Learning convex regularizers satisfying the variational source condition for inverse problems

Subhadip Mukherjee[1], Carola-Bibiane Schönlieb[1], and Martin Burger[2]
[1]Department of Applied Mathematics and Theoretical Physics, University of Cambridge, UK
[2]Department of Mathematics, University of Erlangen-Nuremberg, Germany
Emails: {sm2467, cbs31}@cam.ac.uk, martin.burger@fau.de

1. BACKGROUND ON INVERSE PROBLEMS

- Reconstruct image $x^* \in \mathbb{X}$ from noisy observation (data) $y^\delta = Ax^* + \text{noise} \in \mathbb{Y}$.
  - $A : \mathbb{X} \to \mathbb{Y}$ is the forward operator.
  - $\|\text{noise}\| \leqslant \delta$
- Inverse problems are ill-posed, i.e., A is either non-invertible or poorly conditioned.
- **Variational regularization**:

$$\boxed{x_\lambda \in \underset{x \in \mathbb{X}}{\arg\min}\, \frac{1}{2}\|y^\delta - Ax\|_2^2 + \lambda\, \psi_\theta(x)} \quad (1)$$

  - $\{\psi_\theta\}_{\theta \in \Theta}$ is a convex regularizer.
- **ψ_θ-minimizing solution**:

$$\boxed{x^\dagger \in \underset{x \in \mathbb{X}}{\arg\min}\, \psi_\theta(x) \text{ subject to } Ax = y^0} \quad (2)$$

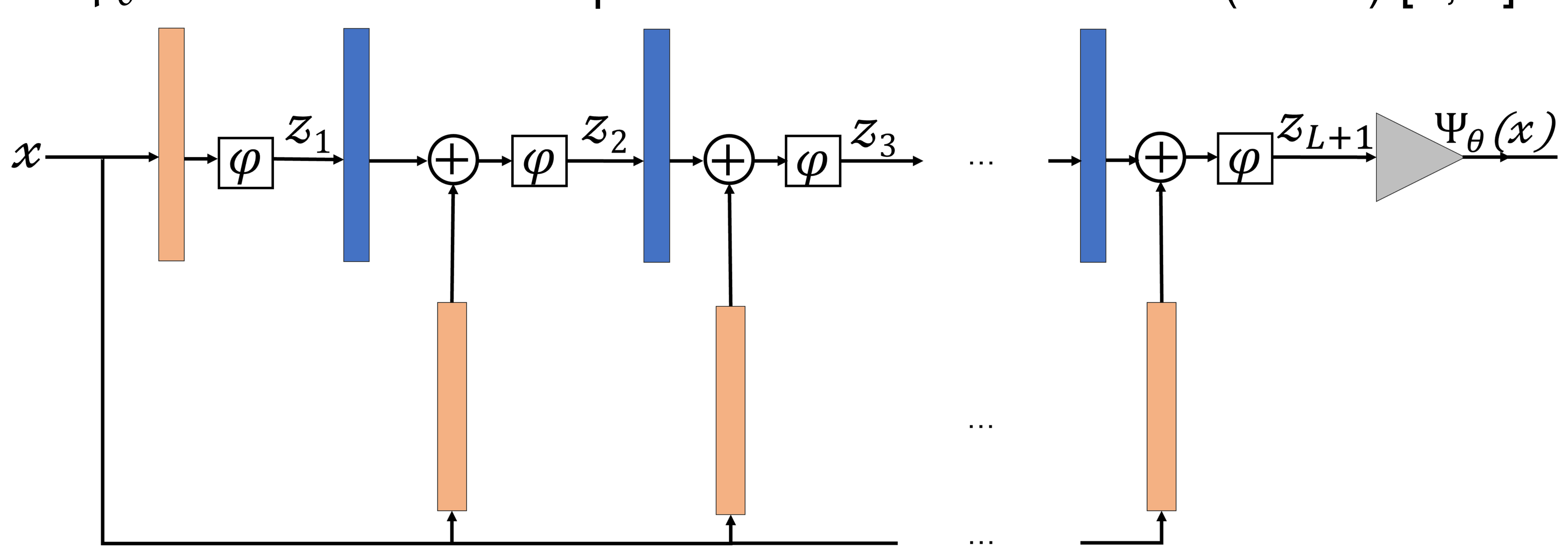
- **Variational source condition**: is satisfied if there exists some $w^\dagger \in \mathbb{Y}$ such that $A^* w^\dagger \in \partial\, \psi_\theta(x^\dagger)$.
- **Bregman distance**: $D_{\psi_\theta}(x_1, x_2) := \left\{\psi_\theta(x_1) - \psi_\theta(x_2) - \langle u, x_1 - x_2\rangle \middle| u \in \partial\psi_\theta(x_2)\right\}$
- **Convergence rate** [1]: If the source condition holds, then for each minimizer x_λ of (1), there exists $d \in D_{\psi_\theta}(x_\lambda, x^\dagger)$ such that

