Convergence guarantees for Deep Equilibrium models with applications to Poisson inverse problems

Christian Daniele, Silvia Villa, Samuel Vaiter and Luca Calatroni

June 5, 2026

SIAM Optimization, Edimburgh

Co-funded by the European Union (ERC, MALIN, 10117133). Views and opinions expressed are however those of the author only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the EU nor the granting authority can be held responsible for them.

Inverse Problems

$$\boxed{y = \mathcal{N}(Ax)}$$

$A \in \mathbb{R}^{m \times n}_{\geq 0}$
$x \in \mathbb{R}^{n}_{\geq 0}, y \in \mathbb{R}^{m}_{\geq 0}$

Gaussian case: $\mathcal{N}(Ax) = Ax + \eta,\ \eta \sim N(0, \sigma^2 I)$

Poisson case: $\mathcal{N}(Ax) =$ Poisson(Ax)

$$\boxed{\underset{x \in C}{\operatorname{argmin}}\, D(y, Ax) + \lambda g(x)}$$

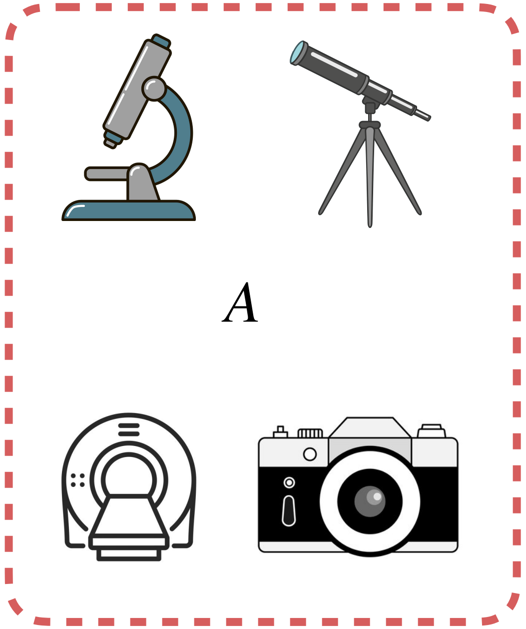
Gaussian case: $D(y, Ax) = \frac{1}{2}\|Ax - y\|_2^2, \quad C = \mathbb{R}^n$

Poisson case: $D(y, Ax) =$ KL$(y, Ax), \quad C = \mathbb{R}^n_+$

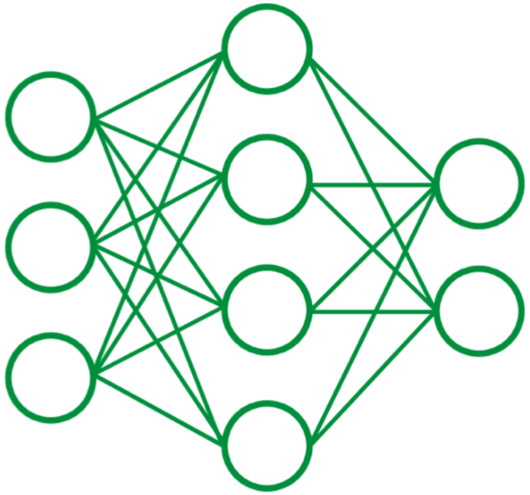
$$g(x) = \begin{cases} \|x\|_2^2 & \text{Tikhonov} \\ \|x\|_1 & \text{L1} \\ \text{TV}(x) & \text{Total Variation} \end{cases}$$

Hybrid data-driven approach

Physics



$+$



Goal:
given $\underset{x\in C}{\operatorname{argmin}}\, D(y, Ax) + \lambda g(x)$, learning λg

Data-driven approaches for IPs

- PnP

[Venkatakrishnan, Bouman, Wohlberg, '13]

$$x^{k+1} = \overset{\mathscr{D}}{\cancel{\mathrm{Prox}_{\tau g}}}(x^k - \tau \nabla D(y, Ax^k))$$

- Unrolling

[Monga, Li and Eldar '21]

$$x^{k+1} = F(x^k, y, A, \theta^k),\ k = 1,\ldots,K, \quad K < +\infty$$

- Deep equilibrium models

$$x^0 \longrightarrow f_\theta \circ f_\theta \circ \ldots \circ f_\theta \circ \ldots \longrightarrow x^\infty$$

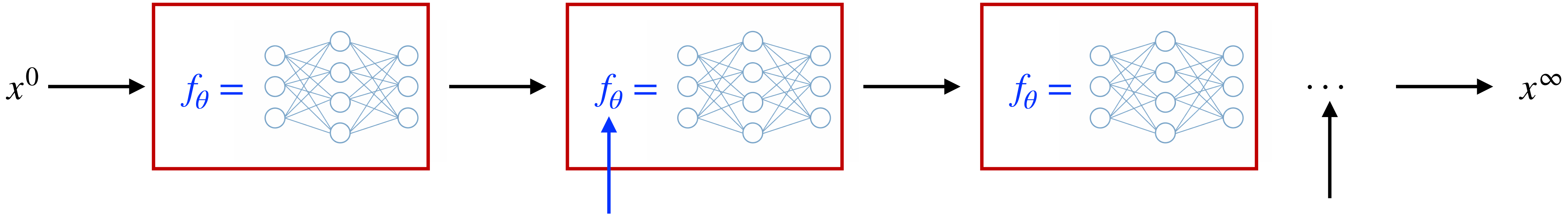
[Bai, Kolter, Koltun '19]

