Transductive Inversion via Deep Transform Learning

Jyoti Maggu
Computer Science and Engineering
Indraprastha Institute of Information Technology
New Delhi, India
jyotim@iiitd.ac.in

Shalini Sharma
Computer Science and Engineering
Indraprastha Institute of Information Technology
New Delhi, India
shalinis@iiitd.ac.in

Angshul Majumdar
Electronics and Communication Engineering
Indraprastha Institute of Information Technology
New Delhi, India
angshul@iiitd.ac.in

***Abstract*— This work addresses the problem of solving a linear inverse problem. Conventional inversion techniques are model based (transductive). The advent of deep learning led the way for data-driven (inductive) inversion techniques. The main issue with inductive inversion is that unless the unseen signal (to be inverted) is similar to the training data, the learnt model fails to generalize rendering poor inversion results. A recent study on deep dictionary learning has shown how it can combine the best of both worlds – deep learning with transductive inversion. In this work, we show how the analysis counterpart of dictionary learning, called transform learning, can be extended deeper for transductive inversion. Results on dynamic MRI reconstruction, show that the proposed technique improves over the state-of-the-art.**

Keywords*— *inverse problem, compressed sensing, deep learning, MRI, reconstruction

I. Introduction (*Heading 1*)

Our interest lies in solving a noisy linear inverse problem. Many problems in machine learning and signal processing such as unmixing, regression, denoising, reconstruction, source separation, etc. fall under this category. Mathematically it is represented as follows,

$$y = Ax + \eta,\ \ \eta \sim N(0, \sigma^2) \tag{1}$$

where y is the observation, A is the linear system of equations, x is the unknown and η is the noise. Problems differ from one another in the nature of A. For example, in denoising, A is an identity, for regression, it is the system of explanatory variables and for magnetic resonance imaging (MRI) reconstruction, it is a Fourier operator.

The most straightforward approach to solving (1) is to find a minimum variance solution. For the commonly assumed Gaussian noise, this turns out to be the pseudo-inverse. For other types of noise, the solution is more sophisticated. In this work, we will assume that the noise is Gaussian.

Later techniques, instead of just solving for the minimum variance solution, assumed some prior. The simplest prior can be the minimum energy solution, which is effectively Tikhonov regularization. Recent approaches, developed over the past one-and-a-half decade, assumed the solution to be sparse. This led to regularization via the l_1-norm leading to the following,

$$\arg\min_x \|y - Ax\|_2^2 + \lambda \|x\|_1 \tag{2}$$

Perhaps the most famous applications of sparse recovery in machine learning are LASSO regression [1, 2]. In signal processing, the field of compressed sensing (CS) [3] started from the idea of sparse recovery.

In this work, we are mainly interested in signal processing aspects of linear inverse problems. CS based techniques became popular in this domain because a large class of signals can be represented sparsely in some fixed transform domain (wavelet, DCT, Gabor, etc.). Many such transforms are either orthogonal[1] or tight-framed[2]. This allows expressing the signal via analysis-synthesis[3] equations. Therefore signal recovery could be framed as a sparse synthesis prior problem,

$$\arg\min_\alpha \|y - A\Psi^T\alpha\|_2^2 + \lambda \|\alpha\|_1 \tag{3}$$

In this formulation, the transform coefficients are solved; the signal is recovered by applying the synthesis equation on the recovered coefficients.

The majority of studies in CS are based on the synthesis prior formulation. It is theoretically well understood and there are many efficient algorithms to solve it. However, in practice the synthesis prior is restrictive; it can accommodate only transforms that follow the analysis-synthesis equations. This precludes many powerful priors such as total variation in image processing tasks. Therefore, in practice, the co-sparse analysis prior formulation [4] is known to yield better results. This is expressed in the following fashion.

