Unrolled Majorization-Minimization Approaches for Sparse Signal Reconstruction in Analytical Chemistry

Presented by: Emilie Chouzenoux
In collaboration with: Mouna Gharbi, Laurent Duval, Jean-Christophe Pesquet, Silvia Villa

October 29, 2025

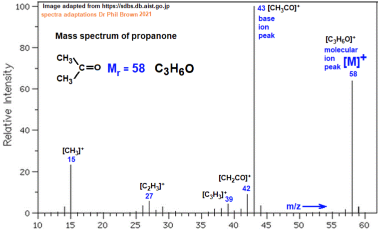
TRAINING DATA-DRIVEN EXPERTS IN OPTIMIZATION
MSCA-ITN 2019

Outline

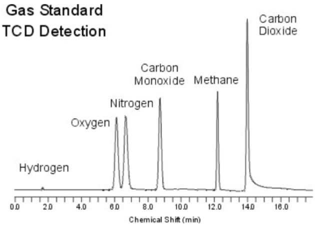
Applicative context: Analytical Chemistry

Applicative context: Collaboration with IFP Energies Nouvelles.
Analytical chemistry: Study of chemical composition or characteristics of compounds.

Mass Spectrometry:
Separation based on mass-to-charge ratio of ions.



Chromatography:
Separation based on the retention time of molecules.



$\Longrightarrow$ Indirect **sparse** measurements + uncertainties: Inverse problems.

Problem formulation

Forward model:

$$\mathbf{z} = \mathbf{H}\overline{\mathbf{x}} + \mathbf{e}, \tag{1}$$

- $\mathbf{z} \in \mathbb{R}^M$: observed acquisition
- $\overline{\mathbf{x}} \in \mathbb{R}^N$: original sparse positive-valued signal
- $\mathbf{H} \in \mathbb{R}^{M \times N}$: measurement degradation, typically a convolution with an application-dependant kernel shape (Gaussian, Voigt, etc.)
- $\mathbf{e}$: corrupting noise, here assumed additive Gaussian iid

$\longrightarrow$ **Our goal is to retrieve an estimate $\widehat{\mathbf{x}} \in \mathbb{R}^N$ of $\overline{\mathbf{x}} \in \mathbb{R}^N$ knowing $\mathbf{H}$ and $\mathbf{z}$.**

Challenges:

- Heterogenous signals
- High dimensionality
- Large databases

State-of-the-art methods

Goal: Find an estimate $\widehat{\mathbf{x}}$ of $\overline{\mathbf{x}}$ from knowledge of $\mathbf{z}$ and $\mathbf{H}$.

1. Model-based

$$\widehat{\mathbf{x}} \in \underset{\mathbf{x}\in\mathbb{R}^N}{\operatorname{argmin}} \; \underbrace{J(\mathbf{Hx},\mathbf{z})}_{\text{Data fidelity}} + \underbrace{\lambda\Psi(\mathbf{x})}_{\text{Regularization}}$$

$\Rightarrow$ Iterative optimization algorithm

2. Learning-based

$$\widehat{\mathbf{x}} = \underbrace{h(\mathbf{z},\widehat{\Theta})}_{\text{Mapping}}$$

$\Rightarrow$ Backpropagation

State-of-the-art methods

Model-based	Learning-based
✓Theoretical guarantees	✓Good accuracy
✓Robustness	✓Easy deployment
✓Explainability	✓Fast (at test phase)
✗High computational cost	✗Lack of robustness
✗Tedious parameter tuning	✗Black-Box: Not explainable

$\longrightarrow$ Get the best out of the two approaches?

Unrolling/Unfolding

❶ Traditional iterative algorithm

1: Init: Choose Θ, $\mathbf{x}_0 \in \mathbb{R}^N$.
2: **for** $k = 0, 1, \ldots$ **do**
3: $\quad \mathbf{x}_{k+1} = \mathcal{I}_k^{(\Theta)}(\mathbf{x}_k, \mathbf{z})$,
4: **end for**
5: Return $\widehat{\mathbf{x}}$.

❷ Reinterpretation to perform unrolling

- Truncate the number of iterations to a fixed value of layers K:

$$\textbf{Iteration } \mathcal{I}_k^{(\Theta)}(\cdot, \mathbf{z}) : \mathbb{R}^N \to \mathbb{R}^N \iff \textbf{Layer } \mathcal{L}_k^{(\Theta_k)}(\cdot, \mathbf{z}) : \mathbb{R}^N \to \mathbb{R}^N.$$

❸ Learning and inferring

- Train: Minimize task-oriented loss ℓ comparing pairs groundtruths/outputs of the unrolled architecture, w.r.t. $(\Theta_k)_{\{0 \leq k \leq K-1\}}$.
- Test: $\widehat{\mathbf{x}} = \mathcal{L}_{K-1}^{(\widehat{\Theta}_{K-1})}(\cdot, \mathbf{z}) \circ \cdots \circ \mathcal{L}_0^{(\widehat{\Theta}_0)}(\cdot, \mathbf{z})(\mathbf{x}_0)$.

This talk

- Unrolled half-quadratic approach for sparse signal recovery.
- Experimental results on Mass Spectrometry data.
- Comprehensive study of unrolling through chromatographic data.

Main references:

M. Gharbi, E. Chouzenoux, and J.-C. Pesquet. An Unrolled Half-Quadratic Approach for Sparse Signal Recovery in Spectroscopy. *Signal Processing*, vol. 218, pp. 109369, May 2024

M. Gharbi, S. Villa, E. Chouzenoux, J.-C. Pesquet, L. Duval Unrolled deep networks for sparse signal restoration in analytical chemistry, In Proceedings of *IEEE MLSP 2024*, London, UK, 22th-25th Sep. 2024

Unrolled half-quadratic approach for sparse signal recovery

Penalized least-squares minimization

Inverse problem: $\boxed{\mathbf{z} = \mathbf{H}\overline{\mathbf{x}} + \mathbf{e}}$

Goal: Recover an estimate $\widehat{\mathbf{x}}$ of $\overline{\mathbf{x}}$, assuming sparsity.

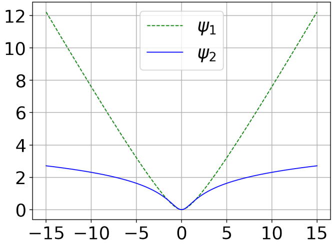
Optimization problem:

$$\widehat{\mathbf{x}} \in \underset{\mathbf{x}\in\mathbb{R}^N}{\operatorname{argmin}} \left(F(\mathbf{x}) = \frac{1}{2}\|\mathbf{H}\mathbf{x} - \mathbf{z}\|^2 + \lambda_1 \Psi_1(\mathbf{x}) + \lambda_2 \Psi_2(\mathbf{x}) \right)$$

$\Longrightarrow$ A **hybrid regularization term** to promote sparsity

$$(\forall i \in \{1,2\})\ (\forall \mathbf{x} \in \mathbb{R}^N)$$

$$\Psi_i(\mathbf{x}) = \sum_{n=1}^{N} \psi_i(x^{(n)}).$$

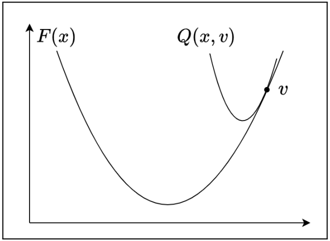


Majorization-Minimization

Majorant tangent function

Let $F : \mathbb{R}^N \longrightarrow \mathbb{R}$. The function $Q : \mathbb{R}^N \times \mathbb{R}^N \longrightarrow \mathbb{R}$ is said majorant tangent to F at $\mathbf{v} \in \mathbb{R}^N$ if

$$(\forall \mathbf{x} \in \mathbb{R}^N) \quad \begin{cases} F(\mathbf{x}) \leq Q(\mathbf{x}, \mathbf{v}), \\ F(\mathbf{v}) = Q(\mathbf{v}, \mathbf{v}). \end{cases}$$



