Deep Equilibrium models for Poisson inverse problems via Mirror Descent

Christian Daniele, Silvia Villa, Samuel Vaiter and Luca Calatroni

SIAM Chapters Meeting for applied mathematicians

January 30, 2026

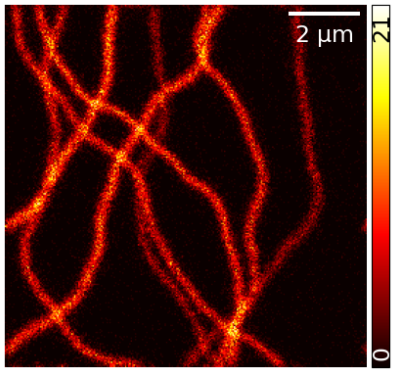
Co-funded by the European Union (ERC, MALIN, 10117133). Views and opinions expressed are however those of the author only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the EU nor the granting authority can be held responsible for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

Poisson Inverse Problems

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

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

Motivation: Poisson noise models **Photon counts** in imaging sensors (e.g. Microscopy) [Bertero and Boccacci '98]



To retrieve x:

$$\underset{x \geq 0}{\operatorname{argmax}}\ \pi_{\mathcal{X}|\mathcal{Y}}(x \mid y) \quad \Longrightarrow \quad \underset{x \geq 0}{\operatorname{argmin}}\ \mathrm{KL}(y, Ax) + R_\theta(x)$$

MAP

$$\pi_{\mathcal{X}|\mathcal{Y}}(x \mid y) \propto \pi_{\mathcal{Y}|\mathcal{X}}(y \mid Ax)\pi_{\mathcal{X}}(x)$$

not available

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

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

- ≥ 0
- convex
- not Lipschitz gradient

UniGe | MaLGa

Variational formulation and regularization

$$\underset{x \geq 0}{\operatorname{argmin}}\ \mathrm{KL}(y, Ax) + R_\theta(x)$$

Model-based

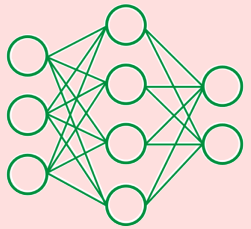
$$R_\theta(x) = \begin{cases} \theta \|x\|_1 & \text{L1} \\ \theta\, \mathrm{TV}(x) & \text{Total Variation} \end{cases}$$

[Rudin, Osher and Fatemi '92]

$\theta > 0$, **regularization parameter**

Goal:

given $\underset{x \geq 0}{\operatorname{argmin}}\ \mathrm{KL}(y, Ax) + R_\theta(x)$, learning $R_\theta =$

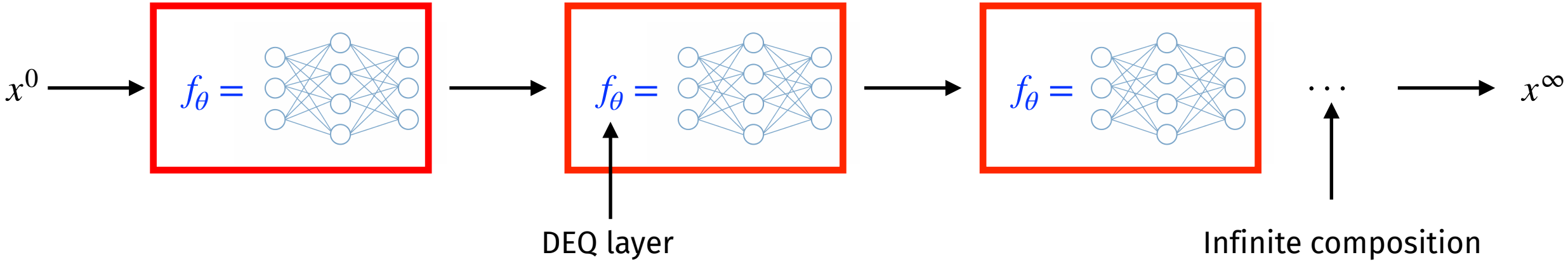


optimization

$$\mathrm{Algo}_\theta(x^0, y, A) \longrightarrow f_\theta(\,\cdot\,; y, A)$$

Challenge: R_θ cannot be learned in an end-to-end fashion

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



Definition (Well-posedness of a DEQ)

The DEQ associated with f_θ: $X \subset \mathbb{R}^n \to X$ is **well-posed** if:

1. $\text{Fix}(f_\theta) \neq \emptyset$;
2. For all $x^0 \in X$, $\{x^k\}_{k\in\mathbb{N}}$ given by:

$$x^k = f_\theta(x^{k-1})$$

converges to $x^\infty \in \text{Fix}(f_\theta)$.

