Unrolled deep networks for sparse signal restoration in analytical chemistry

Mouna Gharbi, Silvia Villa, Emilie Chouzenoux, Jean-Christophe Pesquet, Laurent Duval

2024 IEEE INTERNATIONAL WORKSHOP ON MACHINE LEARNING FOR SIGNAL PROCESSING, SEPT. 22–25, 2024, LONDON, UK

UNROLLED DEEP NETWORKS FOR SPARSE SIGNAL RESTORATION IN ANALYTICAL CHEMISTRY

Mouna GHARBI[†]*, Silvia VILLA*[*]*, Emilie CHOUZENOUX*[†]*, Jean-Christophe PESQUET*[†]*, Laurent DUVAL*[°]

[†] CVN, CentraleSupélec, Inria Saclay, University Paris Saclay, France.

[*] MaLGa, DIMA, Università degli Studi di Genova, Via Dodecaneso 35, 16146 Genova, Italy.

[°] IFP Energies nouvelles, 92852 Rueil-Malmaison, France

ABSTRACT

This paper addresses the problem of sparse signal recovery from linearly transformed and noisy measurements. We propose to adopt the recent 'deep unrolling' paradigm, which consists in creating a deep neural network inspired from an iterative algorithm initially built for penalized loss minimization. The iterations of the algorithm are recast as neural network layers. The use of deep learning frameworks ensures an efficient implementation and the possibility to learn the algorithm native hyperparameters, through the minimization of a task-oriented loss. For a given application, choosing an adequate iterative scheme to unroll, and fine-tuning the architecture is a challenging task. In this work, we present three deep unrolled architectures dedicated to sparse signal recovery. We then perform their comprehensive comparative study, through the motivating application, arising in analytical chemistry, of peak retrieval from blurred and noisy chromatography acquisitions.

1. INTRODUCTION

Sparse signal restoration amounts to retrieving original data from corrupted observations, assuming the sought solution is a sparse vector characterised by few nonzero entries. Measurements $z \in \mathbb{R}^m$ are related to groundtruth spike signal $\bar{x} \in \mathbb{R}^n$ through the model

$$z = H\bar{x} + \varepsilon, \tag{1}$$

where $H \in \mathbb{R}^{m\times n}$ is a linear operator simulating the acquisition model, for instance a convolution, and $\varepsilon \in \mathbb{R}^m$ is the realization of a noise, here assumed additive zero-mean i.i.d. Gaussian. The signal recovery problem consisting of estimating $\bar{x}$ from z is an inverse problem that was tackled thoroughly in literature. A known approach relies on minimizing a designed objective using iterative algorithms. For instance, in explicit regularization [1], a cost function composed of a data fidelity term and a prior term balanced by a regularization hyperparameter is minimized. Regularization is incorporated into the problem to mend the ill-posedness, by enforcing sparsity in the recovered estimate. For instance, ℓ_1 norm, ℓ_0 count measure, or their smooth approximations [2, 3] can be used. Another type of regularization consists in iterative or implicit regularization [4] in which the implicit bias of first-order algorithms is exploited to favor desirable properties of the solution. In this case, the number of iterations plays the role of the regularization parameter. Solving inverse problems like (1) has also been addressed by supervised learning methods. Common deep networks (such as CNNs) have been used for signal deblurring [5], denoising [6], and reconstruction [7]. Such models are trained on large annotated databases, to learn a mapping between inputs (degraded data) and outputs (original data), to be used on test data. Lately, a new approach called unrolling (or unfolding), combining model-based and data-driven approaches has emerged [8]. The idea consists in creating deep networks inspired from iterative algorithms, such that each layer mirrors one iteration of the corresponding algorithm. The benefits are (i) interpretability which contrasts with mainstream black-box supervised learning models, (ii) computation speed provided by deep learning frameworks such as Pytorch and Tensorflow, (iii) efficiency through supervised and task-oriented tuning of the algorithm native hyperparameters. Many works have explored unrolling of various minimization algorithms, such as gradient descent [9], interior point [10], half-quadratic [11], and proximal splitting methods [12, 13, 14, 15]. Bayesian approaches have also recently been combined with unrolling in [16]. The deployment of the unrolling principle involves several prerequisites that might be application-dependent such as the optimization problem formulation, the algorithm to unroll, the parameters to untie, etc. Several recent benchmark studies have been made in the context of image processing, includ-

This work has been supported by the ITN-ETN project TraDE-OPT funded by the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 861137. M.G. and E.C. acknowledge funding support from the European Research Council Starting Grant MAJORIS ERC-2019-STG850925. S.V. acknowledges the support of the project AFOSR (European Office of Aerospace Research and Development) FA8655-22-1-7034, the European Research Council Consolidator Grant SLING 819789 and the EU H2020-MSCA-RISE project NoMADS-777826.