Majorization-Minimization (MM)

$\Rightarrow$ General MM iteration: **m**ajorizing the criterion at the iterate with a (simple) surrogate function, then **m**inimizing the majorant to define the next iterate.

$$\mathbf{x}_{k+1} = \underset{\mathbf{x}\in\mathbb{R}^N}{\operatorname{argmin}}\, Q(\mathbf{x}, \mathbf{x}_k). \tag{2}$$

Assumption on penalties Ψ_1, Ψ_2

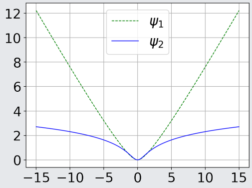
For every $i \in \{1,2\}$, $\psi_i : \mathbb{R} \longrightarrow \mathbb{R}$ is

(i) a differentiable, even function,

(ii) increasing on $[0, +\infty)$,

(iii) such that $\psi_i(\sqrt{\cdot})$ is concave on $[0, +\infty)$.

In addition, ψ_1 is convex.



$\Rightarrow$ Construction of a **quadratic** majorant tangent function.

Half-Quadratic (HQ) algorithm

- **Quadratic majorant function**:

$$(\forall \mathbf{x} \in \mathbb{R}^N, \forall \mathbf{v} \in \mathbb{R}^N) \quad Q(\mathbf{x}, \mathbf{v}) = F(\mathbf{v}) + \nabla F(\mathbf{v})^\top (\mathbf{x} - \mathbf{v}) + \frac{1}{2}(\mathbf{x} - \mathbf{v})^\top \mathbf{A}(\mathbf{v})(\mathbf{x} - \mathbf{v}),$$

where for every $\mathbf{v} \in \mathbb{R}^N$,

$$\mathbf{A}(\mathbf{v}) = \mathbf{H}^\top \mathbf{H} + \lambda_1 \underbrace{\mathrm{Diag}\{(\omega_1(\mathbf{v}^{(n)}))_{1 \le n \le N}\}}_{\Omega_1(\mathbf{v})} + \lambda_2 \underbrace{\mathrm{Diag}\{(\omega_2(\mathbf{v}^{(n)}))_{1 \le n \le N}\}}_{\Omega_2(\mathbf{v})},$$

$$\text{and } (\forall u \in \mathbb{R})\ (\forall i \in \{1,2\})\ \dot{\psi}_i(u) = \varrho_i(u) \text{ and } \omega_i(u) = \frac{\varrho_i(u)}{u}.$$

HQ algorithm

1: Init: Choose $\mathbf{x}_0 \in \mathbb{R}^N$ and $(\gamma_k)_{k \in \mathbb{N}} \in (0,2)$
2: **for** $k = 0, 1, \ldots$ **do**
3: $\quad \mathbf{x}_{k+1} = \mathbf{x}_k - \gamma_k \mathbf{A}(\mathbf{x}_k)^{-1} \nabla F(\mathbf{x}_k)$
4: **end for**

$\longrightarrow$ Convergence to a critical point $\tilde{\mathbf{x}}$ of F and $F(\mathbf{x}_k) \searrow F(\tilde{\mathbf{x}})$ as $k \longrightarrow +\infty$.

Proposed architecture

Reinterpretation: From iteration into layer

For K fixed, $\forall k \in \{0, \dots, K-1\}$,

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \gamma_k \mathbf{A}(\mathbf{x}_k)^{-1} \nabla F(\mathbf{x}_k),$$

$$\Leftrightarrow \mathbf{x}_{k+1} = \mathbf{x}_k - \gamma_k \mathbf{A}(\mathbf{x}_k)^{-1} \left(\mathbf{H}^\top (\mathbf{H}\mathbf{x}_k - \mathbf{z}) + \lambda_1 \Omega_{1,k} \mathbf{x}_k + \lambda_2 \Omega_{2,k} \mathbf{x}_k \right),$$

$$\Leftrightarrow \mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{W}_k \left(R_3 \left(\mathbf{V}_3 \begin{bmatrix} R_0 (\mathbf{V}_0 \mathbf{x}_k + \mathbf{b}_0) \\ R_1 (\mathbf{V}_1 \mathbf{x}_k + \mathbf{b}_1) \\ R_2 (\mathbf{V}_2 \mathbf{x}_k + \mathbf{b}_2) \end{bmatrix} + \mathbf{b}_3 \right) \right),$$

is equivalent to $\mathbf{x}_{k+1} = \mathcal{L}_k^{\mathbf{z}}(\mathbf{x}_k)$.

Weights

$\mathbf{V}_0 = \mathbf{H}^\top \mathbf{H}$

$\mathbf{V}_1, \mathbf{V}_2 = \mathbf{I}_N$

$\mathbf{V}_3 = [\mathbf{I}_N \quad \lambda_1 \mathbf{I}_N \quad \lambda_2 \mathbf{I}_N]$

$\mathbf{W}_k = \gamma_k \mathbf{A}(\mathbf{x}_k)^{-1}$

Biases

$\mathbf{b}_0 = -\mathbf{H}^\top \mathbf{z}$

$\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3 = 0_N$

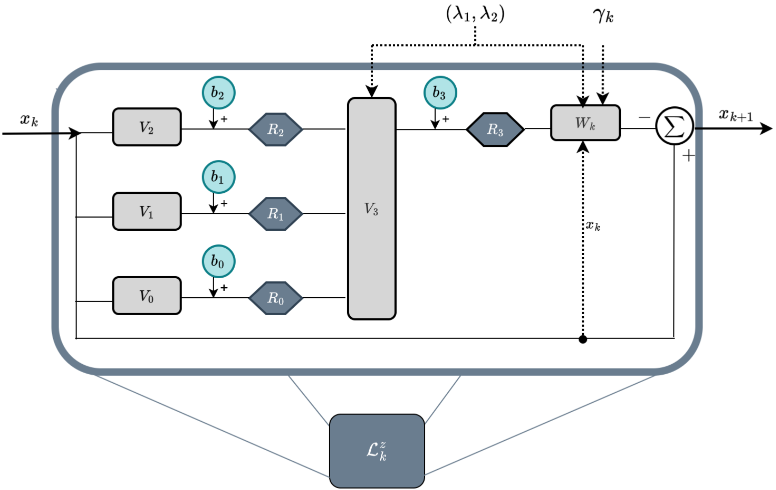
Activations

$R_0(\mathbf{x}) = R_3(\mathbf{x}) = \mathbf{x}$

$R_1(\mathbf{x}) = \left(\varrho_1(\mathbf{x}^{(n)}) \right)_{1 \le n \le N}$

$R_2(\mathbf{x}) = \left(\varrho_2(\mathbf{x}^{(n)}) \right)_{1 \le n \le N}$

Proposed architecture



Mathematical tools

Let $\Gamma_0(\mathbb{R})$ denote the class of proper lower-semicontinuous convex functions from $\mathbb{R}$ to $\mathbb{R} \cup \{+\infty\}$.

- For $\varphi \in \Gamma_0(\mathbb{R})$, its **proximity operator** $\operatorname{prox}_\varphi : \mathbb{R} \to \mathbb{R}$ is defined as

$$\boxed{(\forall x \in \mathbb{R}) \quad \operatorname{prox}_\varphi(x) = \operatorname*{argmin}_{t \in \mathbb{R}} \varphi(t) + \frac{1}{2}(t - x)^2.}$$

Mathematical tools

Let $\Gamma_0(\mathbb{R})$ denote the class of proper lower-semicontinuous convex functions from $\mathbb{R}$ to $\mathbb{R} \cup \{+\infty\}$.

