Algorithm unrolling for solving inverse problems in signal and image processing

Caroline Chaux [a], **Abhijit Singh** [b], **Emmanuel Soubies** [c], **Vincent Tan** [d]

[a] *CNRS, IPAL,* caroline.chaux@cnrs.fr

[b] *CNRS@CREATE,* abhijit.singh@cnrsatcreate.sg

[c] *CNRS, IRIT, Univ Toulouse,* emmanuel.soubies@cnrs.fr

[d] *Department of Mathematics, National University of Singapore,* vtan@nus.edu.sg

* Presenting author

1. Introduction

Inverse problems in signal and image processing are encountered in several scientific domains such as biology, medical imaging, chemistry, and audio signal processing. They require a variety of tasks to be tackled, such as compressed sensing, matrix factorization, deconvolution, and source separation. In these contexts, we assume that we have one (or several) observation(s) of the object(s) of interest and we want to recover the unknown object(s) from its observation(s) [1].

Firstly, it is clear that the modelling of the forward process is very important, and thus the physics induced by the system must be well understood. Secondly, inverting the problem and solving this inverse problem is complicated because usually it is ill-posed. Consequently, we need to use both the information about the direct model as well as the object of interest [2]. Once the forward model is derived, the inverse problem is formulated as an optimization problem (see (2)) which is then solved by a numerical procedure adapted to the shape of the functional to be minimized [3, 4].

In our recent works, we tackled several inverse problems such as deconvolution [5, 6] or robust Principal Component Analysis [7], for which we proposed innovative unrolling strategies that combined physics knowledge and learning. While the formalization of the forward problem is derived in Section 2, the unrolling strategies are developed in Section 3 together with some numerical experiments. Conclusions are finally drawn in Section 4.

2. On solving inverse problems

Solving inverse problems in signal and image processing consists in recovering an estimate $\hat{x} \in \mathbb{R}^N$ of some unknown signal $x \in \mathbb{R}^N$ from its observation $y \in \mathbb{R}^M$. The image formation model (or direct model) can be written as

$$y = \mathcal{N}(Ax), \tag{1}$$

where $A \in \mathbb{R}^{M\times N}$ represents the forward linear operator (e.g., a blur or a sensing matrix) and $\mathcal{N}$ represents the noise (e.g., additive Gaussian or Poisson).

A classical approach to solve the associated inverse problem is to solve the following optimization problem [2]

$$\hat{x} \in \left\{ \underset{x\in\mathbb{R}^N}{\arg\min}\ \mathcal{F}(x) := \mathcal{D}_y(Ax) + \lambda\mathcal{R}(x) \right\}, \tag{2}$$

where the $\mathcal{D}_y(.)$ term measures the discrepancy between the model and the data (the ℓ_2-norm is a natural choice for additive Gaussian noise) and $\mathcal{R}$ is a regularization term introducing prior knowledge on the targeted solution. To ensure a good balance between the data fidelity and regularization terms, we use a regularization parameter $\lambda > 0$.

Several algorithms [3, 4] have been proposed in the literature to solve Problem (2) depending on the mathematical properties (e.g., convexity and differentiability, among others) of the function $\mathcal{F}$. All of them introduce additional hyperparameters (such as the step sizes) that need to be fixed along with the regularization hyperparameter λ, and they play a prominent role in obtaining a good estimate. To automate the determination of these hyperparameters while computing an estimate of the signal $\hat{x}$, one can leverage algorithm unrolling [8].

3. Deep unrolling for image processing

Recently, unrolled (or unfolded) neural networks have been proposed to combine optimization and learning [8]. They integrate information about the direct model within the network architecture, which make them interpretable. In an unrolled network, the iterations of an iterative algorithm are transformed into neural network layers. The resulting network can then be trained and the unknown hyperparameters can be learnt, while information about the direct model is integrated into the weight matrices and the a priori into activation functions.

3.1 Unrolling for regularization parameter estimation

Context and method Let's first consider a deconvolution problem associated with an additive white Gaussian noise and where the signal is expressed in a wavelet basis. The direct model reads

$$y = HW^*x + \epsilon, \tag{3}$$

where H represents a blur operator (here a Gaussian kernel), W^* (resp. W) defines an orthogonal wavelet synthesis (resp. analysis) operator and $\epsilon \in \mathbb{R}^M$ corresponds to a zero-mean Gaussian white noise of variance σ^2. To solve the related inverse problem, we will compute

$$\hat{x} \in \left\{ \arg\min_{x \in \mathbb{R}^N} \mathcal{F}(x) := \frac{1}{2}\|HW^* x - y\|_2^2 + \lambda \|x\|_1 \right\},$$

using an unrolled version of FISTA [9] where the regularization parameter λ is learnt. More precisely, from the maximum a posteriori (MAP) estimator, we obtained that $\lambda \propto \sigma^2/\mu$, with μ being the scale parameter of the a priori. Consequently, instead of learning λ directly, we proposed to learn the parameters $(\theta_1, \theta_2, \vartheta_1, \vartheta_2)$ of linear rectification functions $r_\sigma(s;\theta) = \theta_1 s + \theta_2$ and $r_\mu(u;\vartheta) = \vartheta_1 u + \vartheta_2$ such that $r_\sigma(\hat{\sigma}_y;\theta)$ and $r_\mu(\hat{\mu}_y;\vartheta)$ improves the initial estimates $\hat{\sigma}_y$ and $\hat{\mu}_y$ of σ and μ. This allows us to seek λ as $r_\sigma(\hat{\sigma}_y;\theta)^2/r_\mu(\hat{\mu}_y;\vartheta)$ and use known (or at least easily accessible) information.