Example: Contraction $\Longrightarrow$ Well-posedness

DEQs for image reconstruction

Given:
$y = \text{Poisson}(Ax)$

Design $f_\theta(\,\cdot\,; y)$ s.t. $x^\infty(\theta; y) \approx x$

with $x^\infty(\theta; y) = f_\theta(x^\infty(\theta); y)$

For simplicity:

$f_\theta(\,\cdot\,; y) = f_\theta(\,\cdot\,; y, A)$
$x^\infty(\theta; y) = x^\infty(\theta; y, x^0)$

Ingredients:

$\{(\hat{x}_i, y_i)\}_{i=1}^N$, $y_i = \text{Poisson}(A\hat{x}_i)$, fix the structure of $f_\theta(\,\cdot\,; y_i)$

Training of a DEQ

$$\text{find } \hat{\theta} \in \underset{\theta\in\Theta}{\text{argmin}} \frac{1}{2N} \sum_{i=1}^{N} \|x^\infty(\theta; y_i) - \hat{x}_i\|_2^2$$

$$\text{s.t.} \quad x^\infty(\theta; y_i) = f_\theta(x^\infty(\theta); y_i)$$

Question:
How do we choose f_θ ? ⟶ Optimisation algorithm. Which one?

KL minimisation: Mirror Descent

$$\underset{x \geq 0}{\operatorname{argmin}} \, \mathrm{KL}(y, Ax) + R_\theta(x) := J_\theta(x; y)$$

Given a Legendre function h:

$$x^{k+1} = \nabla h^*(\nabla h(x^k) - \tau(\nabla \mathrm{KL}(y, Ax) + \nabla R_\theta)) =: f_\theta(x^k; y)$$

If $h = \|\cdot\|_2^2$, then $\quad x^{k+1} = x^k - \tau(\nabla \mathrm{KL}(y, Ax^k) + \nabla R_\theta(x^k))$ Gradient Descent

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 property:

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

$\mathrm{KL}(y, A\cdot)$ has a Lipschitz gradient "relative to h"

From Mirror Descent to DEQ-MD

$$f_\theta(x; y) := \nabla h^*(\nabla h(x) - \tau(\nabla \mathrm{KL}(y, Ax) + \nabla R_\theta(x))$$

Proposition (Informal)

Let $R_\theta : \mathbb{R}^n \to \mathbb{R}$ be an **analytic network** and $h(x) = -\sum_{i=1}^{n} \log(x_i)$. Then, for all $x^0 \in (0, +\infty)^n$,

the sequence given by $\{x^k\}_{k\in\mathbb{N}}$ **converges to** $x^\infty(\theta; y, x^0)$ **a critical point of** $J_\theta(\,\cdot\,; y) = \mathrm{KL}(y, A\cdot) + R_\theta$.

In addition it holds: $\mathrm{Fix}(f_\theta(\,\cdot\,; y)) \subset \mathrm{Crit}(J_\theta)$.

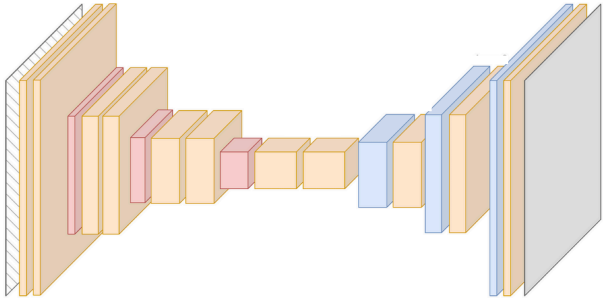
Key point: Learning $f_\theta(\,\cdot\,; y) \implies$ learning the regularization R_θ

Remark: Analyticity is not restrictive!

