Adaptive Neural Gradient Descent Method for Dual Objective Optimization in ISAC Systems

Shayan Zargari†, Diluka Galappaththige⋆, and Chintha Tellambura⋆
†Huawei Canada, 303 Terry Fox Drive, Suite 400, Ottawa, Ontario K2K 3J1, Canada.
⋆Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, T6G 1H9, Canada.
Email: zargari@ualberta.ca, diluka.lg@ualberta.ca, ct4@ualberta.ca

***Abstract*—This paper proposes an adaptive neural gradient descent (ANGD) method for optimizing the transmit precoder matrix in integrated sensing and communication (ISAC) systems. The proposed method jointly maximizes communication capacity and sensing accuracy by learning an adaptive update rule guided by dual-objective gradient information. Unlike conventional convex-concave procedure algorithms (CCPAs) or fixed-step gradient descent methods, ANGD employs a fully connected neural network (FCN) that interprets the instantaneous gradient of a newly defined capacity-sensing balance function (CSBF). This function unifies communication capacity and sensing beampattern gain into a single optimization metric. Through iterative learning, the FCN adaptively modifies both the precoder update direction and its step size, enabling dynamic convergence under highly non-convex conditions without requiring prior training or explicit CSI modeling. The proposed framework thus provides an unsupervised learning-while-solving paradigm that is computationally efficient and highly generalizable. Simulation results demonstrate that the ANGD approach significantly enhances system performance compared to traditional optimization and learning-based beamforming methods. For instance, with ten transmit antennas at the base station, ANGD achieves a $5.96\,\%$ sum rate gain over the traditional CCPA scheme while preserving sensing beampattern integrity, underscoring its potential for scalable and real-time ISAC applications.**

I. Introduction

Integrated sensing and communication (ISAC) systems exploit a unified hardware and spectrum resource to simultaneously support wireless data transmission and radar-like sensing [1]–[5]. A critical design component in ISAC is the transmit beamforming, which determines both the spatial distribution of the communication signals and the resulting sensing beampattern [1]–[5]. Unlike conventional multiple-input multiple-output (MIMO) systems designed solely for either communication or radar, ISAC beamforming design is inherently a dual-objective optimization problem, i.e., maximizing communication capacity while ensuring accurate sensing performance [1]–[5]. These two objectives are tightly coupled and often conflicting. For example, concentrating transmit energy in a particular direction may enhance sensing resolution but reduce the achievable sum rate for communication users, while distributing energy more evenly across users can degrade sensing accuracy [1]–[5].

Conventional ISAC beamforming methods address this challenge through non-convex optimization frameworks, often solved using convex-concave procedure algorithms (CCPAs) or related iterative methods [1]–[4]. While these approaches provide feasible solutions, they face several limitations [6]–[8]. The optimization landscape is highly non-convex, which makes the outcome sensitive to initialization and prone to suboptimal convergence. They also rely on precise channel state information (CSI) and accurate target response models, assumptions that are difficult to satisfy in mobile or cluttered environments. Furthermore, the computational overhead grows rapidly with the number of transmit antennas, since each iteration requires matrix updates and projections, making real-time implementation impractical in large-scale systems [6]–[8].

A. Previous Contributions

To address these challenges, recent studies have turned to machine learning (ML)-based beamforming methods [6]–[8]. For instance, reference [6] proposes an unsupervised deep learning framework based on a convolutional long short-term memory (LSTM) network that predicts beamformers to maximize communication sum rate. The work [7] presents an attention-based LSTM that captures temporal correlations in reflected signals for improved beamforming design. In the automotive ISAC context, reinforcement learning strategies such as the Wolpertinger policy combined with an actor-critic network are proposed to design adaptive beamforming vectors [8]. These ML approaches relax the reliance on exact CSI and demonstrate adaptability to dynamic environments. However, they also suffer from key limitations: they often require large-scale training datasets, involve high computational complexity, and operate as black-box predictors with limited interpretability. Moreover, once trained, their adaptability to unseen operating conditions can be restricted.

To the best of our knowledge, no prior work has investigated an adaptive neural gradient descent (ANGD) technique for ISAC beamforming optimization. Such an approach directly leverages gradient information from the dual-objective formulation while introducing neural adaptability to overcome the rigidity of conventional optimization methods.

B. Motivation and Our Contributions

We thus propose an ANGD method to enhance ISAC performance. To this end, we optimize the transmit beamforming matrix by striking a balance between maximizing communication capacity and optimizing sensing performance through a new metric known as the capacity-sensing balance function (CSBF).

