Adversarial Regularizers in Inverse Problems

Sebastian Lunz[†], Özan Öktem[‡], Carola-Bibiane Schönlieb[†]

[†]University of Cambridge [‡]KTH - Royal Institute of Technology

1 Inverse Problems

- Reconstruct image x from measurement $y = Ax + \epsilon$, A linear operator.
- Difficulty: Applying pseudo-inverse $A^\dagger$ to measurement amplifies noise.
- Examples in medical imaging: Computed Tomography, Magnetic Resonance Imaging, Photo-Acoustic Tomography.

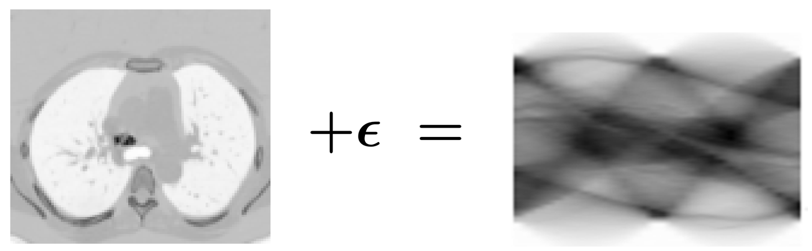
2 The variational approach

- Insert prior knowledge into reconstruction by adding regularization functional R. (Benning&Burger 2018, *Acta Num. 27*) Reconstruct from y by

$$\operatorname{argmin}_x \|Ax - y\| + \lambda R(x)$$

.

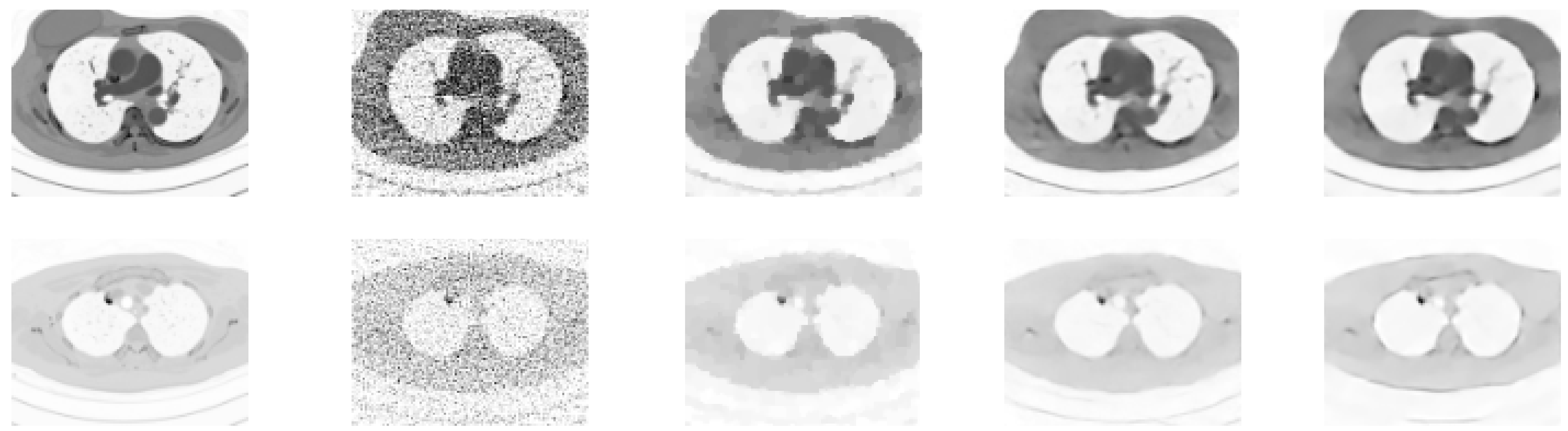
- Popular choice: TV regularization $R = \|\nabla x\|_1$.
- Allows for stability theory and explicit use of the data model, but hand-crafted functionals do not reflect true prior.
- Example of measurements for the Radon transform A used in Computed Tomography.



$A \cdot$ [image] $+\epsilon =$ [sinogram]

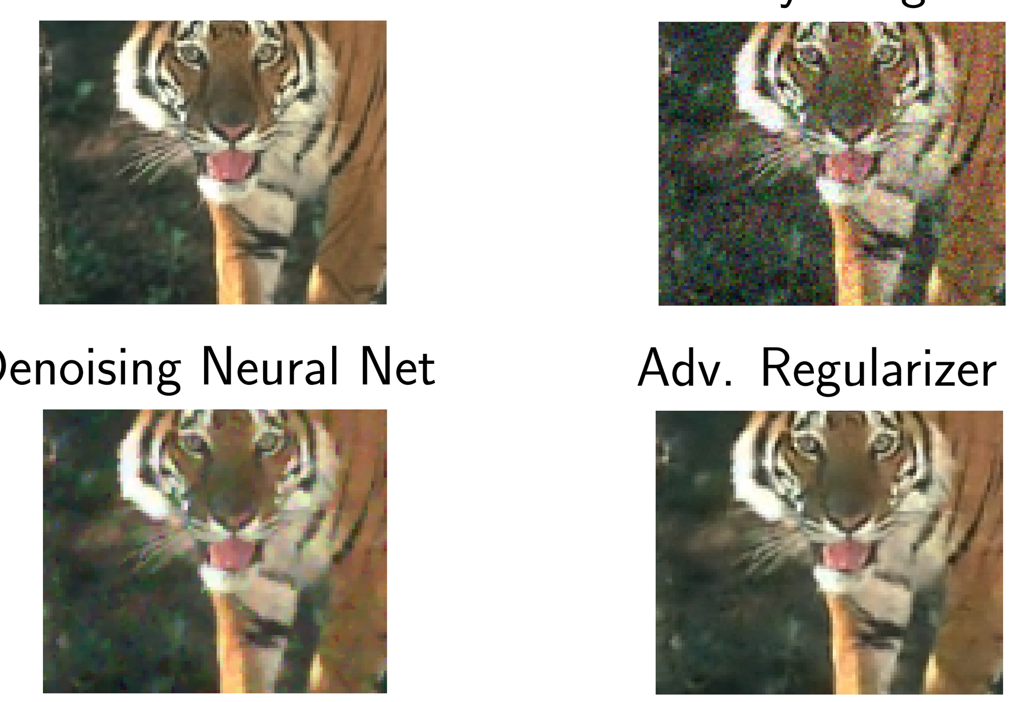
4 Results for Computed Tomography and Denoising

