Tomographic Reconstruction and Wavefront Set

Héctor Andrade Loarca
joint with Gitta Kutyniok, Ozan Öktem, Philipp Petersen

AFG Oberseminar

31st of May, 2018

Applied Functional Analysis Group
Technische Universität Berlin

CT Reconstruction

Forward model

Given by the X-ray transform $\mathcal{R} : L^2(\mathbb{R}^2) \longrightarrow L^2(\mathbb{S}^1 \times \mathbb{R})$:

$$g = \mathcal{R}f(\theta, s) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\delta(x\cos\theta + y\sin\theta - s)dxdy$$

where $f \in \mathcal{D}'(\mathbb{R}^2)$.

CT Reconstruction

Forward model

Given by the X-ray transform $\mathcal{R} : L^2(\mathbb{R}^2) \longrightarrow L^2(\mathbb{S}^1 \times \mathbb{R})$:

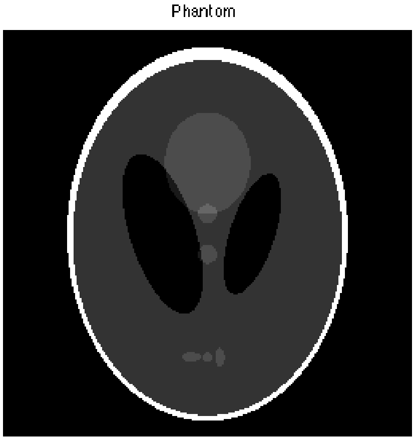
$$g = \mathcal{R}f(\theta, s) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\delta(x\cos\theta + y\sin\theta - s)dxdy$$

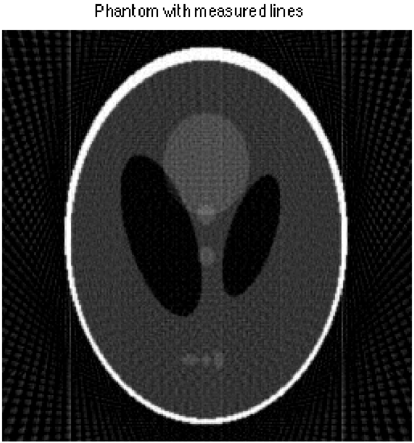
where $f \in \mathcal{D}'(\mathbb{R}^2)$.

Ill-posedness:

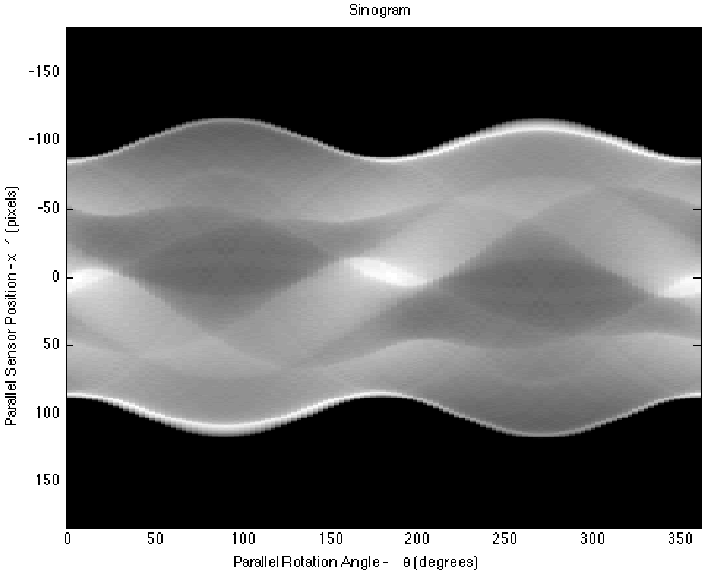
- Filtered back projection (R^{-1}) involves differentiation $\longrightarrow$ increases singularities and noise.
- R^{-1} is unbounded $\longrightarrow$ two far apart images can have very close X-ray transform.

Shepp-Logan phantom





Sinogram


Sinogram
-150
-100
-50
0
50
100
150
Parallel Sensor Position - x ′ (pixels)
0
50
100
150
200
250
300
350
Parallel Rotation Angle - θ (degrees)

Solving inverse problems

First approach

Minimizing the miss-fit against data:

$$\min_f \mathcal{L}(\mathcal{R}(f), g)$$

e.g. $\mathcal{L}(\mathcal{R}(f), g) = ||\mathcal{R}(f) - g||_2^2$. **Downside:** Ill-posedness results on overfitting.