The innovation of the ANGD method lies in its use of a neural network model to adaptively interpret and use gradient information derived from CSBF [9]. Unlike traditional gradient

descent (GD) approaches that follow a predefined path based on the gradient of the objective function, ANGD employs a fully connected network (FCN) as an adaptive mechanism [10], [11]. This FCN dynamically updates its parameters in each iteration based on the current gradient, effectively navigating the complex optimization challenges of ISAC systems. The process begins with the gradient computation of the CSBF over the current transmit precoder matrix. This gradient is then fed into the FCN, which outputs an updated direction for the precoder matrix. In each iteration, the precoder matrix is updated, gradually moving towards an optimal solution that maximizes the CSBF.

With no prior work on ANGD in ISAC, we offer an innovative ANGD-based learning paradigm to enhance ISAC system performance. In particular, we propose a new performance metric, CSBF, to strike a compromise between communication and sensing functionalities. The key contributions of this study are summarized as follows:

- We introduce a new resource allocation framework employing an ANGD-based learning algorithm to outperform classical and modern learning-based optimization methods. We also offer a new ISAC performance metric that balances communication and sensing.
- To the best of our knowledge, this is the first time an ANGD algorithm has been utilized in ISAC systems, effectively dealing with the proposed problem (P1)'s non-convexity.
- Our suggested FCN has a low computational complexity and memory requirements compared to complex DL algorithms, e.g., convolutional neural networks (CNNs). While ensuring both communication and sensing requirements, this technique is applicable in a wide range of practical scenarios without pre-training, owing to its unsupervised learning-while-solving approach.
- Our numerical results demonstrate significant improvements in communication sum rate while maintaining sensing beampattern gain, proving the potential of ANGD for ISAC applications. For instance, with ten transmit at the BS, the suggested ANGD scheme achieves a $5.96\,\%$ sum rate gain over the CCPA.

Notation: Vectors and matrices are expressed by boldface lowercase letters $\mathbf{a}$ and capital letters $\mathbf{A}$, respectively. For a square matrix $\mathbf{A}$, $\mathbf{A}^{\mathrm{H}}$ and $\mathbf{A}^{\mathrm{T}}$ are Hermitian conjugate transpose and transpose of a matrix, respectively. $\mathbf{I}_M$ denotes the M-by-M identity matrix. $\mathbf{0}_M$ is the M-dimensional all-zero vector. The Euclidean norm of a complex vector and the absolute value of a complex scalar are denoted by $\|\cdot\|$ and $|\cdot|$, respectively. Expectation is denoted by $\mathbb{E}[\cdot]$. The distribution of a circularly symmetric complex Gaussian (CSCG) random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$ is denoted by $\sim \mathcal{CN}(\boldsymbol{\mu}, \mathbf{C})$. Besides, $\mathbb{C}^{M\times N}$ and $\mathbb{R}^{M\times 1}$ represent $M \times N$ dimensional complex matrices and $M \times 1$ dimensional real vectors, respectively. Further, $\mathcal{O}$ expresses the big-O notation. Finally, $\mathcal{K} \triangleq \{1,\ldots,K\}$, $\mathcal{N} \triangleq \{1,\ldots,N\}$, and $\mathcal{K}_k \triangleq \mathcal{K}\setminus\{k\}$.

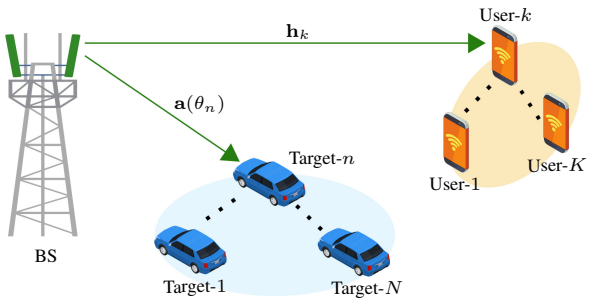

Fig. 1: System model of an ISAC system where a BS utilizes a shared antenna array to simultaneously communicate with K users and N sensing targets.

II. System, Channel, and Signal Models

This section describes the proposed ISAC system, channel, and signal models.

A. System and Channel Models

We consider an ISAC system (Fig. 1), which consists of an M-antenna BS with uniform linear array (ULA) antennas, K single-antenna communication users, and N targets. We assume the BS antennas are spaced at half-wavelengths [12]. During downlink communication with the users, the BS also conducts radar sensing toward N potential targets. It is assumed that the users and targets remain static during a coherence time [3].

We employ block flat-fading channel models. During each fading block, $\mathbf{h}_k \in \mathbb{C}^{M\times 1}$ and $\mathbf{a}(\theta_n) \in \mathbb{C}^{M\times 1}$ denote the channels between the BS and the k-th user and the BS and the n-th target, respectively. The ISAC system can use a time division duplex (TDD) approach for channel estimation by employing techniques such as least squares (LS) or minimum mean square error (MMSE). It is thus assumed that both the BS and users have complete CSI knowledge, while the BS has a general awareness of sensing directions [3], [12]. The pure communication channels are modeled as Rayleigh fading and given by

$$\mathbf{h}_k = \zeta_{h_k}^{1/2}\tilde{\mathbf{h}}_k,\ k \in \mathcal{K}, \tag{1}$$

where ζ_{h_k} is the large-scale path loss, which stays constant for several coherence intervals, and $\tilde{\mathbf{h}}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$ accounts for the small-scale Rayleigh fading.

