Convex Optimization for Inverse Problems Using Convolutional Neural Networks

J. Adler[1, 2], A. Ringh[1], and J. Karlsson[1]

Abstract: In this work we consider a class of optimization problems commonly used for variational regularization of inverse problems, namely optimization problems of the form

$$\min_{f} \; F(T(f), g) + G(f).$$

Here g denotes the data, F the data mismatch term, T the forward operator, and G the regularization term. In particular, total variation (TV) regularization can be given in this form by using

$$F(h, g) \colon = \|h - g\|_2^2, \;\; G(f) \colon = \|\nabla f\|_1.$$

This type of problem is typically solved using some form of iterative scheme, e.g., the Primal Dual Hybrid Gradient (PDHG) by A. Chambolle and T. Pock (JMIV, 40(1):120–145, 2011), and such algorithms are typically run until some convergence criteria is met.

Here we focus on large scale and time-critical applications, where the optimization needs to be finished in a given time window and therefore only allows for a limited number of iterations, $n_{\max}$. In this case, the optimization algorithm can be seen as a neural network which maps the data to the output of the algorithm, $g \mapsto f_{n_{\max}}$. In order to obtain an efficient method, we build on the architecture recently suggested by J. Adler and O. Öktem (arXiv:1707.06474), where a learned primal-dual algorithm is used for solving inverse problems. In this structure a convolutional neural network replaces the proximal operators in a PDHG-like method. The full neural network, denoted by Λ_Θ where Θ are the parameters, is trained using supervised learning minimizing the mean squared error. We propose to use unsupervised learning for training of this network by selecting the parameters based on the value of the objective function $H_g := F(K(f), g) + G(f)$, i.e.,

$$\Theta^* \in \arg\min_{\Theta} E_g \Big[H_g\big(\Lambda_\Theta(g)\big) \Big],$$

This is equivalent to learning an optimization algorithm for the family of objective functions $\{H_g\}_g$. Finally, we consider an example in TV-regularized tomographic reconstruction, and using only 10 iterations, this method considerably outperforms other iterative schemes.

	PDHG	Learned PD
$H_g(f_{10}) - H_g(f^*)$	0.285219	0.009159

[1] Department of Mathematics, KTH, Royal Institute of Technology
Lindstedtsvägen 25, 100 44 Stockholm, Sweden
{*jonasadl,aringh*}*@kth.se, johan.karlsson@math.kth.se*

[2] Research and Physics group
Elekta Instrument AB
Kungstensgatan 18, 103 93 Stockholm, Sweden