UNIVERSITY OF CAMBRIDGE

Adversarial regularization for imaging inverse problems

Subhadip Mukherjee

Department of Applied Mathematics and Theoretical Physics
University of Cambridge, UK

Mathematics for deep learning opening workshop, University of Bath
21 April 2022
Joint work with: M. Carioni, C.-B. Schönlieb, O. Öktem, S. Dittmer, Z. Shumaylov

Inverse problems in imaging

- Reconstruct image $x \in \mathbb{X}$ from observed data

$$y = Ax + \mathrm{w} \in \mathbb{Y}$$

- Main challenges:
  - In general, ill-posed: A is poorly conditioned or non-invertible
  - Multiple x explain y (even with no noise)
  - Instability when the observation is noisy
- Variational regularization: $\boxed{\min_{x} \underbrace{L(y, Ax)}_{\text{depends on imaging physics}} + \lambda \underbrace{R(x)}_{\text{prior belief}}}$
- Desirable properties: well-posedness, convergent regularization

How machine learning can help

Supervised setting:

- Parametric estimator f_θ, $\theta \in \Theta$ (typically, a neural network)
- Training data: i.i.d. samples $(x_i, y_i)_{i=1}^n$ of $(\mathbb{x}, \mathbb{y})$
- Training: $\min\limits_{\theta \in \Theta} L_{\mathbb{X}}\left(x_i, f_\theta(y_i)\right) + \mu\, \varphi(\theta)$

Unsupervised setting:

- Training data: i.i.d. samples $(x_i)_{i=1}^n$ and $(y_j)_{j=1}^{n'}$ of $\mathbb{x}$ and $\mathbb{y}$, respectively
- $\min\limits_x L(y, Ax) + \lambda \underbrace{R_\theta(x)}_{\text{learn from data}}$

Questions:

- How to construct the architecture of f_θ? unrolling
- How to train a regularizer R_θ? adversarial learning

Algorithm unrolling

- Proximal gradient for minimizing the variational objective with a convex R:

$$x_{k+1} = \underbrace{\mathsf{prox}_{\lambda R}}_{\text{learnable}} \underbrace{\left(x_k - \eta_k \nabla L(y, Ax) \Big|_{x=x_k} \right)}_{\text{physics-driven}}$$

- Unroll N times, for a fixed N, and replace the prox with a neural net:

$$x_{k+1} = \phi_{\theta_k} \left(x_k - \eta_k \nabla L(y, Ax) \Big|_{x=x_k} \right), k = 0, 1, \cdots, N-1$$

  - Learned primal-dual: [Adler & Öktem, IEEE-TMI 2018]
- Training: learn $\theta = (\theta_k)_{k=0}^{N-1}$ by $\min_{\theta} \mathbb{E}_{(\mathbb{x},\mathbb{y})\sim\hat{\mu}}\, \mathsf{loss}\left(x_N(\mathbb{y}, \theta), \mathbb{x}\right)$
  - Based on the loss, approximates different statistical estimators (e.g., $\mathsf{loss}(a, b) = \|a - b\|_2^2 \implies x_N(y, \theta) \approx \mathbb{E}[\mathbb{x}|\mathbb{y} = y]$)

Adversarial regularizers (AR)

- Training data:
  - high-quality ground-truth images $(x_i)_{i=1}^{n_1} \sim \pi_x$
  - images with artifacts $(\tilde{x}_i)_{i=1}^{n_2} \sim \pi_{\tilde{x}}$
- Training objective [Lunz et al., NeurIPS-2018]:

$$\min_{\theta} \left(\mathbb{E}_{\mathbb{x} \sim \pi_x} \left[R_\theta \left(\mathbb{x} \right) \right] - \mathbb{E}_{\tilde{\mathbb{x}} \sim \pi_{\tilde{x}}} \left[R_\theta \left(\tilde{\mathbb{x}} \right) \right] \right) \text{ s.t. } R_\theta \in 1 - \text{Lipschitz}$$

- Optimal critic recovers the Wasserstein-1 distance between π_x and $\pi_{\tilde{x}}$

Remarks:

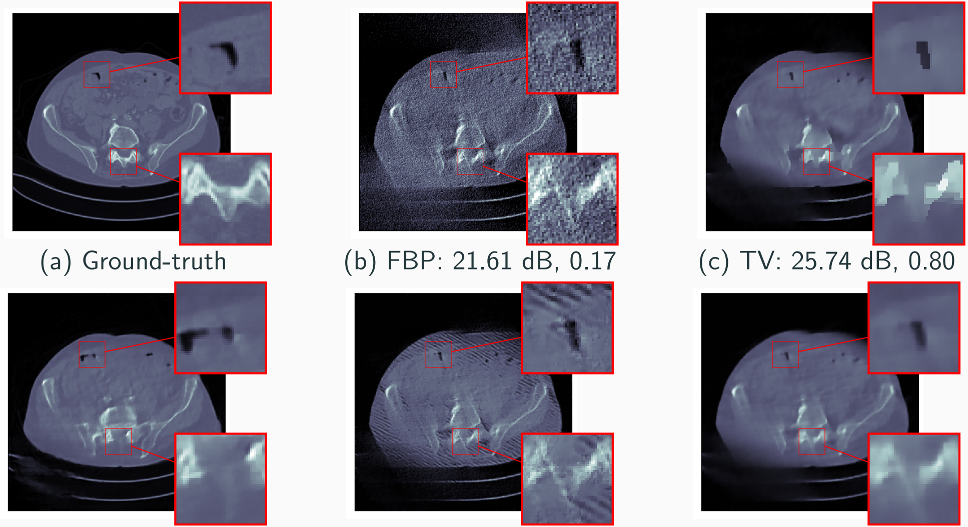
- Weakly stable, but not a convergent regularization scheme
- No global convergence for the variational problem
- Needs 'negative' images for discriminative training
- Solving a variational problem for large images could be expensive

Adversarial convex regularization (ACR)

- Use input-convex neural networks (ICNNs) [Amos et al., ICML-2017] to model R_θ
- Well-posedness (existence, uniqueness, and stability) follows from the classical convex regularization theory
- Convergent regularization: Let $x^\dagger \in \underset{x}{\arg\min}\, R_{\theta^*}(x)$ subject to $Ax = y^0$. Then, for $\delta \to 0$, $\lambda(\delta) \to 0$, and $\dfrac{\delta}{\lambda(\delta)} \to 0$, $\hat{x}_\lambda\left(y^\delta\right)$ converges to $x^\dagger$ in $\|\cdot\|_X$
- Gradient descent converges provably

Ref.: S.M., Dittmer, Shumaylov, Lunz, Öktem, Schönlieb, "Learned convex regularizers for inverse problems," arXiv:2008.02839, code: `github.com/Subhadip-1/data_driven_convex_regularization`

Convexity matters practically



(a) Ground-truth

(b) FBP: 21.61 dB, 0.17

(c) TV: 25.74 dB, 0.80

(d) LPD: 29.51 dB, 0.85

(e) AR: 26.83 dB, 0.71

(f) ACR: 27.98 dB, 0.84

Unrolling vs. data-driven regularization

Algorithm unrolling:

- Fast and high-quality end-to-end reconstruction ✓
- Needs supervision ✗
- No convergence guarantees in general ✗

Adversarial regularization:

- Convergence can be studied using classical regularization theory ✓
- Does not need paired training data ✓
- Reconstruction is expensive ✗
- Generally performs worse than unrolling empirically ✗

Unrolled adversarial regularization (UAR)

$$\boxed{\min_{\phi} \max_{\theta:R_\theta\in 1-\text{Lip.}} \mathbb{E}_{\mathrm{y}\sim\pi_y} \|AG_\phi(\mathrm{y})-\mathrm{y}\|_2^2 + \lambda\left(\mathbb{E}_{\mathrm{y}\sim\pi_y}\left[R_\theta\left(G_\phi(\mathrm{y})\right)\right] - \mathbb{E}_{\mathrm{x}\sim\pi_x}\left[R_\theta\left(\mathrm{x}\right)\right]\right)}$$

- G_ϕ learns to minimize the (expected) variational loss with R_θ as the regularizer
- R_θ learns to tell apart the output of G_ϕ from the ground-truth images
- Key features:
  - unsupervised, no paired data needed for training
  - fast reconstruction, competitive quality with supervised methods
  - end-to-end solution by G can be refined using the variational framework $\implies$ stability guarantees similar to AR