To model the sensing channels between the BS and the targets, we follow the echo signal representation in MIMO radar systems [3]. Thus, we model these channels as line-of-sight (LoS) channels using transmit array steering vectors to the direction θ_n as [3]

$$\mathbf{a}(\theta_n) = \frac{1}{\sqrt{M}}\left[1, e^{j\pi\sin(\theta_n)}, \ldots, e^{j\pi(M-1)\sin(\theta_n)}\right]^{\mathrm{T}},\ n \in \mathcal{N}, \tag{2}$$

where θ_n denotes the direction of the n-th target with respect to the x-axis of the coordinate system.

B. Signal Model

The BS emits a signal denoted by $\mathbf{x} \in \mathbb{C}^{M\times 1}$, which is designed for joint communication and sensing [13]–[15]. The BS transmitted signal is thus expressed as

$$\mathbf{x} = \sum_{k\in\mathcal{K}} \mathbf{w}_k q_k, \tag{3}$$

where $q_k \in \mathbb{C}$ is the intended data symbol for the k-th user with unit power, i.e., $\mathbb{E}\{|q_k|^2\} = 1$, and $\mathbf{w}_k \in \mathbb{C}^{M\times 1}$ is the BS transmit beamforming vector for the k-th user. Furthermore, the BS transmit beamforming is intended to generate an effective beampattern pointing toward prospective target directions of interest, resulting in a higher radar receive signal-to-noise ratio (SNR) and improved sensing performance [13]–[15].

The received signal at the k-th user is given by

$$\begin{aligned} y_k &= \mathbf{h}_k^{\mathrm{H}}\mathbf{x} + z_k \\ &= \mathbf{h}_k^{\mathrm{H}}\mathbf{w}_k q_k + \sum_{i\in\mathcal{K}_k} \mathbf{h}_k^{\mathrm{H}}\mathbf{w}_i q_i + z_k,\ k \in \mathcal{K}, \end{aligned} \tag{4}$$

where $z_k \sim \mathcal{CN}(0, \sigma^2)$ is the k-th user additive white Gaussian noise (AWGN). It is assumed that the noise variance remains consistent across all users.

Remark 1. *The proposed system employs communication beamforming for sensing, thereby preventing excessive interference to communication users while enabling dual functionality [13]–[15]. Although dedicated sensing beamforming can provide additional degrees of freedom at the BS, it also introduces higher interference to communication users who cannot fully suppress the sensing signals. To overcome this limitation, our approach adopts a unified beam for both communication and sensing [13]–[15].*

III. Communication and Sensing Performance

In this section, we derive communication rates of the users and the transmit beampatterns for the targets to evaluate and optimize the ISAC system.

The users exploit the received signal from the BS to decode their intended data. To this end, the received signal-to-noise-to-interference ratio (SINR) at the k-th user can be written as

$$\gamma_k = \frac{|\mathbf{h}_k^{\mathrm{H}}\mathbf{w}_k|^2}{\sum_{i\in\mathcal{K}_k} |\mathbf{h}_k^{\mathrm{H}}\mathbf{w}_i|^2 + \sigma^2},\ k \in \mathcal{K}. \tag{5}$$

Thereby, the rate of the k-th user is approximated as $\mathcal{R}_k = \log_2(1+\gamma_k)$ for $k \in \mathcal{K}$.

For sensing, the transmit beampattern is one of the major performance metrics extensively used in the literature for MIMO radar signal design [16]. We thus focus on its design since an appropriate beampattern design results in improved sensing performance in terms of detection, sensing, or recognition through proper echo wave processing [16]. The transmit beam pattern illustrates the power distribution of the transmit signal relative to the sensing angle $\theta \in [-\pi/2, \pi/2]$. Our ISAC system design uses the BS-transmitted signal for information transmission and target sensing. The beam pattern gain at a certain angular direction, θ_n, is mathematically defined as [2]

$$p(\theta_n) = \mathbf{a}^{\mathrm{H}}(\theta_n)\left(\sum_{k\in\mathcal{K}} \mathbf{w}_k\mathbf{w}_k^{\mathrm{H}}\right)\mathbf{a}(\theta_n),\ n \in \mathcal{N}. \tag{6}$$

A practical sensing system designs the transmit beam pattern depending on the radar target sensing requirements. For example, if the detection task is conducted without knowing the direction of potential targets, it employs a uniformly distributed beampattern. Conversely, if the system has a general idea of the targets' directions, it only maximizes beampattern gains in these potential directions [16].

