Task adapted reconstruction for inverse problems

Jonas Adler
Department of Mathematics
KTH - Royal institute of Technology
Research and Physics
Elekta
jonasadl@kth.se

Sebastian Lunz
Department of Applied Mathematics
and Theoretical Physics
University of Cambridge
lunz@math.cam.ac.uk

Olivier Verdier
Department of Mathematics
KTH - Royal institute of Technology
Department of Computing, Mathematics and Physics
Western Norway University of Applied Sciences
olivierv@kth.se

Carola-Bibiane Schönlieb
Department of Applied Mathematics
and Theoretical Physics
University of Cambridge
cbs31@cam.ac.uk

Ozan Öktem
Department of Mathematics
KTH - Royal institute of Technology
ozan@kth.se

Abstract

We consider the problem of performing a task defined on a model parameter that is only observed indirectly through noisy data in an ill-posed inverse problem. Several such tasks have been approached using deep neural networks, and recent advancements in image reconstruction using learned iterative schemes now enable us to have a fully differentiable, end-to-end trainable, imaging pipeline. The suggested framework is adaptable, with a plug-and-play structure for adjusting to both the inverse problem and the task at hand. The approach is demonstrated on joint tomographic image reconstruction and semantic segmentation.

1 Introduction

The starting point for using techniques from machine learning to solve an inverse problem is to rephrase the latter as a statistical estimation problem. A reconstruction method will then be a (non-randomized) decision rule and statistical decision theory can be used to select an appropriate method.

Statistical inverse problems Let X and Y denote separable Banach spaces and define the *data model* as a mapping $\mathcal{M}\colon X \to \mathscr{P}_Y$ where $\mathscr{P}_Y$ is the set of probability measures on Y. Following [4], a (statistical) inverse problem is the task of estimating the model parameter $x^* \in X$ from data $y \in Y$, which is a single observation generated by Y-valued random variable $\mathsf{y} \sim \mathcal{M}(x^*)$.

A special case is when one has a known forward operator $\mathcal{A}\colon X \to Y$, which models how a model parameter generates data in absence of noise and the Y-valued random variable y is given as

$$\mathsf{y} = \mathcal{A}(x^*) + \mathsf{e} \quad \text{with } \mathsf{e} \sim \mathbb{P}_{\text{noise}}. \tag{1}$$

32nd Conference on Neural Information Processing Systems (NIPS 2018), Montréal, Canada.

Bayesian inversion This is a generic framework for defining reconstruction methods that starts out by assuming that x^* is generated from a X-valued random variable x with known distribution (prior) [4]. It therefore becomes natural to estimate x^* by exploring the posterior distribution of $\mathsf{x} \mid \mathsf{y} = y$.

One needs a criteria to select an appropriate reconstruction method. One popular choice is to maximize the posterior (maximum a posteriori estimator). Another is to minimize Bayes risk, i.e., to consider the reconstruction method $\mathcal{A}^\dagger \colon Y \to X$ that minimizes the averaged expected loss defined as the expectation of some loss function $\ell_X : X \times X \to \mathbb{R}$:

$$\mathcal{R}(\mathcal{A}^\dagger) := \mathbb{E}_{(\mathsf{x},\mathsf{y})}\Big[\ell_X\big(\mathcal{A}^\dagger(\mathsf{y}), \mathsf{x}\big)\Big].$$

The choice of prior is important in Bayesian inversion and Bayesian non-parametric theory [5] provides a large class of priors, e.g., roughness and sparsity priors are commonly used in imaging [6, 3]. Available handcrafted priors are however incomplete in the sense that they only capture a fraction of the a priori information that is available about x^*. As an example, a natural a priori information in medical imaging is that the object being scanned is a human being. It is very difficult, if not impossible, to explicitly construct a prior that captures this information. Next, even if an appropriate prior is explicitly known, exploring the posterior is highly non-trivial due to computational issues and MCMC techniques, albeit efficient, are insufficient for medical imaging applications.

As we shall see next, these challenges are addressed by *learned iterative methods* that use a deep neural network to construct a reconstruction method that minimizes Bayes risk $\mathcal{A}^\dagger \mapsto \mathcal{R}(\mathcal{A}^\dagger)$.