Solving inverse problems

First approach

Minimizing the miss-fit against data:

$$\min_f \mathcal{L}(\mathcal{R}(f), g)$$

e.g. $\mathcal{L}(\mathcal{R}(f), g) = ||\mathcal{R}(f) - g||_2^2$. **Downside:** Ill-posedness results on overfitting.

Approaches to address overfitting?:

- **Knowledge-driven regularization:** Model prescribed beforehand using first principles, data used to calibrate the model.
- **Data-driven regularization:** Model learned from data, without any prior first principles.
- **Hybrid:** Best of both worlds.

Knowledge-driven regularization

Pros ☺

- Guided by first principles (laws encoded by equations), tested and validated independently.
- Not much data required.
- Simple concepts, aiding the understanding.

Knowledge-driven regularization

Pros ☺

- Guided by first principles (laws encoded by equations), tested and validated independently.
- Not much data required.
- Simple concepts, aiding the understanding.

Cons ☹

- Requires explicit description of causal relations, not always a good model exists.
- Not straigtht forward, uncertainty quantification.

Knowledge-driven regularization

Pros ☺

- Guided by first principles (laws encoded by equations), tested and validated independently.
- Not much data required.
- Simple concepts, aiding the understanding.

Cons ☹

- Requires explicit description of causal relations, not always a good model exists.
- Not straigtht forward, uncertainty quantification.

- **Examples:** Analytic pseudoinverse (e.g. FBP), Iterative methods with early stopping (e.g. ART), Variational methods (e.g. TV, ℓ^1).

Data-driven regularization

Pros ☺

- Deep understanding of the problem is not needed, just a lot of data.
- Can capture complicated causal relations without making limiting assumptions.

Data-driven regularization

Pros ☺

- Deep understanding of the problem is not needed, just a lot of data.
- Can capture complicated causal relations without making limiting assumptions.

Cons ☹

- Does not provide any conceptual simplification (not much conceptual understanding is acquired).
- Not easy to incorporate a-priori knowledge.
- Computationally exhaustive.

Data-driven regularization

Pros ☺

- Deep understanding of the problem is not needed, just a lot of data.
- Can capture complicated causal relations without making limiting assumptions.

Cons ☹

- Does not provide any conceptual simplification (not much conceptual understanding is acquired).
- Not easy to incorporate a-priori knowledge.
- Computationally exhaustive.

- **Basic idea:** Parametrized $\mathcal{R}_\theta^\dagger : L^2(\mathbb{S}^1 \times \mathbb{R}) \longrightarrow L^2(\mathbb{R}^2)$ s.t. $\mathcal{R}_\theta^\dagger(g) = f_{\text{true}}\ \forall \theta \in \Theta$ whenever $\mathcal{R}(f_{\text{true}}) = g$. It estimates θ minimizing a loss $L : \Theta \longrightarrow \mathbb{R}_0^+$.

Hybrid methods

Motivation

- If $\mathcal{R}$ is **local** (e.g. deblurring, denoising) $\longrightarrow$ CNN and sufficient known pairs (g, f_{true}) are enough to reconstruct (Jin et al., 2016).
- If $\mathcal{R}$ is **global** (e.g. CT, MRI) $\longrightarrow$ CNN does not work, it becomes unfeasible to work with fully connected layers.

Hybrid methods

Motivation

- If $\mathcal{R}$ is **local** (e.g. deblurring, denoising) $\longrightarrow$ CNN and sufficient known pairs (g, f_{true}) are enough to reconstruct (Jin et al., 2016).
- If $\mathcal{R}$ is **global** (e.g. CT, MRI) $\longrightarrow$ CNN does not work, it becomes unfeasible to work with fully connected layers.

Possible solutions

- **Learned post-processing:** First an initial (not learned) reconstruction (e.g. FBP), and denoise with CNN (Chen et al., 2017).
- **Learned regularizer:** Learn a regularization functional (e.g. dictionary learning) and perform variational regularization (Xu et al. 2012).
- **Learned iterative schemes:** Using as model a classical optimization iterative method and learn the best update in each iteration using a-priori information (Öktem et al. 2017).

