Divergence-Based Regularization for End-to-End Sensing Matrix Optimization in Compressive Sampling Systems

Roman Jacome
Department of Physics
Universidad Industrial de Santander
Bucaramanga,Colombia
roman2162474@correo.uis.edu.co

Henry Arguello
Department of Computer Science
Universidad Industrial de Santander
Bucaramanga,Colombia
henarfu@uis.edu.co

Alejandra Hernandez-Rojas
Department of Physics
Universidad Industrial de Santander
Bucaramanga,Colombia
maria.hernandez26@correo.uis.edu.co

Paul Goyes-Peñafiel
Department of Computer Science
Universidad Industrial de Santander
Bucaramanga,Colombia
ypgoype@correo.uis.edu.co

***Abstract*—Sensing Matrix Optimization (SMO) in Compressed Sensing (CS) systems allows improved performance in the underlying signal decoding. Data-driven methods based on deep learning algorithms have opened a new horizon for SMO. The matrix is designed jointly with a decoder network that performs compressed learning tasks. This design paradigm, named End-to-End (E2E) optimization, comprises two parts: the sensing layer that models the acquisition system and the computational decoder. However, SMO in the E2E network has two main issues: i) it suffers from the vanishing of the gradient since the sensing matrix is the first layer of the network, and ii) there is no interpretability in the SMO, resulting in poorly compressed acquisition. To address these issues, we proposed a regularization function that gives some interpretability to the designed matrix and adds an inductive bias in the SMO. The regularization function is based on the Kullback-Leiber Divergence (KLD), which aims to approximate the distribution of the compressed measurements to a prior distribution. Thus, the sensing matrix can concentrate or spread the distribution of the compressed measurements according to the chosen prior distribution. We obtained optimal performance by concentrating the distribution in the recovery task, while in the classification task, the improvement was obtained by increasing the variance of the distribution. We validate the proposed regularized E2E method in general CS scenarios, such as in the Coded Aperture (CA) design for the Single-Pixel Camera (SPC) and Compressive Seismic Acquisition (CSA) geometry design.**

***Index Terms*—Sensing Matrix Optimization; End-to-End optimization; Compressive Sensing; Compressive Imaging; Compressive Seismic Acquisition.**

I. Introduction

Compressive Sensing (CS) [1] states that a signal $\mathbf{x} \in \mathbb{R}^n$ can be recovered from a small set of observations $\mathbf{y} \in \mathbb{R}^m$ such that $m \ll n$ as $\mathbf{y} = \mathbf{H}_{\boldsymbol{\phi}}\mathbf{x} + \mathbf{w}$, where $\mathbf{H}_{\boldsymbol{\phi}} \in \mathbb{R}^{m \times n}$ is the measurement matrix, ϕ denotes the free-parameters of the system, and $\mathbf{w} \in \mathbb{R}^m$ is additive noise in the acquisition. Decoding the measurements to obtain the underlying signal $\mathbf{x}$ requires additional knowledge to solve this ill-posed problem. While a plethora of decoding algorithms have been proposed, such as those based on sparsity-promoting solution [2] [3], dictionary learning [4], low-rank priors [5], or recent data-driven methods based on deep learning [6] [7]. Complementary to algorithm development, Sensing Matrix Optimization (SMO) has remarkably improved decoding performance [8]. Traditional design methods are based on improving the mutual coherence of the sensing matrix as well as the the representation basis [9], for block-sparse signals [10], joint dictionary and sensing optimization [11] or the restricted isometry property [12]. These designs are mostly based on sparse representations of the desired signal $\mathbf{x}$, which in practice might not be sufficient to describe the signal [13]. Thus, recent data-driven methods have enabled SMO based on data priors, i.e., the SMO is performed depending on the training dataset. The End-to-End (E2E) learning of the sensing matrix and the decoding process by a Deep Neural Network (DNN) has significantly improved the decoding performance. Here, the free parameters of the sensing matrix ϕ are trainable variables jointly optimized with the parameters of DNN that perform the decoding task. This E2E optimization has been successfully applied in CS [14]–[16], computational imaging where the free-parameters are optical coding elements such as Coded Aperture (CA) [17]–[19], diffractive optical elements [20]–[22] or in compressive seismic acquisition geometries, where the design is performed over the receivers or sensors [23]. However, the SMO in the E2E method has the following issues. i) Vanishing of the gradient: Since the sensing is performed in the first layer of the network, the gradient in that layer is smaller than the decoding network. Thus, the performance relies more on the decoder network parameters than the trained sensing matrix. ii) Lack of interpretability: traditional SMO results in interpretable optimization in terms

 ISBN: 978-1-68558-057-5

of the mutual incoherence [24] or the eigenvalues concentration [25]. However, the resulting sensing matrix in the E2E methods does not have an interpretation other than the one adapted to the training data.

