Learning the Proximity Operator in Unfolded ADMM for Phase Retrieval

Pierre-Hugo Vial* , Paul Magron† , Thomas Oberlin‡ , Cédric Févotte*

*IRIT, Université de Toulouse, CNRS †Université de Lorraine, CNRS, Inria, LORIA ‡ISAE-SUPAERO, Université de Toulouse

Contributions

- Unfolding ADMM for phase retrieval (PR) with Bregman divergences.
- Replacing the proximity operator with trainable activation functions.
- Interpretation of the training stage as a metric/proximity operator learning problem.

Phase retrieval

- Data reconstruction from phaseless measurements $\mathbf{r} \in \mathbb{R}_+^K$, traditionally formulated as:

$$\min_{\mathbf{x}\in\mathbb{R}^L} \| |\mathbf{A}\mathbf{x}|^d - \mathbf{r}\|^2.$$

- $\mathbf{A} \in \mathbb{C}^{K\times L}$ is the measurement operator (in audio: the STFT).
- $d = 1$ (magnitude) or $d = 2$ (power spectrograms).

PR with Bregman divergences

- PR problem reformulated with a Bregman divergence in [1]:

$$\min_{\mathbf{x}\in\mathbb{R}^L} \mathcal{D}_\psi(|\mathbf{A}\mathbf{x}|^d \,|\, \mathbf{r}).$$

- Bregman divergences:

$$\mathcal{D}_\psi(\mathbf{p}\,|\,\mathbf{q}) = \sum_{k=1}^{K} [\psi(p_k) - \psi(q_k) - \psi'(q_k)(p_k - q_k)],$$

with ψ strictly-convex, continuously-differentiable generating function.
- They encompass β-divergences, Kullback-Leibler, Itakura-Saito and quadratic loss.

PR via ADMM

- ADMM-based algorithm derived in [1] to solve PR with Bregman divergences.
- Reformulation of the problem with auxiliary variables $\mathbf{u}$ and $\boldsymbol{\theta}$:

$$\min_{\mathbf{x},\mathbf{u},\theta} \mathcal{D}_\psi(\mathbf{u}\,|\,\mathbf{r}) \quad \text{s.t.} \quad (\mathbf{A}\mathbf{x})^d = \mathbf{u} \odot e^{i\boldsymbol{\theta}}.$$

- ADMM iterations:

$$\mathbf{h}_{t+1} = (\mathbf{A}\mathbf{x}_t)^d + \frac{\boldsymbol{\lambda}_t}{\rho}$$
$$\mathbf{u}_{t+1} = \text{prox}_{\rho^{-1}\mathcal{D}_\psi(\cdot\,|\,\mathbf{r})}(|\mathbf{h}_{t+1}|)$$
$$\boldsymbol{\theta}_{t+1} = \angle \mathbf{h}_{t+1}$$
$$\mathbf{x}_{t+1} = \mathbf{A}^{\mathsf{H}}\Big(\mathbf{u}_{t+1} \odot e^{i\boldsymbol{\theta}_{t+1}} - \frac{\boldsymbol{\lambda}_t}{\rho}\Big)^{1/d}$$
$$\boldsymbol{\lambda}_{t+1} = \boldsymbol{\lambda}_t + \rho(\mathbf{A}\mathbf{x}_{t+1} - \mathbf{u}_{t+1} \odot e^{i\boldsymbol{\theta}_{t+1}})$$

- Issue: closed-form of $\text{prox}_{\rho^{-1}\mathcal{D}_\psi(\cdot\,|\,\mathbf{r})}$ not available in general.

