2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)

VAE for Joint Source-Channel Coding of Distributed Gaussian Sources over AWGN MAC

Yashas Malur Saidutta
School of ECE
Georgia Institute of Technology
Atlanta, U.S.A
ysaidutta3@gatech.edu

Afshin Abdi
School of ECE
Georgia Institute of Technology
Atlanta, U.S.A
abdi@gatech.edu

Faramarz Fekri
School of ECE
Georgia Institute of Technology
Atlanta, U.S.A
faramarz.fekri@ece.gatech.edu

***Abstract*—In this paper, we introduce a framework for Joint Source-Channel Coding of distributed Gaussian sources over a multiple access AWGN channel. Although there are prior works that have studied this, they either strongly rely on intuition to design encoders and decoder or require the knowledge of the complete joint distribution of all the distributed sources. Our system overcomes this. We model our system as a Variational Autoencoder and leverage insight provided by this connection to propose a crucial regularization mechanism for learning. This allows us to beat the state of the art by improving the signal reconstruction quality by almost 1dB for certain configurations. The end-to-end learned system is also found to be robust to channel condition variations of ± 5dB and shows a drop in signal reconstruction quality by at most 1dB. Finally, we propose a novel lower bound on the optimal distortion in signal reconstruction and empirically showcase the tightness of the bound in comparison with the existing bound.**

***Index Terms*—joint source-channel coding, distributed encoding, multiple access channels, machine learning, deep learning**

I. Introduction

The beginnings of the research on Joint Source-Channel Coding (JSCC) of analog sources over analog channels can be traced back to the works of Shannon [1] and Kotelnikov [2]. Particularly for Gaussian sources over AWGN channels, there are multiple works performing JSCC in a centralized manner [3], [4]. However, with increasing prevalence of wireless sensor networks, distributed source-channel coding schemes that use Multiple Access Channels (MAC) are becoming very important. In as early as 1980, Cover, El Gamaal and Salehi showed that transmission of distributed correlated sources over a MAC does not obey the separation theorem [5]. With that in mind, we consider the problem of JSCC of Gaussian sources over a linear AWGN MAC. A linear MAC is a wireless channel where the signals of multiple transmitters that are concurrently transmitting add up because of the superposition property of the wireless medium. Such channels are of particular interest because with suitable pre and post-processing any arbitrary function of distributed sources can be recovered [6]. Although, in this paper we do not explore functional compression, the insight gained here will be useful to develop such systems in the future.

This work is supported Jointly by Intel and National Science Foundation under NSF-Intel MLWINS award ID **2003002**.

There have been a couple of works that have explored the problem of JSCC of distributed Gaussian sources over a linear AWGN MAC. Lapidoth et. al. [7] first studied this and derived a lower bound on the optimal distortion achievable. They showed that for a low power region, uncoded transmission is optimal. For the higher power region, they proposed a high dimensional vector quantizer which relies on long-delays. Such systems are complex and unsuitable when the sources have high throughput. Instantaneous or delay free joint source-channel coding of distributed Gaussian sources over AWGN was first explored in [8], [9] for two correlated gaussian sources and subsequently generalized to N sources [10]. However, all these solutions were intuition based. To overcome this reliance on intuition, a purely optimization based method was suggested in [11]. This led to impressive improvements of performance. However, the suggested solution required the complete knowledge of joint distribution of all the source dimensions. To overcome these shortcomings of both the above approaches we propose an optimization driven framework that does not require the knowledge of the joint distribution.

We propose to use machine learning particularly neural networks to solve this optimization problem. Neural networks have been used to perform centralized JSCC in the past [3], [4], [12], [13] and have yielded some impressive insights and results, allowing us to extend to domains beyond the ones supported by human intuition.

Notations: Bold uppercase letters represent random vectors and bold lowercase represent their realization. Uppercase letters represent random variables and lower case letters represent their realization.