[Gilton, Ongie, Willett '21]

[Zou, Liu, Wohlberg and Kamilov '23]

Deep Equilibrium models (DEQs) [Bai, Kolter and Koltun '19, Gilton, Ongie and Willett '21]

Input: $y \in Y$

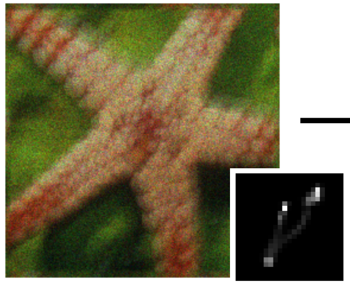


$f_\theta(\,\cdot\,; y)$ DEQ layer

Infinite composition

$$y \mapsto \mathrm{Fix}(f_\theta(\,\cdot\,; y)) \quad \text{DEQ}$$

y



$\xrightarrow{f_\theta(\,\cdot\,; y)}$

x^∞

How to design a well-posed DEQ

$y \in Y \subseteq \mathbb{R}^m$, $f_\theta(\cdot, y): X \subseteq \mathbb{R}^n \to X$

Well-posed DEQ

$\forall x^0 \in X,$ the sequence $\{x^k\}_{k\geq 0},$ given by $x^{k+1} = f_\theta(x^k, y)$ converges to $x^* \in \mathrm{Fix}(f_\theta(\cdot, y))$

Sufficient condition (contraction):

$$\exists\ \rho < 1 \text{ s.t. } \|f_\theta(x_1, y) - f_\theta(x_2, y)\| \leq \rho\|x_1 - x_2\|\ \forall x_1, x_2 \in X$$

Then: $\mathrm{Fix}(f_\theta(\cdot, y)) = \{x^*\}$ and $\forall x^0 \in X$, $x^k \to x^*$

First-order optimisation contractive operators

Gradient Descent (GD) [Briceño-Arias, Pustelnik '23]

Let $F = D(y, A\cdot) + \lambda g$ be L_F-smooth

$$\mathrm{GD}(x) = x - \tau \nabla F(x) = (I - \tau \nabla F)(x)$$

If F is μ_F-strongly convex and $\tau \in (0,2/L_F)$, then GD **is a contraction** with:

$$\rho = \max\{\,|1 - \tau\mu_F|\,,|1 - \tau L_F|\,\} < 1$$

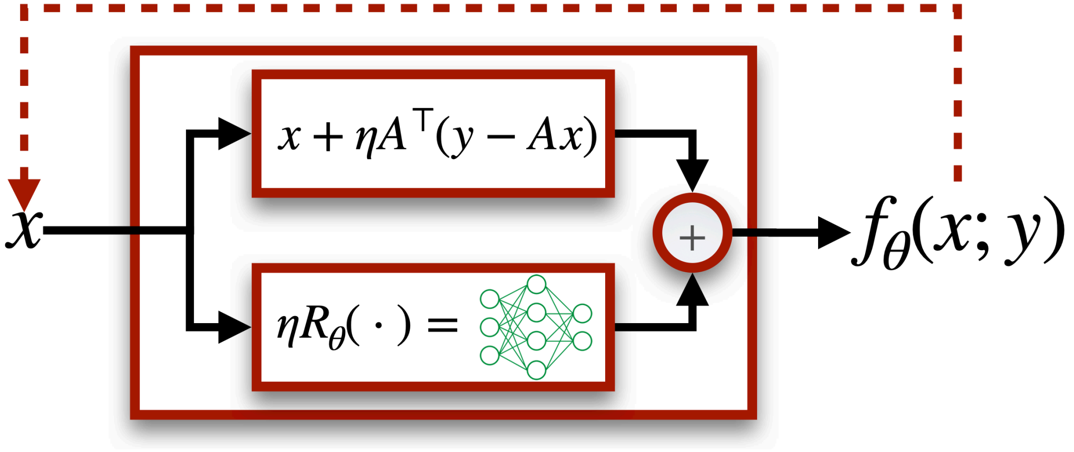
Proximal Gradient Descent (PGD) [Taylor, Hendrickx, Glineur '18]

Let $F = D(y, A\cdot) + \lambda g$ with $D(y, A\cdot)$ L_D-smooth and g convex/non-smooth

$$\mathrm{PGD}(x) = \mathrm{prox}_{\tau g}(x - \tau \nabla D(y, Ax))$$

If $D(y, A\cdot)$ is μ_D-strongly convex, then PGD **is a contraction** for suitable choices of τ