Proposed method

- ADMM as a neural network $\mathbf{U}$ via unfolding:

$$(\mathbf{x}_T, \boldsymbol{\lambda}_T) = \mathbf{U}(\mathbf{x}_0, \boldsymbol{\lambda}_0) = \mathsf{U}_1 \circ \cdots \circ \mathsf{U}_T(\mathbf{x}_0, \boldsymbol{\lambda}_0).$$

- Each layer U_t emulates the ADMM iterations.
- F_t emulates the proximity operator and is built with Adaptive Piecewise Linear (APL) units.

$$\mathsf{F}_t(\mathbf{y}, \mathbf{r}) = \mathsf{APL}_t\left(\gamma_t^{(1)}\mathbf{y} + \gamma_t^{(2)}\frac{\mathbf{r}^{\beta_t - 1}}{\beta_t - 1}\right),$$

$$\mathsf{APL}(\mathbf{y}) = \max(\mathbf{y}, 0) + \sum_{c=1}^{C} w_c \max(-\mathbf{y} + b_c, 0).$$

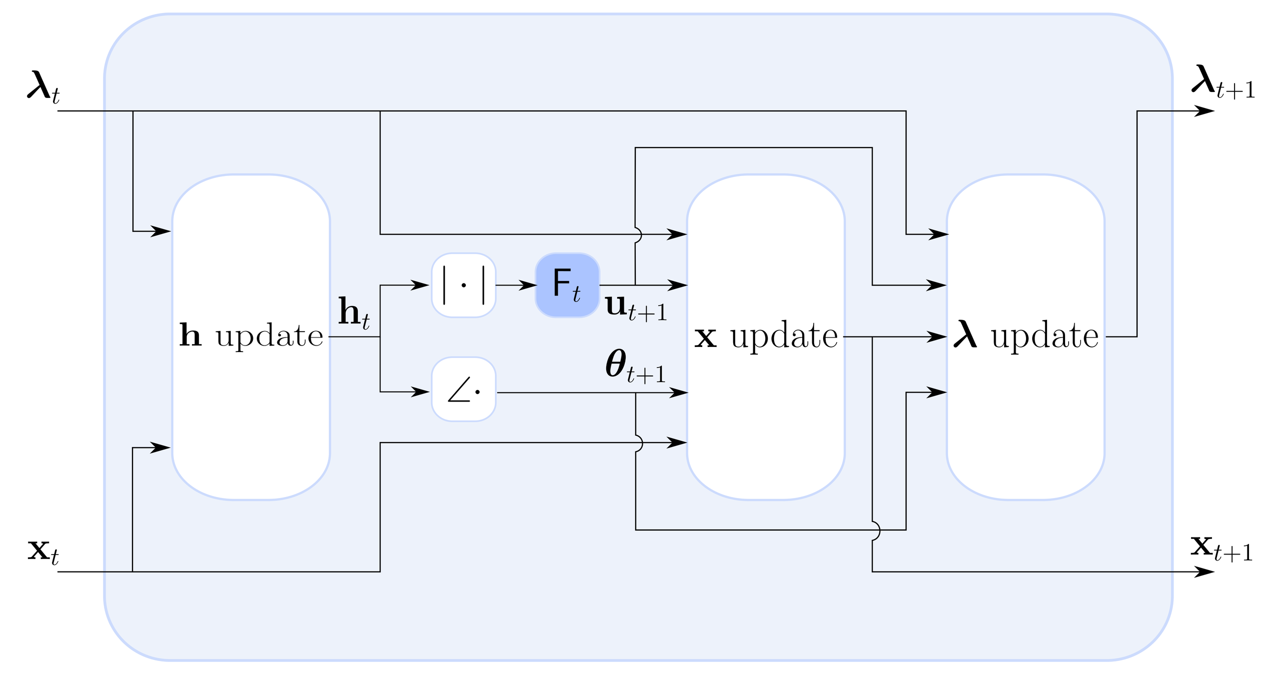
- Trainable parameters:

$$\boldsymbol{\Theta} = \{w_{c,t}, b_{c,t}, \gamma_t^{(1)}, \gamma_t^{(2)}, \beta_t\}.$$

- Two settings: tied and untied parameters.