2 Learned iterative methods

Machine learning, and deep neural networks in particular, have demonstrated a remarkable capacity in capturing intricate relations from example data [7]. Instead of using hand-crafted problem specific models, one uses generic models that are adapted through learning against example data. It is therefore temping to investigate whether one can learn a prior by these techniques.

Instead of providing a hand-crafted prior, we require access to (supervised) training data $(x_i, y_i) \in X \times Y$ generated by (x, y). Next, finding an optimal reconstruction method requires searching over all non-randomized decision rules, which is computationally infeasible. Instead, we restrict our attention to those given by a (deep) neural network architecture since these have large capacity (can approximate any Borel measurable mapping arbitrarily well [8]) and there are computationally feasible implementations. To summarize, we have a parametrized family of reconstruction methods $\mathcal{A}^\dagger_\theta \colon Y \to X$ and a successful training by minimizing the empirical loss

$$\theta^* \approx \operatorname*{arg\,min}_{\theta \in \Theta} \Big[\frac{1}{m}\sum_i \ell_X\big(\mathcal{A}^\dagger_\theta(y_i), x_i\big)\Big] \tag{2}$$

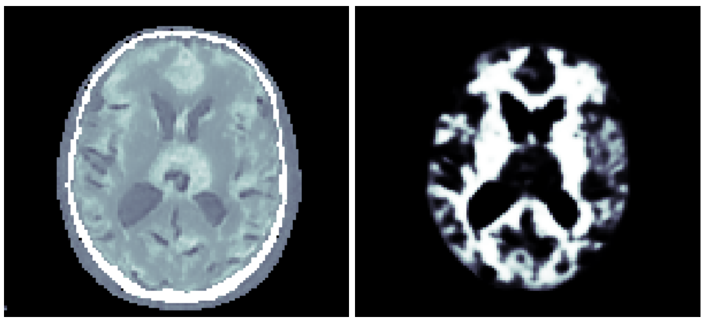
The above is a fully data driven approach for reconstruction, i.e., the (unknown) random variable (x, y) is replaced by its empirical counterpart given by the training data. Hence, the optimal reconstruction method $\mathcal{A}^\dagger_{\theta^*}$ is derived without utilizing knowledge about how data is generated. This is an serious problem for imaging applications where the large number of unknowns require large training datasets that are not available. In many inverse problem, knowledge about how data is generated is contained in the data model $x \mapsto \mathcal{M}(x)$ that is known. The learned iterative schemes [1, 2] construct a deep convolutional neural network architecture that accounts for the data model, or more precisely the data log-likelihood. The idea is to unroll a fixed point iterative scheme relevant for solving the inverse problem and use a CNN to learn the iterative update. The resulting reconstruction method is computationally feasible and outperforms state-of-the-art by a significant margin as shown in [2].

3 Task based reconstruction

A task on X is as an estimator $\mathcal{T}\colon X \to D$ (*task operator*) that maps a model parameter x to a value of some D-valued random variable d. To evaluate how well the task has been solved, we introduce a *task loss* $\ell_D : D \times D \to \mathbb{R}$ and the optimal task operator is the one that minimizes Bayes risk w.r.t. the task loss, i.e.,

$$\mathcal{T}_{\text{opt}} \in \operatorname*{arg\,min}_{\mathcal{T}\colon X \to D} \mathbb{E}_{(\mathsf{x},\mathsf{d})}\Big[\ell_D\big(\mathcal{T}(\mathsf{x}), \mathsf{d}\big)\Big]. \tag{3}$$

True image and segmentation (left and middle) with corresponding tomographic data (right).



C	ℓ_D
0.01	0.113
0.1	0.093
0.5	0.087
0.9	0.108
0.99	0.135
0.999	0.223

Results for $C = 0.9$ (left and middle) and values of the task loss for various C (right).

Figure 1: Joint tomographic reconstruction and segmentation of grey matter. Images shown using [-100, 100] HU window and segmentation using a [0, 1] window.

In practice, we restrict ourselves to a parametrized family of tasks $\mathcal{T}_\phi \colon X \to D$, typically given by deep neural networks. Learned iterative methods for reconstruction can then be combined in a generic manner with any task that can be adequately addressed using such a neural network, thereby resulting in an end-to-end differentiable reconstruction method that is adapted to the task.