Primal-dual algorithm

Minimization problem:

$$\min_{f \in L^2(\mathbb{R}^2)} \mathcal{L}(\mathcal{R}(f), g) + \mathcal{S}(f)$$

Primal-dual algorithm

Minimization problem:

$$\min_{f \in L^2(\mathbb{R}^2)} \mathcal{L}(\mathcal{R}(f), g) + \mathcal{S}(f)$$

Algorithm 2 Non-linear primal-dual algorithm

Given $\sigma, \tau > 0$ s.t. $\sigma\tau||\mathcal{R}|| < 1$, $\gamma \in [0,1]$ and $f_0 \in L^2(\mathbb{R}^2)$, $h_0 \in L^2(\mathbb{S}^1 \times \mathbb{R})$:
for $i = 1, \dots, I$ **do**
  $h_{i+1} \longleftarrow \mathrm{prox}_{\sigma\mathcal{L}}(h_i + \sigma\mathbb{R}(\overline{f}_i))$
  $f_{i+1} \longleftarrow \mathrm{prox}_{\tau\mathcal{S}}(f_i - \tau[\partial\mathbb{R}(f_i)]^*(h_{i+1}))$
  $\overline{f}_{i+1} \longleftarrow f_{i+1} + \gamma(f_{i+1} - f_i)$
end for

Primal-dual algorithm

Minimization problem:

$$\min_{f \in L^2(\mathbb{R}^2)} \mathcal{L}(\mathcal{R}(f), g) + \mathcal{S}(f)$$

Algorithm 3 Non-linear primal-dual algorithm

1: Given $\sigma, \tau > 0$ s.t. $\sigma\tau||\mathcal{R}|| < 1$, $\gamma \in [0,1]$ and $f_0 \in L^2(\mathbb{R}^2)$, $h_0 \in L^2(\mathbb{S}^1 \times \mathbb{R})$:
2: **for** $i = 1, \ldots, I$ **do**
3: $\quad h_{i+1} \longleftarrow \mathrm{prox}_{\sigma\mathcal{L}}(h_i + \sigma\mathbb{R}(\overline{f}_i))$
4: $\quad f_{i+1} \longleftarrow \mathrm{prox}_{\tau\mathcal{S}}(f_i - \tau[\partial\mathbb{R}(f_i)]^*(h_{i+1}))$
5: $\quad \overline{f}_{i+1} \longleftarrow f_{i+1} + \gamma(f_{i+1} - f_i)$
6: **end for**

Proximal operator:

$$\mathrm{prox}_{\tau\mathcal{S}}(f) = \underset{f' \in L^2(\mathbb{R}^2}{\mathrm{argmin}} \left[\mathcal{S}(f') + \frac{1}{2\tau}||f' - f||_2^2 \right]$$

Learned primal-dual algorithm (Öktem et al. 2017)

Algorithm 4 Learned primal-dual algorithm

Given $f_0 \in L^2(\mathcal{R}^2)$, $h_0 \in L^2(\mathcal{S}^1 \times \mathbb{R})$
for $i = 1, \dots, I$ **do**
    $h_{i+1} \longleftarrow \Gamma_{\theta^d_{i+1}}(h_i, \mathcal{R}(f_i), g)$
    $f_{i+1} \longleftarrow \Lambda_{\theta^p_{i+1}}(f_i, [\partial\mathcal{R}(f_i)]^*(h_{i+1}))$
end for

Learned primal-dual algorithm (Öktem et al. 2017)

Algorithm 5 Learned primal-dual algorithm

1: Given $f_0 \in L^2(\mathcal{R}^2)$, $h_0 \in L^2(\mathcal{S}^1 \times \mathbb{R})$
2: **for** $i = 1, \ldots, I$ **do**
3: $\quad h_{i+1} \longleftarrow \Gamma_{\theta^d_{i+1}}(h_i, \mathcal{R}(f_i), g)$
4: $\quad f_{i+1} \longleftarrow \Lambda_{\theta^p_{i+1}}(f_i, [\partial\mathcal{R}(f_i)]^*(h_{i+1}))$
5: **end for**