- For $\varphi \in \Gamma_0(\mathbb{R})$, its **proximity operator** $\mathrm{prox}_\varphi : \mathbb{R} \to \mathbb{R}$ is defined as

$$\boxed{(\forall x \in \mathbb{R}) \quad \mathrm{prox}_\varphi(x) = \underset{t\in\mathbb{R}}{\mathrm{argmin}}\, \varphi(t) + \frac{1}{2}(t-x)^2.}$$

- $\varrho : \mathbb{R} \to \mathbb{R}$ is said α-**averaged** with $\alpha \in [0,1]$ if there exists a 1-Lipschitz function ϑ such that $\boxed{(\forall x \in \mathbb{R})\ \varrho(x) = (1-\alpha)x + \alpha\vartheta(x)}$.
  Moreover, it satisfies

$$\boxed{(\forall (x,t) \in \mathbb{R})^2 \quad |\varrho(x) - \varrho(t)|^2 \leq |x-t|^2 - \frac{1-\alpha}{\alpha}|x - \varrho(x) - t + \varrho(t)|^2}.$$

Mathematical tools

Let $\Gamma_0(\mathbb{R})$ denote the class of proper lower-semicontinuous convex functions from $\mathbb{R}$ to $\mathbb{R} \cup \{+\infty\}$.

- For $\varphi \in \Gamma_0(\mathbb{R})$, its **proximity operator** $\text{prox}_\varphi : \mathbb{R} \to \mathbb{R}$ is defined as

$$\boxed{(\forall x \in \mathbb{R}) \quad \text{prox}_\varphi(x) = \underset{t \in \mathbb{R}}{\text{argmin}}\, \varphi(t) + \frac{1}{2}(t - x)^2.}$$

- $\varrho : \mathbb{R} \to \mathbb{R}$ is said α-**averaged** with $\alpha \in [0, 1]$ if there exists a 1-Lipschitz function ϑ such that $\boxed{(\forall x \in \mathbb{R})\ \varrho(x) = (1 - \alpha)x + \alpha\vartheta(x)}$. Moreover, it satisfies

$$\boxed{(\forall (x, t) \in \mathbb{R})^2 \quad |\varrho(x) - \varrho(t)|^2 \leq |x - t|^2 - \frac{1 - \alpha}{\alpha}|x - \varrho(x) - t + \varrho(t)|^2}.$$

- A function ϱ is said **firmly nonexpansive** if it is α-averaged with $\alpha = 1/2$.

Mathematical tools

Let $\Gamma_0(\mathbb{R})$ denote the class of proper lower-semicontinuous convex functions from $\mathbb{R}$ to $\mathbb{R} \cup \{+\infty\}$.

- For $\varphi \in \Gamma_0(\mathbb{R})$, its **proximity operator** $\mathrm{prox}_\varphi : \mathbb{R} \to \mathbb{R}$ is defined as

$$\boxed{(\forall x \in \mathbb{R}) \quad \mathrm{prox}_\varphi(x) = \underset{t \in \mathbb{R}}{\mathrm{argmin}}\, \varphi(t) + \frac{1}{2}(t-x)^2.}$$

- $\varrho : \mathbb{R} \to \mathbb{R}$ is said α-**averaged** with $\alpha \in [0,1]$ if there exists a 1-Lipschitz function ϑ such that $\boxed{(\forall x \in \mathbb{R})\ \varrho(x) = (1-\alpha)x + \alpha\vartheta(x)}$. Moreover, it satisfies

$$\boxed{(\forall (x,t) \in \mathbb{R})^2 \quad |\varrho(x) - \varrho(t)|^2 \leq |x-t|^2 - \frac{1-\alpha}{\alpha}|x - \varrho(x) - t + \varrho(t)|^2}.$$

- A function ϱ is said **firmly nonexpansive** if it is α-averaged with $\alpha = 1/2$.
- The conjugate of φ is defined as $(\forall u \in \mathbb{R})\ \varphi^*(u) = \sup_{x\in\mathbb{R}} xu - \varphi(x)$.

Activation functions

Proposition

Let $\psi\colon \mathbb{R} \longrightarrow \mathbb{R}$ be even, differentiable, with 1-Lipschitz derivative ϱ.

1. There exists $\alpha \in [1/2, 1]$ and an even function $\varphi \in \Gamma_0(\mathbb{R})$ such that

$$(\forall x \in \mathbb{R}) \quad \varrho(x) = x + 2\alpha(\operatorname{prox}_{\varphi}(x) - x). \tag{3}$$

Activation functions

Proposition

Let $\psi\colon \mathbb{R} \to \mathbb{R}$ be even, differentiable, with 1-Lipschitz derivative ϱ.

1. There exists $\alpha \in [1/2, 1]$ and an even function $\varphi \in \Gamma_0(\mathbb{R})$ such that

$$(\forall x \in \mathbb{R}) \quad \varrho(x) = x + 2\alpha(\mathrm{prox}_{\varphi}(x) - x). \tag{3}$$

2. Let

$$(\forall x \in \mathbb{R}) \quad \widetilde{\varphi}(x) = \varphi(x) + \frac{x^2}{2} \tag{4}$$

and let $\widetilde{\varphi}^* \in \Gamma_0(\mathbb{R})$ be the Fenchel-Young conjugate of $\widetilde{\varphi}$. Then

$$(\forall x \in \mathbb{R}) \quad \psi(x) \stackrel{c}{=} (1 - 2\alpha)\frac{x^2}{2} + 2\alpha\widetilde{\varphi}^*(x), \tag{5}$$

where $\stackrel{c}{=}$ designates equality up to an additive constant.

Activation functions

Proposition

Let $\psi\colon \mathbb{R} \to \mathbb{R}$ be even, differentiable, with 1-Lipschitz derivative ϱ.

1. There exists $\alpha \in [1/2, 1]$ and an even function $\varphi \in \Gamma_0(\mathbb{R})$ such that

$$(\forall x \in \mathbb{R}) \quad \varrho(x) = x + 2\alpha(\mathrm{prox}_{\varphi}(x) - x). \tag{3}$$

2. Let

$$(\forall x \in \mathbb{R}) \quad \widetilde{\varphi}(x) = \varphi(x) + \frac{x^2}{2} \tag{4}$$

and let $\widetilde{\varphi}^* \in \Gamma_0(\mathbb{R})$ be the Fenchel-Young conjugate of $\widetilde{\varphi}$. Then

$$(\forall x \in \mathbb{R}) \quad \psi(x) \overset{c}{=} (1 - 2\alpha)\frac{x^2}{2} + 2\alpha\widetilde{\varphi}^*(x), \tag{5}$$

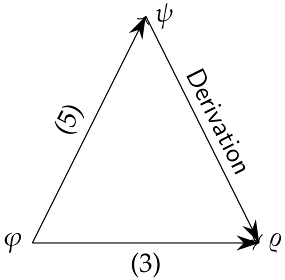
where $\overset{c}{=}$ designates equality up to an additive constant.

3. If ψ is convex, then $\alpha = 1/2$.

Activation functions

Interpretation:

- ψ convex $\rightarrow \exists\, \varphi \in \Gamma_0(\mathbb{R})$, $\varrho = \text{prox}_\varphi \rightarrow \varrho$ firmly nonexpansive.
- ψ non-convex $\rightarrow \varrho$ is an overrelaxation of the proximity operator of a function $\varphi \in \Gamma_0(\mathbb{R}) \rightarrow \varrho$ is $\alpha-$averaged.