A joint task based reconstruction operator is now defined as $\mathcal{T}_\phi \circ \mathcal{A}_\theta^\dagger \colon Y \to D$ and an "optimal" $(\theta, \phi) \in \Theta \times \Phi$ is the one that minimizes the following *joint expected loss (risk)*:

$$(\theta^*, \phi^*) \in \operatorname*{arg\,min}_{(\theta,\phi)\in\Theta\times\Phi} \mathbb{E}_{(\mathsf{x},\mathsf{y},\mathsf{d})}\Big[(1-C)\ell_X\big(\mathcal{A}_\theta^\dagger(\mathsf{y}), \mathsf{x}\big) + C\ell_D\big(\mathcal{T}_\phi \circ \mathcal{A}_\theta^\dagger(\mathsf{y}), \mathsf{d}\big)\Big]. \tag{4}$$

In an actual implementation, one replaces the above expectation with its empirical counterpart given from triplets of training data generated by $(\mathsf{x}, \mathsf{y}, \mathsf{d})$. Next, $C \in [0, 1)$ is a tuning parameter where $C \approx 0$ corresponds to a *sequential* approach and $C \to 1$ to a *fully end-to-end* approach.

The above is a generic approach for task adapted reconstruction with a plug-and-play structure for adapting to the inverse problem and the task at hand. Adapting to the inverse problem is by specifying the data model, e.g. by using a neural network architecture given by a learned iterative method [2], and any task that is given by a trainable neural network can be incorporated.

4 Application

We demonstrate the framework on joint tomographic image reconstruction and segmentation of white brain matter. $\mathcal{A}_\theta^\dagger$ is given by a Learned Primal-Dual method [2], which incorporates a knowledge based model for how data is generated into its architecture, and $\mathcal{T}_\phi$ by a U-Net [9].

Some results are shown in figure 1. Note in particular that (perhaps surprisingly) the "best" segmentation is not obtained by a fully end-to-end approach, instead they are obtained when the reconstruction loss is included as a regularizer. Furthermore, it is clear that the reconstruction obtained for $C = 0.9$ overemphasizes image features relevant for the task, e.g., white-grey matter contrast. This clearly "helps" the task and also visually shows the image features used by the joint approach.

Acknowledgments

The work by Jonas Adler, Olivier Verdier, and Ozan Öktem has been supported by the Swedish Foundation for Strategic Research grant AM13-0049, Industrial PhD grant ID14-0055, and Elekta AB. Carola-Bibiane Schönlieb and Sebastian Lunz acknowledge support from the Engineering and Physical Sciences Research Council (EPSRC) "EP/K009745/1", the EPSRC grant "EP/M00483X/1", the EPSRC centre "EP/N014588/1", the Leverhulme Trust project "Breaking the non-convexity barrier", the CHiPS (Horizon 2020 RISE project grant), the Cantab Capital Institute for the Mathematics of Information, and the Alan Turing Institute "TU/B/000071".

References

[1] J. Adler and O. Öktem. Solving ill-posed inverse problems using iterative deep neural networks. *Inverse Problems*, 33(12):124007 (24pp), 2017. Special issue: 100 Years of the Radon transform.

[2] J. Adler and O. Öktem. Learned primal-dual reconstruction. *IEEE Transactions on Medical Imaging*, 37(6):1322–1332, 2018.

[3] M. Benning and M. Burger. Modern regularization methods for inverse problems. *Acta Numerica*, 27:1–111, 2018.

[4] S. N. Evans and P. B. Stark. Inverse problems as statistics. *Inverse Problems*, 18(4):R1–R55, 2002.

[5] S. Ghosal and A. W. van der Vaart. *Fundamentals of Nonparametric Bayesian Inference*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2017.

[6] J. P. Kaipio and E. Somersalo. *Statistical and Computational Inverse Problems*, volume 160 of *Applied Mathematical Sciences*. Springer Verlag, 2005.

[7] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[8] A. Pinkus. Approximation theory of the MLP model in neural networks. *Acta Numerica*, 8:143–195, 1999.

[9] O. Ronneberger, P. Fischer, and T. Brox. U-Net: Convolutional networks for biomedical image segmentation. In N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III*, volume 9351 of *Lecture Notes in Computer Science*, pages 234–241. Springer Verlag, 2015.