$$\arg\min_x \|y - Ax\|_2^2 + \lambda \|\Psi x\|_1 \tag{4}$$

---

[1] $Orthogonal: \Psi^T\Psi = I = \Psi\Psi^T$

[2] $Tight-frame: \Psi^T\Psi = I \neq \Psi\Psi^T$

[3] $analysis: \Psi x = \alpha$
$synthesis: \Psi^T\alpha = x$

ISBN: 978-1-6654-6798-8

In layman's terms the quality of CS reconstruction is directly proportional to the sparsity of the signal in the transform domain; the sparser the representation better is the recovery. The fixed transforms used in CS are mathematically well defined and generic by nature; they can sparsely represent a wide class of signals. It is well known in signal processing that, to get the best (sparsest) representation, the basis needs to be adaptively learned from the signal itself. This is the reason, dictionary learning based inversion techniques [5, 6] eventually improved over CS. The formulation for dictionary learning based inversion is expressed as follows,

$$\min_{x,D,Z} \|y - Ax\|_2^2 + \lambda\left(\sum_i \|P_i x - Dz_i\|_2^2 \text{ s.t. } \|z_i\|_0 \le \tau\right) \quad (5)$$

The first term is the standard data fidelity term assuming Gaussian noise. The term within the brackets corresponds to dictionary learning. Given the patches of the signal $P_i x$, a dictionary D is learnt such that the corresponding coefficients z_i is sparse. Instead of the standard CS sparsity prior in a fixed transform, (5) learns both the basis (D) and the coefficients (z_i) adaptively from the signal.

In a very recent work [7], it was shown that instead of learning only a single layer of a dictionary, better results can be obtained if multiple layers are learnt. The following formulation is given in [7],

$$\begin{aligned}&\min_{x,D,Z} \|y - Ax\|_2^2 \\ &+\lambda\left(\sum_i \left\|P_i x - D_3\left(\varphi\left(D_2\varphi\left(D_1 z_i\right)\right)\right)\right\|_2^2 + \gamma\|z_i\|_1\right)\end{aligned} \quad (6)$$

Here D_1, D_2, D_3 are three layers of dictionaries. The non-linear activation function φ prevents the collapsing of the three dictionaries into a single one. Note that instead of the l_0-norm, deep dictionary learning employs the l_1-norm to promote sparsity.

Dictionary learning is a synthesis formulation; it learns a basis (dictionary) from the signals such that one can generate the signals from the learnt coefficients. Just as there is an analysis version of compressed sensing, there is an analysis version of dictionary learning called transform learning [8, 9]; it learns an analysis basis (transform) that operates on the signals to generate the corresponding coefficients. The transform learning based inversion formulation is as follows,

$$\begin{aligned}&\min_{x,T,Z} \|y - Ax\|_2^2 + \mu\left(\|T\|_F^2 - \log\det T\right) \\ &+\lambda\left(\sum_i \|TP_i x - z_i\|_F^2 \text{ s.t. } \|z_i\|_0 \le \tau\right)\end{aligned} \quad (7)$$

The term in brackets corresponds to transform learning. Here T operates on the patches of the image to produce sparse coefficients z_i. Note that there is an extra regularization term $\left(\|T\|_F^2 - \log\det T\right)$ in (7); this term is to prevent the trivial solution $T=0$ $Z=0$ and the degenerate solutions where T is very large and Z very small or vice versa.

It has been empirically seen that transform learning yields better results than dictionary learning for inversion tasks [8, 9]. The theoretical reason is not exactly known. In recent times, the authors developed the framework of deep transform learning [10, 11]. However, the deep extension has been used for supervised learning tasks so far, not for inversion. This is the first work that will propose a deep transform learning based inversion. Given that the shallow version of transform learning yields better results than the shallow version of dictionary learning, we expect that our deep transform learning based inversion will improve deep dictionary learning and perhaps over other state-of-the-art inversion techniques. Specifically, in this work, we look into the example of dynamic MRI reconstruction.

II. Proposed Formulation

As mentioned before, transform learning is the analysis equivalent of dictionary learning. Although the technique is known to the signal processing community we review it for the sake of completeness. The model is expressed as

$$TX = Z \quad (8)$$

Here T is the transform, which operates on the data X to generate the representation Z.