Primal and dual operators: Γ_{θ^d} and Λ_{θ^p} are learned CNN-ResNets of the form:

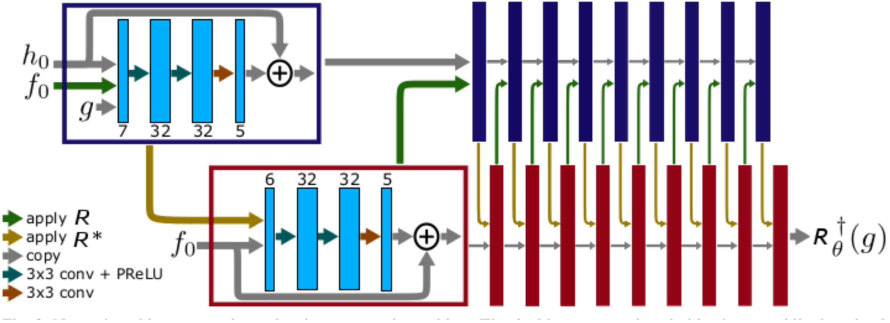
$$Id + \mathcal{W}_{w_d,b_d} \circ \mathcal{A}_{c_d} \circ \ldots \circ \mathcal{W}_{w_2,b_2} \circ \mathcal{A}_{c_2} \circ \mathcal{W}_{w_1,b_1} \circ \mathcal{A}_{c_1}$$

where,

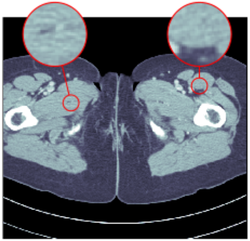
$$\mathcal{W}_{w_i,b_i}(f') = b_i + w_i * f'$$

$$\mathcal{A}_{c_i}(x) = \mathrm{PReLU}(x) = \begin{cases} x & \text{if } x \geq 0 \\ -c_i & \text{else} \end{cases}$$

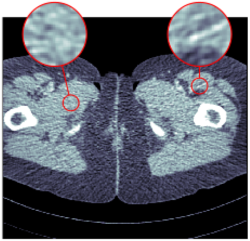
Architecture



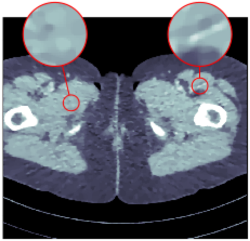
Benchmarks (Adler,Öktem)



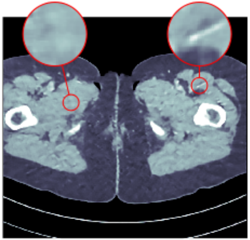
(a) 512 × 512 pixel human phantom



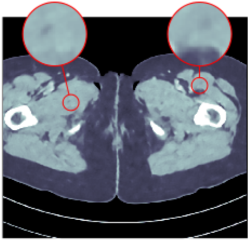
(b) Filtered back-projection (FBP)
PSNR 33.65 dB, SSIM 0.830, 423 ms



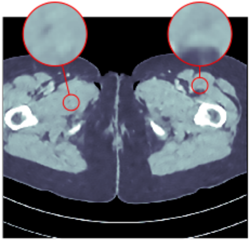
(c) Total variation (TV)
PSNR 37.48 dB, SSIM 0.946, 64 371 ms



(d) FBP + U-Net denoising
PSNR 41.92 dB, SSIM 0.941, 463 ms



(e) Primal-Dual, linear
PSNR 44.10 dB, SSIM 0.969, 620 ms



(f) Primal-Dual, non-linear
PSNR 43.91 dB, SSIM 0.969, 670 ms

Wavefront set as extra information

Definition (N-Wavefront set)

Let $N \in \mathbb{R}$ and f a distribution on $\mathbb{R}^2$. We say $(x, \lambda) \in \mathbb{R}^2 \times \mathbb{R}^2$ is a $N-$regular directed point if there exists a nbd. of U_x of x, a smooth cutoff function Φ with $\Phi \equiv 1$ on U_x and a nbd. V_λ of λ such that:

$$(\Phi f)^\wedge(\eta) = O((1 - |\eta|)^{-N}) \quad \text{for all} \quad \eta = (\eta_1, \eta_2) \quad \text{such that} \quad \frac{\eta_2}{\eta_1} \in V_\lambda$$