In this work, we propose to address these issues by including a regularization in the loss function of the E2E network. Here, we propose a regularization based on the Kullback-Leiber Divergence (KLD) over the distribution of the compressed measurements. The KLD is employed to measure the difference between two probability distributions. Here, we employed the KLD to approximate the distribution of the measurements to a chosen prior distribution. This function has been widely used in DNN to regularize latent representation distribution of the data, as in variational autoencoders [26] or in generative models [27]. One of the reasons for the wide use of this function in DNN regularization is the closed-form solution of the divergence for Gaussian distributions [26] and Laplacian distributions [28], which depends on the mean and variance of the data and prior distribution. We study the effect of the prior distribution for two computational tasks, recovery and classification. We found that smaller variance, i.e., the sensing matrix represents the data in a concentrated distribution, gives better reconstruction performance. While higher variance produces more accurate classification predictions. Thus, the regularizer can be set to obtain optimal performance in different computational tasks. The main interpretation for the recovery case comes from contractive autoencoders [29], which states that the original data is better represented in an invariant low-dimensional manifold. The intuition of the second behavior is that more separated measurements allow better identification of the classes by the decoding network. Preliminary results on this regularizer applied to compressive imaging were presented in [30].

We evaluate the proposed regularized E2E method in three cases. In a general CS setting where $\mathbf{H}_{\boldsymbol{\phi}}$ is a dense matrix, and $\boldsymbol{\phi}$ are all the entries of the sensing matrix. The second case is the Single-Pixel Camera (SPC) [31], which is one of the most common CS systems of imaging applications. Here the sensing matrix entries are binary values representing a CA. The last setting is a Compressive Seismic Acquisition (CSA) model, where the sensing matrix is a diagonal matrix in which entries are binary values denoting the removed receivers.

The rest of the paper is organized as follows. In section II the E2E model is established, III presents the proposed divergence-based regularizers for the E2E training. Section IV shows the mathematical models of the CS systems used to validate the proposed method. Section V contains the numerical simulations of the proposed method and comparisons with non-regularized models. Finally, in section VI the conclusions of this work are presented.

II. End-to-End Optimization

With new developments in data-driven algorithms and deep learning, a method called End-to-End optimization (E2E) has been developed to optimize the sensing procedure and the decoding process jointly. In this approach, the sensing model $\mathbf{H}_{\boldsymbol{\phi}}$ is cast into a differentiable neural network layer, where the free parameters $\boldsymbol{\phi}$ are the weights of this layer, named Sensing Layer (SL). The SL is coupled to a neural network that receives as input the compressive measurements and performs the decoding operator, which is called Computational Decoder (CD), denoted by the operator $\mathcal{N}_\theta$ where θ are the trainable parameters of the network. Considering the dataset $\{\mathbf{x}_k, \mathbf{d}_k\}_{k=1}^K$ where $\mathbf{d}_k$ is the ground-truth, e.g., in the recovery case $\mathbf{d}_k$ is the same input image $\mathbf{x}_k$ and in classification $\mathbf{d}_k$ is the image label. Then, the E2E optimization problem is the following

$$\{\widehat{\boldsymbol{\phi}}, \widehat{\theta}\} = \arg\min_{\boldsymbol{\phi},\theta} \frac{1}{K} \sum_{k=1}^{K} \mathcal{L}(\mathcal{N}_\theta(\mathbf{H}_{\boldsymbol{\phi}}\mathbf{x}_k), \mathbf{d}_k), \tag{1}$$

where $\mathcal{L}$ is the loss function of the computational task. The main goal is to update the sensing matrix and the decoder parameters according to the loss function task. Particularly, following the chain rule, the gradient of the loss function with respect to the SL trainable parameters is

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\phi}} = \frac{1}{K} \sum_{k=1}^{K} \frac{\partial \mathcal{L}}{\partial \theta} \frac{\partial \mathcal{N}_\theta}{\partial \mathbf{y}_k} \frac{\partial \mathbf{y}_k}{\partial \boldsymbol{\phi}}, \tag{2}$$

where $\mathbf{y}_k = \mathbf{H}_{\boldsymbol{\phi}}\mathbf{x}_k$. While the network is training, the gradient of the loss function with respect to the CD parameters $\frac{\partial \mathcal{L}}{\partial \theta}$ is reduced due to the gradient descent optimizer of the network. Consequently, the gradient of the SL parameters decreases even more; thus, the optimization relies more on the CD than on the SL.

