

**Notes on presentation: Learned Tomographic Reconstruction and Wavefront set.**

**Slide 1: Presentation.**

- Contributions by Philipp, Gitta and Ozan. and side contribution of Jonas Adler who has helped me with the training part.

**Slide 2: Slide 3 and 4.**

- Show the forward model, and mention that  $g$  is known as the sinogram in this parametrization. And what you mean by distributions (dual of smooth distributions).
- Ill posedness of the problem: Filtered back projection, involves differentiation, increases singularities and noise. Filtered back projection is unbounded, two far apart images can have very close Radon transform.

**Slide 5: Solving inverse problems.**

- Classical approach when it is not straight forward to compute the inverse. Missfit against data, minimizing an affine transform of the negative loglikelihood. Best probabilistic estimator. Typically use the Square error.  
**Downside:** Ill posed problems typically.
- How to address overfitting, known as regularization theory: You need to rewrite the problem into a more regular one, different methods:
  - Knowledge-driven regularization, uses first principles, and use data to calibrated the parameters of the model (incorporates extra-information).
  - Data-driven regularization: Model learned from data, using methods on statistical estimation.
  - Hybrid, take the best parts from both.

**Slide 6: Knowledge-driven regularization.**

- Ups:
  - Guided by first principles (developed along the history by great scientists), tested and validated independently.
  - Not much data required.
  - Simple concepts, aiding understanding.
- Downs:
  - Requires explicit description of causal relations, not always a good model exists.
  - Not straight forward, uncertainty quantification.

- Examples: Analytic pseudoinverse (good when there is a mollifier in the space, e.g. FBP), Iterative methods with early stopping (e.g. ART, algebraic reconstruction technique, that uses the so called Kaczmarz method for solving the related linear equation); and Variational methods that enforces the preservation of certain a-priori information using a regularization functional (e.g. TV,  $l_1$ ,  $\dots$ , where you impose certain sparsity under some representation).

#### Slide 7: Data-driven regularization.

- Pros:
  - Deep understanding of the problem is not needed, just a lot of sufficiently informative data.
  - Can capture complicated causal relation without making any limiting assumption (so in that sense is less reductionist).
- Cons:
  - Does not provide any conceptual simplification (no much understanding is acquired).
  - Not easy to incorporate a priori knowledge, so one cannot guarantee a reconstruction that make sense.
  - Computationally exhaustive. (and one needs a lot of data to have a reliable solution)
- Basic idea: Parametrized a sequence of pseudoinverse and find the best parameter estimate by minimizing a statistical error against training data.
- Fully trained methods are very dependent on the training set (the sampling set needs to be according to the space, that sometimes is too big and complex in the scientific applications). One can address this issues by combining partially learned methods with a-priori information coming from the

#### Slide 8: Hybrid methods

- Motivation:
  - If the problem is local (like deblurring or denoising), one can use CNN and sufficient training data to reconstruct the image.
  - If the problem is global (like CT or MRI) one cannot use CNN and it becomes unfeasible to use fully connected layers.
- Possible solutions:
  - Learned post-processing: Perform an initial (not learned) reconstruction and then improve the reconstruction (denoise) using CNN (downside it does not give you extra important information than the first approach of non learned reconstruction).

- Learned regularizer: Learn a regularization functional (using for example dictionary learning) and perform variational regularization. (very hard to learn the regularizer, since one needs already an ansatz).
- Learned iterative schemes: Using as model a classical optimization iterative method and learn the best update in each iteration using a-priori information. (in order to get more information than the current reconstructions methods, one needs to work directly from raw data we are gonna centered on than).

**Slide 9: Primal-dual algorithm.**

- One uses the dual problem and minimizes alternatively the dual and the primal problem, the minimization direction replaces the gradient with the proximal operator (for nonsmooth cases, and is the gradient in the smooth case).