The $N-$Wavefront set $WF^N(f)$ is the complement of the $N-$regular directed point. The Wavefront Set $WF(f)$ is defined as

$$WF(f) = \cup_{N>0} WF^N(f)$$

Wavefront set as extra information

Definition (N-Wavefront set)

Let $N \in \mathbb{R}$ and f a distribution on $\mathbb{R}^2$. We say $(x, \lambda) \in \mathbb{R}^2 \times \mathbb{R}^2$ is a $N-$regular directed point if there exists a nbd. of U_x of x, a smooth cutoff function Φ with $\Phi \equiv 1$ on U_x and a nbd. V_λ of λ such that:

$$(\Phi f)^\wedge(\eta) = O((1 - |\eta|)^{-N}) \quad \text{for all} \quad \eta = (\eta_1, \eta_2) \quad \text{such that} \quad \frac{\eta_2}{\eta_1} \in V_\lambda$$

The $N-$Wavefront set $WF^N(f)$ is the complement of the $N-$regular directed point. The Wavefront Set $WF(f)$ is defined as

$$WF(f) = \cup_{N>0} WF^N(f)$$

- How can we compute the wavefront set?
- How can we incorporate it in the learned primal-dual reconstruction?

Answer: Continuous Shearlet frames and Canonical relation

Classical Shearlet Transform (Guo, Kutyniok, Labate; 2006)

$$\langle f, \psi_{a,s,t} \rangle = \int_{\mathbb{R}^2} f(x) \overline{\psi_{a,s,t}(x)} dx$$

where

$$\mathcal{SH}(\psi) = \{\psi_{a,s,t}(x) := a^{-3/4} \psi(S_s A_a x - t) : (a,s,t) \in \mathbb{R}_+ \times \mathbb{R} \times \mathbb{R}^2\}$$

and

$$A_a := \begin{pmatrix} a^1 & 0 \\ 0 & a^{1/2} \end{pmatrix} \quad S_s := \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}$$

Resolution of the WF set with Shearlets

Theorem (Grohs, 2011)

Let ψ be a Schwartz function with infinitely many vanishing moments in x_2-direction and $\{\psi_{a,s,t}\}$ a CSF. Let f be a tempered distribution and $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2$, where

$$\mathcal{D}_1 = \{(t_0, s_0) \in \mathbb{R}^2 \times [-1,1] | \textit{for } (s,t) \textit{ in a neighbourhood } U \textit{ of } (s_0,t_0), \\ |\mathcal{SH}_\psi f(a,s,t)| = O(a^k), \forall k \in \mathbb{N}\}$$

and

$$\mathcal{D}_2 = \{(t_0, s_0) \in \mathbb{R}^2 \times [1,\infty) | \textit{for } (1/s,t) \textit{ in a neighbourhood } U \textit{ of } (s_0,t_0), \\ |\mathcal{SH}_{\tilde{\psi}} f(a,s,t)| = O(a^k), \forall k \in \mathbb{N}\}$$

Then,

$$WF(f) = \mathcal{D}^c$$

Canonical relation (Öktem, Quinto; 2008)

Let $f \in D'(\mathbb{R}^2)$, and assume $\mathbb{R}(\theta, s)$ is given on an open set $U \subset [0, \pi) \times \mathbb{R}$. Let $(\theta_0, s) \in U$, let

$$\xi_0 = \xi_s e_s + \xi_\sigma \sigma(\theta_0) \perp (\cos(\theta_0), \sin(\theta_0))$$

Moreover, let $x_0 \in \ell(\theta_0, s)$. Then $(x_0, \xi_0 dx) \in WF(f)$ if and only if

$$((\theta_0, s), (\xi_\sigma x \cdot \omega(\theta_0))d\theta + \xi_s ds) \in WF(\mathcal{R}(f))$$

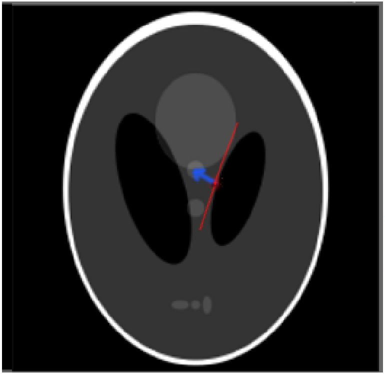
Canonical relation (Öktem, Quinto; 2008)

