Tomographic Reconstruction and Wavefront Set

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

AFG Oberseminar

31st of May, 2018

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 $\mathcal{R} : L^2(\mathbb{R}^2) \longrightarrow L^2(\mathbb{S}^1 \times \mathbb{R})$:

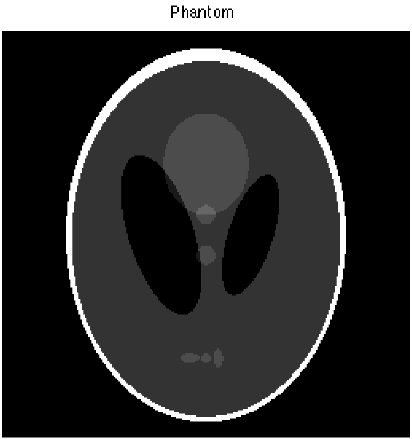
$$g = \mathcal{R}f(\theta, s) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\delta(x \cos\theta + y \sin\theta - s)dxdy$$

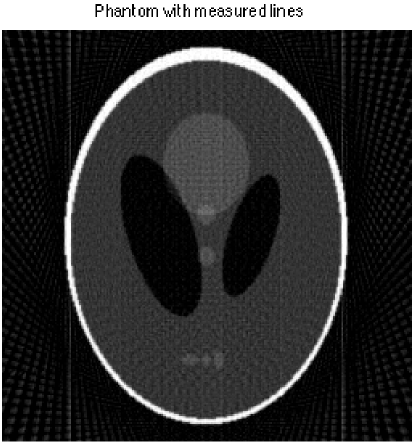
where $f \in \mathcal{D}'(\mathbb{R}^2)$.

Ill-posedness:

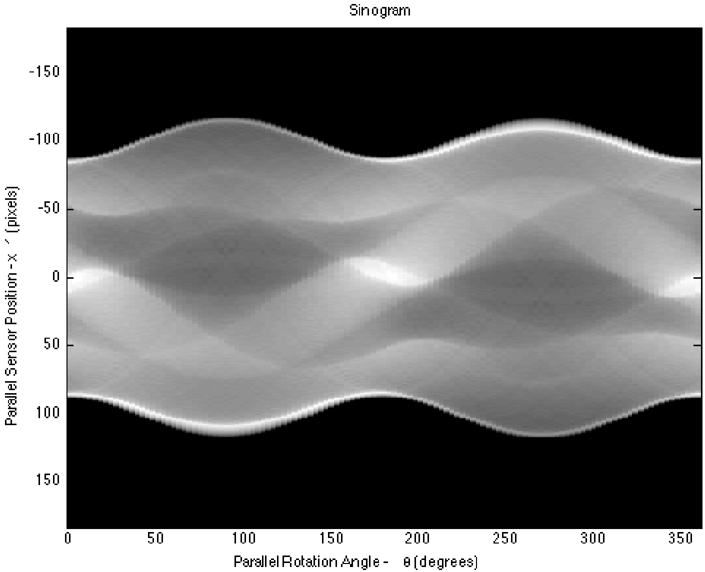
- Filtered back projection (R^{-1}) involves differentiation $\longrightarrow$ increases singularities and noise.
- R^{-1} is unbounded $\longrightarrow$ two far apart images can have very close X-ray transform.

Shepp-Logan phantom





Sinogram


Sinogram
Parallel Sensor Position - x ′ (pixels)
-150
-100
-50
0
50
100
150
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.

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 $L : \Theta \longrightarrow \mathbb{R}_0^+$.

Learning comes into play

One could ask to learn a pseudo-inverse $\mathcal{T}_\Theta(g) \approx f_{\text{true}}$ and learn the parameters $\Theta \in Z$ through a loss functional.

Learning comes into play

One could ask to learn a pseudo-inverse $\mathcal{T}_\Theta(g) \approx f_{\text{true}}$ and learn the parameters $\Theta \in Z$ through a loss functional.

Fully learned method are very dependent on the training set, one would like to combine a partially learned iterative method with known information of the problem (e.g. wavefront set).

Learning comes into play

One could ask to learn a pseudo-inverse $\mathcal{T}_\Theta(g) \approx f_{\text{true}}$ and learn the parameters $\Theta \in Z$ through a loss functional.

Fully learned method are very dependent on the training set, one would like to combine a partially learned iterative method with known information of the problem (e.g. wavefront set).

- If $\mathcal{T}$ is local (e.g. deblurring problem) $\longrightarrow$ convolutional neural network and known pairs (g, f_{true}).
- If $\mathcal{T}$ is global (e.g. Radon transform) $\longrightarrow$ CNN does not work, it becomes unfeasible to work with NN with fully connected layers.

Alternative solutions

1. Recast to image-to-image problem: perform some initial (non machine-learning) reconstruction (e.g. FBP), and then use standard CNN to denoise the initial reconstruction. Upside: it outperforms previous state of the art methods. Donwside: it does not give you more information than using just non-machine learning reconstruction.