Contractive-like DEQ: DE-GRAD



$$D(y, Ax) = \frac{1}{2}\|Ax - y\|_2^2$$

$\lambda \nabla g \to R_\theta$ parameterise the gradient

$$f_\theta(x; y) = (I - \tau \nabla D(y, A\cdot))(x) - \tau R_\theta(x)$$

Deep Equilibrium Gradient Descent (DE-GRAD) [Gilton, Ongie and Willett '21]

$$\|f_\theta(x_1; y) - f_\theta(x_2; y)\| \leq \|(I - \tau \nabla D(y, A\cdot))(x_1 - x_2)\| + \tau\|R_\theta(x_1) + R_\theta(x_2)\|$$

If the following holds:

- $D(y, A\cdot)$ is L_D-smooth and μ_d- strongly convex
- R_θ is $L_{R_\theta}-$Lipschitz

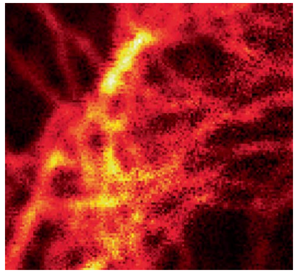
then GD$_\theta$ is a contraction $\forall \tau \in (0, 2/(L_D + L_{R_\theta}))$

Poisson Inverse Problems case

$$\boxed{y = \text{Poisson}(Ax)}$$

$A \in \mathbb{R}^{m\times n}_{\geq 0}$
$x \in \mathbb{R}^{n}_{\geq 0}, y \in \mathbb{R}^{m}_{\geq 0}$

Motivation: Imaging with low photon count (e.g. Microscopy) $\longrightarrow$



[Bertero and Boccacci '98]

$$\boxed{\underset{x\geq 0}{\text{argmin}}\, \mathsf{KL}(y, Ax) + \lambda g(x)}$$

$$\mathsf{KL}(y, Ax) = \sum_{i=1}^{m} d_{\mathsf{KL}}(y_i, (Ax)_i)$$

$$d_{\mathsf{KL}}(y, x) = y \log\left(\frac{y}{x}\right) + x - y$$

- ≥ 0
- convex
- **not L-smooth**

$\longrightarrow$ GD not applicable

$\lambda g \to R_\theta$ parameterise the regularisation function

KL minimisation: Mirror Descent

$$\boxed{\operatorname*{argmin}_{x\geq 0} \mathsf{KL}(y, Ax) + R_\theta(x) + \iota_{[0,a]^n} := \Psi_\theta(\,\cdot\,; y)}$$

$$h = \frac{1}{2}\|x\|_2^2 \implies$$

$$D_h(x,y) = \frac{1}{2}\|x-y\|_2^2$$

Given a Legendre function h:

$$x^{k+1} = \operatorname*{argmin}_{x\in\mathbb{R}^n} \langle x - x^k, \nabla(\mathsf{KL}(y, A\cdot) + R_\theta)(x^k)\rangle + \frac{1}{\tau} D_h(x, x^k) + \iota_{[0,a]^n}(x)$$

Burg's entropy: $h(x) = -\sum_{i=1}^{n} \log(x_i), \quad \mathrm{dom}(h) = (0, +\infty)^n$

[Bauschke, Bolte and Teboulle '17]

Key points:

$x^0 \in \mathrm{int}(\mathrm{dom}(h)) \implies x^k \in \mathrm{int}(\mathrm{dom}(h))\,\forall k \geq 0 \longrightarrow$ no need of projection

KL minimisation: Mirror Descent

$$\boxed{\underset{x\geq 0}{\operatorname{argmin}}\ \mathrm{KL}(y, Ax) + R_\theta(x) + \iota_{[0,a]^n} := \Psi_\theta(\,\cdot\,; y)}$$

$$h = \frac{1}{2}\|x\|_2^2 \implies$$

$$D_h(x,y) = \frac{1}{2}\|x-y\|_2^2$$

Given a Legendre function h:

$$x^{k+1} = \underset{x\in\mathbb{R}^n}{\operatorname{argmin}}\ \langle x - x^k, \nabla(\mathrm{KL}(y, A\cdot) + R_\theta)(x^k)\rangle + \frac{1}{\tau} D_h(x, x^k) + \iota_{[0,a]^n}(x)$$

Burg's entropy: $h(x) = -\sum_{i=1}^{n} \log(x_i), \quad \mathrm{dom}(h) = (0, +\infty)^n$ [Bauschke, Bolte and Teboulle '17]

Key points:

$\tau \propto \frac{1}{L}$, $L > 0$ is s.t. $Lh - (\mathrm{KL} + R_\theta)$ is convex on $\mathrm{int}(\mathrm{dom}(h))$

NoLip property [Bauschke, Bolte and Teboulle '17]

L-SMAD [Bolte, Sabach, Teboulle and Vaisbourd '18]

Relative smoothness [Lu, Freund and Nesterov '18]