ing image restoration [14], and MRI reconstruction [17], but 1D signal restoration, as it occurs in analytical chemistry, remains scarcely studied in that respect. In this work, we provide a comprehensive comparative benchmarking of deep unrolled methods for solving Problem (1). We design three unrolled architectures, relying on distinct optimization problem formulation and resolution schemes. The first one relies on a primal-dual algorithm [18] using an ℓ_1-based constrained formulation inspired from the implicit regularization framework. Up to our knowledge, this is the first instance of unrolling such implicit formulation. The next two are inspired from explicit regularization schemes, as more classically done in unrolling, namely iterative soft thesholding (ISTA) [19] based on ℓ_1 penalty, and half-quadratic (HQ) algorithm [20], that addresses a smoothed penalized least-squares formulation. We compare the three iterative (i.e., unsupervised) and unrolled (i.e., supervised) implementations on synthetic and real datasets, from chromatography compound analysis. The paper is organized as follows: in Section 2, we introduce the mathematical notation and definitions useful for the rest of the work. Section 3 describes our construction for the three deep unrolled methods, devising the algorithms, designing the architectures, and detailing our strategy to learn the desired hyperparameters for each approach. Section 4 summarizes our experimental settings, and presents our comparisons and analysis. Lastly, Section 5 draws conclusions of the work.

2. MATHEMATICAL NOTATION

We adopt the convex analysis notation from [21]. Let $\mathscr{H}$ be a Hilbert space, the subdifferential of a function $f : \mathscr{H} \to \mathbb{R}$ at point x is the set valued operator $\partial f : \mathscr{H} \to 2^{\mathscr{H}}$ such that $\partial f(x) = \{u \in \mathscr{H} | (\forall y \in \mathscr{H}) \langle y - x | u \rangle + f(x) \leq f(y)\}$. The identity operator on $\mathscr{H}$ is denoted by Id. The set of proper, convex and lower semi continuous (l.s.c) functions is denoted by $\Gamma_0(\mathscr{H})$ and the proximity operator is defined, for every $x \in \mathscr{H}$, by $\mathrm{prox}_f(x) = \mathrm{argmin}_{u \in \mathscr{H}} f(u) + \frac{1}{2}\|x-u\|_2^2$. Finally, the convex conjugate of a function is denoted, for every $x \in \mathscr{H}$, by $f^*(x) = \sup_{u \in \mathscr{H}} \langle x | u \rangle - f(u)$. For a given convex closed nonempty set C, ι_C is the indicator function defined, for every $x \in \mathscr{H}$, by $\iota_C(x) = \begin{cases} 0, & \text{if } \quad x \in C \\ +\infty & \text{otherwise} \end{cases}$. The support function $\sigma_C = \iota_C^*$ is defined for all $x \in \mathscr{H}$ by $\sigma_C(x) = \sup_{x' \in C} \langle x' | x \rangle$. We finally denote by $\mathscr{B}(z, \rho)$ the ℓ_2 ball of center z and radius ρ, expressed as $\mathscr{B}(z, \rho) = \{z' \in \mathscr{H} \mid \|z' - z\|_2 \leq \rho\}$.

3. PROPOSED METHOD

In this section, we design three unrolled architectures, built upon proximal primal-dual, ISTA, and HQ iterative schemes to solve (1). For each approach, (i) we formulate the optimization problem to be solved, (ii) we derive the iterative algorithm to solve it, and (iii) we propose the corresponding unrolled neural network by reinterpreting the algorithm iterations into layers, allowing to learn its hyperparameters.

3.1. Unrolled primal-dual (U-PD) architecture

3.1.1. Problem statement

Let us define $\widehat{x} \in \mathbb{R}^n$ an estimate of $\bar{x}$ obtained by solving the constrained minimization problem

$$\widehat{x} \in \arg\min_{x \in \mathbb{R}^n} \|x\|_1 \quad \text{s.t.} \quad \|Hx - z\|_2 \leq \rho. \tag{2}$$

Hereabove, $\|\cdot\|_1$ is ℓ_1 norm which aims to promote a sparsity prior on the estimated solution and $\rho > 0$ is a hyperparameter related to the noise level. In the case of an i.i.d. Gaussian noise with standard deviation σ, a common setting arising from the law of large numbers is $\rho = \sqrt{n}\sigma$. The minimization problem (2) can be solved using a proximal splitting primal-dual algorithm [22], which we describe in the next section. The optimal solution is retrieved by combining the explicit regularization induced by ρ with the implicit regularization given by the early stopping of the iterations [23].