This study assesses communication and sensing efficiency using communication SINR and sensing beampattern gain as metrics. High communication SINR enhances symbol detection and reduces error chances, indicating communication quality. Sensing success relies on beampattern gain, with higher gain increasing target detection probability through effective transmit beamforming [13]–[15].

IV. Problem Formulation

In this section, we propose a new framework designed for ISAC to achieve dual-purpose optimization. Our approach focuses on simultaneously enhancing the overall data rate for communication users and optimizing the effectiveness of the sensing beampattern. We introduce the CSBF to capture this dual objective succinctly. This new objective function is structured to seamlessly integrate the goals of maximizing communication capacity and sensing accuracy into a single metric, achieving an equilibrium between these two functions. Accordingly, the problem is formulated as follows:

$$\text{(P1)}:\ \max_{\mathbf{W}}\quad f(\mathbf{W}) = \sum_{k\in\mathcal{K}} \log_2(1+\gamma_k) + \omega \sum_{n\in\mathcal{N}} p(\theta_n), \tag{7a}$$

$$\text{s.t}\quad \mathrm{Tr}(\mathbf{W}\mathbf{W}^{\mathrm{H}}) \le p_{\max}, \tag{7b}$$

where $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_K] \in \mathbb{C}^{M\times K}$ and (7b) is the BS transmit power constraint with maximum allowable transmit power $p_{\max}$. The parameter ω is a lever to adjust the balance between the communication and sensing priorities.

Note that (P1) is non-convex. However, a generic GD solution can be applied to solve (P1) by iteratively updating $\mathbf{W}$. The update rule at the t-th iteration can be written as

$$\mathbf{W}_{t+1} = \mathbf{W}_t + \eta_{\mathbf{W}_t} \cdot \phi(\nabla_{\mathbf{W}_t} f(\mathbf{W}_t)), \tag{8}$$

where $\nabla_{\mathbf{W}_t} f(\mathbf{W}_t)$ is the gradient of the CSBF over the current beamforming matrix, $\phi(\cdot)$ denotes a custom update function, and $\eta_{\mathbf{W}_t}$ represents the step size at the t-th iteration. However, the GD-based solution can be stuck at saddle points or bad local optima, where the gradient vanishes. The original problem (P1) may be approximated through convex techniques to address this. This might include adopting the successive convex approximation (SCA) strategy to transform the objective function into a simpler linear form, followed by leveraging the semidefinite relaxation (SDR) method for optimizing the beamforming vector.

Algorithm 1 : ANGD Algorithm

Require: Initialize $\mathbf{W}^0, \boldsymbol{\varphi}^0$, learning rates $\eta_{\mathbf{W}}$ and $\eta_{\boldsymbol{\varphi}}$.
for $t = 0$ **to** T **do**
  Compute gradient $\Delta\mathbf{W} = \nabla_{\mathbf{W}_k} f(\mathbf{W}_k)$.
  Apply step size: $\Delta\mathbf{W}_{\text{scaled}} = \eta_{\mathbf{W}} \cdot \Delta\mathbf{W}$.
  Update precoder: $\mathbf{W}_{t+1} = \mathbf{W}_t + \Delta\mathbf{W}_{\text{scaled}}$.
  Project $\mathbf{W}_{t+1}$ onto power constraint based on (11).
  Adjust $\boldsymbol{\varphi}$ using its gradient: $\Delta\boldsymbol{\varphi} = \Lambda(\nabla_{\boldsymbol{\varphi}_t} f(\mathbf{W}_{t+1}))$.
  Apply learning rate and update $\boldsymbol{\varphi}_{t+1}$: $\boldsymbol{\varphi}_{t+1} = \boldsymbol{\varphi}_t + \eta_{\boldsymbol{\varphi}} \cdot \Delta\boldsymbol{\varphi}$.
  Evaluate CSBF using $f(\mathbf{W}_{t+1})$ in (P1).
  If $f(\mathbf{W}_{t+1})$ improves, checkpoint $\mathbf{W}_{t+1}$ and $\boldsymbol{\varphi}_{t+1}$.
end for
Output: Optimized precoder $\mathbf{W}^*$.

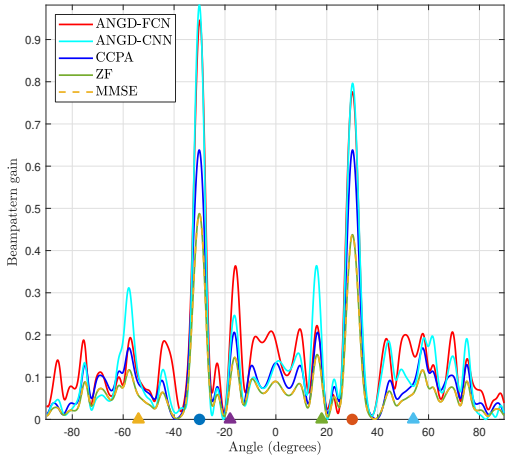

Fig. 2: Beamforming gain comparison with $M = 16$ across a $\pm 90^\circ$ angular spread.