If $h = \frac{1}{2}\|x\|_2^2 \implies L$-smoothness

Deep equilibrium Mirror Descent: DEQ-MD

$$f_\theta(x;y) = \Pi_{[0,a]^n}(\nabla h^*(\nabla h(x) - \tau(\nabla \mathsf{KL}(y, Ax) + \nabla R_\theta(x))))$$

Well-posedness of DEQ-MD $\Longleftrightarrow$ convergence iterates of MD to minimise Ψ_θ

Ψ_θ non-smooth and non-convex functional, convergence relies on [Bolte, Sabach, Teboulle and Vaisbourd '18]:

- **Kurdyka-Łojasiewicz property**
- MD provides a bounded *gradient-like descent* sequence ($\Pi_{[0,a]^n}$ + relative smoothness)

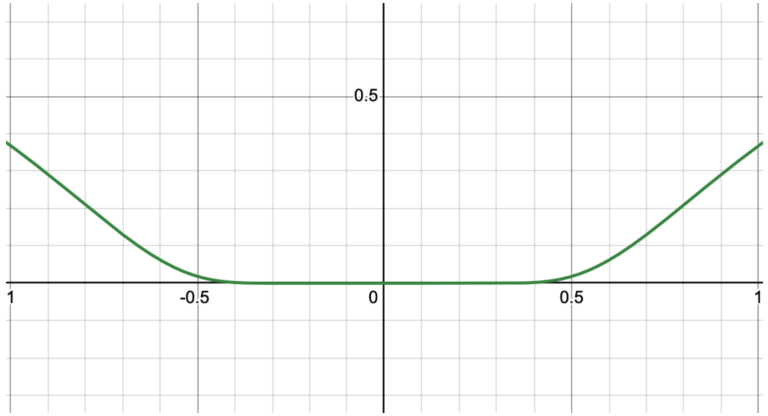
Choose R_θ s.t. Ψ_θ is a KŁ function

Kurdyka-Łojasiewicz property [Attouche, Bolte and Svaiter '13]

A function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is said to have the **Kurdyka-Łojasiewicz property** at $x^* \in \mathrm{dom}(f)$ if there exists a **neighborhood** U of x^* and a continuous concave function $\psi : [0,\delta) \to \mathbb{R}_+$ such that:
$\forall x \in U \cap \{x \mid f(x^*) < f(x) < f(x^*) + \delta\}$, the Kurdyka-Łojasiewicz inequality holds:

$$\psi'(f(x) - f(x^*))\mathrm{dist}(0, \partial f(x)) \geq 1.$$

Idea: *The functional has to be sufficiently sharp around critical points (upon reparameterisation)*



Counterexample: $f(x) = e^{-1/x^2}$

But it is satisfied by a wide class of functions: Analytical, Semialgebraic, Subanalytical...

Ψ_θ is a Kurdyka-Łojasiewicz function

Assumption: R_θ is analytic on $\mathbb{R}^n$

$$\Psi_\theta = \underbrace{\mathsf{KL}(y, Ax)}_{\textit{analytical}} + \underbrace{R_\theta}_{\textit{analytical}} + \underbrace{\iota_{[0,a]^n}}_{\textit{semialgebraic}}$$

[Łojasiewicz '99]

[Łojasiewicz '99]

subanalytical
(stable to sum)

Ψ_θ is a KŁ function

[Bolte, Danilidis and Lewis '07]

subanalytic functions with **closed domain** and continuous upon restriction are KŁ

e.g. $A = I$ (denoising)

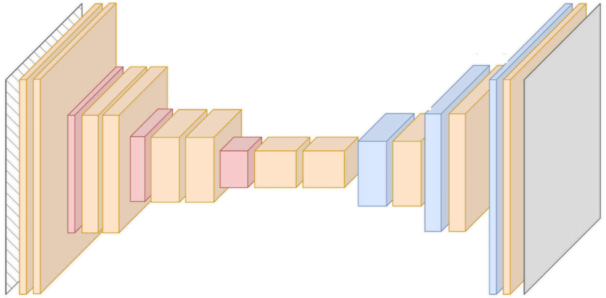
$\implies$ $\mathrm{dom}(\Psi_\theta)$ **not closed**

$A = I \implies \mathrm{dom}(\Psi) = (0,a]^n$

Numerical implementation

Choice of R_θ:

$$R_\theta(x) = \frac{1}{2}\|x - N_\theta(x)\|^2, N_\theta =$$



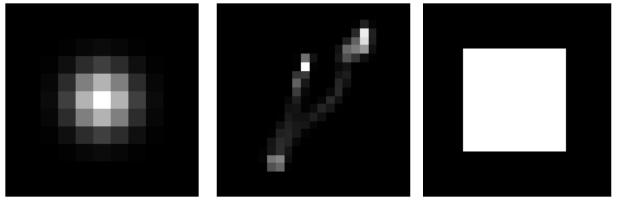
[Romano, Elad, and Milanfar '17]
[Hurault, Leclaire and Papadakis '21]