3.1.2. Primal-dual algorithm

Problem (2) can be recast into an unconstrained form which is the primal problem

$$\min_{x \in \mathbb{R}^n} \quad \Big(\iota_{\mathscr{B}(z,\rho)}(Hx) + \|x\|_1\Big). \tag{3}$$

The associated dual problem [18] is

$$\max_{y \in \mathbb{R}^m} \quad -\Big(\sigma_{\mathscr{B}(z,\rho)}(y) + \| - H^\top y\|_1\Big). \tag{4}$$

The general saddle point problem is

$$\min_{x \in \mathbb{R}^n} \max_{y \in \mathbb{R}^m} \quad \langle Hx, y \rangle + \|x\|_1 - \sigma_{\mathscr{B}(z,\rho)}(y). \tag{5}$$

To solve (5), the primal-dual approach initially proposed in [18] can be used, leading to Algorithm 1. Convergence of the sequence $(x_k)_{k \in \mathbb{N}}$ to a solution to (2) is established, for γ, τ positive stepsizes verifying $\gamma\tau \leq \frac{1}{\|H\|_2^2}$.

Algorithm 1 Primal-dual (PD) algorithm

Init: Choose $\tau, \gamma > 0$, $(x_0, y_0) \in \mathbb{R}^n \times \mathbb{R}^m$ and $y_0 = y_{-1}$.
for $k = 0, 1, \ldots$ **do**
  $\tilde{y}_k = 2y_k - y_{k-1}$
  $x_{k+1} = \mathrm{prox}_{\tau\|\cdot\|_1}(x_k - \tau H^\top \tilde{y}_k)$
  $y_{k+1} = \mathrm{prox}_{\gamma\sigma_{\mathscr{B}(z,\rho)}(\cdot)}(y_k + \gamma H x_{k+1})$
end for

We now give explicitly the expressions of the proximity operators involved in Algorithm 1. First,

$$(\forall x = (x_i)_{1 \leq i \leq n} \in \mathbb{R}^n) \quad \mathrm{prox}_{\tau\|\cdot\|_1}(x) = (\mathrm{sign}(x_i)\max(|x_i| - \tau, 0))_{1 \leq i \leq n}. \tag{6}$$

Second, using Moreau's decomposition in [21, Theorem 14.3 (ii)], for every $\gamma > 0$,

$$(\forall y \in \mathbb{R}^m) \quad \operatorname{prox}_{\gamma\sigma_{\mathscr{B}(z,\rho)}}(y) = y - \gamma P_{\mathscr{B}(z,\rho)}\left(\gamma^{-1}y\right). \tag{7}$$

Hereabove, $P_{\mathscr{B}(z,\rho)}$ is the projection onto the ball $\mathscr{B}(z,\rho)$ of center z and radius ρ. Using the expression of this projection leads, for every $y \in \mathbb{R}^m$, and for every $\gamma > 0$, to

$$\operatorname{prox}_{\gamma\sigma_{\mathscr{B}(z,\rho)}}(y) = \begin{cases} 0, & \text{if} \quad \gamma^{-1}y \in \mathscr{B}(z,\rho) \\ y - \gamma\left(z + \frac{y-\gamma z}{\|y-\gamma z\|_2}\rho\right) & \text{if} \quad \gamma^{-1}y \notin \mathscr{B}(z,\rho) \end{cases}. \tag{8}$$

3.1.3. Unrolled architecture

To create a deep neural network architecture, unrolling starts by converting, for a fixed number of layers K, the algorithm's iterations into layers. In our case, the primal and dual iterations in Algorithm 1 reveal building blocks of feedforward network structures. Namely, we identify the following multi-branch architecture: for a given observed data z, and for $k \in \{0,\dots,K-1\}$,

$$\left(x_{k+1}, y_{k+1}\right) = \left(\mathscr{L}_{p_k}(x_k, y_k, y_{k-1}), \mathscr{L}^z_{d_k}(\mathscr{L}_{p_k}(x_k, y_k, y_{k-1}), y_k)\right). \tag{9}$$

with a primal branch $\mathscr{L}_{p_k}$, and a dual branch $\mathscr{L}^z_{d_k}$ (depending on z), that process elements of the sequences $(x_k)_{0\le k\le K}$ and $(y_k)_{0\le k\le K}$. Each branch can be rewritten as a feedforward layer, such that, for $k \in \{0,\dots,K-1\}$,

$$x_{k+1} = \mathscr{L}_{p_k}(x_k, y_k, y_{k-1}) = \mathscr{A}_{p_k}(W_{p_k}x_k + b_{p_k}), \tag{10}$$

with weight matrix $W_{p_k} = \mathrm{Id}_n$ the identity matrix of $\mathbb{R}^n$, bias $b_{p_k} = -\tau_k H^\top(2y_k - y_{k-1}) \in \mathbb{R}^n$, and activation function

$$(\forall x \in \mathbb{R}^n) \quad \mathscr{A}_{p_k}(x) = \operatorname{prox}_{\tau_k\|\cdot\|_1}(x). \tag{11}$$