Numerical results In Figure 1, we show that the proposed deep unrolled network reaches the performance one would obtain through a costly a grid search procedure over λ values. It is worth mentioning that the resulting network is fully interpretable, and the training is performed on a small dataset ($<$ 50 images) thanks to the physics-informed network which takes 30 minutes to train on a laptop. This results in an inference time of a couple of seconds for each online run versus minutes for an equivalent quality grid-search iterative algorithm.

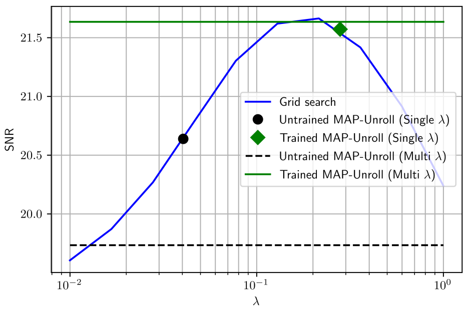

Fig. 1: SNR performances along with λ.

3.2 *Unrolling to mitigate model uncertainties*

Context and method We now consider a deconvolution problem associated to Poisson noise. The direct model is given by (1), where $\mathcal{N}$ stands for a Poisson noise model. The data fidelity term in (2) should be defined (from the MAP) as the Kullback-Leibler divergence [10, 11], which is difficult to optimize. We propose to consider instead a second-order Taylor approximation of $\mathcal{D}_y$, and to define

$$\mathcal{D}_y(Ax) \approx \|Ax - c\|_w^2 := \sum_{m=1}^{M} w_m \left([Ax]_m - c_m\right)^2. \quad (4)$$

In the case of Poisson noise, we can theoretically derive that $w_m = \frac{1}{y_m}$ and $c = y$. For more complex noises, such theoretical results may be difficult or even impossible to derive. Thus, we propose to learn the weights $w \in \mathbb{R}^M$. To do so, we define a weight estimation module and sequentially cascade it together with an unrolled algorithm to reconstruct the image.

Numerical results Looking at the visual results in Figure 3, we can see that the weighted least square (WLS) approach (average PSNR of 27.71 dB) leads to better results than the standard least square (LS) approach (average PSNR of 26.91 dB). In addition, the learnt weights (see Figure 2) are in line with the theoretical values derived for Poisson noise. We would like to underline here that these results can naturally be extended to more complex and general noises.

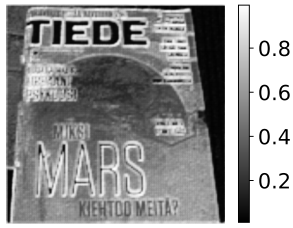

Fig. 2: The learnt w.

Fig. 3: Comparison (with zooms) of visual performance of one image in the test set.

4. Conclusion

Algorithm unrolling is a powerful tool to solve inverse problems in signal and image processing, particularly because it results in fully interpretable physics-informed networks. Unrolled architectures can be derived for various tasks, and they can be used for learning unknown quantities and mitigating errors on the model.

Acknowledgments

This work was supported by the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) Programme.

References

[1] S. Mallat. *A wavelet tour of signal processing: the sparse way*. Academic Press, San Diego, CA, USA, 3rd edition, 2009.

[2] C. Chaux, P. L. Combettes, J.-C. Pesquet, and V. R. Wajs. A variational formulation for frame based inverse problems. *Inverse Problems*, 23(4):1495–1518, Aug. 2007.

[3] P. Combettes and J.-C. Pesquet. *Proximal Splitting Methods in Signal Processing*, pages 185–212. Springer New York, 2011.

[4] N. Komodakis and J.-C. Pesquet. Playing with duality: An overview of recent primal?dual approaches for solving large-scale optimization problems. *IEEE Signal Process. Mag.*, 32(6):31–54, 2015.

[5] P. Nguyen, E. Soubies, and C. Chaux. Map-informed unrolled algorithms for hyper-parameter estimation. In *Proc. Int. Conf. Image Process.*, pages 2160–2164, 2023.

[6] A. Singh, E. Soubies, and C. Chaux. Learning Weighted Least Squares Data Term for Poisson Image Deconvolution. to appear in Proc. Int. Conf. Acoust. Speech Signal Process., January 2025.

[7] E. Z. C. Tan, C. Chaux, E. Soubies, and V. Y. F. Tan. Deep unrolling for nonconvex robust principal component analysis. In *IEEE Int. Workshop Mach. Learn. Signal Process.*, pages 1–6, 2023.

[8] V. Monga, Y. Li, and Y. Eldar. Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing. *IEEE Signal Process. Mag.*, 38(2):18–44, 3 2021.

[9] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.*, 2(1):183–202, 2009.

[10] T. Le, R. Chartrand, and T. J. Asaki. A variational approach to reconstructing images corrupted by poisson noise. *J. Math. Imaging Vision*, 27(3):257–263, 2007.

[11] Z. T. Harmany, R. F. Marcia, and R. M. Willett. This is SPIRAL-TAP: Sparse Poisson Intensity Reconstruction ALgorithms—Theory and Practice. *IEEE Trans. Image Process.*, 21(3):1084–1096, 2011.