Dataset:
500 natural images of size 256 x 256

Data generation:

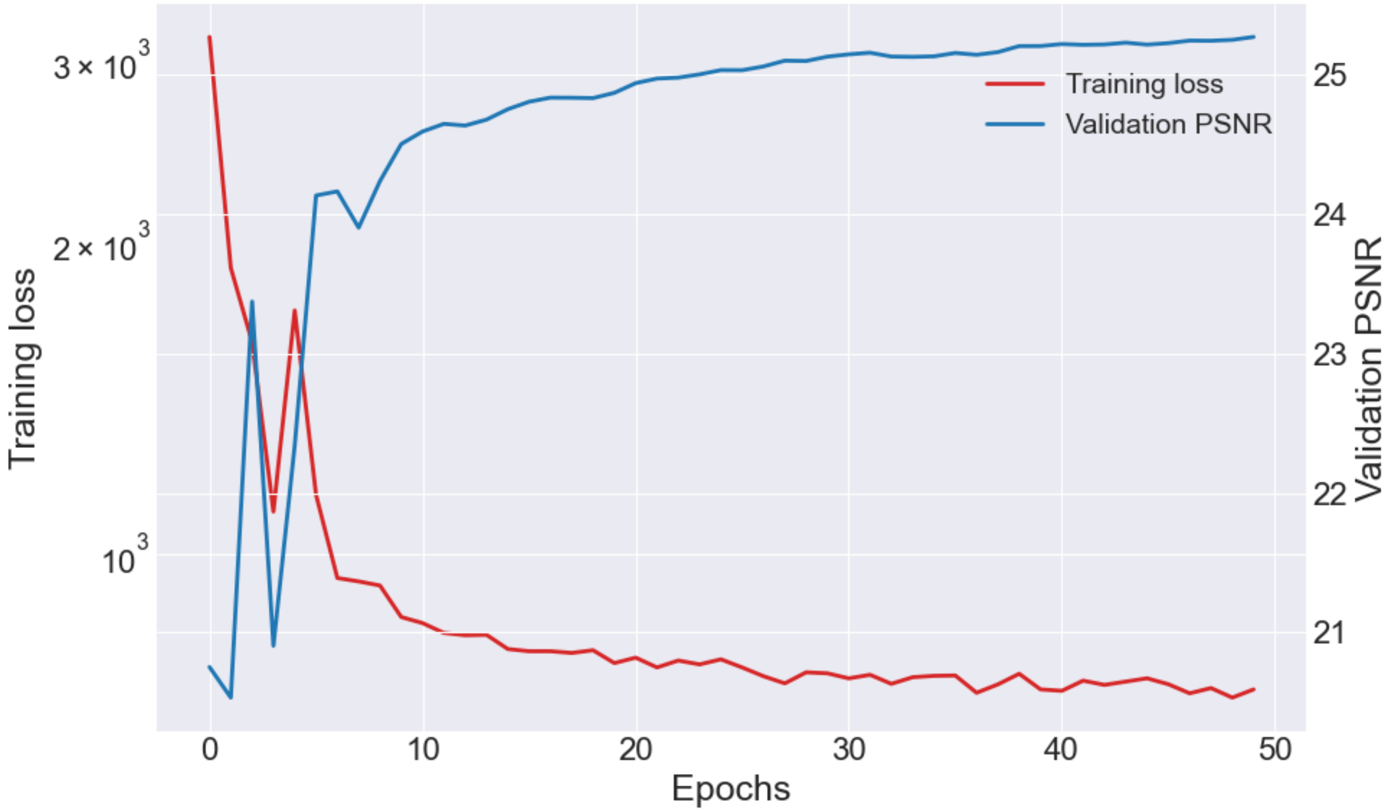
$$y = \text{Poiss}(\alpha A x)$$

$$A =$$



Loss function:

$$\ell(x^\infty, x) = \|x^\infty - x\|^2$$



UniGe | MaLGa

Ψ_θ is a Kurdyka-Łojasiewicz function

Proposition [D., Villa, Vaiter, Calatroni '25]

Subanalytic functions Ψ, which are **l.s.c., coercive**, **continuous** whenever restricted to their domain and such that:

$$S := \{x \in \overline{\mathrm{dom}(\Psi)} \mid \Psi(x) = +\infty\}$$

is closed, are KŁ.