[1] Pierre-Hugo Vial, Paul Magron, Thomas Oberlin, and Cédric Févotte.
Phase retrieval with Bregman divergences and application to audio signal recovery.
IEEE Journal of Selected Topics in Signal Processing, 15(1):51–64, 2021.

[2] Pierre-Hugo Vial, Paul Magron, Thomas Oberlin, and Cédric Févotte.
Learning the proximity operator in unfolded ADMM for phase retrieval.
arXiv preprint arXiv:2204.01360, 2022.

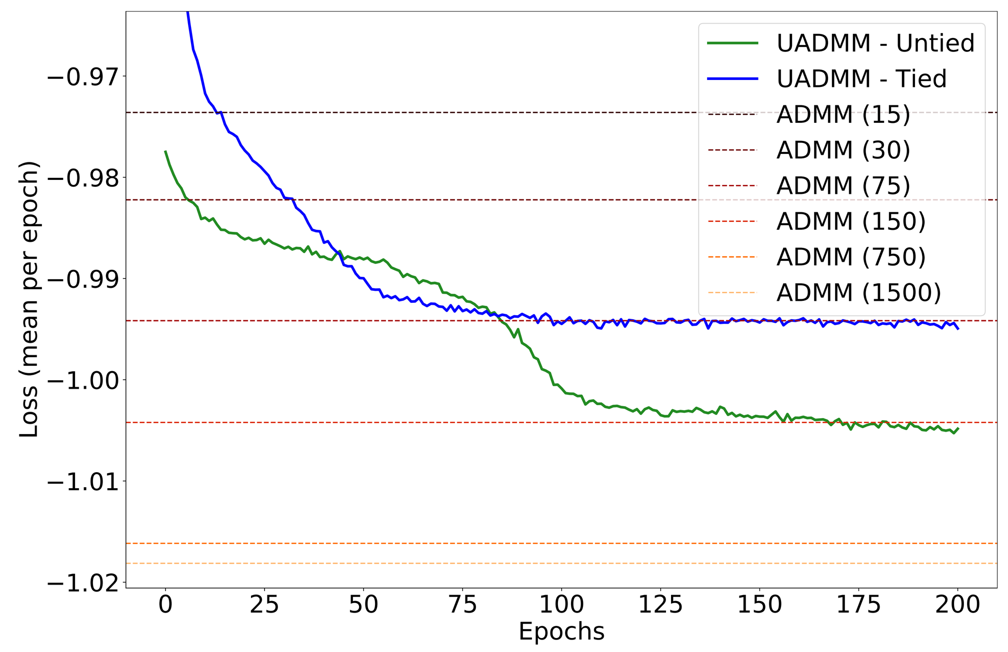


Experiments

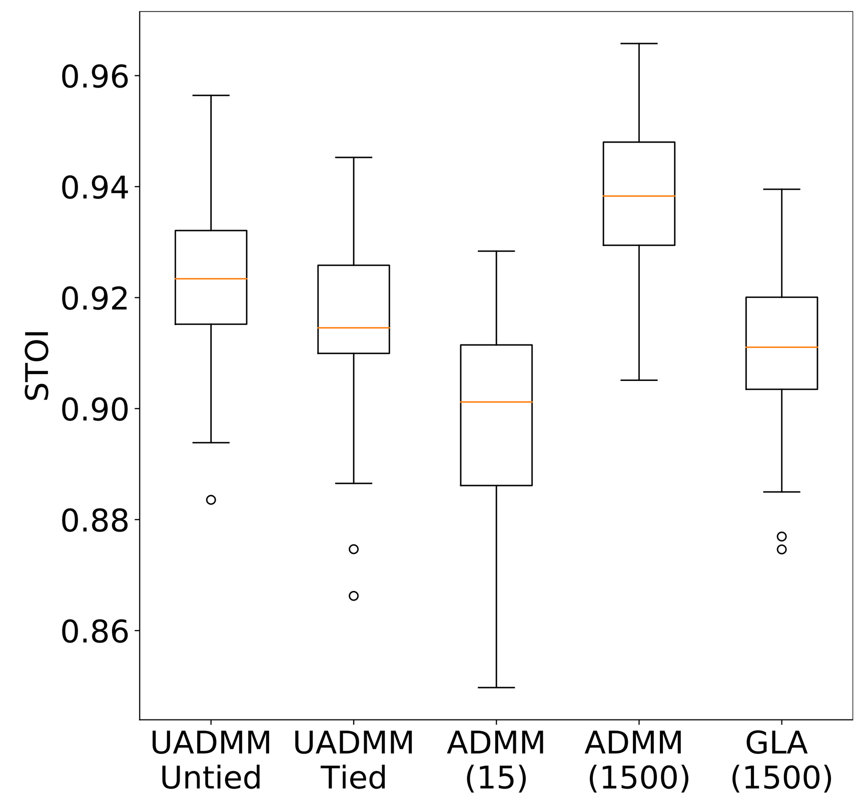
PR with unfolded ADMM (15 layers)