Computed Tomography Reconstruction on LIDC dataset of Lung Scans



Ground truth | Filtered Backprojection | Total Variation | Post-Processing | Adversarial Regularizer

Denoising Results on BSDS



Quantitative Results for Computed Tomography

Method	PSNR (dB)	SSIM
MODEL-BASED		
Filtered Backprojection	14.9	.227
Total Variation	27.7	.890
SUPERVISED		
Post-Processing	31.2	.936
RED	29.9	.904
UNSUPERVISED		
Adversarial Reg.	30.5	.927

5 Distributional Analysis

- Analyze impact of gradient descent over Ψ_Θ on distributions.
- Define $\mathbb{P}_\eta$ to be distribution obtained from $\mathbb{P}_n$ after one step of gradient descent $x \to x - \eta \nabla_x \Psi_\Theta(x)$.
- Assume Ψ_Θ is trained to perfection for $\lambda = \infty$ and $\eta \to \mathrm{Wass}(\mathbb{P}_r, \mathbb{P}_\eta)$ is differentiable at $\eta = 0$, where Wass denotes the Wasserstein distance. Then

$$\frac{d}{d\eta} \mathrm{Wass}(\mathbb{P}_r, \mathbb{P}_\eta)|_{\eta=0} = -\mathbb{E}_{X \sim \mathbb{P}_n}\left[\|\nabla_x \Psi_\Theta(X)\|^2\right].$$

- In particular, if $\|\nabla_x \Psi_\Theta(x)\| = 1$, have explicit control over decay rate of Wasserstein distance.
- The decay in Wasserstein distance induced by Ψ_Θ is the fastest amongst any regularization functional with gradients of unit norm.

6 Analysis and the Data Manifold

- Data Manifold Assumption: Ground truth data $\mathbb{P}_r$ is supported on low-dimensional manifold $\mathcal{M}$.
- Assume projecting noisy data $\mathbb{P}_n$ onto $\mathcal{M}$ gives rise to the distribution of clean images $\mathbb{P}_r$.
- Corresponds to a low-noise assumption in comparison to curvature of Data Manifold.
- Then the optimal critic for $\lambda = \infty$ is given by the distance function to $\mathcal{M}$

$$\Psi_\Theta(x) = \min_{y \in \mathcal{M}} \|x - y\|_2.$$

- Optimal critic can be non-unique. It can in particular be changed arbitrarily outside the support of $\mathbb{P}_r$ and $\mathbb{P}_n$.

3 Learning a Regularization Functional

- Learning regularization functional allows to incorporate prior knowledge on data. Use a neural network Ψ_Θ to parametrize functional.
- Training the neural network by evaluating performance as regularization functional would require solving nested optimization problem (De Los Reyes et al. 2013, *Inv. Prob. & Imag. 7.4*). Computationally unfeasible.
- Instead train the network as a *critic*. Suppress characteristic noise by *discriminating* between distribution of ground truth data $\mathbb{P}_r$ and of unregularized inverse $\mathbb{P}_n$.
- Additional term to enforce a Lipschitz constraint leads to Ψ_Θ parameterizing the Wasserstein distance as in WGANs (Arjovsky et al. 2017, *ICML*; Gulrajani et al. 2017, *NIPS*)

$$\mathbb{E}_{X \sim \mathbb{P}_r}[\Psi_\Theta(X)] - \mathbb{E}_{X \sim \mathbb{P}_n}[\Psi_\Theta(X)] + \lambda \cdot \mathbb{E}\left[(\|\nabla_x \Psi_\Theta(X)\| - 1)_+^2\right].$$

- Once trained, Ψ_Θ is employed to reconstruct from measurement y by solving

$$\operatorname{argmin}_x \|Ax - y\|_2^2 + \lambda \Psi_\Theta(x).$$

The minimization problem is solved using first-order methods like gradient descent.

8 Advantages

- Direct **control over data fidelity** during reconstruction along with **stability theory** allows for qualitative guarantees.
- **Unsupervised data** is sufficient for training, allowing to use real measurement data for training.
- Model-based approach well-suited for **applications in medical imaging**.

7 Stability

- Stability theory for variational problems can be adapted to learned setting. Difference: Regularization functional additionally 1-Lipschitz, but not necessarily bounded below.
- Assumptions: Ψ_Θ lower semi-cts., 1 -Lipschitz, A continuous.
- Either A or Ψ_Θ are coercive.
- Let $y_n \to y$ and let x_n denote x_n corresponding minimizers of variational problem. Then x_n converge to some x weakly and x is a minimizer of

$$\|Ax - y\|_2^2 + \lambda \Psi_\Theta(x).$$

- Conclusion: The solution of the variational problem is weakly stable under changes in the data.