Example: $n = 2,\ A = I, \mathrm{dom}(\Psi_\theta) = (0,a]^2,\ S = ([0,a] \times \{0\}) \cup (\{0\} \times [0,a])$

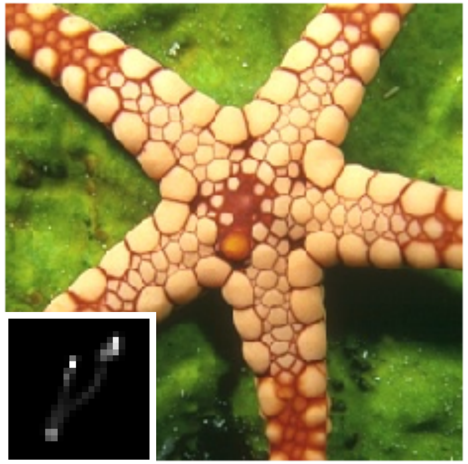
R_θ analytic $\Longrightarrow$ DEQ-MD is well-posed

UniGe | MaLGa

Numerical tests:

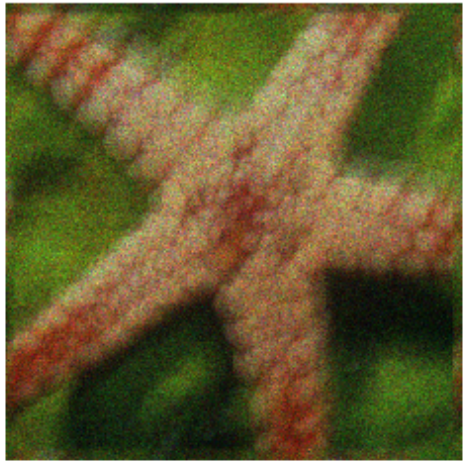
$\alpha = 60$

Ground truth



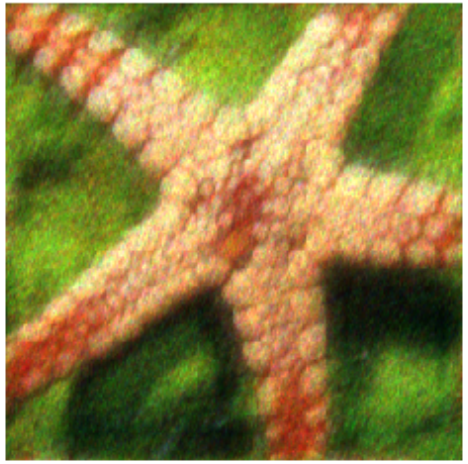
PSNR ↑, SSIM ↑

Measurement

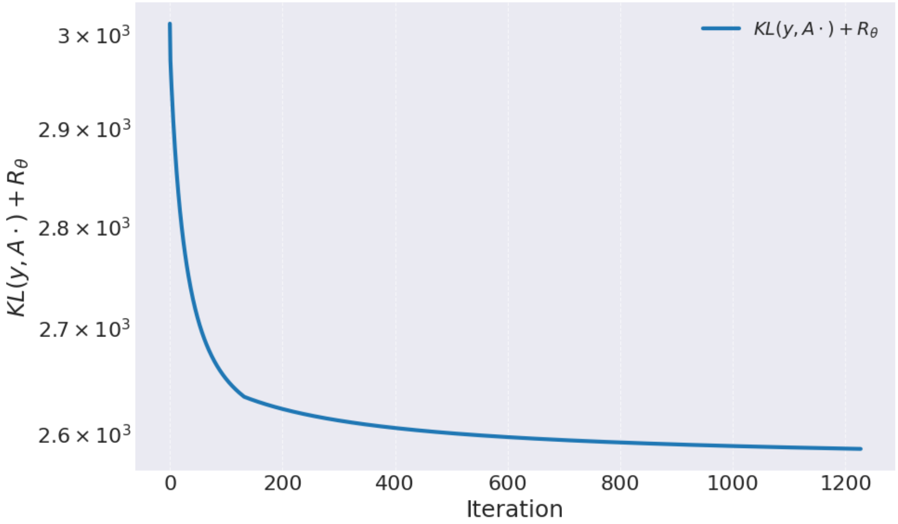


13.3058, 0.1664

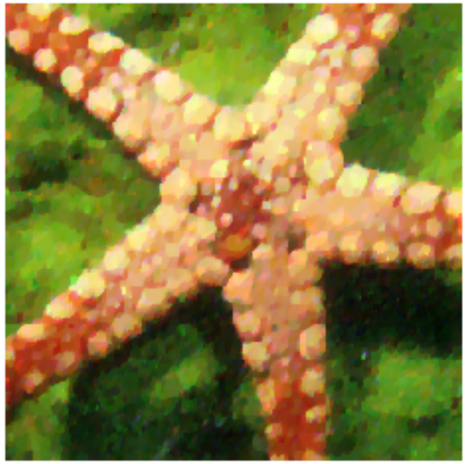
Richardson-Lucy



20.4583, 0.4737



Total Variation



22.3863, 0.5817

B-RED

<u>24.2011, 0.7044</u>

DEQ-RED

24.6214, 0.7255

	$\alpha = 100$			$\alpha = 60$		
Method	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
RL	22.7381	0.5379	0.5570	22.2968	0.5111	0.6017
TV+KL	23.9879	0.6304	<u>0.3888</u>	23.5203	0.6045	<u>0.4254</u>
B-RED	<u>25.0710</u>	<u>0.6939</u>	0.4120	**24.6566**	**0.6723**	0.4384
B-PnP	24.8976	0.6847	0.4417	24.5818	<u>0.6674</u>	0.4401
DEQ-S	24.2080	0.6730	0.3935	23.7252	0.6221	0.4346
DEQ-RED	**25.0955**	**0.6958**	**0.3317**	<u>24.5940</u>	0.6650	**0.3755**

Conclusions & outlook

- DEQ-MD as a convergent learning-the-regularisation framework

  $\longrightarrow$ Bregman MM? [Adly, Chazottes, Chouzenoux, Pesquet, Sureau, '26]

- Convergence beyond L-smoothness and via refined KŁ-like argument

  $\longrightarrow$ General result: other applications?