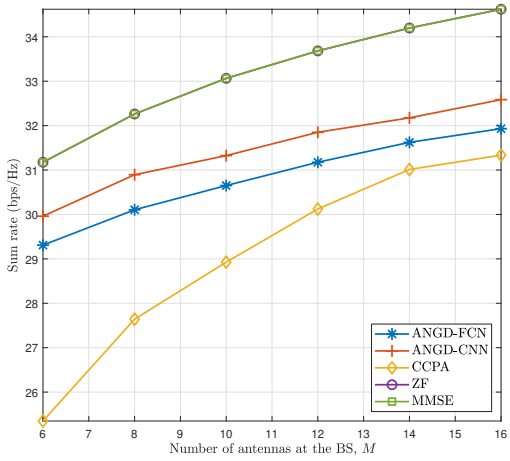

Fig. 3: Sum rate versus the number of BS antennas, M, for various algorithms with $p_{\max} = 30$ dBm.

V. Proposed ANGD for solving CSBF

To solve (P1), we introduce an innovative approach known as the ANGD method, designed for dynamically and adaptively optimizing the transmit precoder matrix. This approach updates the precoder matrix, $\mathbf{W}$, in each iteration using a mechanism defined by an FCN. The parameters guiding this update are also adjusted during each cycle. Unlike the conventional method that relies on a predefined update function $\phi(\cdot)$, our method uses an FCN, represented as $\Phi_{\boldsymbol{\varphi}_t}(\cdot)$, to establish the update rule. Here, $\boldsymbol{\varphi}_t$ represents the parameters used by the function $\Phi_{\boldsymbol{\varphi}_t}(\cdot)$ during the t-th iteration. The FCN takes the gradient $\nabla_{\mathbf{W}_t} f(\mathbf{W}_t)$ as input and produces the modification for the upcoming precoder matrix as follows:

$$\mathbf{W}_{t+1} = \mathbf{W}_t + \Phi_{\boldsymbol{\varphi}_t}(\nabla_{\mathbf{W}_t} f(\mathbf{W}_t)). \tag{9}$$

The ANGD method adjusts its optimization approach on the fly by dynamically updating the FCN parameters, $\boldsymbol{\varphi}_t$, at each step [10], [11]. This update is done through back-propagation, considering the current value of the CSBF. This method allows training and problem-solving simultaneously, removing the need for a pre-defined training dataset [17]. To maximize the CSBF, we use the Adam optimizer to update the FCN parameters represented below:

$$\boldsymbol{\varphi}_{t+1} = \boldsymbol{\varphi}_t + \eta_{\boldsymbol{\varphi}} \cdot \Lambda(\nabla_{\boldsymbol{\varphi}_t} f(\mathbf{W}_{t+1})), \tag{10}$$

where $\eta_{\boldsymbol{\varphi}}$ is the learning rate and Λ denote the Adam optimizer function of the FCN [18]. During each iteration, we perform a projection procedure on the transmit precoder matrix $\mathbf{W}$ to meet the total power limitation. It adjusts $\mathbf{W}$ such that its total power does not exceed the predefined threshold $p_{\max}$, which is given by

$$\text{Proj}_{\text{power}}(\mathbf{W}) = \begin{cases} \mathbf{W}, & \text{if } \text{Tr}(\mathbf{W}\mathbf{W}^{\text{H}}) \le p_{\max}, \\ \dfrac{\sqrt{p_{\max}}\mathbf{W}}{\sqrt{\text{Tr}(\mathbf{W}\mathbf{W}^{\text{H}})}}, & \text{otherwise.} \end{cases} \tag{11}$$

Algorithm 1 outlines the structure of the ANGD method, which has a complexity of $\mathcal{O}(KMT)$. Unlike gradient updates, the ANGD leverages a deep neural model $\Phi_{\boldsymbol{\varphi}_t}(\cdot)$ to interpret the gradient information. This method allows for more refined adjustments, effectively dealing with the complicated optimization

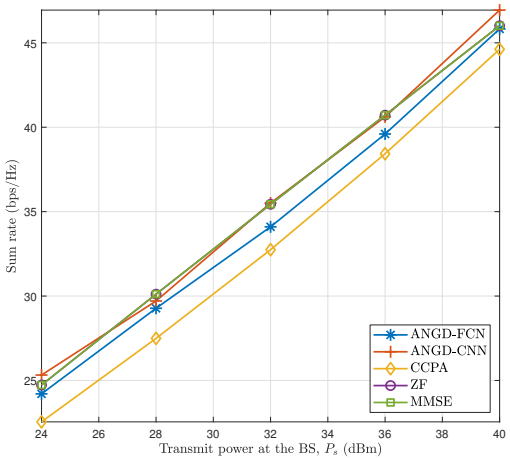

Fig. 4: sum rate versus the transmit power at the BS, P_s, for various algorithms with $M = 10$.

challenges of setting transmit precoders. The process is further refined by incorporating adaptive learning rates and a projection step to ensure compliance with power constraints.