III. Proposed Regularization Function

We propose a regularization function for E2E optimization based on KLD to approximate the distribution of the measurements to a prior distribution. First, define the matrices $\mathbf{X} \in \mathbb{R}^{K\times N}$ and $\mathbf{Y} \in \mathbb{R}^{K\times M}$ containing the set of the high-dimensional signal and compressed measurements, respectively, i.e., $\mathbf{X} = [\mathbf{x}_1^T, \ldots, \mathbf{x}_K^T]^T$ and $\mathbf{Y} = [\mathbf{y}_1^T, \ldots, \mathbf{y}_K^T]^T$. The regularized optimization problem is given by

$$\{\theta^\star, \boldsymbol{\phi}^\star\} = \arg\min_{\theta,\boldsymbol{\phi}} \frac{1}{K} \sum_{k=1}^{K} \mathcal{L}\left(\mathcal{N}_\theta(\mathbf{H}_{\boldsymbol{\phi}}\mathbf{x}_k), \mathbf{d}_k\right) + R(\mathbf{Y}). \tag{3}$$

This type of regularization function is based on the idea behind variational auto-encoders [26]. Particularly, this regularization aims to approximate the probability distribution of the measurements set denoted by the posterior distribution $q_{\boldsymbol{\phi}}(\mathbf{Y}|\mathbf{X})$, to a prior distribution $p_\beta(\mathbf{Y})$ where β is the set of parameters defining the distribution. This regularizer is defined as

$$R_D(\mathbf{Y}) = \mathcal{D}\left(q_{\boldsymbol{\phi}}(\mathbf{Y}|\mathbf{X}) \| p_\beta(\mathbf{Y})\right), \tag{4}$$

where $\mathcal{D}$ denotes the divergence function. Several divergences have been used as loss functions in neural network training. The most common is the Kullback-Leiber Divergence (KLD), employed in variational-autoencoders [26], generative adversarial networks [27], self-supervised learning [32]

 ISBN: 978-1-68558-057-5

among others. Particularly, the KLD is defined as follows, given two probability distributions $P(x)$ and $Q(x)$, we have $\mathcal{D}_{KL}(P\|Q) = \int P(x)\log\left(\frac{P(x)}{Q(x)}\right)dx$. One of the main reasons the KLD is widely used is that it has a closed-form solution when $P(x)$ and $Q(x)$ are Gaussian or Laplacian distributions [26] [28]. In these cases, the parameters for the prior distribution $p_\beta(\mathbf{Y})$ are $\beta = \{\mu_p, \sigma_p\}$, where μ_p is the mean value and σ_p is the variance of the distribution. For the distribution of the measurements $q_{\boldsymbol{\phi}}(\mathbf{Y}\|\mathbf{X})$ we compute statistics of the measurements, where the mean $\boldsymbol{\mu}_{\mathbf{Y}} \in \mathbb{R}^m$ and variance $\boldsymbol{\sigma}_{\mathbf{Y}} \in \mathbb{R}^m_+$ are computed pixel-wise across the measurements training batch. For the Gaussian case, the KLD-based regularizer is defined as

$$R_{KL-G}(\mathbf{Y}) = \log\left(\frac{\sigma_{\mathbf{Y}}}{\sigma_p}\right) - \frac{\boldsymbol{\sigma}^2_{\mathbf{Y}} + (\boldsymbol{\mu}_{\mathbf{Y}} - \mu_p)^2}{2\sigma_p^2} + \frac{1}{2}, \quad (5)$$

and for the Laplacian assumption, the KLD-based regularizer is given by

$$R_{KL-L}(\mathbf{Y}) = \log\left(\frac{\sigma_{\mathbf{Y}}}{\sigma_p}\right) - \frac{\sigma_p + e^{\left(\frac{-|\mu_p - \boldsymbol{\mu}_{\mathbf{Y}}|}{\sigma_p}\right)} + |\mu_p - \boldsymbol{\mu}_{\mathbf{Y}}|}{\sigma_p} - 1. \quad (6)$$

The mean and variance of the prior distribution are hyperparameters that control the effect of the regularizers. Therefore, those hyperparameters must be tuned to obtain the desired goal. The computational complexity of employing these regularization functions in the E2E optimization relies only on computing element-wise logarithm and its corresponding derivative; thus, they do not increase the computational complexity significantly with respect to the baseline E2E

IV. Compressive Sensing System Models

In this section, we present the compressive sensing system models to validate the proposed coding design in the E2E framework.