II. Problem definition and bounds on optimal distortions

A. Problem definition

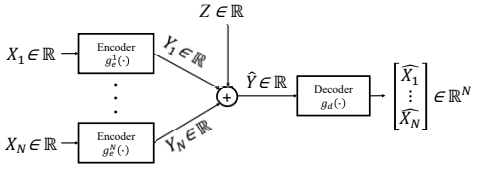


Fig. 1: Distributed encoding framework over a linear MAC AWGN channel

978-1-7281-5478-7/20/$31.00 ©2020 IEEE

Fig. 1 represents the communication system under consideration. There are N distributed Gaussian sources denoted by $X_i \ \forall i \in \{1, ..., N\}$ which are independently encoded by their corresponding encoders $g_e^i(\cdot)$. The output of the encoder $g_e^i(\cdot)$ is denoted as Y_i and its output power is denoted as $P_i^T = E\left[Y_i^2\right]$. The linear MAC AWGN channel is defined as $\hat{Y} = \Sigma_{i=1}^N Y_i + Z$ where, $Z \sim \mathcal{N}(0, \sigma_n^2)$ is the independent AWGN noise. The central decoder receives $\hat{Y}$ and attempts to reconstruct the vector $[X_1, \ldots, X_N]^T$.

Although the neural network framework we propose does not make use of the assumption that $\{X_1, \ldots, X_N\}$ are jointly Gaussian, the assumption is essential to compute the bounds introduced in Sec. II-B. In that vein, $\mathbf{X} \coloneqq [X_1, \ldots, X_N]^T$ and $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma_X})$ where $\mathbf{\Sigma_X}$ is the covariance matrix. Continuing with the same notation, $\hat{\mathbf{X}} \coloneqq \left[\hat{X_1}, \ldots, \hat{X_N}\right]^T$.

B. Lower bound on optimal distortion

In this section we present two lower bounds on the distortion where the second is an extension of the bound in [7]. We assume that all dimensions of $\mathbf{X}$ have the same variance denoted by $\sigma_{\mathbf{X}}^2$ and the correlation coefficient between any two dimensions is the same and is denoted by $\rho_{\mathbf{X}}$. Let R be the total rate to communicate all dimensions of $\mathbf{X}$ and D is the average distortion over all dimensions. D_{opt} represents the average optimal distortion of the system in Fig. 1.

Lemma 1. *For Gaussian sources encoded in a distributed manner to a centralized decoder, provided* $\rho_{\mathbf{X}} \in \left(-\frac{1}{N-1}, 1\right]$ *and* $D < \sigma_{\mathbf{X}}^2$ *we have:*

$$R(D) = \frac{1}{2}\log\frac{N\rho_{\mathbf{X}}\sigma_{\mathbf{X}}^2 + \alpha + \eta}{\eta} + \frac{N-1}{2}\log\frac{\sigma_{\mathbf{X}}^2 + \eta}{\eta} \quad (1)$$

where, $\alpha = (1-\rho_{\mathbf{X}})\sigma_{\mathbf{X}}^2$, $\beta = N(\sigma_{\mathbf{X}}^2 - D)$, $\gamma = N\alpha(N\rho_{\mathbf{X}}\sigma_{\mathbf{X}}^2 + \alpha) - N(N\rho_{\mathbf{X}}\sigma_{\mathbf{X}}^2 + 2\alpha)D$, $\delta = -N\alpha(N\rho_{\mathbf{X}}\sigma_{\mathbf{X}}^2 + \alpha)D$, *and* $\eta = \frac{-\gamma + \sqrt{\gamma^2 - 4\beta\delta}}{2\beta}$.

Proof. This is the same as Lemma 2 [14] when the distributed source dimensions are accessible without noise. □

Lemma 2. *For any two encoder outputs* Y_i *and* Y_j, $\mathbb{E}\left[Y_iY_j\right] \leq |\rho_{\mathbf{X}}|\sqrt{P_i^T P_j^T}$ *if* $\mathbb{E}\left[Y_i\right] = \mathbb{E}\left[Y_j\right] = 0$.