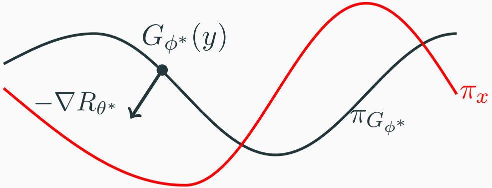
Ref.: S.M., Carioni, Öktem, Schönlieb, "End-to-end reconstruction meets data-driven regularization for inverse problems," NeurIPS-2021, code: `github.com/Subhadip-1/unrolling_meets_data_driven_regularization`

Theoretical results for UAR

- The saddle-point training problem is **well-posed** and **equivalent** to

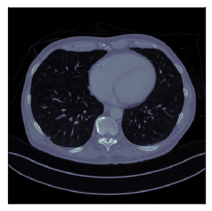
$$\min_{\phi} \mathbb{E}_{\mathrm{y}\sim\pi_y} \|AG_\phi(\mathrm{y})) - \mathrm{y}\|_2^2 + \lambda\, \mathbb{W}_1\left((G_\phi)_{\#}\pi_y, \pi_x\right)$$

- The parameters ϕ of the generator vary continuously w.r.to the noise level.
- Let $x_0 = G_{\phi^*}(y)$, $x_1 = x_0 - \eta\nabla R_{\theta^*}(x_0) \implies \mathbb{W}_1\left(\pi_{x_1}, \pi_x\right) \leq \mathbb{W}_1\left(\pi_{x_0}, \pi_x\right)$

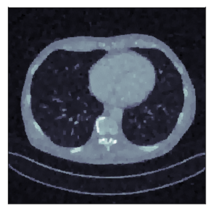


- $\lambda \to 0 \implies G_\phi \to G_{\phi_1^*}$, minimizer of Wasserstein distance s.t. perfect data fit
- $\lambda \to \infty \implies G_\phi \to G_{\phi_2^*}$, minimizer of data fit s.t. perfect distribution matching

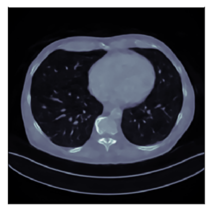
Numerical examples: X-ray CT reconstruction



(a) ground-truth



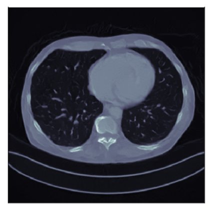
(b) TV: 29.16, 0.77



(c) LPD: 34.05, 0.89

(d) AR: 32.14, 0.84

(e) ACR: 30.14, 0.83



(f) UAR: 33.15, 0.87

Performance evaluation for CT reconstruction

method		PSNR (dB)	SSIM	# param.	time (ms)
FBP		21.28 ± 0.13	0.20 ± 0.02	1	37.0 ± 4.6
TV		30.31 ± 0.52	0.78 ± 0.01	1	28371.4 ± 1281.5
U-Net		34.50 ± 0.65	0.90 ± 0.01	7215233	44.4 ± 12.5
LPD		35.69 ± 0.60	0.91 ± 0.01	1138720	279.8 ± 12.8
AR		33.84 ± 0.63	0.86 ± 0.01	19338465	22567.1 ± 309.7
ACR		31.55 ± 0.54	0.85 ± 0.01	606610	109952.4 ± 497.8
UAR	$\lambda = 0.001$	21.59 ± 0.11	0.22 ± 0.02	20477186	252.7 ± 13.3
	$\lambda = 0.01$	25.25 ± 0.08	0.37 ± 0.01		
	$\lambda = 0.1$	34.35 ± 0.66	0.88 ± 0.01		
	$\lambda = 1.0$	33.27 ± 0.76	0.87 ± 0.01		
UAR with refinement	$\lambda = 0.1$	34.77 ± 0.67	0.90 ± 0.01	–	5863.3 ± 106.1

Summary and conclusions

- Variational framework is a starting point devising hybrid methods
- Imposing convexity on the regularizer $\implies$ theoretical guarantees, but at the expense of numerical performance
  - Better architecture and/or training strategy?
- Unrolled adversarial regularization achieves good empirical performance, and can be analyzed using the variational setting
  - Convergent regularization?
- How to build distributionally robust algorithms?
- Theoretical performance analysis under distribution shifts?