Numerical implementation

Choice of R_θ:

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

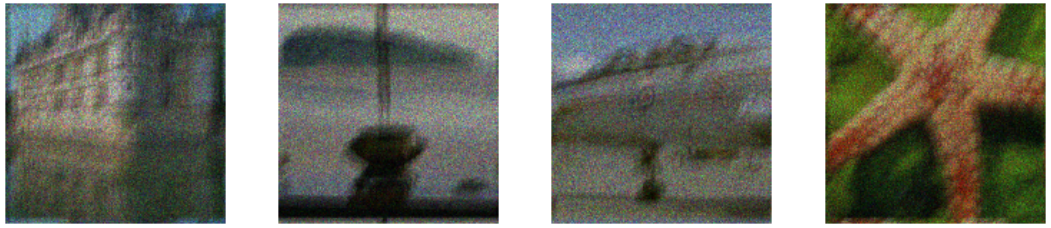


[Romano, Elad, and Milanfar '17]
[Hurault, Leclaire and Papadakis '21]

Dataset:
500 natural images of size 256 x 256

$\hat{x}_i$

y_i



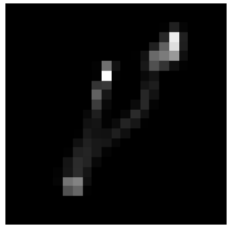
Data generation:

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

α = strength of Poisson noise

$\alpha \in [20,80]$

$A =$

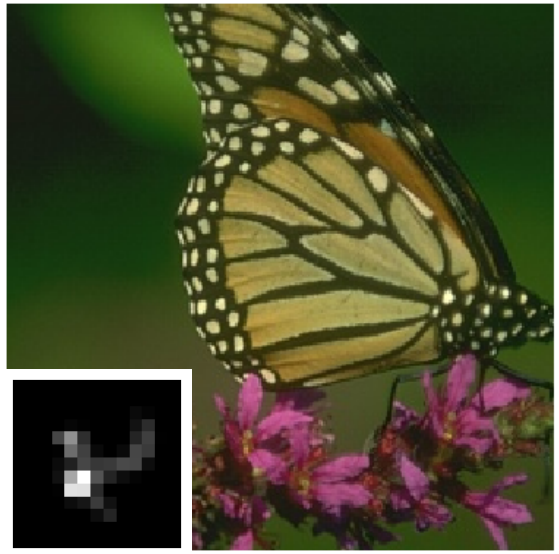


Convolution operator

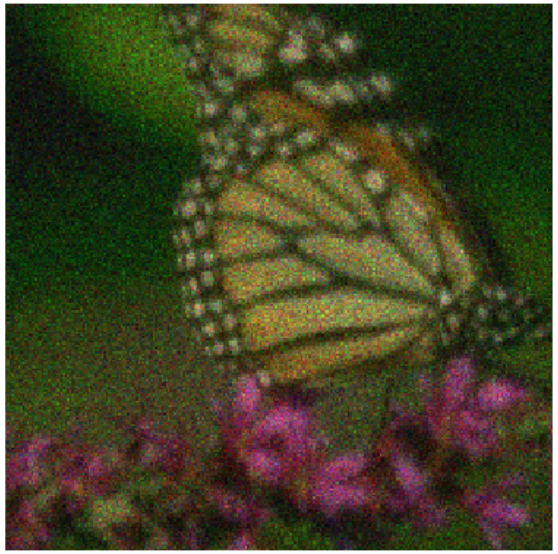
Numerical tests:

$$\alpha = 40$$

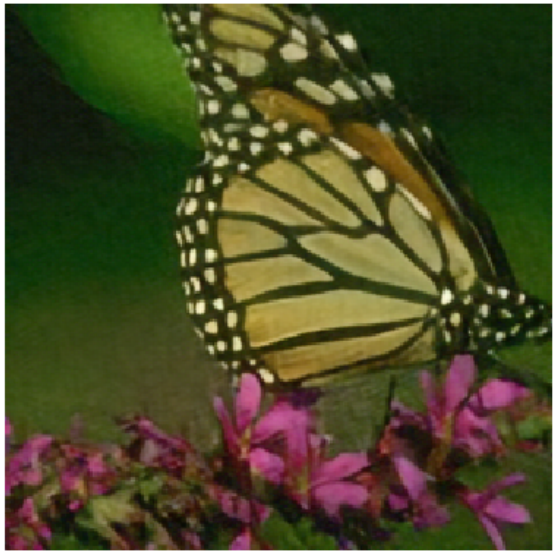
GT



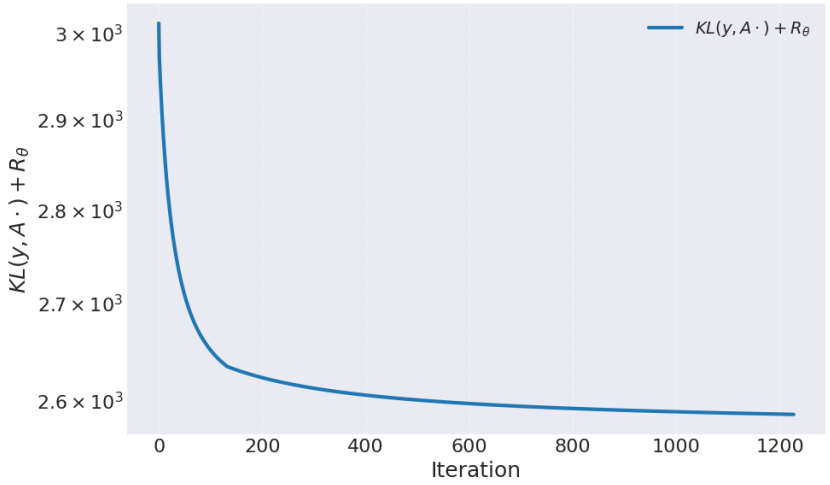
Measurement

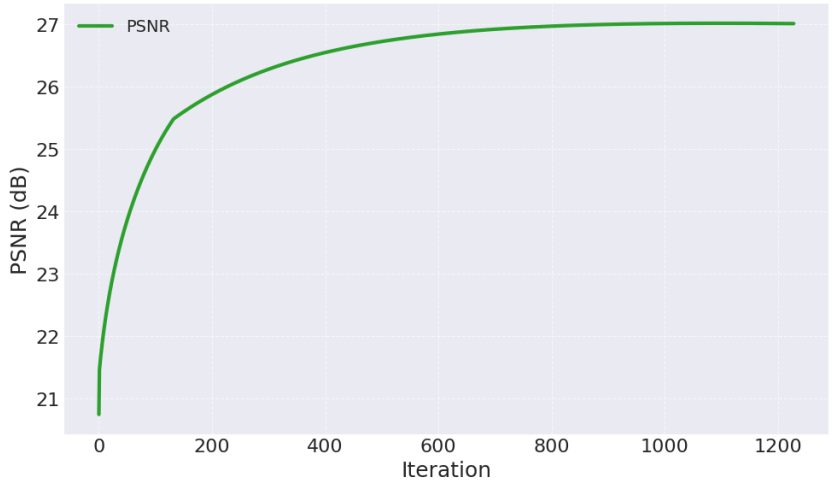


DEQ, PSNR: **28.23**, SSIM: **0.83**



RAM, PSNR: 27.66, SSIM: 0.82



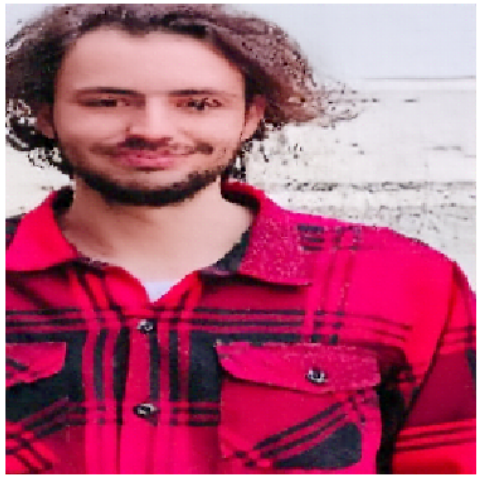


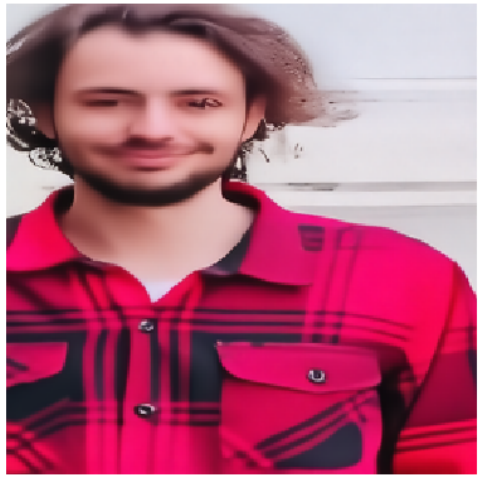
Numerical tests:

$$\alpha = 60$$

President

Y-SIMAI representative

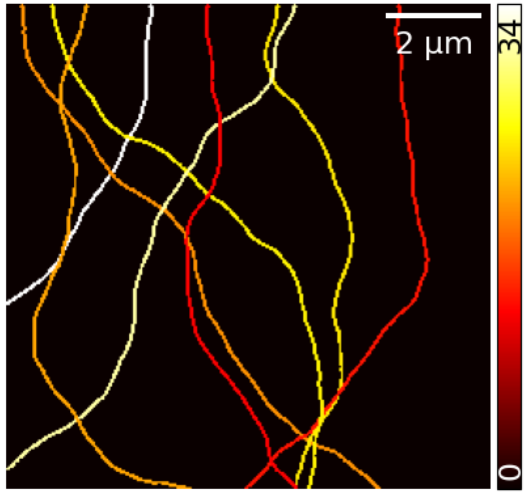




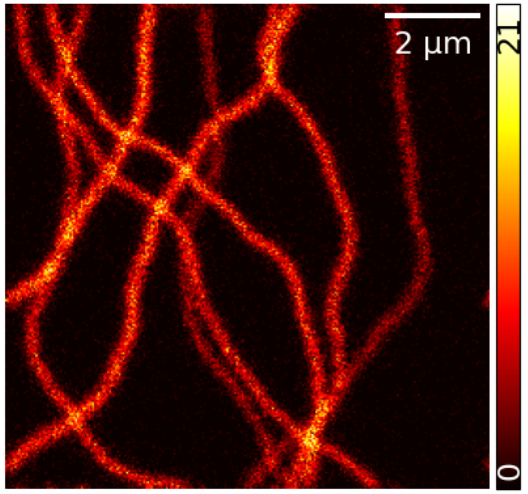
Work in Progress: Testing on microscopy

Supervised training on simulated data

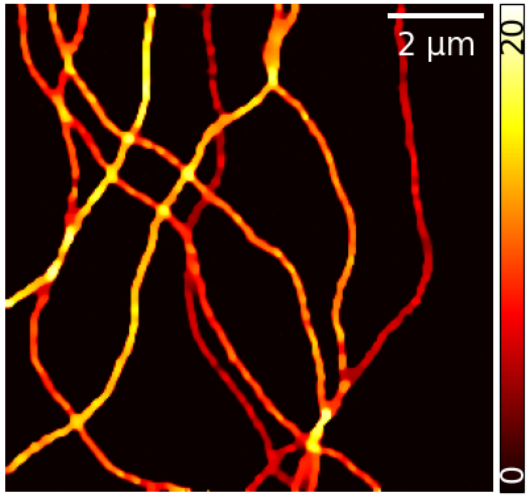
GT



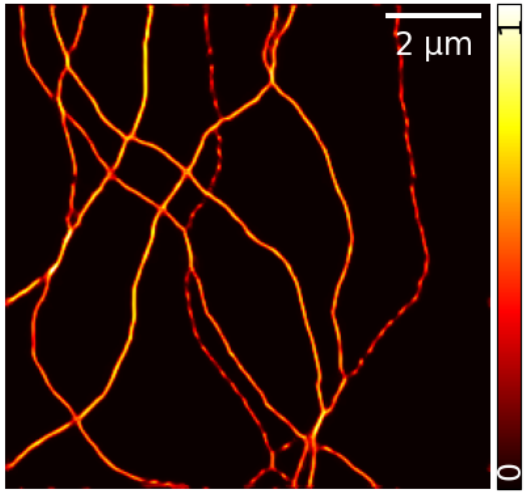
Measurement



TV, PSNR: 17.83, SSIM: 0.78



DEQ, PSNR: **21.05**, SSIM: **0.91**



Conclusion & Outlook

- Learned regularisation for Poisson IPs through MD-type deep equilibrium models
- Interpretable inference and reliable results

Future work:

- Extend training of DEQ-MD in the self-supervised setting
- Extend to CT, MRI

Thanks for your attention!

UniGe

MaLGa

Preprint on Arxiv:

Github:

Weights the action of the regularizer

$$f_{\theta_1,\theta 2}(x;y) = \nabla h^*(\nabla h(x) - \tau(\nabla \mathrm{KL}(y, Ax) + g_{\theta_2}(y)\nabla R_{\theta_1}(x)))$$

More precisely:

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:

$$x^0 \in \mathrm{int}(\mathrm{dom}(h)) \implies x^k \in \mathrm{int}(\mathrm{dom}(h)) \forall k \geq 0$$

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

NoLip property [Bauschke, Bolte and Teboulle '17]

L-SMAD [Bolte, Sabach, Teboulle and Vaisbourd '1

Relative smoothness [Lu, Freund and Nesterov '18]

Gradient computation:

$\{(x_i, y_i)\}_{i=1}^N,\ y_i = \text{Poisson}(Ax_i)$

$$\text{find } \hat{\theta} \in \underset{\theta \in \Theta}{\text{argmin}} \frac{1}{N} \sum_{i=1}^{N} \|x^\infty(\theta; y_i, x_i^0), x_i\|_2^2$$