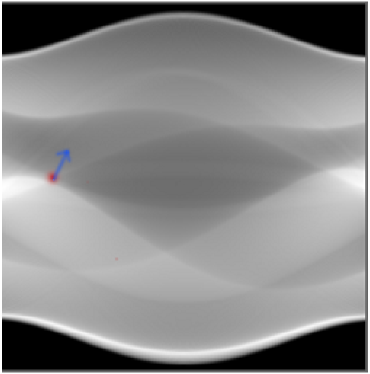
Let $f \in D'(\mathbb{R}^2)$, and assume $\mathbb{R}(\theta, s)$ is given on an open set $U \subset [0, \pi) \times \mathbb{R}$. Let $(\theta_0, s) \in U$, let

$$\xi_0 = \xi_s e_s + \xi_\sigma \sigma(\theta_0) \perp (\cos(\theta_0), \sin(\theta_0))$$

Moreover, let $x_0 \in \ell(\theta_0, s)$. Then $(x_0, \xi_0 dx) \in WF(f)$ if and only if

$$((\theta_0, s), (\xi_\sigma x \cdot \omega(\theta_0))d\theta + \xi_s ds) \in WF(\mathcal{R}(f))$$





Shearlets on the sinogram

- Let $\{\psi_{a,s,t}\}_{(a,s,t)\in\mathbb{R}^+\mathbb{R}\times\mathbb{R}^2}$ continuous compactly compactly supported shearlet, with support on $[0,2\pi)\times\mathbb{R}$, that form a frame of $L^2([0,\pi),\mathbb{R})$.

  Then, the system $\{\tilde{\psi}_{a,s,t}\}_{(a,s,t)\in\mathbb{R}^+\mathbb{R}\times([0,\pi)\times\mathbb{R})}$ given by

$$\tilde{\psi}_{a,s,t}(\theta,s):=\sum_{n\in\mathbb{Z}}\psi_{a,s,t}(\theta+n\pi,s)$$

  is a Continuous Shearlet frame for $L^2(\mathbb{S}^1\times\mathbb{R})$.

Shearlets on the sinogram

- Let $\{\psi_{a,s,t}\}_{(a,s,t)\in\mathbb{R}^+\mathbb{R}\times\mathbb{R}^2}$ continuous compactly compactly supported shearlet, with support on $[0,2\pi)\times\mathbb{R}$, that form a frame of $L^2([0,\pi),\mathbb{R})$.

  Then, the system $\{\tilde{\psi}_{a,s,t}\}_{(a,s,t)\in\mathbb{R}^+\mathbb{R}\times([0,\pi)\times\mathbb{R})}$ given by

$$\tilde{\psi}_{a,s,t}(\theta,s) := \sum_{n\in\mathbb{Z}} \psi_{a,s,t}(\theta+n\pi,s)$$

  is a Continuous Shearlet frame for $L^2(\mathbb{S}^1\times\mathbb{R})$.

- We have all the tools!

Modified learned primal-dual

Modified problem:

$$\min_{f \in L^2(\mathbb{R}^2)} \mathcal{L}(\mathcal{R}(SH_\psi^{-1}\tilde{f}), g) \quad \text{s.t.} \ \mathbf{C}(WF(SH_\psi^{-1}\tilde{f}))) = WF(g)$$

Algorithm 6 Modified learned primal-dual algorithm

1: Given $\tilde{f}_0 \in L^2(\mathbb{R}^2)$, $h_0 \in L^2(\mathbb{S}^1 \times \mathbb{R})$
2: **for** $i = 1, \ldots, I$ **do**
3: $\quad h_{i+1} \longleftarrow \Gamma_{\theta^d_{i+1}}(h_i, \mathcal{R}(SH_\psi^{-1}\tilde{f}_i)), g)$
4: $\quad \tilde{f}_{i+1} \longleftarrow \Lambda_{\theta^p_{i+1}}(f_i, [\partial\mathcal{R}(SH_\psi^{-1}\tilde{f}_i))]^*(h_{i+1}))$
5: $\quad \mathbf{C}(WF(SH_\psi^{-1}\tilde{f}_{i+1}))) = WF(h_{i+1})$
6: **end for**

