Robust Hybrid Beamforming for Integrated Sensing and Communications via Learned Optimization

Lei Wang, Sergiy A. Vorobyov, and Esa Ollila

April 24, 2025

Aalto University, Espoo, Finland

This research was supported by the Research Council of Finland grant #359848

Motivation

- **Integrated sensing and communications (ISAC)** has attracted enormous research interest from both academia and industry.
- Deep learning has been widely applied to beamforming designs recently. The black-box nature makes DNNs hard to deploy in practical scenarios [**?**].
- To eliminate this issue, model-based learning framework (e.g., **algorithm unrolling**) has been proposed to make neural networks interpretable [**?**].
- **Steering vector mismatches** arise in practice because of imperfections in array calibration, distorted antenna array shape, etc [**?**].

This work:

Robust hybrid beamforming for ISAC system under bounded uncertainties in sensing reception using algorithm unrolling

Introduction to Algorithm Unrolling (AU)

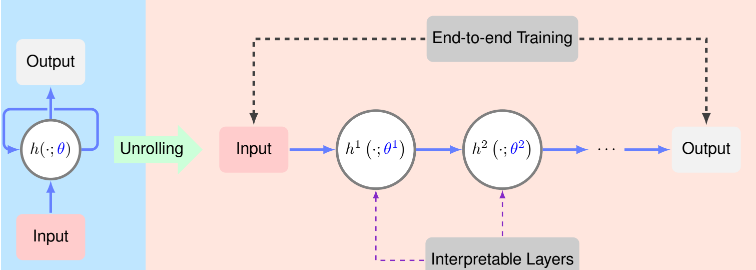


Figure from [?]

Procedure of AU

- pick an iterative algorithm
- unroll it to a neural network (NN)
- select a set of NN parameters to learn

Advantages of AU

- (can) achieve better performance
- naturally inherit interpretability
- require fewer training data

System Model

System composition

- A dual-function radar-communication base station (BS)
  N_t transmit antennas, N_r receive antennas
- M single-antenna communication users
- T point-like targets
- J interferences

Transmit signal

$$\mathbf{x} = \mathbf{F}_{\mathrm{a}}\mathbf{F}_{\mathrm{d}}^{\mathrm{c}}\mathbf{c} + \mathbf{F}_{\mathrm{a}}\mathbf{F}_{\mathrm{d}}^{\mathrm{s}}\mathbf{s}$$

where

- $\mathbf{F}_{\mathrm{a}} \in \mathbb{C}^{N_t \times L}$: analog beamforming matrix (L is # of RF chains)
- $\mathbf{F}_{\mathrm{d}}^{\mathrm{s}} \in \mathbb{C}^{L \times T}$: digital beamforming matrix for sensing
- $\mathbf{F}_{\mathrm{d}}^{\mathrm{c}} \in \mathbb{C}^{L \times M}$: digital beamforming matrix for communications
- $\mathbf{c} \in \mathbb{C}^{M \times 1}$: communication data symbols satisfying $\mathbf{c} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$
- $\mathbf{s} \in \mathbb{C}^{T \times 1}$: radar waveforms satisfying $\mathbb{E}[\mathbf{s}\mathbf{s}^{\mathsf{H}}] = \mathbf{I}_T$

Communication performance metric

Achievable sum-rate (SR):

$$\gamma_{\mathrm{c}} = \sum_{m=1}^{M} \log(1 + \mathsf{SINR}_m^{\mathrm{c}}),$$

where

$$\mathsf{SINR}_m^{\mathrm{c}} = \frac{|\mathbf{h}_m^{\mathsf{H}} \mathbf{F}_{\mathrm{a}} \mathbf{f}_{\mathrm{d},m}|^2}{\sum_{j=1,\, j\neq m}^{M+T} |\mathbf{h}_m^{\mathsf{H}} \mathbf{F}_{\mathrm{a}} \mathbf{f}_{\mathrm{d},j}|^2 + \sigma_m^2}$$

σ_m^2 = noise power at m^{th} user equipment

$\mathbf{h}_m$ = communication channel from BS to the m^{th} user ($\in \mathbb{C}^{N_t \times 1}$)

Sensing performance metric

At BS, t^{th} target's echo signal is filtered by a **receive combiner** $\mathbf{w}_t \in \mathbb{C}^{N_r \times 1}$ to obtain y_t^s.