Interplay between penalty ψ, activation ϱ and convex function φ.

Activation functions

Examples:

Penalty Name	Penalization $\psi(t)$	Activation $\varrho(t)$	α
Convex penalties			
Fair potential	$\delta(\lvert t\rvert - \delta\log(\frac{\lvert t\rvert}{\delta}+1))$	$\frac{\delta t}{\lvert t\rvert+\delta}$	$\frac{1}{2}$
Green	$\log(\cosh(t))$	$\tanh(t)$	$\frac{1}{2}$
Nonconvex penalties			
Hyperbolic tangent	$\delta^2\tanh(\frac{t^2}{2\delta^2})$	$\frac{t}{\cosh(\frac{t^2}{2\delta^2})^2}$	0.9581
Cauchy	$\frac{\delta^2}{2}\log(1+\frac{t^2}{\delta^2})$	$\frac{\delta^2 t}{t^2+\delta^2}$	$\frac{9}{16}$

Table 1: Examples of penalties ψ satisfying our assumption, their derivatives (i.e., nonlinear activation in our architecture) ϱ and the averaging constants α. All expressions are valid for every $t \in \mathbb{R}$, $(\lambda, \delta) \in]0, +\infty[^2$ and $\kappa \in [1, 2]$.

Deriving α-averaging constants: **Robustness study**.

Learning Strategy

For $k \in \{0, \ldots, K-1\}$, $\mathcal{L}_k^{(\theta_k)}$ learns:

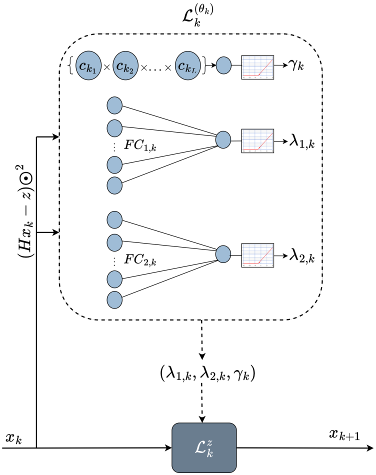
- **Regularization parameters** $\forall i \in \{1, 2\}$:

$$\lambda_{i,k} = \mathrm{ReLU}\left(\mathrm{FC}_{i,k}\left(([\mathbf{H}\mathbf{x}_k - \mathbf{z}]_m)^2)_{1 \leq m \leq M}\right)\right).$$

- **Stepsize parameter**

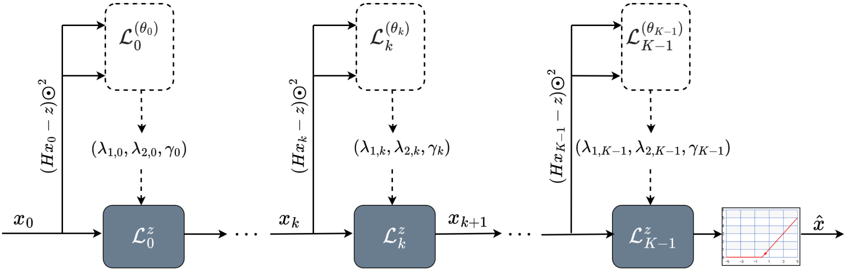
$$\gamma_k = \mathrm{ReLU}(c_{k,1} \times c_{k,2} \times \ldots \times c_{k,L}),$$

with $L \geq 1$ (typically, $L = 10$)



Learning Strategy

- Proposed architecture: **U-HQ**



Supervised **U-HQ** network: For every $k \in \{0, \dots, K-1\}$, the learning block $\mathcal{L}_k^{\theta_k}$ feeds the layer $\mathcal{L}_k^{\mathbf{z}}$ learnt hyperparameters. ReLU at the end of U-HQ ensures the positivity of the restored signal.

Learning Strategy

1. Training Set: $\{(\overline{\mathbf{x}}^{(i)}, \mathbf{z}^{(i)}),\ i \in \{1, \dots, S\}\}$
2. Feedforward model:

$$f_\Theta(\mathbf{x}_0^{(i)}; \mathbf{z}^{(i)}) = \mathcal{L}_{K-1}^{\mathbf{z}^{(i)}} \circ \cdots \circ \mathcal{L}_k^{\mathbf{z}^{(i)}} \circ \cdots \circ \mathcal{L}_0^{\mathbf{z}^{(i)}}(\mathbf{x}_0^{(i)}) \tag{6}$$

3. Backpropagation

$$\widehat{\Theta} = \underset{\Theta \in \mathbb{R}^{K(2M+L)}}{\operatorname{argmin}} \quad \frac{1}{S} \sum_{i=1}^{S} \ell(f_\Theta(\mathbf{x}_0^{(i)}; \mathbf{z}^{(i)}), \overline{\mathbf{x}}^{(i)}) \tag{7}$$

4. Inference/test: $\widehat{\mathbf{x}} = f_{\widehat{\Theta}}(\mathbf{x}_0; \mathbf{z})$

Experimental results on Mass Spectrometry data

Here, we showcase the efficiency of our proposed architecture **U-HQ** for the reconstruction of sparse mass spectrometry signals.

- First, we build several challenging realistic datasets.
- Second, we compare U-HQ to model-based iterative, deep learning and unrolled methods.
- Third, we justify our architectural design through an ablation study.

Experimental Settings

- **Datasets:** Variable noise levels + kernel shape + sparsity levels.

Name	Signal model	Blur Model	Noise	Data split
Dataset 1	MassBank	Ricker (v)	$(0, 0.5)$	900/100/100
Dataset 2	MassBank	Ricker (v)	$(0.5, 1.0)$	900/100/100
Dataset 3	MassBank	Fraser Suzuki (v)	$(0, 0.5)$	900/100/100
Dataset 4	MassBank	Ricker (c)	$(0, 0.5)$	900/100/100
Dataset 5	MassBank	Gaussian (c)	$(0, 0.5)$	900/100/100
Dataset 6	MassBank	Fraser Suzuki (c)	$(0, 0.5)$	900/100/100
Dataset 7	Averagine	Gaussian (c)	2	1000/200/200
Dataset 8	Averagine	Gaussian (c)	$(0, 2)$	1000/200/200

- **Penalties:**
  - ψ_1 : Fair potential
  - ψ_2 : Cauchy penalty
- **Initialization:** $\mathbf{x}_0 = 0$.
- **Evaluation metrics:** Avg (STD) SNR and Avg (STD) TSNR.

Results: U-HQ vs optimization-based methods

- **HQ-SC**: Half Quadratic algorithm with Stopping Criteria.
- **HQ-ES**: Half Quadratic algorithm with Early Stopping rule.

