedom of detectors, in terms of placement, technology, and modelling needs to be expanded, to allow for e.g. detectors to be placed on the side of the passive volume, or otherwise rotated. Finally, we will be looking to expand the repertoire of benchmark scenarios used for demonstration and testing.

We invite those interested to consider reading our full paper, available at Ref. [8], in which we delve more into the mathematics of our optimisation, the modelling of our detectors, and give a greater overview of our software package.

Acknowledgements

This work was partially supported by the EU Horizon 2020 Research and Innovation Programme under grant agreement No. 101021812 ("SilentBorder"), and by the Fonds de la Recherche Scientifique - FNRS under Grants No. T.0099.19 and J.0070.21. The collaboration of Aitor Orio was co-funded by the Spanish Ministry of Science and Innovation through the program "Ayudas para contratos para la formación de doctores en empresas (Doctorados Industriales) 2018" (Grant reference: DIN2018-009886). Pietro Vischia's work was partially supported by the FNRS under Grant No. 40000963 and by the Ramón y Cajal program under the Project No. RYC2021-033305-I. Max Lamparth was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC-2094 – 390783311.

References

[1] O. S. Brüning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole, and P. Proudlock, *LHC Design Report*. CERN Yellow Reports: Monographs. CERN, Geneva, 2004. https://cds.cern.ch/record/782076.

[2] K. N. Borozdin *et al.*, *Radiographic imaging with cosmic-ray muons*, Nature **422** (2003) 277.

[3] International Atomic Energy Agency, *Muon imaging: Present Status and Emerging Applications*, IAEA TECDOC 2012, IAEA, Vienna, 2022. https://www.iaea.org/publications/15182/muon-imaging.

[4] M. Weekes, A. Alrheli, D. Barker, D. Kikoła, A. K. Kopp, M. Mhaidra, J. Stowell, L. Thompson, and J. J. Velthuis, *Material identification in nuclear waste drums using muon scattering tomography and multivariate analysis*, Journal of Instrumentation **16** no. 05, (2021) P05007.

[5] S. Barnes, A. Georgadze, A. Giammanco, M. Kiisk, V. A. Kudryavtsev, M. Lagrange, and O. L. Pinto, *Cosmic-Ray Tomography for Border Security*, Instruments **7** no. 1, (2023) 13.

[6] C. L. Morris, J. Bacon, K. Borozdin, J. Fabritius, H. Miyadera, J. Perry, and T. Sugita, *Horizontal cosmic ray muon radiography for imaging nuclear threats*, Nucl. Instr. Meth B **330** (2014)42–46.

[7] P. Martinez Ruiz del Arbol, P. Gomez Garcia, C. Diez Gonzalez, and A. Orio Alonso, *Non-destructive testing of industrial equipment using muon radiography*, Philosophical Transactions of the Royal Society A **377** no. 2137, (2019) 20180054.

[8] G. C. Strong, M. Lagrange, A. Orio, A. Bordignon, F. Bury, T. Dorigo, A. Giammanco, M. Heikal, J. Kieseler, M. Lamparth, P. Martínez Ruíz del Árbol, F. Nardi, P. Vischia, and H. Zaraket, *TomOpt: Differential optimisation for task- and constraint-aware design of particle detectors in the context of muon tomography*, 2023.

[9] A. Paszke *et al.*, *PyTorch: An Imperative Style, High-Performance Deep Learning Library*, in *Advances in Neural Information Processing Systems 32*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, eds., pp. 8024–8035. Curran Associates, Inc., 2019. http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

[10] P. Shukla and S. Sankrith, *Energy and angular distributions of atmospheric muons at the Earth*, Int. J. Mod. Phys. A **33** no. 30, (2018) 1850175, arXiv:1606.06907 [hep-ph].

[11] **Particle Data Group** Collaboration, R. L. Workman and Others, *Review of Particle Physics*, PTEP **2022** (2022) 083C01.

[12] L. Schultz, K. Borozdin, J. Gomez, G. Hogan, J. McGill, C. Morris, W. Priedhorsky, A. Saunders, and M. Teasdale, *Image reconstruction and material Z discrimination via cosmic ray muon radiography*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **519** no. 3, (2004)687–694. https://www.sciencedirect.com/science/article/pii/S0168900203028808.

[13] L. N. Smith and N. Topin, *Super-convergence: very fast training of neural networks using large learning rates*, in *Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications*, T. Pham, ed., vol. 11006, pp. 369 – 386, International Society for Optics and Photonics. SPIE, 2019. arXiv:1708.07120. https://doi.org/10.1117/12.2520589.

[14] L. N. Smith, *A disciplined approach to neural network hyper-parameters: Part 1 - learning rate, batch size, momentum, and weight decay*, CoRR **abs/1803.09820** (2018), arXiv:1803.09820. http://arxiv.org/abs/1803.09820.