A. Single Pixel Camera

The SPC uses a set of CA $\boldsymbol{\phi} = \{\boldsymbol{\phi}_p\}_{p=1}^P$ that spatially modulate all the information of the scene, where the index p denotes each captured snapshot. In particular, it is a binary pattern in which we employ values $\{-1, 1\}$ as suggested in [33]. Mathematically, the sensing matrix is built as the concatenation of the vectorized CA of each shot $\mathbf{H}_{\boldsymbol{\phi}} = [\boldsymbol{\phi}_1^T, \ldots, \boldsymbol{\phi}_P^T]^T$ where P denotes the total number of snapshots. Then, the sensing model is given by

$$\mathbf{y} = \mathbf{H}_{\boldsymbol{\phi}}\mathbf{x} + \mathbf{w}, \quad (7)$$

where $\mathbf{y} \in \mathbb{R}^P$ is the compressed SPC measurements. An important factor in the SPC is the compression ratio γ defined as $\gamma = \frac{P}{N}$. Here, the optimized parameters are the CA. Since its entries are binary-valued, we add the regularization term proposed in [17] in the optimization problem in (3), which promotes this physical constraint. The E2E optimization for this case is the following

$$\{\theta^\star, \phi^\star\} = \arg\min_{\theta,\phi} \frac{1}{K}\sum_{k=1}^{K} \mathcal{L}\left(\mathcal{N}_\theta(\mathbf{H}_\phi \mathbf{x}_k), \mathbf{d}_k\right) + R(\mathbf{Y}) + \rho R_i(\boldsymbol{\phi}), \quad (8)$$

where ρ is a regularization parameter and the regularization $R_i(\boldsymbol{\phi}) = \sum_{ij}(1-\phi_{ij})^2(1+\phi_{ij})^2$.

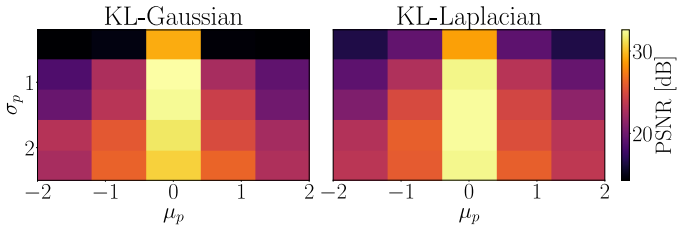

Fig. 1. Recovery performance for the general CS scenario employing the KLD regularizers with the Gaussian (left) and Laplacian (right) cases.

B. Compressive seismic acquisition

The cross-spread is a fundamental seismic acquisition geometry involving one linear arrangement of shot points and receivers perpendicular to each other [34] [35]. To mathematically represent the seismic data acquired by a cross-spread, let $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ be a data cube where each dimension represents I_1 time samples, I_2 receivers, and I_3 number of shots. However, due to different reasons, such as economic limitations and environmental constraints, the observed seismic field data is irregular and incomplete along the receiver dimension, leading to a recovery task. To simulate the undersampled data, let $\boldsymbol{\phi} \in \{0,1\}^{I_2}$ be a sampling vector with dimensions equal to the number of receivers. The entries of $\boldsymbol{\phi}$, denoted as ϕ_i, define whether the information is acquired. If $\phi_i = 0$, the receiver is removed; otherwise, $\phi_i = 1$, and it is acquired. The diagonalization of the sampling vector derives the diagonal sampling matrix as $\mathbf{H}_{\boldsymbol{\phi}} = \mathrm{diag}(\boldsymbol{\phi})$. Once $\mathbf{H}_{\boldsymbol{\phi}}$ is built, the undersampled measurements are obtained via n-mode product($\times_n$) defined in [36]

$$\mathcal{Y} = \mathcal{X} \times_2 \mathbf{H}_{\boldsymbol{\phi}}, \quad (9)$$

where Eq. 9 represents the 2-mode product between the full data $\mathcal{X}$ and $\mathbf{H}_{\boldsymbol{\phi}}$. The undersampled measurements $\mathcal{Y} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ contains the removed receivers as columns in zero for each shot.

A conventional relation that determines the number of acquired receivers by the sensing matrix is the transmittance, calculated as

$$\delta_{\boldsymbol{\phi}} = \sum_{i=1}^{M} \frac{\phi_i}{I_2}. \quad (10)$$

For instance, when $\delta_{\boldsymbol{\phi}} = 0.7$, the 70% of the total receivers are acquired. The E2E optimization is mathematically expressed as

$$\{\hat{\phi}, \hat{\theta}\} = \arg\min_{\phi,\theta} \mathcal{L}\left(\mathcal{N}_\theta\left(\mathcal{X} \times_2 \mathbf{H}_\phi\right), \mathcal{X}\right) + \rho R(\phi), \quad (11)$$

where the regularization $R(\phi) = (\delta_0 - \delta_\phi)^2$ controls the transmittance to converge to a desired value δ_0, and ρ represents a weight parameter.

