# Comparing Analysis and Synthesis in Deep Prior Learning for Inverse Problems Resolution

Benoît Malézieux[1,2], Thomas Moreau[1], and Matthieu Kowalski[2]

[1] Université Paris-Saclay, Inria, CEA, Palaiseau, 91120, France
[2] L2S, Université Paris-Saclay–CNRS–CentraleSupelec

**Abstract.** Inverse problems consist in recovering a signal given a noisy transformation when no ground truth is available. Some prior knowledge about the signal is required and must be provided to the reconstruction algorithm to solve the problem. In this work, our goal is to learn the prior in the unsupervised setting by leveraging the sparsity property of natural signals in order to find a simpler representation of the data. We design two methods, derived from two different models and based on Deep Learning and sparse coding algorithms.

**Keywords:** Inverse Problems · Sparse Coding · Deep Learning · Optimization · Dictionary Learning · Unsupervised Learning.

## 1 Motivation

The resolution of inverse problems is a major challenge in many fields, from astrophysics to bio-imaging. They are usually ill-posed and it is rare that we have a ground truth to evaluate them. The most classical resolution approach is to formulate them as optimization problems integrating a prior on the structure of the solutions. The computational efficiency and the quality of the solutions depend strongly on the prior choice.

Another approach consists in learning a prior on clean or noisy data, as it is done in Dictionary Learning [6]. In this work, we propose to extend this method to the case where only degraded transformations of the data are available. We discuss the usage of deep learning to learn directly an adapted prior using a parametric model on a class of signals, based on the seminal work [4].

**Contributions:** we introduce two methods to learn a parametric prior in inverse problems context, using only the degraded measurements and without access to the ground truth (unsupervised). These methods are evaluated using three different classes of priors.

## 2 Problem and state of the art

We consider the case where the observed signal $\mathbf{y} \in \mathbb{R}^m$ is obtained as the sum of a linear transformation $A \in \mathbb{R}^{m \times n}$ of an unknown signal $\mathbf{x} \in \mathbb{R}^n$ and some additive Gaussian white noise $\mathbf{b} \in \mathbb{R}^m$:

$$\mathbf{y} = A\mathbf{x} + \mathbf{b} \ . \tag{1}$$

Linear inverse problems problems aim at recovering $\mathbf{x}$ from the observation of $\mathbf{y}$. In unsupervised signal reconstruction, we only have access to the measurements $\mathbf{y}$. When $m < n$, $A$ is not invertible, and the problem can't be solved without adding some information about the unknown signal. Two formulations leveraging sparsity are generally used to inject prior knowledge: Analysis and Synthesis [3].

*Analysis:* The idea of Analysis is to find a transform $\Gamma \in \mathcal{C}_A$, which changes the signal into a sparse representation, where $\mathcal{C}_A$ is a class of priors. The optimization problem to solve is the following, where $\lambda > 0$ is a regularization parameter:

$$\min_{\mathbf{x}\in\mathbb{R}^n, \Gamma\in\mathcal{C}_A} F_A(\mathbf{x}, \Gamma) \triangleq \frac{1}{2} \|A\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \left\|\Gamma^T \mathbf{x}\right\|_1 \ . \tag{2}$$

*Synthesis:* Synthesis makes the assumption that the signal is a linear combination of a few atoms, which are the columns of a dictionary $\Lambda \in \mathcal{C}_S$, where $\mathcal{C}_S$ is a class of priors. The optimization problem becomes:

$$\min_{\mathbf{z}\in\mathbb{R}^L, \Lambda\in\mathcal{C}_S} F_S(\mathbf{z}, \Lambda) \triangleq \frac{1}{2} \|A\Lambda\mathbf{z} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{z}\|_1 \ . \tag{3}$$

The challenging part compared to previous works is that we aim at learning $\Gamma$ or $\Lambda$ using only the corrupted measurements $\mathbf{y}$ and not the ground truth. We propose to solve the problem using bi-level optimization and deep learning.