Modified learned primal-dual

Modified problem:

$$\min_{f \in L^2(\mathbb{R}^2)} \mathcal{L}(\mathcal{R}(SH_\psi^{-1}\tilde{f}), g) \quad \text{s.t. } \mathbf{C}(WF(SH_\psi^{-1}\tilde{f})) = WF(g)$$

Algorithm 7 Modified learned primal-dual algorithm

1: Given $\tilde{f}_0 \in L^2(\mathbb{R}^2)$, $h_0 \in L^2(\mathbb{S}^1 \times \mathbb{R})$
2: **for** $i = 1, \ldots, I$ **do**
3: $\quad h_{i+1} \longleftarrow \Gamma_{\theta^d_{i+1}}(h_i, \mathcal{R}(SH_\psi^{-1}\tilde{f}_i)), g)$
4: $\quad \tilde{f}_{i+1} \longleftarrow \Lambda_{\theta^p_{i+1}}(f_i, [\partial\mathcal{R}(SH_\psi^{-1}\tilde{f}_i))]^*(h_{i+1}))$
5: $\quad \mathbf{C}(WF(SH_\psi^{-1}\tilde{f}_{i+1})) = WF(h_{i+1})$
6: **end for**

Are we done?

Modified learned primal-dual

Modified problem:

$$\min_{f \in L^2(\mathbb{R}^2)} \mathcal{L}(\mathcal{R}(SH_{\psi}^{-1}\tilde{f}), g) \quad \text{s.t. } \mathbf{C}(WF(SH_{\psi}^{-1}\tilde{f}))) = WF(g)$$

Algorithm 8 Modified learned primal-dual algorithm

1: Given $\tilde{f}_0 \in L^2(\mathbb{R}^2)$, $h_0 \in L^2(\mathbb{S}^1 \times \mathbb{R})$
2: **for** $i = 1, \ldots, I$ **do**
3: $h_{i+1} \longleftarrow \Gamma_{\theta^d_{i+1}}(h_i, \mathcal{R}(SH_{\psi}^{-1}\tilde{f}_i)), g)$
4: $\tilde{f}_{i+1} \longleftarrow \Lambda_{\theta^p_{i+1}}(f_i, [\partial\mathcal{R}(SH_{\psi}^{-1}\tilde{f}_i))]^*(h_{i+1}))$
5: $\mathbf{C}(WF(SH_{\psi}^{-1}\tilde{f}_{i+1})))=WF(h_{i+1})$
6: **end for**

Are we done? **Of course not!. Where is the code?!**

Implementation

- Learned primal dual on the Shearlet coefficients. ✓

Implementation

- Learned primal dual on the Shearlet coefficients. ✓
- Wavefront set computation:

Implementation

- Learned primal dual on the Shearlet coefficients. ✓
- Wavefront set computation:
  - Theory works on continuous distributions.

Implementation

- Learned primal dual on the Shearlet coefficients. ✓
- Wavefront set computation:
  - Theory works on continuous distributions.
  - One needs infinite number of shearlet coefficients.

Implementation

- Learned primal dual on the Shearlet coefficients. ✓
- Wavefront set computation:
  - Theory works on continuous distributions.
  - One needs infinite number of shearlet coefficients.
  - There is no faithful digital Wavefront set version (Petersen, 2018).

Implementation

- Learned primal dual on the Shearlet coefficients. ✓
- Wavefront set computation:
  - Theory works on continuous distributions.
  - One needs infinite number of shearlet coefficients.
  - There is no faithful digital Wavefront set version (Petersen, 2018).
  - Learned Wavefront set extractor (in progress).

Implementation

- Learned primal dual on the Shearlet coefficients. ✓
- Wavefront set computation:
  - Theory works on continuous distributions.
  - One needs infinite number of shearlet coefficients.
  - There is no faithful digital Wavefront set version (Petersen, 2018).
  - Learned Wavefront set extractor (in progress).
- **Lesson:** Don't assume the continuous theory will work in the computer ☹.

Thanks!

Questions?