$$\boxed{d \leqslant \lambda \frac{\|w^\dagger\|^2}{2} + \frac{\delta^2}{2\lambda}}. \quad (3)$$

Therefore, choosing $\lambda \propto \delta$ leads to an $\mathcal{O}(\delta)$ convergence rate of the variational reconstruction x_λ to $x^\dagger$.

2. PARAMETRIZING THE REGULARIZER

- ψ_θ is taken to be an input-convex neural network (ICNN) [2, 3].



- ψ_θ is constructed recursively by taking non-negative sums of convex functions (starting from affine), and then applying a (point-wise) monotonically-increasing convex activation.
- The filter weights in the orange layers need to be $\geqslant 0$, whereas the blue layers can have any real-valued filters.
- The activation functions are taken to be leaky-ReLU with negative slope 0.2 (convex and monotone).

3. LEARNING THE REGULARIZER

- **Main idea**: If A is invertible, source condition dictates that for any solution x of (1), the following holds:

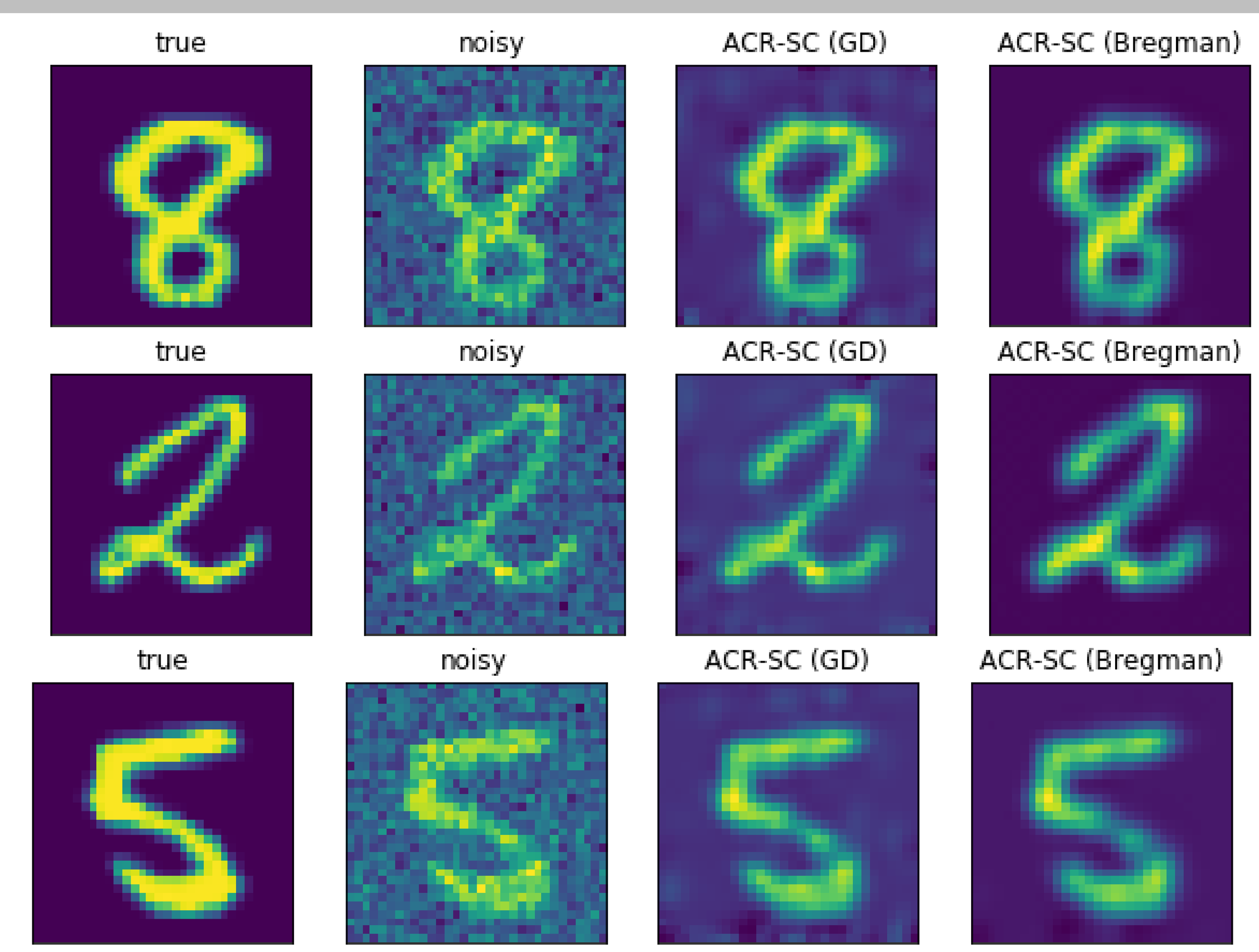
$$\boxed{\ell_{\text{sc}}(x;\theta) = \|(A^*)^{-1}\nabla_x \psi_\theta(x)\| < \infty}$$