Mutual information (MI) for multiple (T) targets:

$$\gamma_{\mathrm{s}} = \sum_{t=1}^{T} \log(1 + \mathsf{SINR}_t^{\mathrm{s}}),$$

where the **sensing SINR** for the t^{th} target:

$$\mathsf{SINR}_t^{\mathrm{s}} = \frac{\|\mathbf{w}_t^{\mathsf{H}} \mathbf{H}_t^{\mathrm{s}} \mathbf{F}_{\mathrm{a}} \mathbf{F}_{\mathrm{d}}\|_2^2}{\sum_{p=1,\, p\neq t}^{T} \|\mathbf{w}_t^{\mathsf{H}} \mathbf{H}_p^{\mathrm{s}} \mathbf{F}_{\mathrm{a}} \mathbf{F}_{\mathrm{d}}\|_2^2 + \sum_{j=1}^{J} \|\mathbf{w}_t^{\mathsf{H}} \mathbf{H}_j^{\mathrm{s}} \mathbf{F}_{\mathrm{a}} \mathbf{F}_{\mathrm{d}}\|_2^2 + \sigma_t^2 \|\mathbf{w}_t\|_2^2},$$

$\mathbf{H}_l^{\mathrm{s}} = \alpha_l \mathbf{a}_{\mathrm{r}}(\theta_l) \mathbf{a}_{\mathrm{t}}^{\top}(\theta_l)$ is the target steering matrix (radar round-trip channel)

$\theta_l =$ direction of the point target

$\alpha_l =$ propagation reflection coefficients $(\sim \mathcal{N}(0, \beta_l))$

System Model

- **Uncertainty model**

$$\mathbf{a}_{\mathrm{r}} = \hat{\mathbf{a}}_{\mathrm{r}} + \boldsymbol{\delta}_{\mathrm{r}},$$

where $\hat{\mathbf{a}}_{\mathrm{r}}$ = presumed receive steering vector, $\boldsymbol{\delta}_{\mathrm{r}}$ = error bounded by ε_r:

$$\boldsymbol{\delta}_{\mathrm{r}} \in \mathcal{A}_{\mathrm{r}} \triangleq \left\{ \boldsymbol{\delta}_{\mathrm{r}} \mid \frac{1}{N_r} \|\boldsymbol{\delta}_{\mathrm{r}}\|_2 \leq \varepsilon_r \right\}.$$

- **Problem** (maximize the communication SR and worst-case sensing MI over $\mathcal{A}_{\mathrm{r}}$):

$$\begin{aligned} \max_{\mathbf{F}_{\mathrm{a}}, \mathbf{F}_{\mathrm{d}}, \mathbf{w}_t} \quad & \left(\rho \gamma_{\mathrm{c}} + (1-\rho) \min_{\boldsymbol{\delta}_{\mathrm{r}} \in \mathcal{A}_{\mathrm{r}}} \gamma_{\mathrm{s}} \right) \\ \text{s.t.} \quad & |[\mathbf{F}_{\mathrm{a}}]_{m,n}| = 1, \\ & \|\mathbf{w}_t\|_2^2 = 1, \\ & \|\mathbf{F}_{\mathrm{a}} \mathbf{F}_{\mathrm{d}}\|_{\mathrm{F}}^2 = P_t, \end{aligned}$$

where ρ denotes fixed weight, and P_t is the power budget of BS.

Projected Gradient Descent and Ascent (PGDA)

PGD-based minimization over the uncertainty set $\mathcal{A}_{\rm r}$

- Gradient step

$$\tilde{\boldsymbol{\delta}}_{\rm r}^{(i+1)} = \boldsymbol{\delta}_{\rm r}^{(i)} - \mu_{\rm r}^{(i)} \frac{\partial \gamma_{\rm s}\left(\mathbf{F}_{\rm a}^{(i)}, \mathbf{F}_{\rm d}^{(i)}, \mathbf{w}_t^{(i)}, \boldsymbol{\delta}_{\rm r}^{(i)}\right)}{\partial \boldsymbol{\delta}_{\rm r}^{*}},$$

where $\mu_{\rm r}$ is the step size for $\boldsymbol{\delta}_{\rm r}$.