V. Simulations and Results

The implementation of the method was performed on Tensorflow and Keras libraries [37]. We trained the E2E network for 100 epochs for all the experiments, halving the learning rate every 40 epochs. The Adam optimizer [38] was employed, setting its hyperparameter with the default values. The input of each network was the transpose operation of the sensing matrix to the measurements, i.e., $\mathbf{H}_\phi^T \mathbf{y}_k$. To evaluate the performance on the classification task, we employ the accuracy metric defined as

$$A = \frac{1}{C} \sum_{c=1}^{C} \frac{\text{TP}_c}{\text{Total}_c},$$

where C is the number of the classes and TP are the True Positive. For the recovery task, we employ the Peak-Signal-to-Noise-Ratio (PSNR) defined as

$$\text{PSNR} = 10 \log_{10} \left(\frac{\max(\mathbf{x})}{\text{MSE}(\mathbf{x}, \hat{\mathbf{x}})} \right),$$

where max returns the maximum value of $\mathbf{x}$ and $\text{MSE}(\cdot, \cdot)$ is the mean squared error.

A. General compressive sensing case

In the first experiment to validate the performance of the proposed regularized E2E network, we study a general compressive imaging scenario, not imposing any physical and structural meaning on the sensing matrix $\mathbf{H}_\phi$. Here we use a compression ratio of 10%. The MNIST dataset of handwritten digits was employed. This dataset contains 60000 training examples and 10000 for testing. We upscale the image to 32×32. For the CD model, we employed a fully connected layer.

We analyze the effect of the prior distribution's mean and variance (μ_p, σ_p) on the network performance. Here, we vary μ_p from -2 to 2, and σ_p was changed from 0.1 to 2.0, taking five equispaced values. The results of this experiment are shown in Figure 1 where the test set reconstruction PSNR is plotted in terms of μ_p and σ_p. The optimal reconstruction PSNR values are obtained at variances close to 1.0 and for means close to 0. These results suggest better reconstruction performance is obtained by concentrating on the measurement distribution. The main interpretation is that reducing the representation space can improve the CD performance since the variability of the data is reduced. Some visual results of the reconstructions test set examples are shown in Figure 2 employing the best models for each regularization function, where an improvement is presented in regularized models compared with the non-regularized ones.

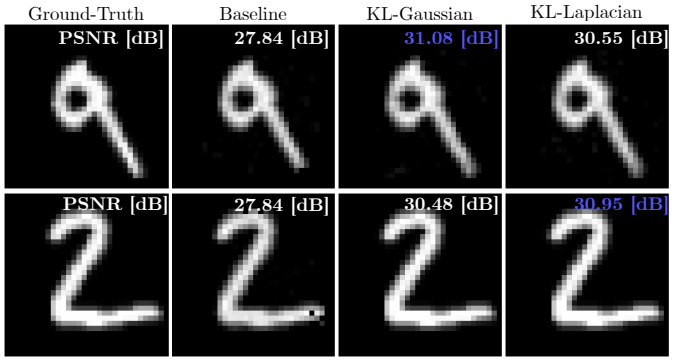

Fig. 2. Visual results of two reconstructed MNIST test images for the non-regularized model and the models trained with the KL-Gaussian and KL-Laplacian regularizers.

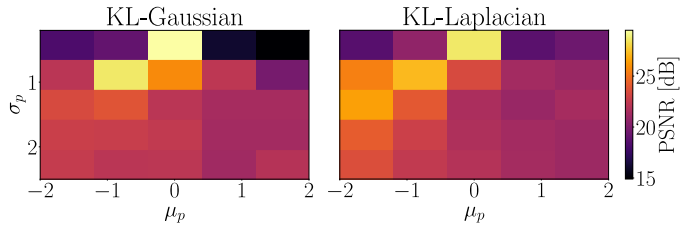

Fig. 3. Recovery performance for the SPC system the KLD regularizers with the Gaussian (left) and Laplacian (right) cases.

B. Single Pixel Camera Setting

For the SPC, we performed experiments on classification and recovery tasks. The classification is performed directly from the compressed measurements without reconstructing the underlying scene. During the training of the E2E network, the parameter of the physical constraint regularizer ρ was dynamically updated during training as suggested in [17], which in the first epochs the ρ is very low, thus not constraining the training of the SL and it is increased to obtain a binary CA. For both the recovery and classification tasks, we employed the Fashion MNIST dataset with 60000 images for training and 10000 for testing. All images were resized to 32×32.

Recovery experiments: For this experiment, we vary the values of μ_p from -2 to 2, and σ_p was changed from 0.1 to 2.0, taking five equispaced values. The CD in this experiment is a UNET [39] with five downsampling and five upsampling blocks. The results of this experiment are shown in Figure 3. Here, the performance obtained is similar to that obtained in the CS case, where lower variance yields better reconstruction performance. Also, similar to the results in Figure 1, the optimal performance is obtained in $\mu_p = 0$, following the concept of batch normalization where the centered output distribution yields more stable training and better performance [40]. Figure 4 presents visual results of two reconstructed test images where the regularized models outperform the baseline model.