Proof. For any two random variables Y_i, Y_j, the maximum correlation coefficient is defined as [15],

$$R(X_i, X_j) = sup_{\psi_i, \psi_j}\rho\left(\psi_i(X_i), \psi_j(X_j)\right) \quad (2)$$

where $\rho(\cdot)$ represents the pearson correlation coefficient and $\psi_i(\cdot)$ and $\psi_j(\cdot)$ are non-constant functions. Further, when $E\left[X_i|X_j\right] = aX_j$ and $\mathbb{E}\left[X_j|X_i\right] = bX_i$, where a and b are constants, it can be shown that $R(X_i, X_j) = |\rho_{\mathbf{X}}|$ [15]. Since X_i and X_j are jointly Gaussian and their marginal means are zero, the conditions $E\left[X_i|X_j\right] = aX_j$ and $\mathbb{E}\left[X_j|X_i\right] = bX_i$ hold [16]. If $\mathbb{E}\left[Y_i\right] = \mathbb{E}\left[Y_j\right] = 0$, then $\mathbb{E}\left[Y_iY_j\right] \leq |\rho_{\mathbf{X}}|\sqrt{P_i^T P_j^T}$. □

Proposition 1. *The average optimal distortion* D_{opt} *is lower bounded by* D' *where* D' *is the solution of the equation*

$$R(D) = \frac{1}{2}\log_2\left(1 + \frac{\Sigma_{i=1}^N P_i^T + 2\Sigma_{i=1,j>i}^N |\rho_{\mathbf{X}}|\sqrt{P_i^T P_j^T}}{\sigma_n^2}\right) \quad (3)$$

where $R(D)$ *is as given in Lemma 1. This is true provided that the following assumptions hold,*

1) $\mathbb{E}\left[g_e^i(X_i)\right] = 0$.
2) *Assumptions of Lemma 1.*

Proof. We split the linear MAC AWGN into two components the linear MAC and the AWGN component as shown in Fig. 2.

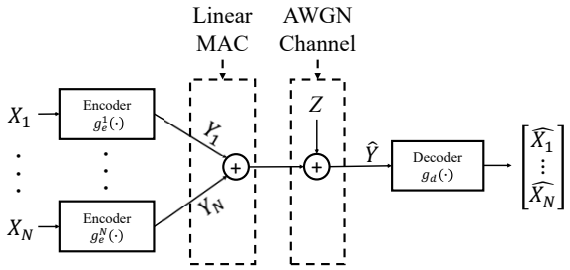


Fig. 2: Distributed communication system broken into subparts.

To compute the capacity of the AWGN component, we compute the input power to the AWGN channel as $P_{AWGN} = \mathbb{E}\left[(Y_1 + ... + Y_N)^2\right]$. Applying Lemma 2, the capacity of the AWGN channel denoted as C is bounded by,

$$C \leq \frac{1}{2}\log_2\left(1 + \frac{\Sigma_{i=1}^N P_i^T + 2\Sigma_{i=1,j>i}^N |\rho_{\mathbf{X}}|\sqrt{P_i^T P_j^T}}{\sigma_n^2}\right) \quad (4)$$

The optimal distortion is the solution to the equation $R(D) = C$. Since, we know only an upper bound on the capacity, solving (3) we get a lower bound on the optimal distortion. □

Proposition 2. *The optimal distortion* D_{opt} *is lower bounded by* D''*, where* D'' *is given by*

$$D'' = \left(\frac{\sigma_n^2 det(\Sigma)}{\Sigma_{i=1}^N P_i^T + 2\Sigma_{i=1,j>i}^N |\rho_{\mathbf{X}}|\sqrt{P_i^T P_j^T} + \sigma_n^2}\right)^{\frac{1}{N}} \quad (5)$$

under the following assumptions

1) $\mathbb{E}\left[g_e^i(X_i)\right] = 0$.
2) *The distortions across all dimensions are the same i.e.* $D_1 = ... = D_N = D$.
3) $\Sigma_{\mathbf{X}} \succeq DI$ *where* I *is the identity matrix of size* N.

Proof. The proof is similar to proof of Theorem $IV.1$ of [7] with an extension to N dimensional sources.

Similar to Prop. 1, we get the same limit on C as in (4).A centralized system, where all dimensions are encoded at a single encoder can achieve as good or better distortion by exploiting the knowledge provided by the other inputs.