Penalty	Dataset 1	Dataset 2	Dataset 3
HQ-SC			
Convex	28.28 (7.37)/34.48 (7.18)	19.93 (3.79)/25.36 (5.27)	26.20 (5.49)/30.85 (5.01)
Non-convex	27.99 (6.26)/36.39 (7.36)	22.13 (4.12)/28.07 (5.62)	30.45 (5.04)/31.20 (5.02)
Hybrid	28.77 (6.49)/35.96 (7.22)	22.58 (4.13)/27.85 (5.62)	30.41 (4.96)/31.07 (4.95)
HQ-ES			
Convex	28.23 (7.79)/34.52 (7.79)	18.86 (3.64)/25.89 (5.35)	25.40 (5.81)/30.03 (6.00)
Non-convex	28.04 (6.33)/36.56 (7.55)	22.16 (4.10)/28.07 (5.77)	30.18 (5.22)/30.78 (5.36)
Hybrid	28.27 (6.47)/36.48 (7.58)	22.20 (4.12)/28.05 (5.78)	30.18 (5.22)/30.78 (5.36)
U-HQ			
Tikhonov	7.24 (3.86)/15.64 (7.97)	6.93 (3.27)/15.56 (7.89)	2.14 (1.01)/4.34 (3.30)
Convex	28.83 (5.40)/33.75 (6.06)	19.98 (3.14)/25.36 (4.94)	25.42 (5.65)/29.87 (5.19)
Non-convex	31.13 (5.37)/32.89 (5.26)	24.27 (3.19)/27.17 (4.67)	32.19 (5.67)/33.91 (5.79)
Hybrid	**31.56 (5.37)/34.63 (5.26)**	**25.27 (3.57)/27.04 (4.65)**	**33.75 (7.28)/35.87 (7.65)**

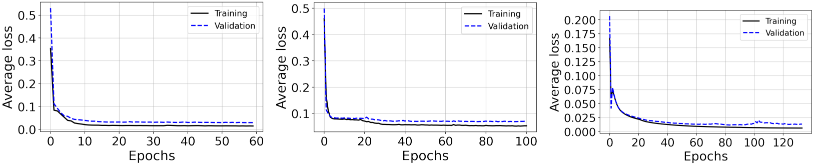
Table 2: Avg(std) SNR /Avg(std) TSNR.

Results: Computational complexity and stability of U-HQ

- **Complexity analysis**

	Dataset 1	Dataset 2	Dataset 3
HQ-SC			
Iterations (averaged)	6.38	8.33	10.4
CPU time (s)	5.47	7.26	8.47
GPU time (s)	0.29	0.37	0.47
U-HQ			
Number of layers	8	8	8
GPU time (s)	0.17	0.17	0.17

- **Loss monitoring**



(a) Dataset 1 (b) Dataset 2 (c) Dataset 3

Results: Ablation Study on U-HQ

- **U-HQ-DE** : Deep Equilibrium.
- **U-HQ-FixS**: Fixed Stepsize.
- **U-HQ-FixN**: Fixed Noise model.
- **U-HQ-FixN-OverP**: Fixed Noise+over-parametrization.

Penalty	Dataset 1	Dataset 2	Dataset 3
U-HQ-DE			
Hybrid	28.16 (5.57)/34.43 (6.34)	22.26 (3.46)/27.95 (5.03)	31.66 (5.85)/33.04 (5.63)
U-HQ-FixS			
Convex	28.27 (6.41)/34.26 (6.78)	20.64 (3.57)/26.66 (4.72)	26.72 (5.96)/30.25 (5.97)
Non-convex	29.44 (4.98)/32.96 (5.62)	24.34 (3.47)/26.67 (4.29)	32.22 (6.76)/34.43 (6.12)
Hybrid	29.44 (4.98)/32.99 (5.62)	24.50 (3.52)/26.40 (4.31)	32.34 (5.60)/33.41 (5.33)
U-HQ-FixN			
Convex	28.88 (6.70)/36.27 (7.74)	22.56 (3.47)/26.10(4.26)	26.51 (6.39)/31.84 (6.25)
Non-convex	29.89 (5.03)/33.19 (5.65)	25.08 (3.81)/26.46 (4.25)	32.17 (5.88)/33.53 (5.54)
Hybrid	30.07 (4.86)/32.78 (5.40)	25.22 (3.91)/26.57 (4.34)	32.36 (5.73)/33.65 (5.41)
U-HQ-FixN-OverP			
Convex	30.26 (7.23)/36.13 (7.47)	24.52 (3.70)/27.08 (4.41)	31.07 (5.53)/33.22 (5.30)
Non-convex	29.96 (5.24)/33.44 (5.78)	25.03 (3.85)/26.25 (4.23)	32.17 (5.22)/32.88 (4.97)
Hybrid	30.38 (4.88)/32.30 (5.28)	25.18 (3.87)/26.52 (4.27)	32.48 (5.41)/33.29 (5.11)
U-HQ			
Tikhonov	7.24 (3.86)/15.64 (7.97)	6.93 (3.27)/15.56 (7.89)	2.14 (1.01)/4.34 (3.30)
Convex	28.83 (5.40)/33.75 (6.06)	19.98 (3.14)/25.36 (4.94)	25.42 (5.65)/29.87 (5.19)
Non-convex	31.13 (5.37)/32.89 (5.26)	24.27 (3.19)/27.17 (4.67)	32.19 (5.67)/33.91 (5.79)
Hybrid	**31.56 (5.37)/34.63 (5.26)**	**25.27 (3.57)/27.04 (4.65)**	**33.75 (7.28)/35.87 (7.65)**

Table 4: Avg(std) SNR /Avg(std) TSNR.

Results: U-HQ vs state of the art benchmarks

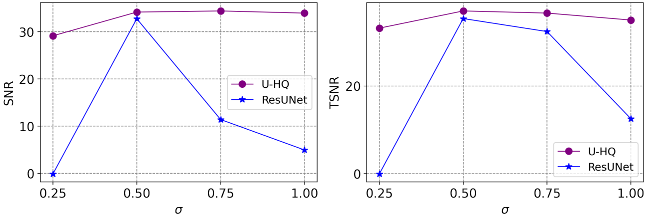
- **U-ISTA**: Unrolled iterative soft threshholding algorithm.
- **U-PD**: Unrolled primal-dual algorithm.
- **AE**: Autoencoder.
- **FCNet**: Fully connected network.
- **ResUNet**: Residual UNet network.

Penalty	Dataset 1	Dataset 2	Dataset 3
	U-ISTA		
Convex	20.55 (6.89)/21.25 (6.60)	21.62 (5.85)/23.03 (5.83)	17.03 (6.39)/17.92 (5.72)
	U-PD		
Convex	22.13 (6.76)/25.16 (6.42)	21.55 (4.87)/22.88 (4.80)	24.42 (4.60)/26.21 (4.61)
	U-HQ		
Tikhonov	7.24 (3.86)/15.64 (7.97)	6.93 (3.27)/15.56 (7.89)	2.14 (1.01)/4.34 (3.30)
Convex	28.83 (5.40)/33.75 (6.06)	19.98 (3.14)/25.36 (4.94)	25.42 (5.65)/29.87 (5.19)
Non-convex	31.13 (5.37)/32.89 (5.26)	24.27 (3.19)/27.17 (4.67)	32.19 (5.67)/33.91 (5.79)
Hybrid	**31.56 (5.37)/34.63 (5.26)**	**25.27 (3.57)/27.04 (4.65)**	**33.75 (7.28)/35.87 (7.65)**
	DL		
FCNet	1.97 (1.83)/2.29 (2.11)	1.90 (1.93)/2.17 (2.26)	1.83 (1.71)/2.11 (1.91)
AE	0.32 (0.43)/0.49 (0.56)	0.31 (0.45) /0.46 (0.59)	0.35 (0.45)/0.50 (0.55)
ResUNet	29.97 (6.13)/31.84 (6.36)	25.05 (5.28)/ 26.23 (5.60)	28.67 (3.67)/29.83 (3.71)

Table 5: Avg(std) SNR /Avg(std) TSNR.