Classification experiments: Here, we evaluate the proposed regularization functions on the classification high-level task. The CD is a Mobilnet-V2 [41], which is a lightweight classification network. The same values in the experiment of Figure

 ISBN: 978-1-68558-057-5

TABLE I. OVERALL TEST PERFORMANCE FOR EVERY SETTING. IN BOLD AND UNDERLINED ARE SHOWN THE BEST RESULTS OF EACH EXPERIMENT.

System	Dataset	Task	Metric	Model		
				No Regularized	KL-Gaussian	KL-Laplacian
General Compressive Sensing	MNIST	Recovery	PSNR	31.87	**32.56**	32.42
SPC	Fashion MNIST	Recovery	PSNR	28.35	**29.49**	28.60
	Fashion MNIST	Classification	Accuracy	0.866	**0.886**	0.881
CSA	SEAM Phase II	Recovery	PSNR	34.27	37.38	**41.22**

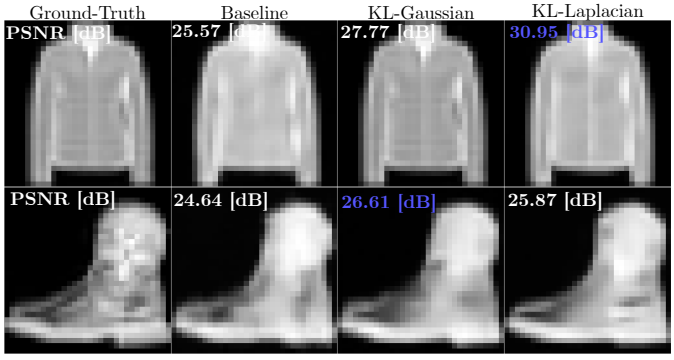

Fig. 4. Visual results of the reconstructed image of the Fashion MNIST dataset in SPC setting for the non-regularized model and the models trained with the KL-Gaussian and KL-Laplacian regularizers.

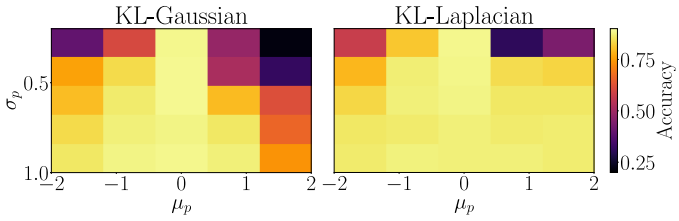

Fig. 5. Classification performance for the SPC system the KLD regularizers with the Gaussian (left) and Laplacian (right) cases.

3 of μ_p and σ_p were used in this scenario. The results are shown in Figure 5, where an opposite performance is obtained compared to the recovery case. Higher variance gives better classification performance.

C. Compressive seismic acquisition setting

For the compressive seismic acquisition, we employed the synthetic dataset SEAM Phase II built by the SEG Advanced Modeling Program (SEAM) during its second project, named "SEAM Phase II–Land Seismic Challenges". The Foothills model is focused on mountainous regions with sharp topography at the surface and high geological complexity at depth, which makes this data set a challenge for seismic data reconstruction [42]. The seismic survey covers a rectangular patch of 1.5×1.2 km with a total sampled depth of 4100 ms. The training and testing datasets comprise 381 images of 128×128. The transmittance value was set to $\delta_0 = 0.6$. The CD network is a convolutional neural network with 5 convolutional layers with 128 filters each. Here we set for both regularizers $\mu_p = 0.5$ and $\sigma_p = 1.6$. Figure 6 shows the reconstruction of two seismic test data, where the best results are obtained by the KL-Laplacian regularization. Nevertheless, the KL-Guassian model outperforms the non-regularized model. Also, it is shown the subsampling vector for each model.

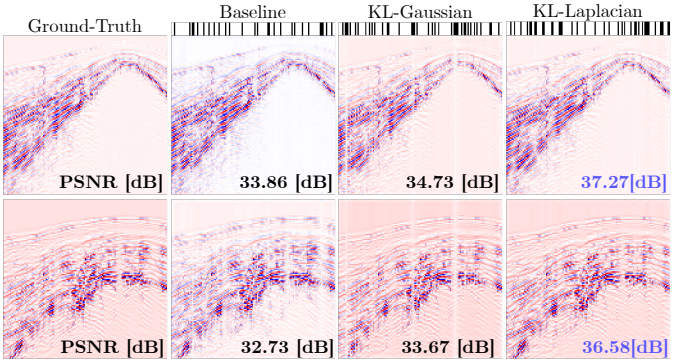

Fig. 6. Visual results of the reconstructed seismic data for the non-regularized model and the models trained with the KL-Gaussian and KL-Laplacian regularizers.

Finally, summarizing the performance of the aforementioned experiments, for the general CS scenario, the SPC and the CSA, Table I presents the test set performance for every experiment. In the CS case, the KL-Gaussian performs better at obtaining almost 1 dB than the non-regularized training. Both regularizers improve the baseline model for the SPC in the recovery task. Similarly, in classification, the optimal performance was obtained by the KL-Gaussian, gaining up to 2% respect to the base E2E model. Finally, in CSA, the KL-Lapacian significantly improved up to 7 [dB] in recovery performance.