## 3 Proposed approach

Both Analysis and Synthesis prior learning models can be rewritten as bi-level optimization problems to minimize only over $\Lambda$ or $\Gamma$ (see [6] for example):

$$\min_{\Gamma\in\mathcal{C}_A} F_A(\mathbf{x}^*(\Gamma), \Gamma) \quad \text{s.t.} \quad \mathbf{x}^*(\Gamma) = \operatorname*{argmin}_{\mathbf{x}\in\mathbb{R}^n} F_A(\mathbf{x}, \Gamma) \ , \tag{4}$$

$$\min_{\Lambda\in\mathcal{C}_S} F_S(\mathbf{z}^*(\Lambda), \Lambda) \quad \text{s.t.} \quad \mathbf{z}^*(\Lambda) = \operatorname*{argmin}_{\mathbf{z}\in\mathbb{R}^L} F_S(\mathbf{z}, \Lambda) \ . \tag{5}$$

However, the gradients cannot be computed in closed-form given that the exact values of $\mathbf{z}^*(\Lambda)$ or $\mathbf{x}^*(\Gamma)$ are not available. As done in [4], we propose to use automatic differentiation to approximate these gradients. Similar ideas have been used in [5] and [7] for supervised learning.

A neural network is built by unrolling N iterations of FISTA [1] for Synthesis or Condat-Vu Algorithm [2] for Analysis. Its output is an approximation of $\mathbf{z}^*(\Lambda)$ or $\mathbf{x}^*(\Gamma)$. The network is parameterized by the prior, which can be learned by projected gradient descent. The gradients of (4) and (5) are computed using automatic differentiation.

We consider three classes of priors: Unit Norm (UN) priors for columns (resp. rows) in the synthesis prior $\Lambda$ (resp. analysis prior $\Gamma^T$), UN Tight Frame (UNTF) for the analysis prior as proposed in [8] where $\Gamma^T \Gamma = I$ and convolutional priors where $\Lambda$ and $\Gamma$ are made of several filters with small supports.

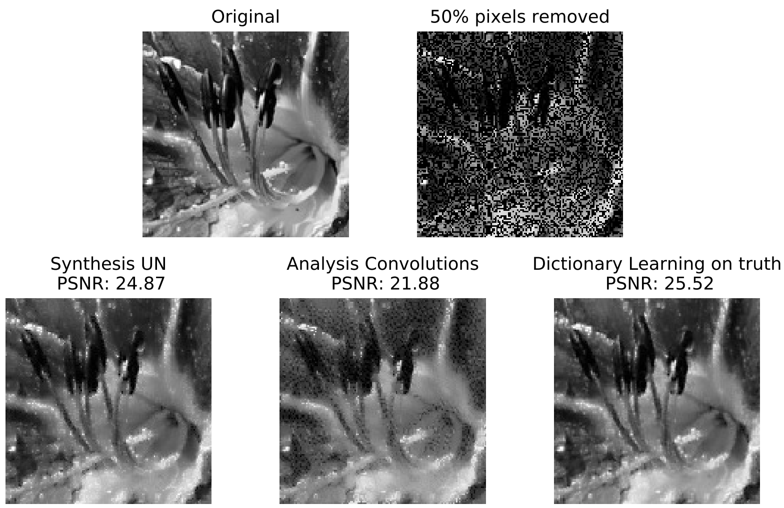


**Fig. 1.** Inpainting on a $128 \times 128$ image. Our algorithms use the degraded image, while the baseline learns a prior on the ground truth first. Synthesis outperforms analysis and recovers the image up to a level close to the baseline.

The main interest of these methods compared to classical deep learning approaches is that they can be analyzed using the theory from dictionary learning while still retaining an architecture trainable end-to-end. Our future works will consist in providing theoretical guaranties on their level of performance.

## References

1. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imaging Sciences **2**, 183–202 (2009)
2. Condat, L.: A primal–dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms. Journal of Optimization Theory and Applications **158**(2), 460–479 (2013)
3. Elad, M., Milanfar, P., Rubinstein, R.: Analysis versus synthesis in signal priors. Inverse Problems **23**, 947 (2007)
4. Gregor, K., LeCun, Y.: Learning fast approximations of sparse coding. International conference on machine learning pp. 399–406 (2010)
5. Jiu, M., Pustelnik, N.: A deep primal-dual proximal network for image restoration. preprint (2020)
6. Mairal, J., Bach, F., Ponce, J., Sapiro, G.: Online learning for matrix factorization and sparse coding. Journal of Machine Learning Research **11** (2009)
7. Scetbon, M., Elad, M., Milanfar, P.: Deep k-svd denoising. preprint (2019)
8. Yaghoobi, M., Nam, S., Gribonval, R., Davies, M.E.: Constrained overcomplete analysis operator learning for cosparse signal modelling. IEEE Transactions on Signal Processing **61**(9), 2341–2355 (2013)