The FCN designed for ISAC optimization integrates an input layer for gradients, two hidden layers with rectified linear unit (ReLU) activations for processing, and an output layer for updates. This setup transforms the input into a higher-dimensional space, enhancing the network's ability to learn from gradient changes and ensuring non-linearity. The output layer, aligned with the input's dimension, finalizes the update process.

The ANGD strategy significantly advances solving the non-convex CSBF maximization problem. It goes beyond the conventional GD and convex optimization algorithms by providing a more refined and less aggressive approach to optimization. This approach prevents the algorithm from being stuck in suboptimal local minima and saddle points. It achieves this by integrating $\Phi_{\boldsymbol{\varphi}_t}(\cdot)$ into the process, enabling updates to proceed even in the absence of gradients [17].

VI. Simulation Results

In this section, we present simulation results to evaluate the efficacy of the proposed ANGD algorithm within ISAC systems. The simulation parameters include a noise power (σ^2) of $-80\,\mathrm{dBm}$, a maximum transmit power ($p_{\max}$) of $30\,\mathrm{dBm}$, a learning rate for the FCN parameters ($\eta_{\boldsymbol{\varphi}}$) and step size for the transmit precoder ($\eta_{\mathbf{W}}$) of 10^{-4} and 1, respectively. The setup involves $M = 16$ antennas, $K = 2$ users, $N = 4$ targets, and runs for $T = 500$ steps. Additional parameters include $\omega = 10^4$. To establish a centralized communication point, the BS is positioned at coordinates $\{0, 0\}$. The targets and users are randomly distributed within circular regions centered at $\{50, 0\}$ and $\{50, 30\}$, respectively, each with a radius of $20\,\mathrm{m}$. Users are positioned at angles of -30° and 30° while the sensing directions from the BS are set at -54°, -18°, 18°, and 54°, covering a wide angular range. The simulation consists of 10^3 Monte Carlo trials to ensure robust statistical validity. Furthermore, the path loss model is given by

$$L(d) = C_0 \left(\frac{d}{D_0} \right)^{-\nu}, \tag{12}$$

where $C_0 = -30\,\mathrm{dB}$ represents the path loss at the reference distance $D_0 = 1\,\mathrm{m}$, d is the individual link distance, and $\nu = 2$ is the path loss exponent. We name our method ANGD-FCN and compare it against the following benchmarks.

1) *ZF beamforming:* The zero-forcing (ZF) beamforming, denoted as $\mathbf{W}_{\mathrm{ZF}} = \mathbf{H}(\mathbf{H}^{\mathrm{H}}\mathbf{H})^{-1}$, completely nullifies inter-user interference by projecting each user's signal onto the orthogonal subspace of the others. This scheme thus achieves interference-free transmission when the BS has at least as many antennas as users ($M \geq K$). However, ZF focuses solely on communication performance and does not account for sensing targets, resulting in limited control over the beampattern and reduced sensing accuracy in ISAC systems.

2) *MMSE beamforming:* The MMSE beamforming aims to minimize the total mean squared error between the estimated (transmitted) and actual (received) user signals, achieving a trade-off between interference suppression and noise amplification. The beamforming matrix is expressed as $\mathbf{W}_{\mathrm{MMSE}} = \mathbf{H}\left(\mathbf{H}^{\mathrm{H}}\mathbf{H} + \sigma^2\mathbf{I}_K\right)^{-1}$. Unlike ZF, MMSE maintains better robustness in low-SNR or highly correlated channel conditions, balancing interference cancellation and power efficiency. However, similar to ZF, it is optimized purely for communication objectives and neglects the sensing aspect, offering no explicit control over the transmit beampattern toward sensing targets.

Furthermore, $\mathbf{H} \in \mathbb{C}^{M\times K}$ represents the channel matrix that contains the channel vectors of all communication users, with the k-th column vector denoted as $\mathbf{h}_k$ for $k \in \mathcal{K}$.

3) *CCPA method:* We also evaluate beamforming performance using a CCPA that iteratively solves a series of convex approximations of the non-convex problem (P1). The CCPA framework employs SDR and SCA techniques to reformulate the original optimization. Specifically, by defining $\mathbf{W}_k = \mathbf{w}_k\mathbf{w}_k^{\mathrm{H}}$ and noting that $\mathbf{W}_k$ is positive semidefinite with $\mathrm{Rank}(\mathbf{W}_k) = 1$, the rank constraint is relaxed to obtain a tractable SDP [19], [20]. This relaxed problem can be efficiently solved using convex optimization tools such as CVX [21]. Although the CCPA method guarantees convergence to a stationary point, it generally incurs high computational complexity and may converge to suboptimal local solutions, especially in large-scale ISAC scenarios.