Thus, the distortion achieved by it (denoted $D_{opt,cent}$) is a lower bound on the distortion achieved by the separate encoder system i.e. $D_{opt,cent} \leq D_{opt}$. The rate-distortion function for the centralized case is [17]

$$R(D) = \frac{1}{2}\log_2 \frac{det(\Sigma)}{D^N} \Leftrightarrow \Sigma \succeq DI \tag{6}$$

The optimal centralized distortion is achieved by equating the rate and the capacity. Since only a bound on capacity is known we get

$$D_{opt,cent} \geq \left(\frac{\sigma_n^2 det(\Sigma)}{\Sigma_{i=1}^N P_i^T + 2\Sigma_{i=1,j>i}^N |\rho_{\mathbf{X}}| \sqrt{P_i^T P_j^T} + \sigma_n^2} \right)^{\frac{1}{N}} \tag{7}$$

□

III. Connecting to VAEs

We model the encoders and decoders using neural networks. A straightforward loss function to train them is obtained by the Lagrangian of minimizing mean square error under an average power constraint

$$\min_{\Phi_1,\ldots,\Phi_N,\Theta} \frac{1}{N}\mathbb{E}\left[||\mathbf{x} - g_d\left(\Sigma_{i=1}^N g_e^i(x_i) + z\right)||_2^2\right] + \lambda\Sigma_{i=1}^N \mathbb{E}[||g_e^i(x_i)||_2^2] \tag{8}$$

where the encoder neural networks $g_e^i(\cdot)$ are parametrized by Φ_i and the decoder network $g_d(\cdot)$ is parametrized by Θ.

A. VAE Formulation

However, we take a different approach by looking at this problem as a Variational Autoencoder (VAE). VAEs were proposed as generative models trained to produce realistic samples from a latent distribution [18]. The objective of the VAE is to maximize the log-likelihood of the source symbols under the generative model modeled by a neural network which is parametrized by Θ'. Let, $\mathbf{V}$ represent the source symbols and $\mathbf{W}$ represent the latent variable.

$$\begin{aligned}&\mathbb{E}_{\mathbf{V}}\left[\log(p_{\Theta'}(\mathbf{v}))\right] \\ &\geq \mathbb{E}_{\mathbf{V},\mathbf{w}\sim q_{\Phi'}(\mathbf{w}|\mathbf{v})}\left[\log\left(p_{\Theta'}(\mathbf{v}|\mathbf{w})\right) - \log\left(\frac{q_{\Phi'}(\mathbf{w}|\mathbf{v})}{p(\mathbf{w})}\right)\right]\end{aligned} \tag{9}$$

To efficiently train the generative model, another neural network called the encoder that approximates the true posterior distribution over the latent variable given data sample $p_{\Theta'}(\mathbf{w}|\mathbf{v})$ by $q_{\Phi'}(\mathbf{w}|\mathbf{v})$ is used. The entire system is trained by maximizing the Variational Lower bound (VLBO) denoted in the RHS of (9).

In VAEs, $q_{\Phi'}(\mathbf{w}|\mathbf{v})$ is assumed to be of the form $\mathcal{N}(\mathbf{w}; \boldsymbol{\mu}_{\mathbf{W}}(\mathbf{v}), diag(\boldsymbol{\Sigma}_{\mathbf{W}}(\mathbf{v})))$, where $\boldsymbol{\mu}_{\mathbf{W}}(\mathbf{v}), \boldsymbol{\Sigma}_{\mathbf{W}}(\mathbf{v})$ are outputs of the encoder neural network $f_e(\mathbf{v};\Phi')$ parametrized by Φ'. The VLBO in its current form is not amenable to end-to-end gradient descent because of the sampling operation to generate $\mathbf{w}$. However, the assumption that $q_{\Phi'}(\mathbf{w}|\mathbf{v})$ is normally distributed allows us to rewrite the VLBO using the reparametrization trick [18] and results in the setup shown in Fig. 3, where $\odot$ represents elementwise multiplication.

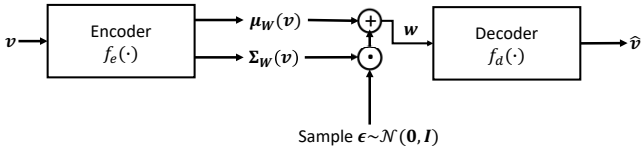


Fig. 3: VAE with reparametrization

This is similar to the original problem that we are solving in Fig. 1, provided that we make certain changes. The covariance matrix of $q_{\Phi'}(\mathbf{w}|\mathbf{v})$ is not generated by the encoder, it is a characteristic of the channel which is constant. This leads to a special architecture of the VAEs called the Constant Variance VAE [19]. We then make the following equivalences,

- The encoder of the VAE models the encoders of the JSCC and the linear part of the MAC i.e. $f_e(\cdot) \equiv \Sigma_{i=1}^N g_e^i(\cdot)$.
- The data sample input to the VAE is equivalent to the symbol we are trying to encode i.e. $\mathbf{x} \equiv \mathbf{v}$.
- The latent variable is equivalent to the noisy received codeword i.e. $\mathbf{w} \equiv \hat{y}$.
- $\Phi' \equiv \Phi := \{\Phi_1, ..., \Phi_N\}$ and $\Theta' \equiv \Theta$.