Similarly, for $k \in \{0,\dots,K-1\}$,

$$y_{k+1} = \mathscr{L}^z_{d_k}(x_{k+1}, y_k) = \mathscr{A}_{dk}(W_{dk}y_k + b_{dk}), \tag{12}$$

with $W_{d_k} = \mathrm{Id}_m$, bias $b_{dk} = \gamma_k H x_{k+1} \in \mathbb{R}^m$, and activation function

$$(\forall y \in \mathbb{R}^m) \quad \mathscr{A}_{dk}(y) = \operatorname{prox}_{\gamma_k\sigma_{\mathscr{B}(z,\rho_k)}(\cdot)}(y). \tag{13}$$

For the sake of conciseness, we denote by $\mathscr{L}^z_k$, for $k \in \{0,\dots,K-1\}$, a global layer of our proposed architecture, such that $(x_{k+1}, y_{k+1}) = \mathscr{L}^z_k(x_k, y_k)$. We then propose to untie sequences $(\tau_k)_{0\le k\le K-1}$, $(\gamma_k)_{0\le k\le K-1}$ and $(\rho_k)_{0\le k\le K-1}$, involved in activation and bias terms of primal and dual branches, allowing them to vary along layers to be learnt automatically. To this end, we design network architectures which will be attached to each layer $\mathscr{L}^z_k$, enforcing their positivity through the ReLU activation function

$$(\forall z = (z_\ell)_{1\le\ell\le m} \in \mathbb{R}^m) \quad \mathrm{ReLU}(z) = (\max(0, z_\ell))_{1\le\ell\le m}.$$

Specifically, let $\theta = \{t_k, g_k, r_k\}_{0\le k\le K-1}$ be the vector of parameters to be learned then, for every $k \in \{0,\dots,K-1\}$,

$$\tau_k = \mathrm{ReLU}(t_k), \quad \gamma_k = \mathrm{ReLU}(g_k), \quad \rho_k = \mathrm{ReLU}(r_k). \tag{14}$$

The global layer architecture is showcased on Fig.1. Once the multi-layer architecture is defined, there remains to train it, to estimate the learnable weights θ. Let $\mathscr{S} = \{(\bar{x}_s, z_s) \mid s \in \{1,\dots,S\}\}$ be a training set comprised of S pairs of groundtruth and degraded data samples. The (primal-dual) network output, for a given data z and initialization (x_0, y_0) reads $g^z_\theta(x_0, y_0) = \mathscr{L}^z_{K-1} \circ \dots \circ \mathscr{L}^z_k \circ \dots \circ \mathscr{L}^z_0(x_0, y_0) \in \mathbb{R}^n \times \mathbb{R}^m$. For a predefined loss $\ell : \mathbb{R}^n \times \mathbb{R}^n \mapsto \mathbb{R}$, acting on the primal space, the optimal hyperparameter setting $\widehat{\theta}$ is the solution to the optimization problem

$$\min_\theta \quad E(\theta) = \frac{1}{S}\sum_{s=1}^{S} \ell(\widetilde{g}^{z_s}_\theta(x_{s,0}, y_{s,0}), \bar{x}_s), \tag{15}$$

where $\widetilde{g} \in \mathbb{R}^n$ denotes the primal component of g. A common choice for the loss is the mean squared error (MSE) loss obtained by setting $\ell(x, \bar{x}) = \frac{1}{n}\|x - \bar{x}\|^2$. Standard neural network training algorithms such as stochastic gradient or Adam can be used to solve (15).

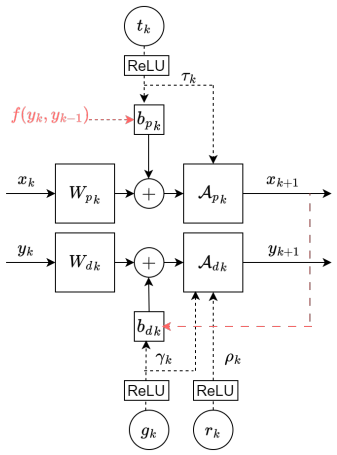

Fig. 1. Overview of one layer $\mathscr{L}^z_k$ of the proposed U-PD architecture. $(W_{p_k})_{0\le k\le K-1}$, $(W_{dk})_{0\le k\le K-1}$, $(b_{p_k})_{0\le k\le K-1}$, $(b_{dk})_{0\le k\le K-1}$, $(\mathscr{A}_{p_k})_{0\le k\le K-1}$ and $(\mathscr{A}_{dk})_{0\le k\le K-1}$ are respectively primal/dual weight operators, biases and activations. $(g_k)_{0\le k\le K-1}$, $(r_k)_{0\le k\le K-1}$, and $(t_k)_{0\le k\le K-1}$ are learnable weights and ReLU is an activation enforcing positivity.