- Application to Poisson IPs

Thanks for your attention!

christian.daniele@edu.unige.it

- C. Daniele, S. Villa, S. Vaiter, L. Calatroni, *Deep equilibrium models for Poisson Imaging Inverse Problems via Mirror Descent*, to appear in **SIAM Journal of Imaging Sciences**, 19 (2), (2026).
- C. Daniele, S. Villa, S. Vaiter, L. Calatroni, *Noise and task-adaptive Deep-Equilibrium Mirror-Descent for Poisson Image reconstruction*, submitted (2026).

Forward pass

Challenges:

Fixed point iteration $x^{k+1} = f_\theta(x^k; y)$ slow (sublinear rate)

$\tau \propto \dfrac{1}{L}$ with $L > 0$ s.t. $Lh - (\mathsf{KL} + R_\theta)$ is convex, hard to compute

Backtracking for DEQ-MD [Hurault, Kamilov, Leclaire and Papadakis '23]

Input: $\gamma \in (0,1), \eta \in (0,1)$
Define: $T_\tau(x) = \nabla h^*(\nabla h(x) - \tau(\nabla \mathsf{KL}(y, Ax) + \nabla R_\theta(x))$
while $\Psi_\theta(x^k) - \Psi_\theta(T_\tau(x^k)) < \dfrac{\gamma}{\tau} D_h(T_\tau(x^k), x^k)$ **do**
$\tau \leftarrow \eta\tau$

Update: $x^{k+1} = T_\tau(x^k)$

UniGe | MaLGa

Backward pass:

$\{(x_i, y_i)\}_{i=1}^N,\ y_i = \text{Poisson}(Ax_i), \ell : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_{\geq 0}$

$$\text{find } \hat{\theta} \in \underset{\theta \in \Theta}{\operatorname{argmin}} \frac{1}{N} \sum_{i=1}^{N} \ell(x^\infty(\theta; y_i, x_i^0), x_i)$$

$$\bar{\theta} \in \Theta, \nabla_\theta \ell(x^\infty(\bar{\theta}; y, x^0), x) = \frac{\partial x^\infty(\bar{\theta}; y, x^0)}{\partial \theta}^\top \frac{\partial \ell(x^\infty(\bar{\theta}; y, x^0))}{\partial x} \qquad \frac{\partial x^\infty(\bar{\theta}; y, x^0)}{\partial \theta} = \frac{f_{\bar{\theta}}(x^\infty(\bar{\theta}; y\,.\,x^0))}{\partial \theta}$$

$$\frac{\partial x^\infty(\bar{\theta}; y, x^0)}{\partial \theta} = \left(I - \frac{\partial f_{\bar{\theta}}(x)}{\partial x} \bigg|_{x = x^\infty} \right)^{-1} \frac{\partial f_{\bar{\theta}}(x^\infty(\bar{\theta}; y, x^0))}{\partial \theta}$$

Constant memory

Backward pass:

$\{(x_i, y_i)\}_{i=1}^N,\ y_i = \text{Poisson}(Ax_i), \ell : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_{\geq 0}$

$$\boxed{\text{find } \hat{\theta} \in \underset{\theta \in \Theta}{\text{argmin}} \frac{1}{N} \sum_{i=1}^{N} \ell(x^\infty(\theta; y_i, x_i^0), x_i)}$$

$$\bar{\theta} \in \Theta, \nabla_\theta \ell(x^\infty(\bar{\theta}; y, x^0), x) = \frac{\partial x^\infty(\bar{\theta}; y, x^0)}{\partial \theta}^\top \frac{\partial \ell(x^\infty(\bar{\theta}; y, x^0))}{\partial x}$$

$$\frac{\partial x^\infty(\bar{\theta}; y, x^0)}{\partial \theta} = \frac{f_{\bar{\theta}}(x^\infty(\bar{\theta}; y\,.\,x^0))}{\partial \theta}$$

$$\boxed{\frac{\partial x^\infty(\bar{\theta}; y, x^0)}{\partial \theta} \approx \left(I - \cancel{\frac{\partial f_{\bar{\theta}}}{\partial x}\Big|_{x^\infty}} \right)^{-1} \frac{\partial f_{\bar{\theta}}(x^\infty(\bar{\theta}; y, x^0))}{\partial \theta}}$$

Jacobian Free Backpropagation/
One step differentiation

[Fung, Heaton, Li, McKenzie, Osher and Yin '22]
[Bolte, Pauwels and Vaiter '23]

$$\|\frac{\partial f_{\bar{\theta}}}{\partial x}\| < 1$$

Our case: $\|\frac{\partial f_{\bar{\theta}}}{\partial x}\| \leq 1$

UniGe | MaLGa

Convergence proof structure: idea

- **Issue:** classical KŁ results require a *closed domain*, but $\text{dom}(\Psi)$ is not closed.