- Dataset: speech signals from TIMIT.
- Training with negative STOI and ADAM optimizer.
- Evaluation with the STOI metric.

- Training loss:



- Evaluation on the test set:

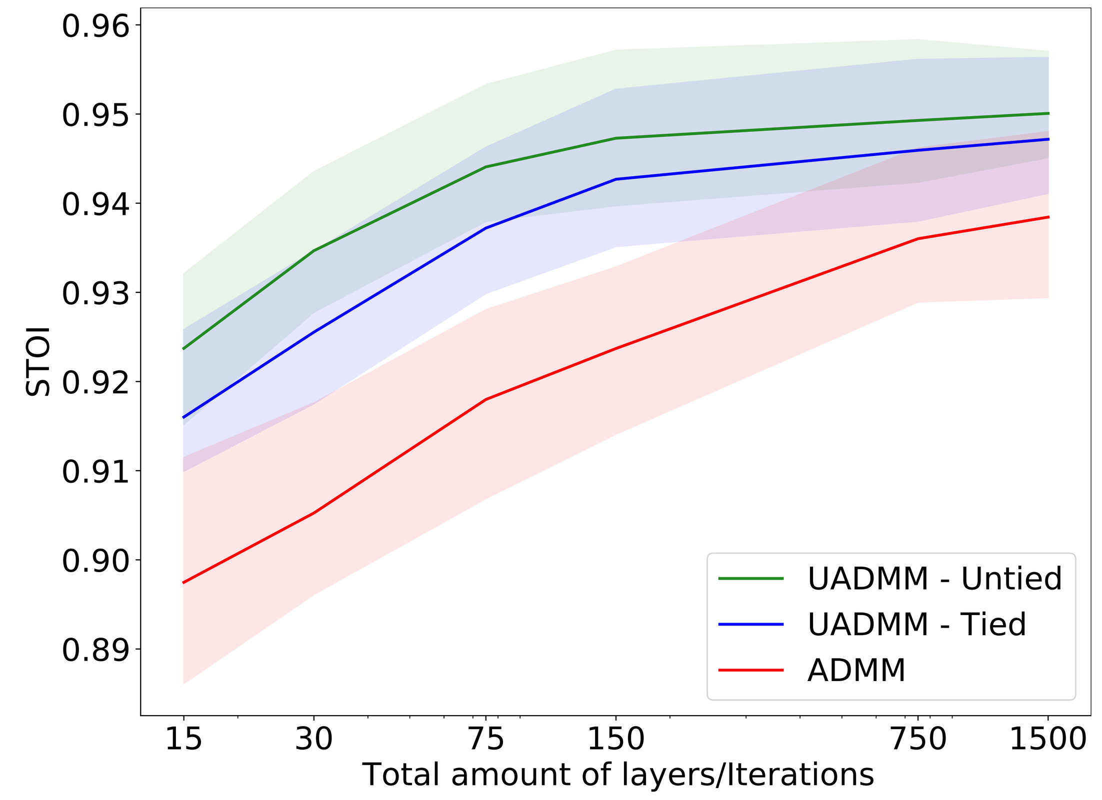


PR with iterated unfolded ADMM

- Iterated model ($n \times 15$ layers):