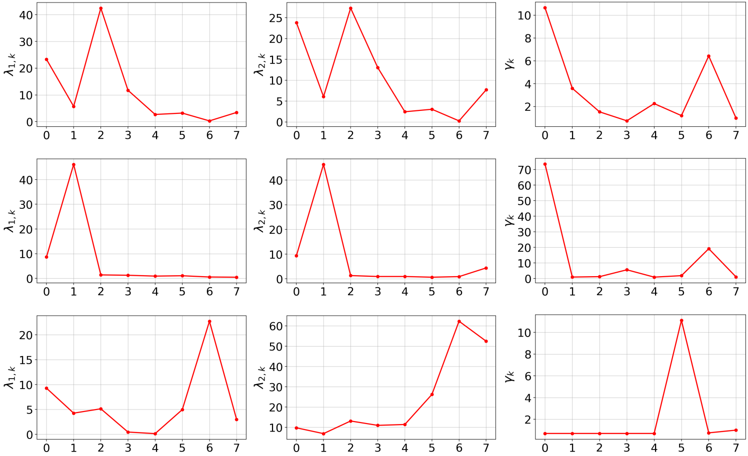
Results: U-HQ vs state of the art benchmarks

- **Generalization capacity on Dataset 4**
  1. Training set with a fixed kernel shape.
  2. Different test sets with varying kernel shapes.



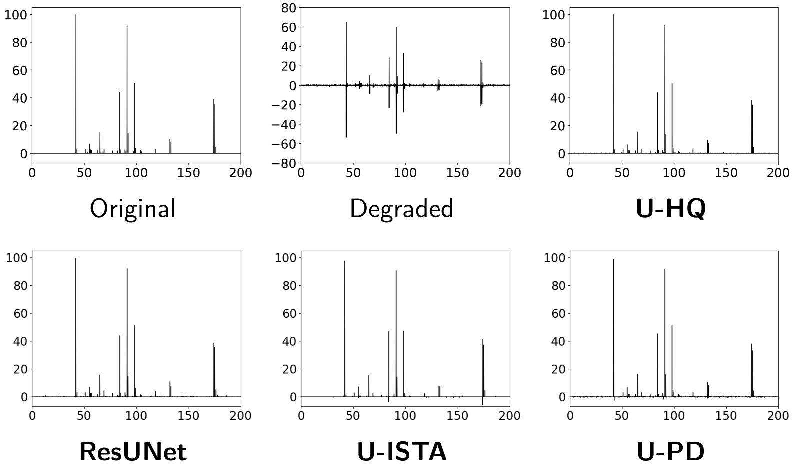
⟶ **U-HQ** generalizes well on mismatched test data, unlike **ResUNet**.

Results: Learnt parameters trends



Learnt $(\lambda_{1,k})_{0\leq k\leq K-1}$ (left), $(\lambda_{2,k})_{0\leq k\leq K-1}$ (middle) and $(\gamma_k)_{0\leq k\leq K-1}$ (right), averaged on test set, wrt k (x-axis) for **U-HQ**, for Dataset 1, 2 and 3 (top to bottom respectively).

Results: Reconstructed signal example

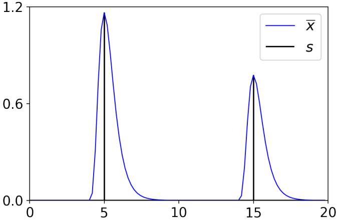


Benchmark of unrolled architectures on chromatographic data

Problem formulation

- **Inverse problem**

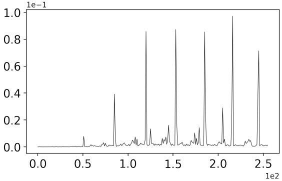
$$\mathbf{z} = \mathbf{H}(\underbrace{\pi * \mathbf{s}}_{\bar{\mathbf{x}}}) + \mathbf{e}$$



- **Goal:**
  Benchmarking HQ-SC, ISTA, PD, **U-HQ**, **U-ISTA** and **U-PD** on simulated chromatographic parametric datasets using chemically-driven evaluation metrics to recover an estimate $\widehat{\mathbf{x}}$ of $\bar{\mathbf{x}}$.

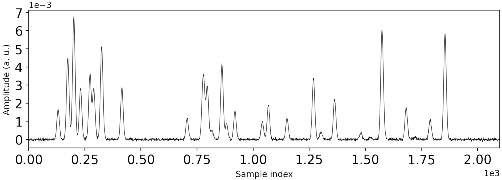
Parametric model inspired from real data

- **Actual chromatogram shape**



- **Peak modelling: Fraser-Suzuki chromatogram simulation**

$$\left(\forall x > m - \frac{\sigma_f}{a}\right): \quad \pi(x) \propto \exp\left(-\frac{1}{2a^2}\log\left(1 + a\frac{(x-m)}{\sigma_f}\right)^2\right).$$



Parametric databases: parameters

- For spiky signal $\mathbf{s}$:
  - number of samples N;
  - number of spikes or sparsity level P; expressed relatively as P/N;
  - peak separation limit $d_{\text{min}} \in \{1, \ldots, N\}$;
  - spike intensities: $|\mathcal{N}(0,1)|$ (absolute value of a standard normal distribution);
- For peak kernel π:
  - peak width $\sigma_f > 0$;
  - asymmetric tailing coefficient $a > 0$ ($a \to 0$: Gaussian peak);
- For external disturbance sources:
  - additive "noise": zero-mean Gaussian, standard deviation $\sigma_{\mathbf{e}} > 0$;
  - potential instrument response blur kernel with width $\sigma_{\mathbf{H}} > 0$.

Parametric databases

- **Parameters**
  - Signal $\mathbf{s}$: N, P, d_{min}.
  - Peak kernel π: σ_f, a.
  - External disturbances: $\sigma_{\mathbf{e}}$, $\sigma_{\mathbf{H}}$.
- **Datasets Summary**

Parameter\Dataset	D0	D1	D2	D3	D4	D5	D6
P/N	1.5%	3%	4.5%	1.5%	1.5%	3%	3%
d_{min}	5	3	1	5	5	3	3
σ_f	0.5	0.5	0.5	0.5	0.5	0.5	0.5
a	0.2	0.2	0.2	0.4	0.6	0.2	0.2
$\sigma_{\mathbf{e}}$	0.02	0.02	0.02	0.02	0.02	0.04	0.06
$\sigma_{\mathbf{H}}$	1	1	1	1	1	1	1

$\longrightarrow$ Creation of several datasets with different chemical properties: sparsity, peak shape and noise level.

Parametric databases: design of experiments (DoE)

- **Parameters**
  - Signal $\mathbf{s}$: N, P, d_{min}.
  - Peak kernel π: σ_f, a.
  - External disturbances: $\sigma_{\mathbf{e}}$ ($\sigma_{\mathbf{H}}$).
- **Dataset DoE difficulty**

Variation\Difficulty	Low	Mid	High
Sparsity ($P/N, d_{\text{min}}$)	D0 (1.5%, 5)	D1 (3%, 3)	D2 (4.5%, 1)
Asymmetry (a)	D0 (0.2)	D3 (0.4)	D4 (0.6)
Noise ($\sigma_{\mathbf{e}}$)	D1 (0.02)	D5 (0.04)	D6 (0.06)

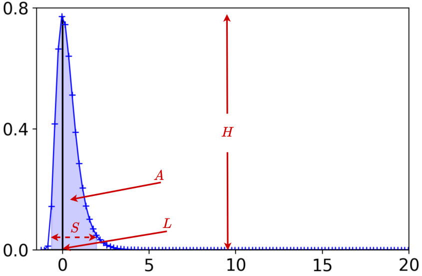
Table 6: DoE. Parametric variations on sparsity, asymmetry and noise.