In JSCC, $q_\Phi(\hat{y}|\mathbf{x})$ becomes $\mathcal{N}\left(\hat{y}; \Sigma_{i=1}^N g_e^i(x_i), \sigma_n^2\right)$ by construction. The choice of $p_\Theta(\mathbf{x}|\hat{y})$ depends on the type of data and the distortion we are attempting to minimize. In our case, we are trying to minimize MSE. Since, maximizing the likelihood of the data under a normal distribution is equivalent to minimizing the MSE, we assume $p_\Theta(\mathbf{x}|\hat{y}) = \mathcal{N}\left(\mathbf{x}; g_d(\hat{y}), \sigma^2 I\right)$, where σ is a hyperparameter. If the data was binary and we were looking to minimize the Hamming distortion, then we would assume $p_\Theta(\mathbf{x}|\hat{y})$ to be a Bernoulli distribution which would lead to cross-entropy loss. By incorporating these we get the loss function for minimization as

$$\begin{aligned}\mathcal{L}(\Phi,\Theta) = \frac{1}{\sigma^2}\mathbb{E}_{\mathbf{X},Z}\left[||\mathbf{x} - g_d\left(\Sigma_{i=1}^N g_e^i(x_i) + z\right)||_2^2\right] \\ - \mathbb{E}_{\mathbf{X},Z}\left[log\left(p_{\hat{Y}}(\Sigma_{i=1}^N g_e^i(x_i) + z)\right)\right]\end{aligned} \tag{10}$$

Eqn. (10) follows by rewriting (9) and dropping the term $\mathbb{E}_{\mathbf{X},Z}\left[log\left(q_\Phi(\Sigma_{i=1}^N g_e^i(x_i) + z|\mathbf{x})\right)\right]$ which represents the differential entropy of the normal distribution $q_\Phi(\hat{y}|\mathbf{x})$. This is because the differential entropy depends only on σ_n^2 which is constant w.r.t. Φ and Θ. Since, we do not know the form of $p_{\hat{Y}}(\cdot)$ we hypothesize that it can be modeled as a Gaussian Mixture Model (GMM) whose parameters are learned during training. Since minimizing the loss function (10) also minimizes the negative log-likelihood of observing the samples of $\hat{Y}$ conditioned on the parameters of the GMM; Φ, Θ and the GMM parameters can all be learned by gradient descent in one unified training step. The value of σ^2 is chosen according to the average power constraint at which we want to design the system.

IV. Implementation

A. Training

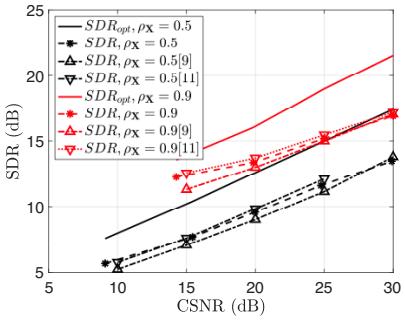
The loss function described in (10) is prone to overfitting because of the GMM training component. Hence we introduce an extra regularization term as,

$$\mathcal{L}_{reg}(\Phi,\Theta,\Omega) = \mathcal{L}(\Phi,\Theta) + \lambda_1 \mathbb{E}_{\mathbf{X}}\left[\Sigma_{i=1}^{N} ||g_e^i(x_i)||_2^2\right] \quad (11)$$

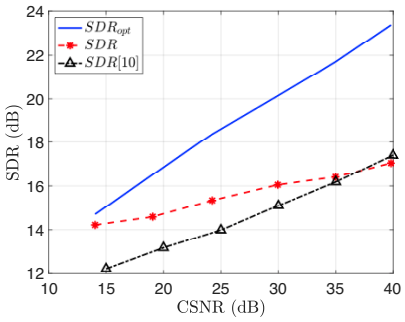
where Ω is the set of parameters of the GMM and λ_1 is the weighting factor. Each of the encoders and the decoders are implemented using fully connected neural networks with three hidden layers of 10 and $10N$ neurons, respectively. The test results are reported over 10^6 samples.