- The smaller the quantity $\ell_{\text{sc}}(x;\theta)$ is, the more suitable x would be as a variational solution.
  - Encourages the ground-truth images to the solution of the resulting variational problem.
- For non-invertible A, replace the inverse with the Moore-Penrose pseudo-inverse.
- **Training loss**:

$$\boxed{L(\theta) = \frac{1}{n}\sum_{i=1}^{n} \psi_\theta(x_i) - \frac{1}{n}\sum_{i=1}^{n} \psi_\theta(z_i) + \lambda_{\text{gp}} L_{\text{gp}}(\theta) + \lambda_{\text{sc}} L_{\text{sc}}(\theta)} \quad (4)$$

  - x_i: ensemble of clean images
  - $z_i = A^\dagger y_i^\delta$: ensemble of noisy images
  - L_{gp}: soft gradient penalty to enforce 1-Lipschitz bound on ψ_θ
- Can solve the variational problem with the learned regularizer via sub-gradient or Bregman iterations [4].

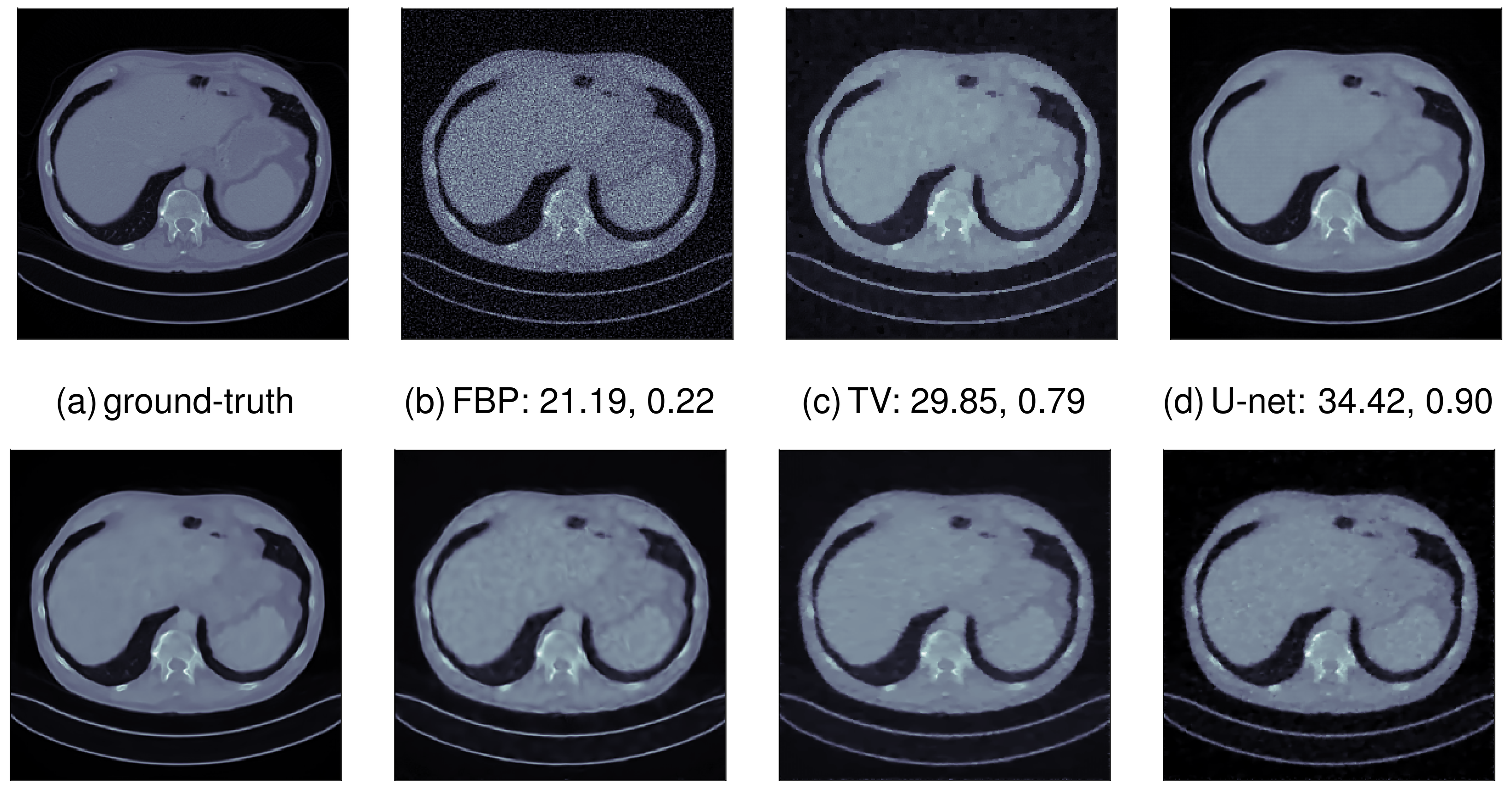
4. NUMERICAL EXAMPLES: DENOISING



- $\lambda_{\text{sc}} = 2.0$, $\lambda_{\text{gp}} = 10.0$
- The average PSNR and SSIM over 100 randomly chosen test images:
  1. noisy: 13.93 ± 0.13 dB, 0.51 ± 0.08
  2. ACR-SC (GD): 22.72 ± 0.64, 0.77 ± 0.04
  3. ACR-SC (Bregman): 20.29 ± 0.88, 0.86 ± 0.03
- Took $\lambda = 25$ for the Bregman technique and $\lambda = 5$ for vanilla gradient-descent.
- Bregman iterations perform better in terms of recovering the contrast while yielding effective denoising.

5. CT RECONSTRUCTION EXPERIMENTS

- **Experiments on Mayo-clinic low-dose CT data (2016)**:
  1. Extracted 2D slices of size 512×512 from 3D scans.
  2. Trained on 9 patients (2250 slices), evaluated on one (128 slices).