VI. CONCLUSION AND FUTURE WORK

We proposed two regularizations based on the KLD end-to-end joint sensing matrix optimization and decoding. The proposed regularizations approximate the distribution of the measurements set to a prior distribution. We show that the low-variance and zero-mean prior distributions yield optimal recovery since they concentrate the training data in the low-dimensional space, thus easing the decoding process. While for the classification task, high-variance and zero-mean priors provide improved classification performance since spreading the distribution allows easier class identification by the decoding network. We validate the performance of the proposed design in a general compressed-sensing case (unconstrained and unstructured sensing matrix), in the single-pixel camera, obtaining up to 1 [dB] and 2% gain in recovery and classification, and for compressive seismic acquisition showing improvements of up to 7 [dB].

 ISBN: 978-1-68558-057-5

ACKNOWLEDGMENT

This work was supported by project 110287780575 through the agreement 785-2019 between the Agencia Nacional de Hidrocarburos and the Ministerio de Ciencia, Tecnología e Innovacion and Fondo Nacional de Financiamiento para la Ciencia, la Tecnología y la Innovacion Francisco José de Caldas.

REFERENCES

[1] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.

[2] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Applied and computational harmonic analysis*, vol. 27, no. 3, pp. 265–274, 2009.

[3] J. M. Bioucas-Dias and M. A. Figueiredo, "A new twist: Two-step iterative shrinkage/thresholding algorithms for image restoration," *IEEE Transactions on Image processing*, vol. 16, no. 12, pp. 2992–3004, 2007.

[4] G. Chen and D. Needell, "Compressed sensing and dictionary learning," *Finite Frame Theory: A Complete Introduction to Overcompleteness*, vol. 73, p. 201, 2016.

[5] W. Dong, G. Shi, X. Li, Y. Ma, and F. Huang, "Compressive sensing via nonlocal low-rank regularization," *IEEE transactions on image processing*, vol. 23, no. 8, pp. 3618–3632, 2014.

[6] J. Zhang, B. Chen, R. Xiong, and Y. Zhang, "Physics-inspired compressive sensing: Beyond deep unrolling," *IEEE Signal Processing Magazine*, vol. 40, no. 1, pp. 58–72, 2023.

[7] Y. Wu, M. Rosca, and T. Lillicrap, "Deep compressed sensing," in *International Conference on Machine Learning*, pp. 6850–6860, PMLR, 2019.

[8] V. Abolghasemi, S. Ferdowsi, B. Makkiabadi, and S. Sanei, "On optimization of the measurement matrix for compressive sensing," in *2010 18th European Signal Processing Conference*, pp. 427–431, IEEE, 2010.

[9] G. Li, Z. Zhu, D. Yang, L. Chang, and H. Bai, "On projection matrix optimization for compressive sensing systems," *IEEE Transactions on Signal Processing*, vol. 61, no. 11, pp. 2887–2898, 2013.

[10] L. Zelnik-Manor, K. Rosenblum, and Y. C. Eldar, "Sensing matrix optimization for block-sparse decoding," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4300–4312, 2011.

[11] J. M. Duarte-Carvajalino and G. Sapiro, "Learning to sense sparse signals: Simultaneous sensing matrix and sparsifying dictionary optimization," *IEEE Transactions on Image Processing*, vol. 18, no. 7, pp. 1395–1408, 2009.

[12] C. F. Gaumond and G. F. Edelmann, "Sparse array design using statistical restricted isometry property," *The Journal of the Acoustical Society of America*, vol. 134, no. 2, pp. EL191–EL197, 2013.

[13] M. Grasmair and V. Naumova, "Conditions on optimal support recovery in unmixing problems by means of multi-penalty regularization," *Inverse Problems*, vol. 32, no. 10, p. 104007, 2016.

[14] L. Baldassarre, Y.-H. Li, J. Scarlett, B. Gözcü, I. Bogunovic, and V. Cevher, "Learning-based compressive subsampling," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 4, pp. 809–822, 2016.

[15] S. Wu, A. Dimakis, S. Sanghavi, F. Yu, D. Holtmann-Rice, D. Storcheus, A. Rostamizadeh, and S. Kumar, "Learning a compressed sensing measurement matrix via gradient unrolling," in *International Conference on Machine Learning*, pp. 6828–6839, PMLR, 2019.

[16] A. Adler, M. Elad, and M. Zibulevsky, "Compressed learning: A deep neural network approach," *arXiv preprint arXiv:1610.09615*, 2016.