For all experiments we assume that all dimensions of $\mathbf{X}$ have the same variance denoted by $\sigma_{\mathbf{X}}^2$ and the correlation coefficient between any two dimensions is the same denoted by $\rho_{\mathbf{X}}$. This is not a requirement of the system design but is done so that the bounds in Prop. 1 can be used.

B. Simulations



(a) $N = 2$, $\rho_{\mathbf{X}} = 0.5$ and $\rho_{\mathbf{X}} = 0.9$



(b) $N = 3$, $\rho_{\mathbf{X}} = 0.95$

Fig. 4: Simulation results for various N and $\rho_{\mathbf{X}}$ values in comparison with the works of [9]–[11].

Fig. 4 showcases the performance of trained systems in comparison with the works of [9]–[11]. $CSNR$ stands for the Channel Signal to Noise Ratio in dB, defined as $10\log_{10}\frac{(1/N)\Sigma_{i=1}^{N}\mathbb{E}\left[Y_i^2\right]}{\sigma_n^2}$. SDR stands for the Signal to Distortion Ratio in dB, defined as $10\log_{10}\frac{\sigma_{\mathbf{X}}^2}{D}$, where D is the average distortion across all dimensions of $\mathbf{X}$. Our learned system comfortably outperforms the works of [9], [10] for most of the powers tested on, which we consider as SOTA since they do not require knowledge of the joint distribution of $\mathbf{X}$. For example at medium powers like $CSNR = 30$dB for $N = 3$ and $\rho_{\mathbf{X}} = 0.95$, our learned system has an $SDR = 16.05$dB and [10] has an $SDR = 15.10$dB, which showcases that our learned encoders and decoder outperform the existing SOTA by almost 1dB. However, it underperforms at extremely high powers. The following configurations are presented here because they were also tested in [9]–[11] which serve as a basis for our comparison.

One of the key observations in our learned systems and prior work [9]–[11] is the tendency of the system to assign unequal powers to the transmitters in some cases, for example when the correlation is low. Thus, the optimal distortion obtained by solving Prop. 1, which is computed with respect to the power assignments specific to our learned system cannot be compared against the simulated SDR of [9]–[11]. Forcing the system to assign equal powers to the transmitters in such cases leads to performance degradation.

C. Robustness

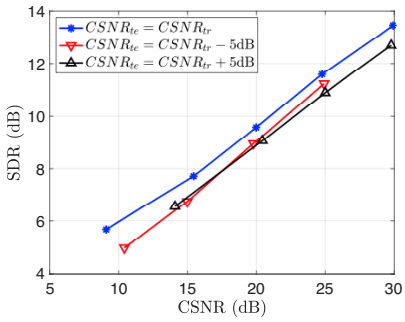


Fig. 5: Robustness of $N = 2$, $\rho_{\mathbf{X}} = 0.5$ with $CSNR_{te} = CSNR_{tr} \pm 5$dB

Fig. 5 portrays the robustness of the system to changes in the $CSNR$ during deployment. $CSNR_{tr}$ represents the $CSNR$ used during the training of the system and $CSNR_{te}$ represents the $CSNR$ used during the test phase. We perturb the $CSNR_{te}$ by $\pm 5dB$ w.r.t. the $CSNR_{tr}$ and plot the changes in SDR. We compare against the SDR of $CSNR_{tr} = CSNR_{te}$. We observe that the SDR drops by 1dB for the worst case and showcases that the system is robust to changes in the channel conditions during deployment without any need for retraining.