**Slide 10: Learned primal-dual algorithm.**

- The idea here is to use as a priori information the forward operator (relation between the data and the ground truth), we want to learn from the data, and the most powerful prior information we have is the forward operator  $\mathbf{R}$ , but this just goes from images to data, we need something that goes from data to images, and this is the adjoint operator, this is reflected in the learned primal dual which is inspired (although is not the same principle) in the proximal primal dual algorithm.
- One first uses a CNN to update the data (dual step), then applies the dual and use the result as input to another CNN which updates the reconstruction (primal step), and then apply the transform and update it using another NN and so on, normally with 10 times of this procedure is enough.
- This can be trained with initial raw data and reconstruction.
- The learned primal dual operators are given by the sequential application of an affine operator defined by convolutional filters  $\mathbf{w}$ , and biases  $\mathbf{b}$ , and a nonlinearity given by Parametric Rectified Linear Units (PReLU).
- The good thing about this is that one separates the global aspects of the problem into the forward model and its adjoint and only needs to learn the local aspects (the forward operator and its adjoint make sure that the global features get reconstructed), the downside is that one needs to perform back propagation through this neural network that among other things, contains 10 calls to the forward operator and 10 calls to the adjoint operator and 20 small neural networks in between.
- The improvements in benchmarking are also the runtime which is comparable with a normal FBP and 100 times faster than a TV of similar PSNR,

of course it uses way more parameter therefore it requires more memory (gpu).

- Oektem et al. learned using ellipses models (as Shepp Locan phantom) and Mayo dataset of human phantoms.
- Now, one still wants to improve the reconstruction, and no more things can be done from the learning part, one should incorporate more information about the ground truth. One very important information that one would like to recover is the wavefront set.

#### **Slide 11: Architecture**

- Show the bimodal ResNet architecture with convolutional sublayers.

#### **Slide 12: Benchmarks**

- Show some results given by Ozan and Jonas, and mention the time of processing.
- The question is, can we add more a-priori information in our procedure, which one is important in CT, how to add it.

#### **Slide 13: Can we add more a-priori information.**

- The CT reconstruction community have used the Wavefront set as a powerful theoretical tool to characterize directions that can be reconstructed faithfully and directions that cannot.
- If you are able to reconstruct the wavefront set exactly, you dont need to care anymore of weird artifacts that are not in it.
- But the wavefront set is a very abstract concept, that characterizes the singularities of a distribution and the orientation of this singularities. It is characterize by decay rato of the a localized Fourier transform in some specific direction.
- How can we incorporate information of the WF set

#### **Slide 14: Answer: Continuous Shearlet frames and canonical relation**

- The Wavefront set is a abstract concept that cannot be easily computed. The direction resolution of the Shearlet frames let us do the work.
- Introduce the Classical Shearlet transform, that everybody knows.

#### **Slide 15: Resolution of the WF set with Shearlets**

- Theorem of the Resolution of the wavefront set using continuous shearlet frames.
- Positions and directions where the shearlet coefficients decay rapidly are regular directed points.
- How can we then compute the wavefront set of a distribution if we just know its X-ray transform: Canonical relation.

#### **Slide 16:**

- Show the canonical relation: Explain, if we have a point in the singular domain with certain orientation, there is a point (given by that relation) in the image domain in the line associated with the point in the sinogram will be in the singular support and its orientation will be orthogonal to the orientation in the sinogram.
- How to compute the wavefront set in the sinogram if we don't have a continuous shearlet frame in the sinogram? we construct it

#### **Slide 17: Shearlets on the sinogram.**

- Present the form as a periodic expansion of compactly supported shearlet frames in the angle direction, say that it also forms a Continuous shearlet frame and is already proven.
- Now we have all the tools.

#### **Slide 18: Modified learned primal-dual.**

- We will do the reconstruction over the shearlet coefficients of the image, so some directional information will be preserved (and until some point as well the wavefront set). Now the constraint of the preservation of the wavefront set using the canonical transform.
- Add in each iteration a projection of the both wavefront set (data and image) to coincide up to the canonical relation.

#### **Slide 19: Implementation**

- The learned primal dual on the Shearlet coefficients (non wavefront set preservation) is already implemented, show some results in Jupyter.
- The Wavefront set computation is the hard part:
  - The theory just works on the continuous setting.

- One needs infinite number of shearlet coefficients to compute the wavefront set.
- Philipp indeed proved that for general distributions there is no way to do a faithful wavefront set definition on digital setting (you will always need to zoom in infinitely).
- But we know that until some level the shearlet coefficients give us some information of the wavefront set (we saw it also in the results of max).
- Inspired by some edges extractors using NN we are now developing a wavefront set extractor using CNNs classifiers (show the general idea in jupyter).