- Projection step

$$\boldsymbol{\delta}_{\rm r}^{(i+1)} = \min\left\{\frac{\varepsilon_r N_r}{\|\tilde{\boldsymbol{\delta}}_{\rm r}^{(i+1)}\|_2}, 1\right\} \tilde{\boldsymbol{\delta}}_{\rm r}^{(i+1)}.$$

Projected Gradient Descent and Ascent (PGDA)

PGA-based maximization over $\mathbf{F}_\mathrm{a}$, $\mathbf{F}_\mathrm{d}$, and $\mathbf{w}_t$

- Gradient step ($\gamma = \rho\gamma_\mathrm{c} + (1-\rho)\gamma_\mathrm{s}$)

$$\tilde{\mathbf{F}}_\mathrm{a}^{(i+1)} = \mathbf{F}_\mathrm{a}^{(i)} + \mu_\mathrm{a}^{(i)} \frac{\partial\gamma\left(\mathbf{F}_\mathrm{a}^{(i)}, \mathbf{F}_\mathrm{d}^{(i)}, \mathbf{w}_t^{(i)}, \boldsymbol{\delta}_\mathrm{r}^{(i+1)}\right)}{\partial\mathbf{F}_\mathrm{a}^*},$$

$$\tilde{\mathbf{F}}_\mathrm{d}^{(i+1)} = \mathbf{F}_\mathrm{d}^{(i)} + \mu_\mathrm{d}^{(i)} \frac{\partial\gamma\left(\mathbf{F}_\mathrm{a}^{(i)}, \mathbf{F}_\mathrm{d}^{(i)}, \mathbf{w}_t^{(i)}, \boldsymbol{\delta}_\mathrm{r}^{(i+1)}\right)}{\partial\mathbf{F}_\mathrm{d}^*},$$

$$\tilde{\mathbf{w}}_t^{(i+1)} = \mathbf{w}_t^{(i)} + \mu_\mathrm{w}^{(i)} \frac{\partial\gamma\left(\mathbf{F}_\mathrm{a}^{(i)}, \mathbf{F}_\mathrm{d}^{(i)}, \mathbf{w}_t^{(i)}, \boldsymbol{\delta}_\mathrm{r}^{(i+1)}\right)}{\partial\mathbf{w}_t^*},$$

where μ_a, μ_d, and μ_w are the step sizes for $\mathbf{F}_\mathrm{a}$, $\mathbf{F}_\mathrm{d}$, and $\mathbf{w}_t$, respectively.

- Projection: $[\mathbf{F}_\mathrm{a}^{(i+1)}]_{m,n} = \frac{[\tilde{\mathbf{F}}_\mathrm{a}^{(i+1)}]_{m,n}}{|[\tilde{\mathbf{F}}_\mathrm{a}^{(i+1)}]_{m,n}|}$, $\mathbf{F}_\mathrm{d}^{(i+1)} = \frac{\sqrt{P_t}\tilde{\mathbf{F}}_\mathrm{d}^{(i+1)}}{\left\|\mathbf{F}_\mathrm{a}^{(i+1)}\tilde{\mathbf{F}}_\mathrm{d}^{(i+1)}\right\|_\mathrm{F}}$, $\mathbf{w}_t^{(i+1)} = \frac{\tilde{\mathbf{w}}_t^{(i+1)}}{\|\tilde{\mathbf{w}}_t^{(i+1)}\|_2}$.

Unrolled PGDA

- **Aim**: use algorithm unrolling to tune step sizes in PGDA algorithm based on data.
- **Trainable parameters**:

$$\mathbf{\Theta} \triangleq [\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_I]$$

  with $\boldsymbol{\mu}_i \triangleq [\mu_{\mathrm{a}}^{(i)}, \mu_{\mathrm{d}}^{(i)}, \mu_{\mathrm{w}}^{(i)}, \mu_{\mathrm{r}}^{(i)}]^\top$ = step size vector for i^{th} iteration.
- **Loss function** ($\gamma = \rho\gamma_{\mathrm{c}} + (1-\rho)\gamma_{\mathrm{s}}$):

$$\mathcal{L}(\mathbf{\Theta}) = -\frac{1}{|\mathcal{D}|}\sum_{d=1}^{|\mathcal{D}|}\frac{1}{I}\sum_{i=1}^{I}\log(1+i)\gamma(\mathbf{h}_m^{(d)}, \mathbf{H}_l^{\mathrm{s},(d)}, \mathbf{F}_{\mathrm{a}}^{(i)}, \mathbf{F}_{\mathrm{d}}^{(i)}, \mathbf{w}_t^{(i)}, \boldsymbol{\delta}_{\mathrm{r}}^{(i)}),$$