$$\bar{\theta} \in \Theta, \nabla_\theta \ \|x^\infty(\bar{\theta}; y, x^0), x\|_2^2 = \frac{\partial x^\infty(\bar{\theta}; y, x^0)}{\partial \theta}^\top (x^\infty(\bar{\theta}; y, x^0) - x) \qquad \frac{\partial x^\infty(\bar{\theta}; y, x^0)}{\partial \theta} = \frac{f_{\bar{\theta}}(x^\infty(\bar{\theta}; y\,.\,x^0))}{\partial \theta}$$

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

Constant memory

Gradient computation:

$\{(x_i, y_i)\}_{i=1}^N,\ y_i = \text{Poisson}(Ax_i)$

$$\text{find } \hat{\theta} \in \underset{\theta \in \Theta}{\text{argmin}} \frac{1}{N} \sum_{i=1}^{N} \|x^\infty(\theta; y_i, x_i^0), x_i\|_2^2$$

$$\bar{\theta} \in \Theta, \nabla_\theta \ \|x^\infty(\bar{\theta}; y, x^0), x\|_2^2 = \frac{\partial x^\infty(\bar{\theta}; y, x^0)}{\partial \theta}^\top (x^\infty(\bar{\theta}; y, x^0) - x)$$

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

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

Jacobian Free Backpropagation/
One step differentiation

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

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

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

Forward pass

Challenges:

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

$\tau \propto \frac{1}{L}$ with $L > 0$ s.t. $Lh - (\mathrm{KL} + R_\theta)$ 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: $T_\tau(x) = \nabla h^*(\nabla h(x) - \tau(\nabla \mathrm{KL}(y, Ax) + \nabla R_\theta(x))$

while $\Psi_\theta(x^k) - \Psi_\theta(T_\tau(x^k)) < \frac{\gamma}{\tau} D_h(T_\tau(x^k), x^k)$ **do**

$\tau \leftarrow \eta\tau$

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

ISM