D. Effect of GMM regularization

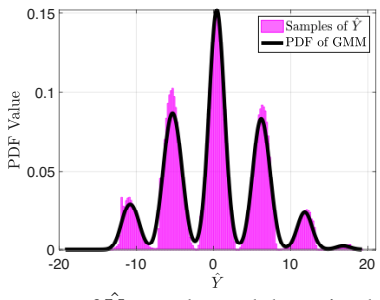
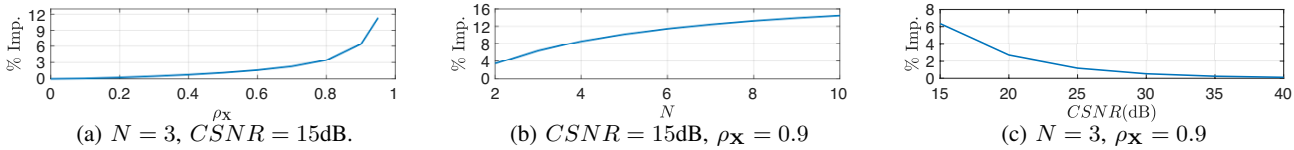


Fig. 6: Histogram of $\hat{Y}$ samples and the trained GMM to model it for $N = 2$, $\rho_{\mathbf{X}} = 0.5$ at $CSNR = 25$dB.



(a) $N = 3$, $CSNR = 15$dB. (b) $CSNR = 15$dB, $\rho_{\mathbf{X}} = 0.9$ (c) $N = 3$, $\rho_{\mathbf{X}} = 0.9$

Fig. 7: Comparison of bounds D' and D''.

Fig. 6 showcases the histogram of the samples of $\hat{Y}$ and the pdf of the GMM that has learned to model it. This extra learned regularization in (11) helps us get better performance of $SDR = 11.59$dB than the standard loss function (8) which gave $SDR = 11.14$dB at $N = 2$, $\rho_{\mathbf{X}} = 0.5$ and $CSNR = 25$dB. Similarly, at $N = 3$, $\rho_{\mathbf{X}} = 0.95$ and $CSNR = 30$dB, it gave a 1dB improvement in the simulated SDR.

E. Comparison of the bounds

In Fig. 7 we compare the bounds from Prop. 1 (D') and Prop. 2 (D''). For these simulations we assumed that all encoders have equal power i.e. $P_1^T = \cdots = P_N^T$. The figures plot the percentage of improvement in the bound, defined as, $\%\ Imp. = 100\frac{D'-D''}{D''}$. Prop. 2 is derived by assuming that all the dimensions of $\mathbf{X}$ are available at central encoder. Since D' does not use that, it is expected and empirically found valid that $D' \geq D''$. A rigorous proof of this is put off for later. In Fig. 7a and Fig. 7b as $\rho_{\mathbf{X}}$ and N increase respectively, $\%\ Imp.$ increases. Without loss of generality we assume that the centralized encoder compresses the sources in the order of X_1 to X_N. When compressing some dimension X_i conditioned on $X_1, ..., X_{i-1}$, the uncertainty in X_i is reduced when either i is large or when $\rho_{\mathbf{X}}$ increases. The centralized encoding is thus able to leverage the lower uncertainty and able to more efficiently compress the symbols when compared to the distributed case which has to encode X_i independently. This is also clear when $\rho_{\mathbf{X}} = 0$ and D' becomes equal to D''. Thus, D'' decays faster than D' and using Prop. 1 provides a much tighter lower bound for performance of system in Fig. 1. Finally, in Fig. 7c, we see that as the $CSNR$ increases both bounds predict approximately the same i.e. $D'' \approx D'$. This is because at higher powers the available rate is high enough that the loss due to compression is negligible and the dominating loss caused by the channel noise affects both systems equally.

V. Conclusion

In this paper, we explored a neural network based solution for designing encoders and decoders for distributed joint source-channel coding of distributed Gaussian sources over a linear AWGN MAC. We have modeled our system as a Variational Autoencoder and leveraged this insight to propose crucial learned regularization that helped us improve the performance beyond the state of the art by sometimes as much as 1dB better reconstruction quality. The learned system was also found to be robust to changes in the channel conditions. Finally, we derived a novel lower bound on the optimal distortion for distributed Gaussian sources over a linear AWGN MAC and empirically showcased its improvement over existing bounds by testing it over various correlations, power constraints and number of sources.

References

[1] C. E. Shannon, "Communication in the Presence of Noise," *Proceedings of the IRE*, vol. 37, no. 1, pp. 10–21, 1949.

[2] V. A. Kotelnikov, *The Theory of Optimum Noise Immunity*. McGraw-Hill, 1959, vol. 34.