Evaluation: classical signal metrics

$$\mathrm{MSE}(\mathbf{p}, \widehat{\mathbf{p}}) = \frac{1}{n}\|\mathbf{p} - \widehat{\mathbf{p}}\|^2,$$

$$\mathrm{SNR}(\mathbf{p}, \widehat{\mathbf{p}}) = 20\log_{10}\left(\frac{\|\mathbf{p}\|}{\|\mathbf{p} - \widehat{\mathbf{p}}\|}\right),$$

$$\mathrm{TSNR}(\mathbf{p}, \widehat{\mathbf{p}}) = 20\log_{10}\left(\frac{\sum_i |\mathbf{p}^{(i)}|^2}{\sum_i |\mathbf{p}^{(i)} - \widehat{\mathbf{p}^{(i)}}|^2}\right), \quad \forall i \in \bigcup_{1\leq j\leq P} \mathcal{S}_j\,.$$



Evaluation: peak HALmetric quantities

- Ground-truth quantities

$$\overline{L_j} = \arg \max_{i \in \{1,\dots,N\}} \Big((s \odot \delta_{S_j}) * \pi \Big)_{1 \le i \le N},$$

$$\overline{H_j} = p_{\overline{L_j}},$$

$$\mathcal{S}_j = \{i \in \{1, \dots, N\} \quad \text{s.t.} \quad (s \odot \delta_{S_j}) * \pi \ge \vartheta \overline{H_j}\} = [aj \cdot \cdot bj],$$

$$\overline{A_j} = \sum_{i \in \mathcal{S}_j \backslash b_j} (p_i + p_{i+1})/2.$$

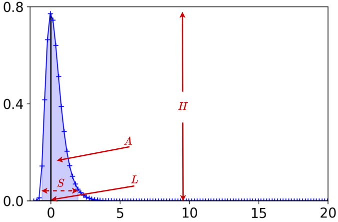
- Estimates using oracle $\mathcal{S}_j$

$$\widehat{L_j} = \arg \max_{i \in \mathcal{S}_j} \hat{p}_i,$$

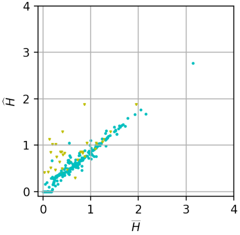
$$\widehat{H_j} = \hat{p}_{\widehat{L_j}},$$

$$\widehat{A_j} = \sum_{i \in \mathcal{S}_j \backslash b_j} (\hat{p}_i + \hat{p}_{i+1})/2.$$

Evaluation metrics: HALmetrics



- Subjective HALmetrics: Scatter plots.



- Objective HALmetrics: Given $\widehat{\mathbf{H}}$, $\overline{\mathbf{H}}$, we define:

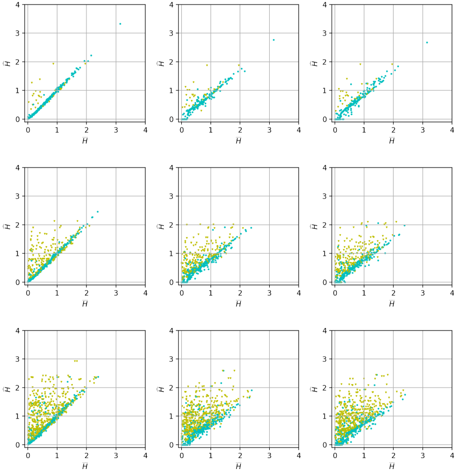
$$\boxed{\mathrm{NMAE}(\overline{\mathbf{H}}, \widehat{\mathbf{H}}) = \frac{\sum_{j=1}^{P} |\overline{H_j} - \widehat{H_j}|}{\sum_{j=1}^{P} |\overline{H_j}|}.}$$

Similarly, for $\mathrm{NMAE}(\overline{\mathbf{L}}, \widehat{\mathbf{L}})$ and $\mathrm{NMAE}(\overline{\mathbf{A}}, \widehat{\mathbf{A}})$.

Results: classical and objective HALmetrics

		MSE	SNR	TSNR	NMAE($\bar{\mathbf{H}}, \hat{\mathbf{H}}$)	NMAE($\bar{\mathbf{A}}, \hat{\mathbf{A}}$)	NMAE($\bar{\mathbf{L}}, \hat{\mathbf{L}}$)
D0	**U-HQ**	**4.7×10^{-4} (1.2×10^{-4})**	**19.6 (0.9)**	**20.0 (1.0)**	**0.1 (0.0)**	**0.012 (0.003)**	**3.1×10^{-6} (1.5×10^{-6})**
	U-PD	3.7×10^{-3} (1.6×10^{-3})	10.9 (1.3)	10.9 (1.3)	0.2 (0.0)	0.107 (0.025)	8.0×10^{-6} (1.8×10^{-6})
	U-ISTA	3.0×10^{-3} (1.0×10^{-3})	11.7 (1.0)	11.7 (1.0)	0.2 (0.0)	0.110 (0.020)	8.0×10^{-6} (1.7×10^{-6})
D1	**U-HQ**	**1.1×10^{-3} (2.3×10^{-4})**	**19.4 (0.7)**	**19.6 (0.7)**	**0.2 (0.1)**	**0.014 (0.002)**	**3.8×10^{-6} (9.5×10^{-7})**
	U-PD	10.0×10^{-3} (2.5×10^{-3})	9.7 (0.7)	9.7 (0.7)	0.4 (0.0)	0.107 (0.016)	6.3×10^{-6} (8.7×10^{-7})
	U-ISTA	8.0×10^{-3} (2.1×10^{-3})	10.6 (0.7)	10.6 (0.7)	0.3 (0.0)	0.109 (0.019)	6.1×10^{-6} (8.6×10^{-7})
D2	**U-HQ**	**1.7×10^{-3} (3.4×10^{-4})**	**19.6 (0.8)**	**19.7 (0.8)**	**0.4 (0.1)**	**0.015 (0.002)**	**5.2×10^{-6} (9.5×10^{-7})**
	U-PD	1.7×10^{-2} (3.3×10^{-3})	9.5 (0.5)	9.5 (0.5)	0.5 (0.1)	0.101 (0.012)	5.8×10^{-6} (7.2×10^{-7})
	U-ISTA	1.4×10^{-2} (2.6×10^{-3})	10.4 (0.6)	10.4 (0.6)	0.5 (0.1)	0.102 (0.012)	5.8×10^{-6} (7.5×10^{-7})
D3	**U-HQ**	**8.7×10^{-4} (2.1×10^{-4})**	**16.9 (0.7)**	**17.3 (0.8)**	**0.1 (0.1)**	**0.015 (0.002)**	**3.7×10^{-6} (1.7×10^{-6})**
	U-PD	4.4×10^{-3} (1.4×10^{-3})	10.0 (0.9)	10.1 (0.9)	0.3 (0.0)	0.107 (0.020)	9.0×10^{-6} (1.7×10^{-6})
	U-ISTA	3.7×10^{-3} (1.1×10^{-3})	10.7 (0.7)	10.7 (0.7)	0.2 (0.0)	0.104 (0.027)	8.7×10^{-6} (1.7×10^{-6})
D4	**U-HQ**	**1.4×10^{-3} (2.8×10^{-4})**	**14.9 (0.6)**	**15.1 (0.6)**	**0.1 (0.1)**	**0.012 (0.002)**	**4.5×10^{-6} (1.9×10^{-6})**
	U-PD	5.0×10^{-3} (1.4×10^{-3})	9.3 (0.5)	9.3 (0.5)	0.3 (0.0)	0.1 (0.0)	9.6×10^{-6} (1.6×10^{-6})
	U-ISTA	4.5×10^{-3} (1.2×10^{-3})	9.8 (0.6)	9.8 (0.6)	0.3 (0.0)	0.083 (0.011)	9.3×10^{-6} (1.6×10^{-6})
D5	**U-HQ**	**1.5×10^{-3} (3.3×10^{-4})**	**17.8 (0.7)**	**18.0 (0.8)**	**0.2 (0.1)**	**0.018 (0.003)**	**4.2×10^{-6} (1.0×10^{-6})**
	U-PD	10.0×10^{-3} (2.3×10^{-3})	9.6 (0.7)	9.6 (0.7)	0.4 (0.0)	0.108 (0.015)	6.5×10^{-6} (9.1×10^{-7})
	U-ISTA	7.9×10^{-3} (1.8×10^{-3})	10.6 (0.6)	10.6 (0.6)	0.4 (0.0)	0.106 (0.014)	6.3×10^{-6} (9.1×10^{-7})
D6	**U-HQ**	**1.9×10^{-3} (4.0×10^{-4})**	**16.8 (0.8)**	**17.1 (0.8)**	**0.2 (0.1)**	**0.021 (0.003)**	**4.2×10^{-6} (9.8×10^{-7})**
	U-PD	9.9×10^{-3} (2.3×10^{-3})	9.7 (0.7)	9.7 (0.7)	0.4 (0.0)	0.108 (0.016)	6.3×10^{-6} (8.7×10^{-7})
	U-ISTA	7.9×10^{-3} (1.7×10^{-3})	10.6 (0.6)	10.7 (0.6)	0.3 (0.0)	0.107 (0.014)	6.1×10^{-6} (8.6×10^{-7})