4) *ANGD-CNN:* We also explore the utilization of CNN as an alternative to FCNs in Algorithm 1. The CNN begins with an input handling layer designed to accommodate two channels, effectively processing the real and imaginary components of the input signals. Within the model, two one-dimensional (1D) layers are employed to extract features from the input data sequentially. Each of these layers is configured with 16 filters, utilizing kernels of size 3, and employs padding to ensure that the output dimensionality matches that of the input. After each convolutional layer, the ReLU activation function is applied, introducing non-linearity to the model and enabling it to capture intricate patterns within the data. Finally, the model incorporates a fully connected layer that maps the extracted features to the output space. The complexity of ANGD-CNN is approximately given by $\mathcal{O}(TKM\sum_{l=1}^{L} n_{l-1}s_l^2 n_l)$, where L is the number of filters, s_l denotes the size of the filter in the l-th Conv layer, while n_{l-1} and n_l denote the size of the input and the output of l-th Conv layer, respectively.

In Fig. 2, the circles and triangles on the x-axis represent communication users and sensing targets. The location of these markers indicates the respective positions or angles at which users and targets are experiencing the beamforming gain from the BS's signals. It provides a comparative analysis of the beamforming gains achieved using an antenna array with $M = 8$ elements at the BS across different beamforming algorithms. The plots provide a visual assessment of the directivity and performance of each beamforming strategy by illustrating their directional gain patterns. This comparative study is instrumental in determining each algorithm's ability to effectively focus emitted power toward intended directions – a key factor for enhancing spatial filtering and reducing interference. The sharpness of the main lobe peaks in the plots indicates an algorithm's beam steering accuracy. In contrast, the depth of the dips reflects

each algorithm's capability to suppress signal reception from undesired directions, as evidenced by the sidelobe attenuation. The ANGD-FCN and ANGD-CNN algorithms demonstrate superior directivity, which is evident in both diagrams. Although the CCPA algorithm shows improved directivity with the increased array size, it slightly lags behind the proposed ANGD-FCN and ANGD-CNN methods regarding peak sharpness and sidelobe suppression.

Fig. 3 shows a sum rate comparison across various beamforming algorithms as a function of the number of BS antennas for $p_{\max} = 30\,\mathrm{dBm}$. It shows that an increase in the number of antennas at the BS leads to a proportional enhancement in the sum rate for all considered algorithms, capitalizing on the advantages of spatial multiplexing that a larger antenna array affords. The ZF and MMSE algorithms significantly show parallel performance curves, consistently outperforming other methods, regardless of the number of antennas employed. ZF and MMSE algorithms outperform the CCPA method by achieving higher signal peaks directed at users and having lower sidelobes than ANGD-FCN and ANGD-CNN, indicating more focused transmission and less interference. Although the ANGD-FCN and ANGD-CNN methods fall short of minimizing the sidelobes demonstrated by ZF and MMSE, they still achieve a notable gain over the CCPA algorithm. For instance, with $M = 10$, ANGD-FCN and ANGD-CNN, respectively, archive $5.96\,\%$ and $8.29\,\%$ sum rate gain over CCPA. CCPA performs inferior to all compared algorithms, indicating its limits in maximizing throughput.

Fig. 4 depicts how the sum rate scales with increasing BS transmit power P_s for $M = 10$. As observed, the ZF and MMSE algorithms exhibit identical performance, indicating that they are equally proficient at maximizing the sum rate with no significant difference in spectral efficiency. Furthermore, they exhibit superior sum rate efficiency compared to the ANGD-FCN and ANGD-CNN algorithms. Meanwhile, these algorithms' performance is intermediate, positioned between the CCPA and the leading ZF/MMSE. It records a more pronounced sum rate increase in response to a boost in transmit power than the CCPA approach. For example, the ANGD-FCN and ANGD-CNN schemes outperform the CCPA scheme by $4.14\,\%$ and $8.44\,\%$, respectively, with $P_s = 32\,\mathrm{dBm}$. Conversely, although positive, CCPA's progression in sum rate is less marked with higher transmit power, implying it may not be as adept at translating additional power into throughput gains compared to the proposed algorithms.

VII. Conclusion

The proposed ANGD algorithm is a significant breakthrough in optimizing ISAC systems. It uniquely integrates communication capacity and sensing accuracy by employing the CSBF. Unlike traditional methods, ANGD leverages a deep-based update mechanism to dynamically adjust the transmit precoder matrix, allowing it to navigate complex optimization landscapes effectively. Through adaptive learning, ANGD ensures efficient convergence towards efficient solutions, outperforming conventional CCPA beamforming algorithms. Simulation results confirm ANGD's superiority over existing methods, highlighting its potential to revolutionize ISAC system design by enabling enhanced performance and efficiency across various applications.