3.2. Unrolled ISTA (U-ISTA) architecture

3.2.1. Problem formulation

Let $\widehat{x}$ be the minimizer of the following regularized cost function F for estimating $\bar{x}$,

$$\widehat{x} \in \arg\min_{x\in\mathbb{R}^n}\left(F(x) = \frac{1}{2}\|Hx - z\|_2^2 + \chi\|x\|_1\right). \tag{16}$$

Hereabove, χ is a regularization hyperparameter balancing the data fidelity term and the non-differentiable ℓ_1 norm aiming to promote the sparsity prior.

3.2.2. Algorithm

To solve (16), ISTA [19], summarized in Algorithm 2 can be used, γ is a stepsize in $]0, 2/\|H\|_2^2[$.

Algorithm 2 ISTA algorithm

Init: Choose $\gamma, \chi > 0$, $x_0 \in \mathbb{R}^n$.
for $k = 0, 1, \ldots$ **do**
    $x_{k+1} = \text{prox}_{\gamma\chi\|\cdot\|_1}(x_k - \gamma H^\top (Hx_k - z))$
end for

3.2.3. Unrolled architecture

As described in Section 3.1.3, after reinterpreting the algorithm initial K iterations as feedforward structures, sequences $(\gamma_k)_{0\leq k\leq K-1}$ and $(\chi_k)_{0\leq k\leq K-1}$ are untied and learnt by enforcing nonnegativity using ReLU, yielding U-ISTA architecture.

3.3. Unrolled half-quadratic (U-HQ) architecture

3.3.1. Problem formulation

Lastly, we propose minimizing a differentiable cost function to obtain the estimate $\widehat{x}$ of $\bar{x}$. The objective function is composed of a data fidelity term and a sparsity-inducing regularization function, so yielding

$$\widehat{x} \in \underset{x\in\mathbb{R}^n}{\text{argmin}} \left(F(x) = \frac{1}{2}\|Hx - z\|_2^2 + \Psi(x) \right), \qquad (17)$$

with $(\forall x \in \mathbb{R}^n)\ \Psi(x) = \sum_{i=1}^n \psi(x_i)$. The i-th component of a vector x in $\mathbb{R}^n$ is denoted by x_i with $i \in \{1, \ldots, n\}$ and $\psi : \mathbb{R} \to [0, +\infty)$ is defined as the hybrid potential,

$$(\forall t \in \mathbb{R}) \quad \psi(t) = \lambda_1 \delta_1 \left(|t| - \delta_1 \log\left(\frac{|t|}{\delta_1} + 1 \right) \right) + \lambda_2 \frac{\delta_2^2}{2} \log\left(1 + \frac{t^2}{\delta_2} \right). \qquad (18)$$

The above penalty, introduced in [11], mixes the Fair potential, that approximates the ℓ_1 norm, and the Cauchy potential, approaching the ℓ_0 count measure.

3.3.2. Half-quadratic algorithm

The optimization problem (17) is minimized by the half-quadratic (HQ) algorithm [20], given in Algorithm 3. For every $k \in \mathbb{N}$, $A(x_k)$ is a symmetric definite positive matrix ensuring a majorizing property on the curvature of F, and $\gamma_k > 0$ is a stepsize. Explicit expressions of $\nabla F(x_k)$ and $A(x_k)$ for Problem (1), and convergence guarantees for HQ method, can be found for instance, in [3].

Algorithm 3 Half-quadratic (HQ) algorithm

Init: Choose $x_0 \in \mathbb{R}^n$.
for $k = 0, 1, \ldots$ **do**
    Build majorant metric $A(x_k)$,
    $x_{k+1} = x_k - \gamma_k A(x_k)^{-1} \nabla F(x_k)$.
end for

3.3.3. Unrolled architecture

As for U-HQ, we learn $(\lambda_{1,k}, \lambda_{2,k}, \gamma_k)_{0\leq k\leq K-1}$ according to the architecture described in [11, Sec.4.3]. Namely, we retained the variant U-HQ-FixS, from [11, Sec.4.3], for its simplicity and low complexity.

4. EXPERIMENTAL RESULTS

In this section, we present our experimental results, through an application arising from chromatography. Chromatography is a separative technique in analytical chemistry, aiming to identify chemical components of a mixture, given their selective adsorption and retention time. The acquisition step, given the chromatogram $\bar{x}$, reads as Problem (1), where H models the instrument response (e.g., Gaussian blur) [24, 25].

