Learning Structured Sparse Matrices for Signal Recovery via Unrolled Optimization

Jonathan Sauder[1], Martin Genzel[2], Peter Jung[13]
sauder@campus.tu-berlin.de, m.genzel@uu.nl, peter.jung@tu-berlin.de
https://openreview.net/forum?id=IxKSqOq1TKQ

Short Summary

In this paper, we propose a method for learning sparse, structured measurement operators for signal recovery in linear inverse problems. Our method fuses unrolled optimization with Gumbel reparametrizations. We demonstrate the flexibility of our method and show that it can improve upon the standard regimes based on randomness in two compressed sensing scenarios.

Problem Statement

We consider linear inverse problems, where we have:

1. An unknown signal x drawn from a distribution, of which we have access to a database (training set)
2. A set of admissible linear operators $\mathbf{\Phi} \subset \mathbb{C}^{m \times n}$ which follow the hardware/physics requirement of the measurement process
3. A parametrized reconstruction function $f_\theta : \mathbb{R}^m \to \mathbb{R}^n$ that takes measurements and returns the recovered signal
4. A differentiable loss function $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ in signal space

Want to find measurement operator and parameters of reconstruction function that optimize signal recovery over the expectation of the signals distribution:

$$\min_{\theta, \mathbf{\Phi}} \mathbb{E}_x \Big[\mathcal{L}(f_\theta(\mathbf{\Phi} x), x) \Big]$$

Background & Related Work

Unrolled optimization is a technique in which the computational graph of iterative optimization algorithm is unrolled to yield a neural network. This allows back-propagating with regards to parameters involved in the convex optimization scheme. We use unrolled IHT [1] and NA-ALISTA [2] in our experiments.

Gumbel reparametrizations [3,4,5] are a technique for estimating gradients in computational graphs which include nodes that are categorical random variables.

Our Method:

We propose the use of unrolled optimization in conjunction with Gumbel reparametrizations to learn sparse structured measurement matrices for signal recovery.

Consider the set of possible indices of the measurement operator $\mathcal{I} := \{1 \dots m\} \times \{1 \dots n\}$. We partition this index set and apply a Gumbel top-K reparametrization on each set in the partition. This results in a binary matrix that can be used in any autograd framework to construct the measurement operator.

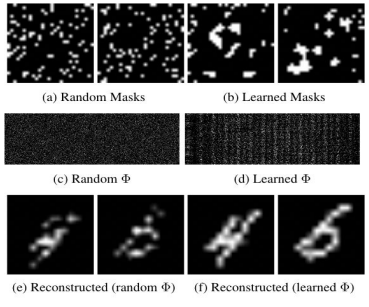
Algorithm 1: Learning a binary matrix with d_i ones per set I_i of the partition.

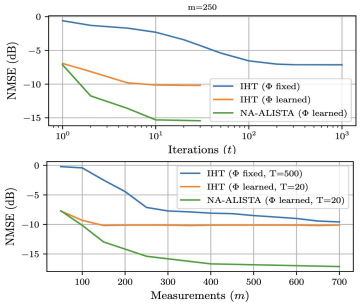
Input: signal x (training data), temperature τ, top-K-keeps $d_1, \dots, d_l \in \mathbb{N}$, differentiable reconstruction algorithm $f : \mathbb{R}^m \to \mathbb{R}^n$, index partition $\mathcal{P}(\mathcal{I}) = \{I_1, \dots, I_l\}$
Learnable Parameters: Parameters of measurement matrix $\varphi \in \mathbb{R}^{m \times n}$, parameters of reconstruction algorithm θ

$G \sim_{i.i.d.} Gumbel(0,1)^{m \times n}$
for $i \in \{1 \dots l\}$ **do**
  logits := $(\varphi[I_i] + G[I_i])/\tau$
  probs := softmax(logits)
  hard := topk(probs, d_i)
  $\Phi[I_i]$:= hard.detach() + probs - probs.detach()
end for
$y := \Phi x$
loss := $\mathcal{L}(f_\theta(y), x)$
loss.backward()

Application: Single Pixel Imaging

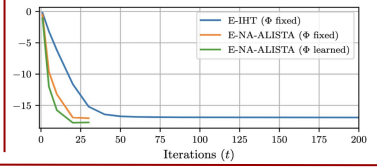
Following Algorithm 1, we partition the indices into the rows of the measurement operator, and select d pixels per row. We unroll algorithms of the form: $\hat{x}^{(t+1)} = \text{prox}_{\lambda \|\cdot\|_1}\Big(\hat{x}^{(t)} + \gamma \nabla_{\hat{x}^{(t)}} \big(\|y - \Phi \hat{x}^{(t)}\|_2^2\big)\Big)$
By learning a measurement operator that adapts to the dataset structure, our method improves the reconstruction (left). Our method decreases the number of iterations required (top right) and the number of measurements required (bottom right) compared to a random Φ

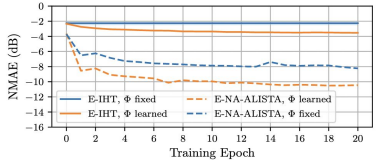




Application: Compressed Sensing with left-d-regular Graphs

Following Algorithm 1, we partition the indices into the columns. We select d ones per column using Gumbel top-K operators. We unroll algorithms of the form: $\hat{x}^{(t+1)} = \mathcal{H}_s\Big(\hat{x}^{(t)} + \mathcal{M}(y - \Phi \hat{x}^{(t)})\Big)$, where $\mathcal{H}_s$ denotes hard thresholding and $\mathcal{M}$ denotes the median operation. Our method leads to faster convergence (left) and better reconstruction (right) than when using a random Φ .





[1]Communications and Information Theory Chair, Technical University of Berlin
[2]Utrecht University
[3]Data Science in Earth Observation, Technical University of Munich

References
[1] Blumensath, Thomas, and Mike E. Davies. "Iterative hard thresholding for compressed sensing." Applied and computational harmonic analysis 27.3 (2009): 265-274.
[2] Behrens, Freya, Jonathan Sauder, and Peter Jung. "Neurally augmented alista." ICLR 2020.
[3] Beck, Amir, and Marc Teboulle. "A fast iterative shrinkage-thresholding algorithm for linear inverse problems." SIAM journal on imaging sciences 2.1 (2009): 183-202.
[4] Maddison, Chris J., Andriy Mnih, and Yee Whye Teh. "The concrete distribution: A continuous relaxation of discrete random variables." arXiv preprint arXiv:1611.00712 (2016).
[5] Vieira, T. 2014. Gumbel-max trick and weighted reservoir sampling.