References

[1] X. Liu, T. Huang, N. Shlezinger, Y. Liu, J. Zhou, and Y. C. Eldar, "Joint transmit beamforming for multiuser MIMO communications and MIMO radar," *IEEE Trans. Signal Process.*, vol. 68, pp. 3929–3944, Jul. 2020.

[2] Z. He, W. Xu, H. Shen, Y. Huang, and H. Xiao, "Energy efficient beamforming optimization for integrated sensing and communication," *IEEE Wireless Commun. Lett.*, vol. 11, no. 7, pp. 1374–1378, Jul. 2022.

[3] Z. He, W. Xu, H. Shen, D. W. K. Ng, Y. C. Eldar, and X. You, "Full-duplex communication for ISAC: Joint beamforming and power optimization," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 9, pp. 2920–2936, Sept. 2023.

[4] D. Galappaththige, M. Mohammadi, G. A. Baduge, and C. Tellambura, "Cell-free integrated sensing and communication: Principles, advances, and future directions," *arXiv*, 2025.

[5] A. Hakimi, D. Galappaththige, and C. Tellambura, "A roadmap for NF-ISAC in 6G: A comprehensive overview and tutorial," *Entropy*, vol. 26, no. 9, Sept. 2024.

[6] C. Liu *et al.*, "Learning-based predictive beamforming for integrated sensing and communication in vehicular networks," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 8, pp. 2317–2334, Aug. 2022.

[7] Z. Wang and V. W. Wong, "Deep learning for ISAC-enabled end-to-end predictive beamforming in vehicular networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2023, pp. 5713–5718.

[8] L. Xu, R. Zheng, and S. Sun, "A deep reinforcement learning approach for integrated automotive radar sensing and communication," in *Proc. IEEE 12th Sensor Array Multichannel Signal Process. Workshop (SAM)*, Jun. 2022, pp. 316–320.

[9] F. Mehmood, S. Ahmad, and T. K. Whangbo, "An efficient optimization technique for training deep neural networks," *Math.*, vol. 11, no. 6, Mar. 2023.

[10] J. Kim, H. Lee, S.-E. Hong, and S.-H. Park, "Deep learning methods for universal MISO beamforming," *IEEE Wireless Commun. Lett.*, vol. 9, no. 11, pp. 1894–1898, Nov. 2020.

[11] S. Zargari, A. Hakimi, C. Tellambura, and A. Maaref, "Enhancing AmBC systems with deep learning for joint channel estimation and signal detection," *IEEE Trans. Commun.*, vol. 72, no. 3, pp. 1716–1731, Mar. 2024.

[12] S. Zargari, D. Galappaththige, C. Tellambura, and H. V. Poor, "A Riemannian manifold approach to constrained resource allocation in ISAC," *IEEE Trans. Commun.*, vol. 73, no. 5, pp. 3655–3670, 2025.

[13] N. Zhao, Y. Wang, Z. Zhang, Q. Chang, and Y. Shen, "Joint transmit and receive beamforming design for integrated sensing and communication," *IEEE Commun. Lett.*, vol. 26, no. 3, pp. 662–666, Mar. 2022.

[14] Z. Wang, Y. Liu, X. Mu, Z. Ding, and O. A. Dobre, "NOMA empowered integrated sensing and communication," *IEEE Commun. Lett.*, vol. 26, no. 3, pp. 677–681, Mar. 2022.

[15] Z. Huang, K. Wang, A. Liu, Y. Cai, R. Du, and T. X. Han, "Joint pilot optimization, target detection and channel estimation for integrated sensing and communication systems," *IEEE Trans. Wireless Commun.*, vol. 21, no. 12, pp. 10 351–10 365, Dec. 2022.

[16] P. Stoica, J. Li, and Y. Xie, "On probing signal design for MIMO radar," *IEEE Trans. Signal Process.*, vol. 55, no. 8, pp. 4151–4161, Jul. 2007.

[17] Z. Yang, J.-Y. Xia, J. Luo, S. Zhang, and D. Gündüz, "A learning-aided flexible gradient descent approach to MISO beamforming," *IEEE Wireless Commun. Lett.*, vol. 11, no. 9, pp. 1895–1899, Sept. 2022.

[18] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Rep. (ICLR)*, May 2015.

[19] S. Zargari, S. Farahmand, and B. Abolhassani, "Joint design of transmit beamforming, IRS platform, and power splitting SWIPT receivers for downlink cellular multiuser MISO," *Phys. Commun.*, vol. 48, p. 101413, Oct. 2021.

[20] A. Hakimi, S. Zargari, C. Tellambura, and S. Herath, "Sum rate maximization of MIMO monostatic backscatter networks by suppressing residual self-interference," *IEEE Trans. Commun.*, vol. 71, no. 1, pp. 512–526, Jan. 2023.

[21] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.