$$(\mathbf{x}_{nT}, \boldsymbol{\lambda}_{nT}) = \mathbf{U}^n(\mathbf{x}_0, \boldsymbol{\lambda}_0) = \mathbf{U} \circ \cdots \circ \mathbf{U}(\mathbf{x}_0, \boldsymbol{\lambda}_0).$$

- Evaluation over the test set with iterated model:



- With identical number of layers/iterations, unfolded ADMM outperforms ADMM for PR.

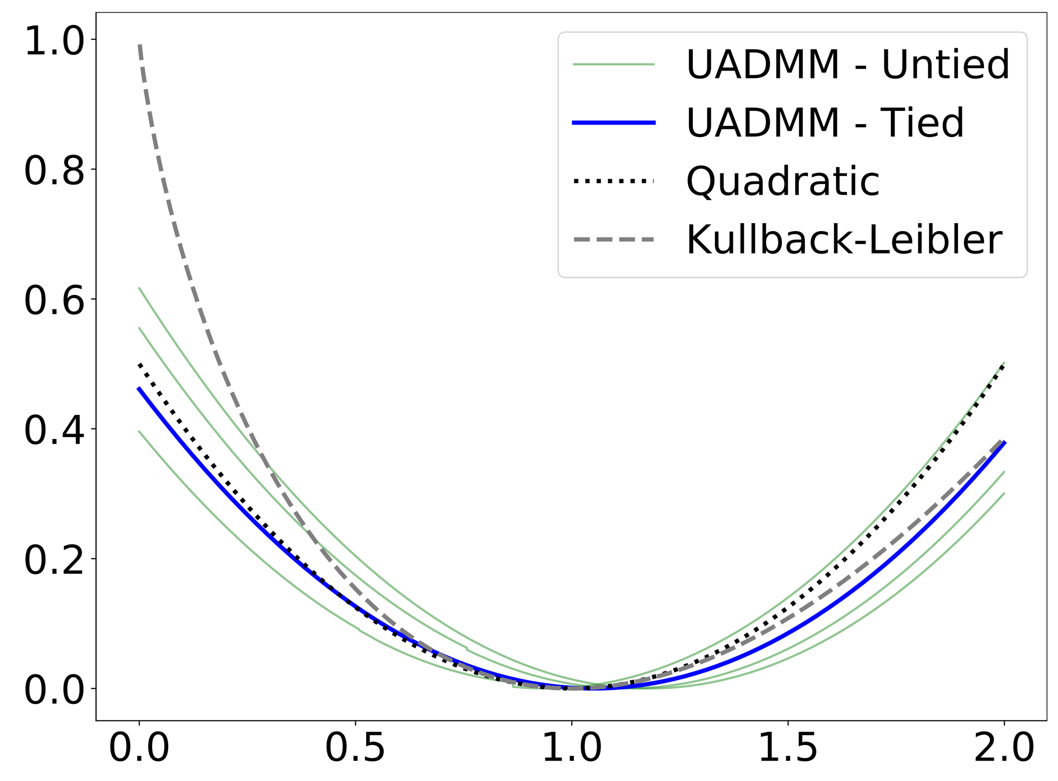
Metric learning

- Existence and characterization of $f_{\mathbf{r},t}$ s.t. :

$$\mathsf{F}_t(\mathbf{y}, \mathbf{r}) = \text{prox}_{f_{\mathbf{r},t}}(\mathbf{y}).$$

- Learning tied parameters → learning $\mathcal{D}_\psi(\cdot\,|\,\mathbf{r})$.

- Learned metrics $f_{r,t}(y)$ with $r = 1$:



- Interpretable and light architecture.