The optimization problem for transform learning is expressed as –

$$\min_{T,Z} \|TX - Z\|_F^2 + \lambda\left(\|T\|_F^2 - \log\det T\right) + \mu\|Z\|_1 \quad (9)$$

Here l_1-norm enforces sparsity on the representation. The factor $-\log\det T$ imposes a full rank on the learned transform; this prevents the degenerate solution ($T=0$, $Z=0$). The additional penalty $\|T\|_F^2$ is to balance scale; without this $-\log\det T$ can keep on increasing producing degenerate results in the other extreme.

The minimization problem (9) is solved by alternately updating the two variables [8, 9].

$$Z \leftarrow \min_Z \|TX - Z\|_F^2 + \mu\|Z\|_1$$

$$T \leftarrow \min_T \|TX - Z\|_F^2 + \lambda\left(\varepsilon\|T\|_F^2 - \log\det T\right)$$

Updating the coefficients is straightforward. It can be updated via one step of soft thresholding.

$$Z \leftarrow signum(TX) \cdot \max\left(0, |TX| - \frac{\mu}{2}\right) \quad (10)$$

There is a closed-form update for the Transform as well. This is given by –

$$\begin{aligned}&XX^T + \lambda\varepsilon I = LL^T \\ &L^{-1}XZ^T = USV^T \\ &T = 0.5R\left(S + (S^2 + 2\lambda I)^{1/2}\right)Q^T L^{-1}\end{aligned} \quad (11)$$

Our proposed deep transform learning [10, 11] is the multi-layer extension of the shallow one. It can be thought of as the application of multiple levels of transforms to generate the coefficients. Mathematically this is expressed as follows –

$$T_N(\varphi...(T_2(\varphi(T_1X))) = Z \quad (12)$$

Here φ denotes the activation function; without which all the transforms will collapse into a single one. Previous studies [10, 11] used tanh or sigmoid as activation functions. In this work, we will be using rectified linear unit (ReLU) type activations.

Before going into the formulation, we discuss the reason for the deep extension. The shallow inversion formulation is linear. It is known in neural network theory about the function approximation capability of non-linear networks with ReLU [12]. The approximation capacity improves when one goes deeper with ReLU [13]. This is the prime reason behind the extension to deeper layers of transform with ReLU activation. We extend the basic transform learning based inversion formulation (9) to accommodate multiple layers of transforms. Mathematically this is expressed as follows,

$$\min_{x,T_1,T_2,T_3,Z} \|y-Ax\|_2^2 + \lambda\sum_i \|T_3T_2T_1P_ix - z_i\|_F^2 + \lambda\left(\mu\sum_{j=1}^{3}\left(\|T_j\|_F^2 - \log\det T_j\right) + \gamma\|z_i\|_1\right) \quad (13)$$

In (13) we have shown the formulation for three layers. The derivation for the solution will be generic enough to solve for more. Also, the coefficients after the application of each layer of transform should be non-negative to impose ReLU activation. This means that $T_1P_ix > 0$ and $T_2T_1P_ix > 0$.

To solve (13) we resort to the variable splitting technique [14]. We introduce two sets of proxy variables $w_i = T_1P_ix$ and $h_i = T_2T_1P_ix$. This leads to the following augmented Lagrangian formulation.

$$\min_{x,T_1,T_2,T_3,H,W,Z} \|y-Ax\|_2^2 + \lambda\left(\sum_i \|T_3h_i - z_i\|_F^2 + \|T_2w_i - h_i\|_F^2 + \|T_1P_ix - w_i\|_F^2\right) + \lambda\left(\mu\sum_{j=1}^{3}\left(\|T_j\|_F^2 - \log\det T_j\right) + \gamma\|z_i\|_1\right) \quad (14)$$

H and *W* are formed by stacking the h_i's and w_i's as columns. Owing to the non-negativity constraints the proxy variables w_i and h_i need to be greater than 0. Ideally, we would require to have two multiplicative hyper-parameters corresponding to the two newly introduced terms $\|T_2w_i - h_i\|_F^2$ and $\|T_1P_ix - w_i\|_F^2$; these hyper-parameters would have to be gradually increased with passing iterations to impose strict equality between the proxies and the variables; however, we argue that since these two terms correspond to two intermediate layers of deep transform learning, there is no reason to give them selective importance over others. Hence we keep the multiplicative hyper-parameter to be unity.