Results: subjective HALmetrics evaluation, height (H)



Scatter plots: $(\overline{\mathbf{H}}, \hat{\mathbf{H}})$. Columns: **U-HQ**, **U-ISTA**, **U-PD**. Rows: sparsity ($D0$, $D1$, $D2$).

Baseline results: K arbitrary fixed

What if we do not tune the number of layers of unrolled architectures?

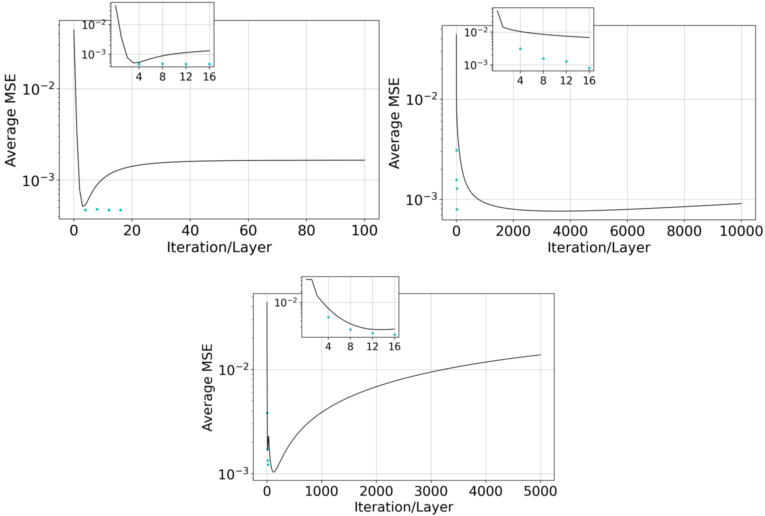
- **Reconstruction Quality: U-HQ** is better in terms of global and local metrics. E.g.,

		NMAE($\bar{\mathbf{H}}, \hat{\mathbf{H}}$)	NMAE($\bar{\mathbf{A}}, \hat{\mathbf{A}}$)	NMAE($\bar{\mathbf{L}}, \hat{\mathbf{L}}$)
D1	**U-HQ**	**0.217 (0.051)**	**0.014 (0.002)**	$\mathbf{3.812 \times 10^{-6}}$ **(**$\mathbf{9.459 \times 10^{-7}}$**)**
	U-PD	0.361 (0.043)	0.107 (0.016)	6.329×10^{-6} (8.687×10^{-7})
	U-ISTA	0.350 (0.044)	0.109 (0.019)	6.121×10^{-6} (8.587×10^{-7})

- **Execution Time:** All unrolled methods prove to be fast with **U-ISTA** being the fastest.
- **Effect of complexifying datasets:** Datasets with higher sparsity, lower tailing parameter and lower noise are easier to reconstruct.

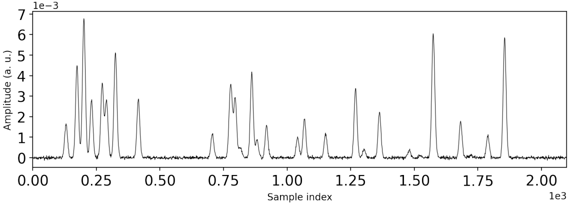
Results: Unrolling tuning

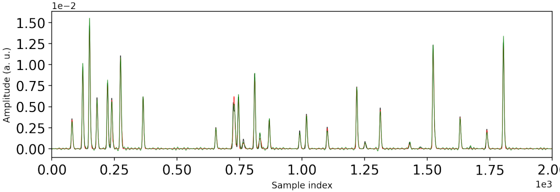
What is the relation between an unrolled architecture and its iterative version?



Average MSE of iterative/unrolled HQ (top left), ISTA (top right) and PD (bottom).

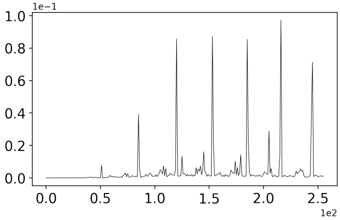
Reconstruction on Simulated Data



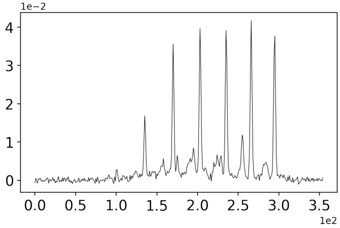


Example from D0: Degraded spectrum $\mathbf{z}$ (top), ground truth and restored signals (bottom). Black line depicts the ground truth signal $\mathbf{p}$, red line is the restoration using **U-ISTA**, and green line is for the restoration using **U-HQ**.

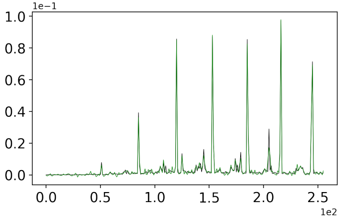
Reconstruction on Real Data U-HQ vs U-ISTA



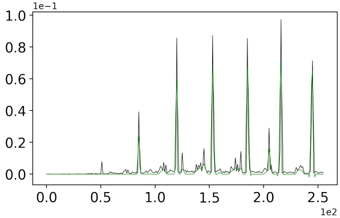
Original



Degraded



U-HQ (17.99 dB)



U-ISTA (4.90 dB)

In a nutshell

- Neural network architecture based on unrolling an MM half-quadratic algorithm
- Proximal properties of resulting activations functions
- Application to mass spectrometry data recovery
- Benchmarking of iterative vs unrolled schemes on chromatographic data retrieval

References:

M. Gharbi, E. Chouzenoux, and J.-C. Pesquet. An Unrolled Half-Quadratic Approach for Sparse Signal Recovery in Spectroscopy. *Signal Processing*, vol. 218, pp. 109369, May 2024

M. Gharbi, S. Villa, E. Chouzenoux, J.-C. Pesquet, L. Duval Unrolled deep networks for sparse signal restoration in analytical chemistry, In Proceedings of *IEEE MLSP 2024*, London, UK, 22th-25th Sep. 2024

https://github.com/GHARBIMouna/Unrolled-Half-Quadratic

Future work

- **Theoretical questions**
  - Stability study
  - Convergence properties of learned iterative schemes
  - Enforcing positivity and other desirable properties in the objective function
- **Architecture Design**
  - Learn $\mathbf{H}$, blind deconvolution context.
  - Higher more general number of penalty branches
  - Postprocessing blocks
  - Full flexibility of architecture's weights and biases.
  - Learning smoothing parameters
- **Application to chemistry**
  - More realistic databases.
  - Integration of our method into a pipeline, coupled with baseline removal and chemical identification modules.

Thank you for your attention!