4.1. Datasets

We build three synthetic datasets, D0, D1, and D2, with pairs $(\bar{x}, z)$, of size $n = 2000$ and $m = 2099$, using the generative model (1). Matrix H is a Toeplitz convolution matrix, associated to a centered Gaussian blur with standard deviation equals one, and ε is a Gaussian i.i.d. noise, with standard deviation 0.02. Each ground truth signal $\bar{x}$ mimics a realistic 1D chromatography diagram, using $\bar{x} = \pi * \bar{s}$, with π a Fraser-Suzuki kernel [26], with peak width 0.5 and asymmetry coefficient 0.2. Moreover, the ground truth $\bar{s}$ is a randomly generated spiky signal whose sparsity pattern depends on the dataset, namely $\bar{s}$ has 1.5%, 3%, and 4.5% non-zero entries, and 5, 3, and 1, as its minimum distance between two consecutive peaks, for datasets D0, D1, and D2, respectively. We create training, validation, and test sets of sizes 1000, 200 and 200 for the three datasets. Finally, we also present inference results, for each method, on the real spectra from [24].

4.2. Iterative methods

We compare U-PD, U-ISTA, and U-HQ, as described in Section 3, to their respective iterative optimization approaches PD, ISTA, and HQ. The stepsizes γ and τ of PD algorithm are set following the strategy proposed in [23], namely $\tau\gamma = \frac{0.99}{\|H\|_2^2}$ and $\tau = \gamma$. The bound ρ is finetuned manually, using a predefined grid of values proportionally to the noise level. For initialization we use $x_0 = y_0 = y_{-1}$. The set of parameters providing the lowest MSE on 200 samples randomly picked from the training set is retained, to be used on the test set. For

ISTA, we also directly finetuned χ proportionally to the noise level. The last iterative-based method we compare with is the HQ algorithm, where parameters $(\lambda_i, \delta_i)_{i\in\{1,2\}}$ are determined by grid search. PD, ISTA, and HQ are run until stabilization of the MSE, on an Intel@ Xeon(R) W-2135 CPU @ 3.70 GHz with 12 cores, using Python 3.8.5 code.

4.3. Training settings

The MSE loss is used for training all unrolled methods as well as to finetune parameters of optimization-based methods. In all experiments, training and validation batch sizes are set to 5. Learnable weights are tuned through backpropagation using Adam, where the learning rate is set according to each dataset, to maintain a stable training. As entry for the multiple architectures, we use the null vector, namely for U-HQ and U-ISTA, $x_0 = 0$ and for U-PD, $x_0 = y_0 = y_{-1} = 0$. To set the number of layers, we train the unrolled architectures for several choices of K, and retain the number of layers returning a minimal loss on validation set while maintaining a reasonable training execution time. Unrolled methods are implemented using Pytorch 1.8.0, and run on an Nvidia DGX server with GPU V100 SXM2, 32 GB RAM.

4.4. Numerical results

Table 1 summarizes all our results. Unrolled architectures outperform their iterative counterparts in terms of average restoration quality, measured in MSE, and truncated SNR in dB (TSNR, as defined, for instance, in [11]), notably for HQ and PD. U-ISTA seems to be slightly disadvantaged, which might show that learning only the regularization and stepsize parameters was not enough to reach full representation capacity for this network. PD and its unrolled version are (sometimes far) below their competitors, in quantitative scores. This is probably due to the slow convergence of PD iterations, as it can be seen in Figure 2(left). Inference time of unrolled methods is significantly lower than that of iterative methods, since the number of layers is considerably lower than typical iteration numbers, and a GPU implementation is used. Among all unrolled approaches, U-HQ exhibits a higher execution time, due to costly matrix inversions, while U-ISTA is the fastest. We finally test our trained networks, U-PD, U-ISTA, and U-HQ, on the restoration of a real signal from [25], degraded in a similar fashion than our synthetic signals. We retrieve SNR values of 14.31 dB, 14.92 dB, and 21.68 dB, respectively. On this signal, U-HQ outperforms ℓ_1-based methods U-PD and U-ISTA, by a large margin, due to its smoothed hybrid penalty (18). The latter helps to better retrieve the non-null signal baseline, as well as to recover precisely its peak intensities without bias effect, as can shown in Figure 2(right).