[3] Y. M. Saidutta, A. Abdi, and F. Fekri, "M to 1 Joint Source-Channel Coding of Gaussian Sources via Dichotomy of the Input Space Based on Deep Learning," in *Data Compression Conference (DCC)*, 2019.

[4] ——, "Joint Source-Channel Coding of Gaussian Sources Over AWGN Channels via Manifold Variational Autoencoders," in *57th Annual Allerton Coference*, 2019.

[5] T. Cover, A. E. Gamal, and M. Salehi, "Multiple Access Channels with Arbitrarily Correlated Sources," *IEEE Transactions on Information theory*, vol. 26, no. 6, pp. 648–657, 1980.

[6] M. Goldenbaum, H. Boche, and S. Stańczak, "Analog Computation via Wireless Multiple-Access Channels: Universality and Robustness," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2012, pp. 2921–2924.

[7] A. Lapidoth and S. Tinguely, "Sending a Bivariate Gaussian Over a Gaussian MAC," *IEEE Transactions on Information Theory*, vol. 56, no. 6, pp. 2714–2752, 2010. [Online]. Available: https://ieeexplore.ieee.org/document/5466544/

[8] P. A. Floor, A. N. Kim, N. Wernersson, T. A. Ramstad, M. Skoglund, and I. Balasingham, "Distributed Zero-Delay Joint Source-Channel Coding for a Bi-variate Gaussian on a Gaussian MAC," in *2011 19th European Signal Processing Conference*, Conference Proceedings, pp. 2084–2088.

[9] ——, "Zero-Delay Joint Source-Channel Coding for a Bivariate Gaussian on a Gaussian mac," *IEEE Transactions on Communications*, vol. 60, no. 10, pp. 3091–3102, 2012. [Online]. Available: https://ieeexplore.ieee.org/document/6253208/

[10] P. A. Floor, A. N. Kim, T. A. Ramstad, I. Balasingham, N. Wernersson, and M. Skoglund, "On Joint Source-Channel Coding for a Multivariate Gaussian on a Gaussian MAC," *IEEE Transactions on Communications*, vol. 63, no. 5, pp. 1824–1836, 2015.

[11] J. Kron, F. Alajaji, and M. Skoglund, "Low-Delay Joint Source-Channel Mappings for the Gaussian Mac," *IEEE Communications Letters*, vol. 18, no. 2, pp. 249–252, 2014. [Online]. Available: https://ieeexplore.ieee.org/document/6682636/

[12] K. Choi, K. Tatwawadi, T. Weissman, and S. Ermon, "NECST: neural joint source-channel coding," *CoRR*, vol. abs/1811.07557, 2018.

[13] Y. M. Saidutta, A. Abdi, and F. Fekri, "Joint Source-Channel Coding for Gaussian Sources Over AWGN Channels Using Variational Autoencoders," in *IEEE Symposium on Information Theory (ISIT)*, 2019.

[14] Y. Wang, L. Xie, S. Zhou, M. Wang, and J. Chen, "Asymptotic Rate-Distortion Analysis of Symmetric Remote Gaussian Source Coding: Centralized Encoding vs. Distributed Encoding," *Entropy*, vol. 21, no. 2, p. 213, 2019.

[15] A. Dembo, A. Kagan, L. A. Shepp *et al.*, "Remarks on the Maximum Correlation Coefficient," *Bernoulli*, vol. 7, no. 2, pp. 343–350, 2001.

[16] C. K. Williams and C. E. Rasmussen, *Gaussian Processes for Machine Learning*. MIT press Cambridge, MA, 2006, vol. 2, no. 3.

[17] J.-J. Xiao and Z.-Q. Luo, "Compression of Correlated Gaussian Sources Under Individual Distortion Criteria," in *43rd Allerton Conference on Communication, Control, and Computing*, 2005, pp. 438–447.

[18] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," *ArXiv e-prints (Published in ICLR 2014)*, 2013. [Online]. Available: http://arxiv.org/abs/1312.6114

[19] P. Ghosh, M. S. Sajjadi, A. Vergari, M. Black, and B. Schlkopf, "From Variational to Deterministic Autoencoders," *arXiv preprint arXiv:1903.12436*, 2019.