[17] J. Bacca, T. Gelvez-Barrera, and H. Arguello, "Deep coded aperture design: An end-to-end approach for computational imaging tasks," *IEEE Transactions on Computational Imaging*, vol. 7, pp. 1148–1160, 2021.

[18] R. Jacome, J. Bacca, and H. Arguello, "D 2 uf: Deep coded aperture design and unrolling algorithm for compressive spectral image fusion," *IEEE Journal of Selected Topics in Signal Processing*, pp. 1–11, 2022.

[19] E. Vargas, J. N. Martel, G. Wetzstein, and H. Arguello, "Time-multiplexed coded aperture imaging: Learned coded aperture and pixel exposures for compressive imaging systems," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 2692–2702, 2021.

[20] H. Arguello, J. Bacca, H. Kariyawasam, E. Vargas, M. Marquez, R. Hettiarachchi, H. Garcia, K. Herath, U. Haputhanthri, B. Singh Ahluwalia, *et al.*, "Deep optical coding design in computational imaging," *arXiv e-prints*, pp. arXiv–2207, 2022.

[21] V. Sitzmann, S. Diamond, Y. Peng, X. Dun, S. Boyd, W. Heidrich, F. Heide, and G. Wetzstein, "End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging," *ACM Transactions on Graphics (TOG)*, vol. 37, no. 4, pp. 1–13, 2018.

[22] H. Arguello, S. Pinilla, Y. Peng, H. Ikoma, J. Bacca, and G. Wetzstein, "Shift-variant color-coded diffractive spectral imaging system," *Optica*, vol. 8, no. 11, pp. 1424–1434, 2021.

[23] A. Hernandez-Rojas and H. Arguello, "3d geometry design via end-to-end optimization for land seismic acquisition," in *2022 IEEE International Conference on Image Processing (ICIP)*, pp. 4053–4057, IEEE, 2022.

[24] J. Xu, Y. Pi, and Z. Cao, "Optimized projection matrix for compressive sensing," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, pp. 1–8, 2010.

[25] Y. Mejia and H. Arguello, "Binary codification design for compressive imaging by uniform sensing," *IEEE Transactions on Image Processing*, vol. 27, no. 12, pp. 5775–5786, 2018.

[26] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[27] T. Nguyen, T. Le, H. Vu, and D. Phung, "Dual discriminator generative adversarial nets," *Advances in neural information processing systems*, vol. 30, 2017.

[28] C. A. Metzler, H. Ikoma, Y. Peng, and G. Wetzstein, "Deep optics for single-shot high-dynamic-range imaging," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1375–1385, 2020.

[29] S. Rifai, G. Mesnil, P. Vincent, X. Muller, Y. Bengio, Y. Dauphin, and X. Glorot, "Higher order contractive auto-encoder," in *Joint European conference on machine learning and knowledge discovery in databases*, pp. 645–660, Springer, 2011.

[30] R. Jacome, A. Hernandez-Rojas, and H. Arguello, "Probabilistic regularization for end-to-end optimization in compressive imaging," in *Computational Optical Sensing and Imaging*, pp. CW1B–1, Optica Publishing Group, 2022.

[31] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE signal processing magazine*, vol. 25, no. 2, pp. 83–91, 2008.

[32] W.-C. Hung, V. Jampani, S. Liu, P. Molchanov, M.-H. Yang, and J. Kautz, "Scops: Self-supervised co-part segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 869–878, 2019.

[33] J. Bacca, L. Galvis, and H. Arguello, "Coupled deep learning coded aperture design for compressive image classification," *Optics express*, vol. 28, no. 6, pp. 8528–8540, 2020.

[34] O. Yilmaz, *Seismic Data Analysis: Processing, Inversion, and Interpretation of Seismic Data*, vol. 1. Society of Exploration Geophysicists, 2008.

[35] C. L. Liner, *Elements of 3D Seismology*. Society of Exploration Geophysicists, jan 2016.

[36] L. D. Lathauwer, B. D. Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM Journal on Matrix Analysis and Applications*, vol. 21, pp. 1253–1278, 1 2000.

[37] F. Chollet *et al.*, "Keras: The python deep learning library," *ascl*, pp. ascl–1806, 2018.

[38] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[39] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pp. 234–241, Springer, 2015.

[40] S. Santurkar, D. Tsipras, A. Ilyas, and A. Madry, "How does batch normalization help optimization?," *Advances in neural information processing systems*, vol. 31, pp. 2488–2498, 2018.

[41] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4510–4520, 2018.

 ISBN: 978-1-68558-057-5

[42] C. Regone, J. Stefani, P. Wang, C. Gerea, G. Gonzalez, and M. Oristaglio, "Geologic model building in seam phase ii — land seismic challenges," *The Leading Edge*, vol. 36, pp. 738–749, 9 2017.

 ISBN: 978-1-68558-057-5