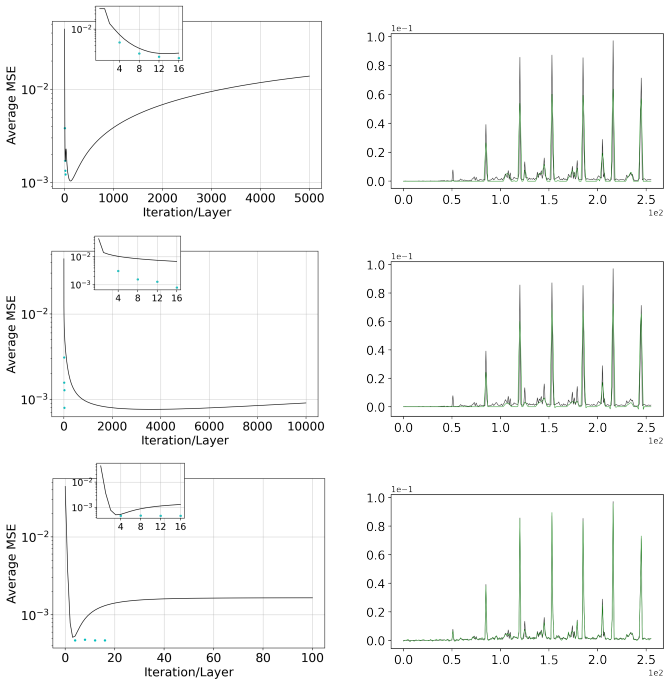

Fig. 2. Left: Average MSE loss (and zoom) vs iterations/layers, on validation set of D0, for iterative schemes (continuous black) and trained unrolled schemes (cyan markers), using PD, ISTA, and HQ (from top to bottom. Right: Real chromatogram from [25] (black line) and its restoration (green line), using U-PD, U-ISTA, and U-HQ (from top to bottom)

5. CONCLUSIONS

This work proposes a comprehensive study of three deep unrolling approaches for solving a sparse inverse problem arising in analytical chemistry. Our experiments show the competitiveness of all unrolling paradigms with respect to their respective iterative counterparts. The fastest iterative methods seem to yield optimal unrolling architectures, which confirms previous studies in the field of image processing. The final choice of method remains highly dataset and task dependent, and benchmarking remains necessary. To that end, reproducible code is made available by the authors at `https://github.com/GHARBIMouna`.

6. REFERENCES

[1] J. Honerkamp and J. Weese, "Tikhonovs regularization method for ill-posed problems: A comparison of different methods for the determination of the regularization parameter," *Continuum Mechanics and Thermodynamics*, vol. 2, pp. 17–30, 1990.

[2] P. Zheng, E. Chouzenoux, and L. Duval, "PENDANTSS: penalized norm-ratios disentangling additive noise, trend and sparse spikes," *IEEE Signal Processing Letters*, vol. 30, pp. 215–219, 2023.

[3] E. Chouzenoux, A. Jezierska, J.-C. Pesquet, and H. Talbot, "A

	D0			D1			D2		
Name	MSE	TSNR (dB)	Time (s.)	MSE	TSNR (dB)	Time (s.)	MSE	TSNR (dB)	Time (s.)
PD	$1.37 \cdot 10^{-2}$	4.93	5.17	$2.63 \cdot 10^{-2}$	5.45	5.17	$6.99 \cdot 10^{-2}$	3.42	5.17
ISTA	$\mathbf{8.92 \cdot 10^{-4}}$	**16.90**	2.63	$\mathbf{1.85 \cdot 10^{-3}}$	**17.17**	2.63	$\mathbf{2.78 \cdot 10^{-3}}$	**17.44**	2.63
HQ	$1.28 \cdot 10^{-3}$	15.56	**0.74**	$2.52 \cdot 10^{-3}$	15.77	**0.73**	$3.77 \cdot 10^{-3}$	16.15	**0.73**
U-PD	$1.21 \cdot 10^{-3}$	15.62	$3 \cdot 10^{-2}$	$2.88 \cdot 10^{-3}$	15.10	$3 \cdot 10^{-2}$	$4.77 \cdot 10^{-3}$	15.09	$2 \cdot 10^{-2}$
U-ISTA	$8.17 \cdot 10^{-4}$	17.47	$\mathbf{7 \cdot 10^{-3}}$	$3.16 \cdot 10^{-3}$	14.78	$\mathbf{6 \cdot 10^{-3}}$	$5.55 \cdot 10^{-3}$	14.41	$\mathbf{6 \cdot 10^{-3}}$
U-HQ	$\mathbf{4.70 \cdot 10^{-4}}$	**19.99**	$9 \cdot 10^{-3}$	$\mathbf{1.05 \cdot 10^{-3}}$	**19.64**	0.33	$\mathbf{1.66 \cdot 10^{-3}}$	**19.79**	0.33

Table 1. Comparative performance: Averaged MSE, TSNR, and inference time, over test sets of D0, D1, and D2 datasets.

majorize-minimize subspace approach for l2-l0 image regularization," *SIAM Journal on Imaging Science*, vol. 6, no. 1, pp. 563–591, 2013.