We employ the alternating direction method of multipliers (ADMM) [15] to solve (14). We update each of the variables as sub-problems.

$$\text{P1:}\min_{T_1}\sum_i \|T_1P_ix - w_i\|_F^2 + \mu\left(\|T_1\|_F^2 - \log\det T_1\right)$$

$$\text{P2:}\min_{T_2}\sum_i \|T_2w_i - h_i\|_F^2 + \mu\left(\|T_2\|_F^2 - \log\det T_2\right)$$

$$\text{P3:}\min_{T_3}\sum_i \|T_3h_i - z_i\|_F^2 + \mu\left(\|T_3\|_F^2 - \log\det T_3\right)$$

$$\text{P4:}\min_{x}\|y-Ax\|_2^2 + \lambda\sum_i \|T_1P_ix - w_i\|_F^2$$

$$\text{P5:}\min_{h_i}\|T_3h_i - z_i\|_F^2 + \|T_2w_i - h_i\|_F^2 \ \forall i$$

$$\text{P6:}\min_{w_i}\|T_2w_i - h_i\|_F^2 + \|T_1P_ix - w_i\|_F^2 \ \forall i$$

$$\text{P7:}\min_{z_i}\|T_3h_i - z_i\|_F^2 + \gamma\|z_i\|_1 \ \forall i$$

All the sub-problems have closed-form solutions. P1 to P3 are standard transform updates. This has been discussed in (11). P4 to P6 are least-square problems that have a closed-form solution in the form of pseudoinverse. However, here we solve it using conjugate gradient since *A* may not always be available as an explicit matrix. P7 is a sparse coefficient update, we have already discussed its closed-form update in (10). It must be noted that for the updates of P4 and P5, one must need to ensure that the solution is non-negative; ideally one needs Forward-Backward splitting type iterative algorithms to solve this. However, these iterations make the overall algorithm computationally complex; our simple fix is to enforce non-negativity on h_i and w_i by putting all the negative values to zeroes after the pseudo-inverse solution.

The problem is non-smooth and non-convex. There is a recent work that shows the convergence of ADMM (to local minima) for such a class of problems [16] especially when each of the sub-problems has a closed-form update. The convergence is local, in practice, we stop the iterations when the objective function does not change much over consequent iterations.

The computational complexity of each iteration is mainly dictated by transform updates. Since they require computing singular value decompositions, their complexity is of $O(n^3)$. The complexity for solving the least square problems is $O(n^2)$ by conjugate gradient. The last sub-problem (sparse update) costs $O(n)$.

III. Experimental Evaluation

We address the problem of dynamic MRI reconstruction. Experiments are conducted on two publicly available datasets[4]. The two sequences will be called the Cardiac Perfusion Sequences 1 and 2. The data was collected on a 3T Siemens scanner. In this work we simulated a radial sampling with 24 lines that were acquired for each time frame; this corresponds to an under-sampling ratio of 0.21. The full resolution of the dynamic MR images is 128 x 128. About 6.7 samples were collected per second. The scanner parameters for the radial

[1] [4] http://www.sci.utah.edu/bisti.html

acquisition were TR=2.5–3.0 msec, TE=1.1 msec, flip angle = 12° and slice thickness = 6 mm. The reconstructed pixel size varied between 1.8 mm^2 and 2.5 mm^2. Each image was acquired in a ~ 62-msec read-out, with a radial field of view (FOV) ranging from 230 to 320 mm.