Idea: isolate region around $\mathbf{x}^* \in \text{Crit}(\Psi)$ avoiding $+\infty$

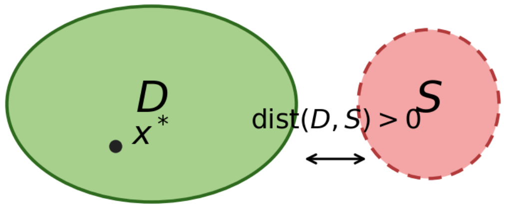
- **Step 1**: localise around $\mathbf{x}^* \in \text{Crit}(\Psi)$ considering $D = \Psi^{-1}(-\infty, M]$, where $M = \Psi(\mathbf{x}^*)$.
- $D \in \mathbb{R}^n$ is compact and separated from $S = \left\{\mathbf{x} : \Psi(\mathbf{x}) = +\infty\right\}$? Then $\varepsilon = \text{dist}(D, S) > 0$
- $\mathbf{x}^* \in \text{int}(D)$, set $D' = D$.
- If $\mathbf{x}^* \in \partial D$, enlarge slightly to D_ε and define $B_\varepsilon = D_\varepsilon \cap \text{dom}(\Psi) \to$ find $D' = \Psi(-\infty, M_\varepsilon], M_\varepsilon < +\infty$
- **Step 2:** Having $D' \subset \text{dom}(\Psi)$, construct $U(\mathbf{x}^*)$: $U(\mathbf{x}^*) \cap \text{dom}(\Psi) \subset D'$

Idea: apply classical results

- D' is subanalytic.
- **Step 3:** Define $\tilde{\Psi} = \Psi + \iota_{D'}$. It is subanalytic with closed domain. Theory applies.
- KŁ holds for $\tilde{\Psi}$ at $\mathbf{x}^*$
- **Conclusion:** $\Psi \equiv \tilde{\Psi}$ on $U(\mathbf{x}^*)$

Convergence proof structure: idea

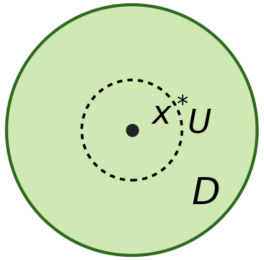
1. Geometry



$D = \Psi^{-1}((-\infty, M])$

$S = \left\{ x \in \overline{\mathrm{dom}(\Psi)} : \Psi(x) = +\infty \right\}$

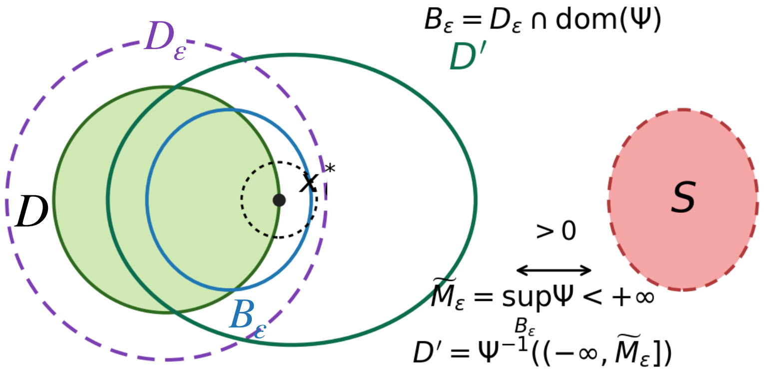
2a. Case $x^* \in \mathrm{int}(D)$



Take $D' = D$.

$U \cap \mathrm{dom}(\Psi) \subset D'$.

2b. Case $x^* \in \partial D$



$U \cap \mathrm{dom}(\Psi) \subset D'$

3. Reduction

$$\widetilde{\Psi} = \Psi + \iota_{D'}$$

$\mathrm{dom}(\widetilde{\Psi}) = D'$ is closed

D' is subanalytic as

$D' = \Psi^{-1}([-m, M])$

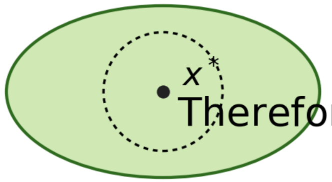
$\tilde{\Psi}$ is subanalytic

4. Apply classical KL

Apply the classical KL theorem to $\widetilde{\Psi}$ at x^*.

$\tilde{\Psi}$ satisfies KŁ at x^*.

5. Transfer back locally



On U, one has $\widetilde{\Psi} = \Psi$.

Therefore Ψ satisfies the KL property at x^*.