  - weighted sum of losses (enhancing the system to improves along the iterations [?]).
  - $\mathcal{D}$ = data set containing communication and sensing channel realizations.
- **Tune** the hyperparameter matrix:

$$\mathbf{\Theta}^\star = \arg\min_{\mathbf{\Theta}} \mathcal{L}(\mathbf{\Theta})$$

Numerical Evaluation

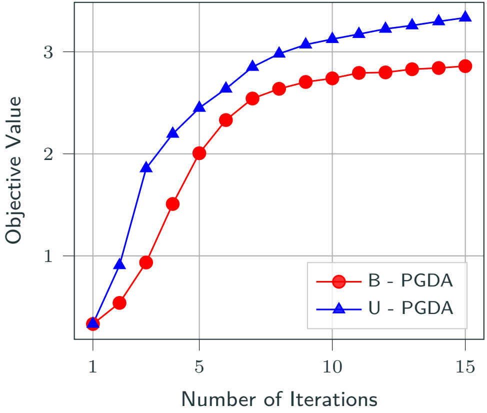
Parameter settings

- $M = 4,\ T = 8,\ J = 4,\ N_t = 6,\ N_r = 8,\ L = 4$
- $P_t = 5$ dB (power budget)
- weight factor $\rho = 0.8 \quad (\gamma = 0.8\gamma_\mathrm{c} + 0.2\gamma_\mathrm{s})$

Training settings:

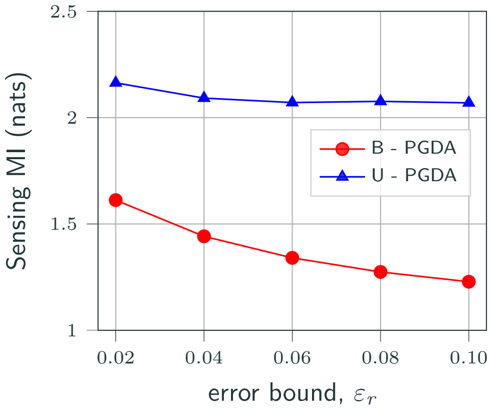
- Channel vector $\mathbf{h}$, reflection coefficients α_t (for target) and α_j (interfer) are circular Gaussian noises with variances 0 dB , $\beta_t = 2$ dB and $\beta_j = 3$ dB, respectively
- Dataset sizes for training and testing: 1000 and 100
- SGD optimizer with learning rate of 0.01
- Fixed step size for PGDA and initial step size for unrolled PGDA: 0.1

Convergence evaluation



- The benchmark PGDA may easily get trapped in local optima when solving non-convex problem.
- The unrolled PGDA is capable of **learning the update rules from data**, allowing it to fit the objective function and escape local optima more effectively than the benchmark.

Robustness evaluation



- The sensing MI curve achieved by the unrolled PGDA has lower slope than that corresponding to the benchmark PGDA.
- The unrolled PGDA is less sensitive than the benchmark PGDA in the face of increased level of uncertainties.
- The unrolled PGDA is more robust than the benchmark PGDA.

Conclusion

- We formulated an optimization problem aimed at maximizing the communication SR and worst-case sensing MI with bounded uncertainty.
- A PGDA algorithm was developed for solving the formulated optimization problem.
- Unrolled PGDA algorithm was proposed, and step sizes tuned based on the data.
- unrolled PGDA exhibits **faster convergence** and **better robustness** than the PGDA

Future work (in progress):

- The uncertainties of communication channel and transmit steering vector will be explored.
- A "safeguard" mechanism could be explored to ensure algorithm unrolling performs no worse than traditional algorithms on out-of-distribution tasks.

References i

- Ortal Lavi and Nir Shlezinger, *Learn to rapidly and robustly optimize hybrid precoding*, IEEE Trans. Commun. **71** (2023), no. 10, 5814–5830.
- Vishal Monga, Yuelong Li, and Yonina C. Eldar, *Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing*, IEEE Signal Processing Magazine **38** (2021), no. 2, 18–44.
- Nir Shlezinger, Yonina C. Eldar, and Stephen P. Boyd, *Model-based deep learning: On the intersection of deep learning and optimization*, IEEE Access **10** (2022), 115384–115398.
- S.A. Vorobyov, A.B. Gershman, and Zhi-Quan Luo, *Robust adaptive beamforming using worst-case performance optimization: a solution to the signal mismatch problem*, IEEE Trans. Signal Process. **51** (2003), no. 2, 313–324.