Alternative solutions

1. Recast to image-to-image problem: perform some initial (non machine-learning) reconstruction (e.g. FBP), and then use standard CNN to denoise the initial reconstruction. Upside: it outperforms previous state of the art methods. Donwside: it does not give you more information than using just non-machine learning reconstruction.
2. Incorporate enough a-priori information to make the problem tractable and learn the rest (Learned Primal-dual algorithm): First use CNN to update the data (dual step), then apply $\mathcal{T}^*$ and use the result as input to another neural network which updates the reconstruction (primal step), then apply $\mathcal{T}$ and use it as input to a neural network that updates the data, and so on. Upside: it separates the global aspect of the problem into the forward model and its adjoint and only need to learn local aspects. Downside: to train the NN one needs to perform back-propagation through this NN severak times.

Wavefront set as extra information

Definition (N-Wavefront set)

Let $N \in \mathbb{R}$ and f a distribution on $\mathbb{R}^2$. We say $(x,\lambda)\mathbb{R}^2 \times \mathbb{R}^2$ 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

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

Question? How can one incorporate extra information from the $N-$Wavefront set of an image by knowing just its Radon Transform.

Answer: Canonical shearlet transform of the sinogram

Classical Shearlet Transform

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

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

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

Theorem (Resolution of the Wavefront set by continuous shearlet frames; Grohs, 2011)

Let Ψ be a Schwartz function with infinitely many vanishing moments in x_2-direction. Let f be a tempered distribution and $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2$, where
$\mathcal{D}_1 =\{(t_0, s_0) \in \mathbb{R}^2 \times [-1, 1]$ *: for* (s, t) *in a nbd.* U *fo* (s_0, t_0)*,*
$|\mathcal{SH}_\psi f(a, s, t)| = O(a^k)$ *for all* $k \in \mathbb{N}$*, with the implied constant uniform over* $U\}$ *and*
$\mathcal{D}_2=\{ (t_0, s_0) \in \mathbb{R}^2 \times (1, \infty]$ *: for* $(1/s, t)$ *in a nbd.* U *of* (s_0, t_0)*,*
$|\mathcal{SH}_{\psi^\nu} f(a, s, t)| = O(a^k)$ *for all* $k \in \mathcal{N}$*, with the implied constant uniform over* $U\}$*. Then*

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

Theorem (Resolution of the Wavefront set by continuous shearlet frames; Grohs, 2011)

Let Ψ *be a Schwartz function with infinitely many vanishing moments in* x_2-*direction. Let* f *be a tempered distribution and* $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2$, *where*
$\mathcal{D}_1 = \{(t_0, s_0) \in \mathbb{R}^2 \times [-1, 1]$: *for* (s, t) *in a nbd.* U *fo* (s_0, t_0),
$|\mathcal{SH}_\psi f(a, s, t)| = O(a^k)$ *for all* $k \in \mathbb{N}$, *with the implied constant uniform over* $U\}$ *and*
$\mathcal{D}_2 = \{ (t_0, s_0) \in \mathbb{R}^2 \times (1, \infty]$: *for* $(1/s, t)$ *in a nbd.* U *of* (s_0, t_0),
$|\mathcal{SH}_{\psi^\nu} f(a, s, t)| = O(a^k)$ *for all* $k \in \mathcal{N}$, *with the implied constant uniform over* $U\}$. *Then*

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

Theorem (O. Ötkem et al., 2008)

Broadly speaking, a point on the $N-$*Wavefront set of a distribution corresponds to a point on the* $N + 1/2-$*Wavefront set of its Radon transform, with the corresponding directions.*

Only thing left: Shearlets on the sinogram

- Using results of compactly supported shearlets and shearlets on bounded domains, one can construct a shearlet frame on the space of the sinogram $L^2_{x_1-2\pi}([0,2\pi)\times\mathbb{R})$, given by

$$\psi^{x_1-2\pi}_{a,s,t}(x_1,x_2) := \sum_{\ell\in\{-1,0,1\}} \psi_{a,s,t}(x_1+2\pi\ell, x_2)$$

  where $\psi_{a,s,t}$ is a shearlet compactly supported on $[0,2\pi)\times\mathbb{R}$, whose corresponding system form a frame for $L^2([0,2\pi)\times\mathbb{R})$.

Only thing left: Shearlets on the sinogram

- Using results of compactly supported shearlets and shearlets on bounded domains, one can construct a shearlet frame on the space of the sinogram $L^2_{x_1-2\pi}([0,2\pi)\times\mathbb{R})$, given by

$$\psi^{x_1-2\pi}_{a,s,t}(x_1,x_2) := \sum_{\ell\in\{-1,0,1\}} \psi_{a,s,t}(x_1+2\pi\ell, x_2)$$

  where $\psi_{a,s,t}$ is a shearlet compactly supported on $[0,2\pi)\times\mathbb{R}$, whose corresponding system form a frame for $L^2([0,2\pi)\times\mathbb{R})$.
- Then in the learned primal-dual algorithm one can incorporate as extra information the $N-$Wavefront set of the image by pulling back the $N+1/2-$Wavefront set captured by the proposed shearlet frame. This will let us to get a solution with minimum lost of the important features of the images.

Thanks!

Questions?