We have compared our method with a few recent transductive reconstruction techniques. The first one is based on bi-linear modeling based recovery (BLM) [17]. The second one is deep unrolling [18]. We have also compared with DDL based reconstruction [7]. Both [17] and [18] are relatively recent techniques.

As before, for our proposed method we tuned the parameters using grid search on a separate dynamic MRI data not used here. The obtained parametric values are - $\lambda = .2$, $\mu = .5$ and $\gamma = .2$. 3D patches of size 16 x 16 x 4 were used. We obtained the best results for 4 layers; the number of basis elements in different layers is 256-128-128-64.

The experimental results are shown in Table I. MRI reconstruction quality is usually measured by Normalized Mean Squared Error (NMSE). We use the same metric. From the numerical results, we find that our method yields the best results. BLM which is a transductive technique yields the worst results. Both DDL and BCS* yield similar results. Our proposed method is worse than DDL and BCS* for 2 layers but is better than these for 3, 4 and 5 layers.

The numerical results do not give the complete picture for MRI reconstruction. Therefore it is customary to show reconstructed and difference (between ground-truth and original) images. We will only show the results with 4 layers from our proposed deep transform learning.

One frame each from the reconstructed Cardiac Perfusion 1 sequence is shown in Fig. 1. One can see that even though BLM shows poor NMSE, its reconstruction quality is actually at par with DDL and Deep Unrolling; BLM shows a lot of reconstruction artifacts but can preserve the edges. DDL and Deep Unrolling on the other hand overtly smooth the tissue boundaries. Our proposed technique preserves tissue boundaries with minimal artifacts.

The difference images for Cardiac Perfusion 1 are shown in Fig. 2. These are obtained by taking the absolute difference between the fully sampled ground truth and the reconstructed images. The thus obtained difference images are contrast enhanced uniformly for visual clarity. From these difference images, we can see that BLM indeed generates considerable reconstruction artifacts; the artifacts are much less pronounced in Deep Unrolling. DDL improves over Deep Unrolling. Our method yields the best reconstruction; the artifacts are negligible.

TABLE I. RECONSTRUCTION PERFORMANCE IN TERMS OF NMSE

Method	BLM	Deep Unrolling	DDL 3 layer	Proposed 2 layer	Proposed 3 layer	Proposed 4 layer	Proposed 5 layer
Cardiac 1	0.0586	0.0356	0.0315	0.0408	0.0219	**0.0184**	0.0307
Cardiac 2	0.0474	0.0312	0.0298	0.0400	0.0202	**0.0149**	0.0286

Fig. 1. Reconstructed Images 1. Left to Right – Ground-truth, BLM, Deep Unrolling, DDL and Proposed

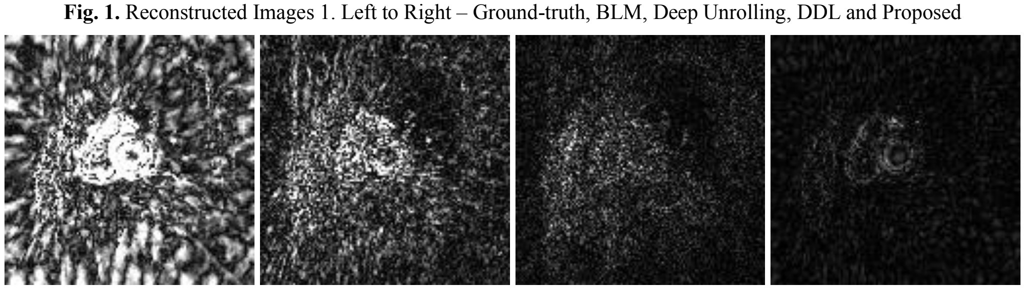

Fig. 2. Difference Image. Left to Right – BLM, Deep Unrolling, DDL and Proposed

IV. CONCLUSION

This is the first work that shows that proposes a generic inversion approach based on deep transform learning. In particular, we have addressed the problem of dynamic MRI reconstruction. For this problem, we have compared it with the state-of-the-art. We improve upon the rest by a considerable margin.