  3. Parallel-beam projection, 200 angles, 400 lines/angle, Gaussian noise with $\sigma_e = 2.0$ (25 dB of signal-to-noise ratio in the data space).
  4. No. of learnable parameters in ψ_θ: 590928.
  5. $\lambda_{\text{sc}} = 2.0$ and $\lambda_{\text{gp}} = 5.0$, trained for 10 epochs, batch-size was four. Adam optimizer with $\eta, \beta_1, \beta_2 = 10^{-5}, 0.90, 0.99$.



(a) ground-truth (b) FBP: 21.19, 0.22 (c) TV: 29.85, 0.79 (d) U-net: 34.42, 0.90

(e) LPD: 35.76, 0.92 (f) AR: 33.52, 0.86 (g) ACR: 31.24, 0.86 (h) ACR-SC: 30.93, 0.85

- ACR-SC is only marginally inferior to ACR, while it still outperforms classical model-based methods such as total-variation (TV).

6. SUMMARY

- Developed a novel training loss for learning a data-driven convex regularizer satisfying the source condition.
- Unsupervised learning, no paired data needed.
- Possible to derive convergence rate estimates for the resulting variational problem.
- Theoretical grounding for Bregman iterations.
- Enforcing the source condition does not lead to any significant drop in reconstruction quality.

REFERENCES

1. M. Burger and S. Osher, "Convergence rates of convex variational regularization," Inverse Problems, 2004.
2. B. Amos, L. Xu, and J. Z. Kolter, "Input convex neural networks," ICML, 2017.
3. S. Mukherjee, S. Dittmer, Z. Shumaylov, S. Lunz, O. Öktem, and C.-B. Schönlieb, "Learned convex regularizers for inverse problems," arXiv:2008.02839v2, 2021.
4. M. Benning and M. Burger, "Modern regularization methods for inverse problems," Acta Numerica, 2018.