[4] B. Kaltenbacher, A. Neubauer, and O. Scherzer, "Iterative regularization methods for nonlinear ill-posed problems," in *Iterative Regularization Methods for Nonlinear Ill-Posed Problems*. de Gruyter, 2008.

[5] C.-J. Schuler, M. Hirsch, S. Harmeling, and B. Schölkopf, "Learning to deblur," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 7, pp. 1439–1451, 2015.

[6] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process*, vol. 26, no. 7, pp. 3142–3155, 2017.

[7] M.-T. McCann, K.-H. Jin, and M. Unser, "Convolutional neural networks for inverse problems in imaging: A review," *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 85–95, 2017.

[8] V. Monga, Y. Li, and Y.-C. Eldar, "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing," *IEEE Signal Process. Mag.*, vol. 38, no. 2, pp. 18–44, 2021.

[9] D. Ren, W. Zuo, D. Zhang, L. Zhang, and M.-H. Yang, "Simultaneous fidelity and regularization learning for image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 1, pp. 284–299, 2019.

[10] C. Bertocchi, E. Chouzenoux, M.C. Corbineau, J.-C. Pesquet, and M. Prato, "Deep Unfolding of a Proximal Interior Point Method for Image Restoration," *Inverse Problems*, vol. 36, no. 3, pp. 034005, Feb. 2020.

[11] M. Gharbi, E. Chouzenoux, and J.-C. Pesquet, "Unrolled half-quadratic approach for sparse signal recovery in spectroscopy," *Signal Processing*, vol. 218, pp. 109369, May 2014.

[12] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *Proceedings of the 27th International Conference on International Conference on Machine Learning (ICML 2010)*, 2010, pp. 399–406.

[13] M. Savanier, E. Chouzenoux, J.-C. Pesquet, and C. Riddell, "Deep unfolding of the DBFB algorithm with application to ROI CT imaging with limited angular density," *IEEE Trans. Comp. Imaging*, vol. 9, pp. 502–516, 2023.

[14] H.T.-Vy. Le, N. Pustelnik, and M. Foare, "The faster proximal algorithm, the better unfolded deep learning architecture ? The study case of image denoising," in *Proceedings of the 30th European Signal Processing Conference (EUSIPCO 2022)*, 2022, pp. 947–951.

[15] M. Jiu and N. Pustelnik, "A deep primal-dual proximal network for image restoration," *IEEE J. Sel. Top. Signal Process.*, vol. 15, no. 2, pp. 190–203, 2021.

[16] Y. Huang, E. Chouzenoux, and J.-C. Pesquet, "Unrolled variational Bayesian algorithm for image blind deconvolution," *IEEE Trans. Image Process*, vol. 32, pp. 430–445, 2022.

[17] Z. Ramzi, P. Ciuciu, and J.-L. Starck, "Density compensated unrolled networks for non-Cartesian MRI reconstruction," in *Proceedings of 18th IEEE International Symposium on Biomedical Imaging (ISBI 2021)*, 2021, pp. 1443–1447.

[18] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imaging Vision*, vol. 40, no. 1, pp. 120–145, 2011.

[19] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm with application to wavelet-based image deblurring," in *Proceedings of 34th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2009)*, 2009, pp. 693–696.

[20] M. Allain, J. Idier, and Y. Goussard, "On global and local convergence of half-quadratic algorithms," *IEEE Trans. Image Process*, vol. 15, no. 5, pp. 1130–1142, 2006.

[21] H.-H. Bauschke and Patrick L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer Publishing Company, 2nd edition, 2017.

[22] 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*, vol. 32, no. 6, pp. 31–54, 2014.

[23] C. Molinari, M. Massias, L. Rosasco, and S. Villa, "Iterative regularization for low complexity regularizers," in *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics (AISTATS 2021)*, San Diego, CA, 2021.

[24] C. Vendeuvre, R. Ruiz-Guerrero, F. Bertoncini, L. Duval, D. Thiébaut, and M.-C. Hennion, "Characterisation of middle-distillates by comprehensive two-dimensional gas chromatography (GC × GC): A powerful alternative for performing various standard analysis of middle-distillates," *Journal of Chromatography A*, vol. 1086, no. 1-2, pp. 21–28, 2005.

[25] X. Ning, I.W. Selesnick, and L. Duval, "Chromatogram baseline estimation and denoising using sparsity (BEADS)," *Chemometrics and Intelligent Laboratory Systems*, vol. 139, pp. 156–167, 2014.

[26] A. Felinger, *Data Analysis and Signal Processing in Chromatography*, Elsevier, 1998.