In the future, we would like to incorporate structured inverse problems, for example, trees, graphs, group-sparsity, etc. into the deep transform learning based inversion framework.

REFERENCES

[1] R. Tibshirani, "Regression shrinkage and selection via the lasso," Journal of the Royal Statistical Society, Series B (Methodological), pp.267-288, 1996.

[2] J. Duan, C. Soussen, D. Brie, J. Idier and Y. Wang, "On LARS/Homotopy Equivalence Conditions for Over-Determined LASSO," IEEE Signal Processing Letters, vol. 19, no. 12, pp. 894-897, 2012.

[3] R. G. Baraniuk, "Compressive Sensing [Lecture Notes]," IEEE Signal Processing Magazine, vol. 24, no. 4, pp. 118-121, 2007.

[4] M. Elad, P. Milanfar and R. Rubinstein, "Analysis versus synthesis in signal priors," Inverse problems, vol. 23, no. 3, p.947, 2007.

[5] Y. Li, S. Ding and Z. Li, "Dictionary learning with the cosparse analysis model based on summation of blocked determinants as the sparseness measure," Digital Signal Processing, vol. 48, pp. 298-309, 2016.

[6] G. Li, Z. Zhu, X. Wu and B. Hou, "On joint optimization of sensing matrix and sparsifying dictionary for robust compressed sensing systems", Digital Signal Processing, vol. 73, pp. 62-71, 2018.

[7] J. Lewis D., V. Singhal and A. Majumdar, "Solving Inverse Problems in Imaging via Deep Dictionary Learning," IEEE Access, vol. 7, pp. 37039-37049, 2019.

[8] S. Ravishankar and Y. Bresler, "Learning sparsifying transforms", IEEE Transactions on Signal Processing, vol. 61, no. 5, pp. 1072-1086, 2013.

[9] S. Ravishankar, B. Wen and Y. Bresler. "Online sparsifying transform learning-Part I: Algorithms", IEEE Journal of Selected Topics in Signal Processing, vol. 9, no. 4, pp. 625-636, 2015.

[10] J. Maggu and A. Majumdar, "Greedy deep transform learning," IEEE International Conference on Image Processing, 2017, pp. 1822-1826, 2017.

[11] J. Maggu and A. Majumdar, "Unsupervised Deep Transform Learning," IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 6782-6786, 2018.

[12] P. Petersen and F. Voigtlaender, "Optimal approximation of piecewise smooth functions using deep ReLU neural networks," Neural Networks, vol. 108, pp. 296-330, 2018.

[13] D. Yarotsky, "Error bounds for approximations with deep ReLU networks," Neural Networks, vol. 94, pp.103-114, 2017.

[14] P.L. Combettes and J.C. Pesquet, "Proximal splitting methods in signal processing," Fixed-point algorithms for inverse problems in science and engineering, pp. 185-212, 2011.

[15] E. Ghadimi, A. Teixeira, I. Shames and M. Johansson, "Optimal Parameter Selection for the Alternating Direction Method of Multipliers (ADMM): Quadratic Problems," IEEE Transactions on Automatic Control, vol. 60, no. 3, pp. 644-658, 2015.

[16] Y. Wang, W. Yin and J. Zeng, "Global convergence of ADMM in nonconvex nonsmooth optimization," Journal of Scientific Computing, pp.1-35, 2015.

[17] G. N. Shetty, K. Slavakis, A. Bose, U. Nakarmi, G. Scutari and L. Ying, "Bi-Linear Modeling of Data Manifolds for Dynamic-MRI Recovery," IEEE Transactions on Medical Imaging, vol. 39, no. 3, pp. 688-702, 2020.

[18] X. Zhang, Q. Lian, Y. Yang and Y. Su, "A deep unrolling network inspired by total variation for compressed sensing MRI," Digital Signal Processing, vol. 107, 102856, 2020.