Robust Optimization Methods and Applications to Transmit/Receive Beamforming in Radar and Communications

Yongwei Huang
Guangdong Polytechnic Normal University, Guangzhou, China

Sergiy A. Vorobyov
Aalto University, Finland

Tutorial, EUSIPCO 2025

Sicily, Italy

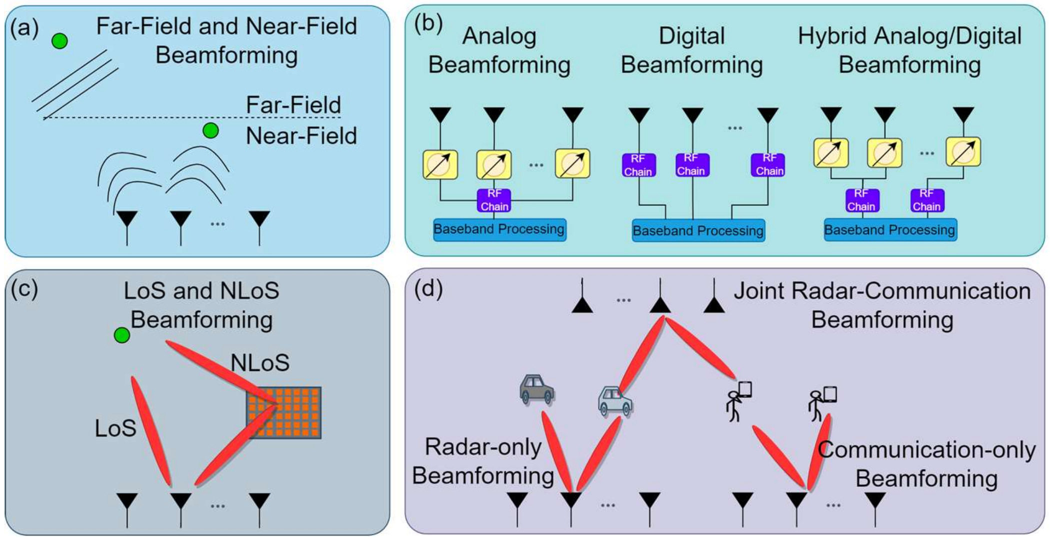
Tree Ways to Order the Material

- By Characterization of Uncertainty Sets
  - Deterministic uncertainty modeling
  - Stochastic uncertainty modeling
  - Distributional uncertainty modeling
- By Solution Approaches
  - Convex optimization
  - Non-convex optimization
  - Learning to optimize
- By Applications
  - Receive Beamforming
  - Transmit Beamforming

Part I: Introduction to Robust Beamforming

Beamforming Definition

- Signal processing technique that exploits arrays of sensors improve one of many possible application-driven performance objectives
- Steer, shape, and focus an electromagnetic wave
- Applications: radar, sonar, wireless communications, acoustics, astronomy, seismology, and medical imaging.



Beamforming techniques by (a) transmission range (b) transceiver architectures (c) paths and (d) applications

Evolution of Beamformers

1905: Braun
Early demonstration of gains provided by a phased array to direct radio waves

1950s
Beamforming in sonar

1990s-early 2000s
Convex beamformers

Late 2010s
Learning-based beamformers

1901: Marconi
Circular array with 4 antennas to improve the gain of trans-Atlantic Morse code transmission

1940s
Antenna diversity

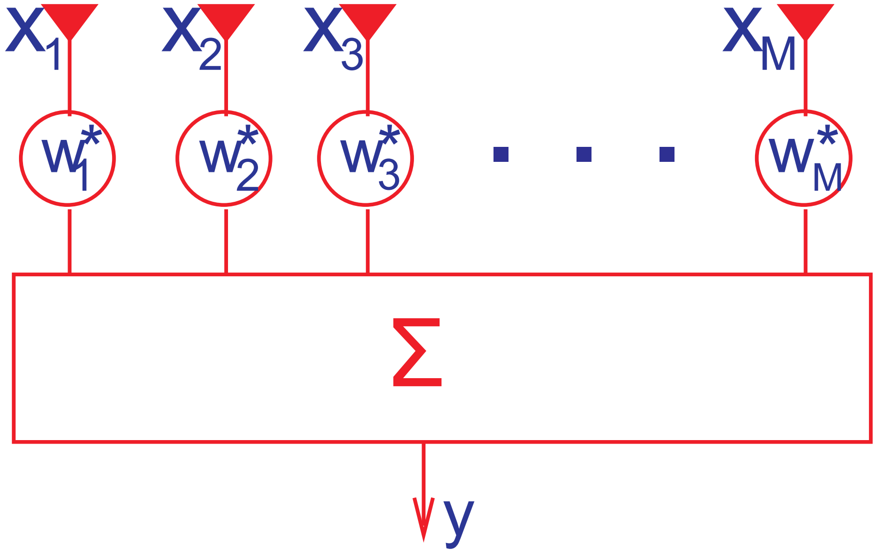
1960s
Adaptive beamforming (Widrow) & Capon Beamformer to select the weight vectors or beamformers to minimize the array output power.

Late 2000s
Nonconvex beamformers

Robust Beamforming Optimization: Background

The output of a narrowband beamformer

$$
\begin{aligned}
y(k) &= \boldsymbol{w}^H \boldsymbol{x}(k) \\
\boldsymbol{w} &= [w_1, \ldots, w_M]^T \\
\boldsymbol{x}(k) &= [x_1(k), \ldots, x_M(k)]^T
\end{aligned}
$$



Background

The received data vector

$$\boldsymbol{x}(k) = \underbrace{s(k)\boldsymbol{a}}_{\text{signal}} + \underbrace{\boldsymbol{i}(k)}_{\text{interference}} + \underbrace{\boldsymbol{n}(k)}_{\text{noise}}$$

Max signal-to-interference-plus-noise ratio (SINR) criterion for a point source

$$\max_{\boldsymbol{w}} \mathrm{SINR}\,, \qquad \mathrm{SINR} = \frac{\sigma_{\mathrm{s}}^2 \, |\boldsymbol{w}^H \boldsymbol{a}|^2}{\boldsymbol{w}^H \boldsymbol{R}_{\mathrm{i+n}} \boldsymbol{w}}$$

Interference-plus-noise covariance matrix

$$\boldsymbol{R}_{\mathrm{i+n}} = \mathrm{E}\left\{ \left(\boldsymbol{i}(k) + \boldsymbol{n}(k)\right) \left(\boldsymbol{i}(k) + \boldsymbol{n}(k)\right)^H \right\}$$

Background

Equivalent MVDR beamforming problem [Capon; 1968]

$$\min_{\boldsymbol{w}} \boldsymbol{w}^H \boldsymbol{R}_{\mathrm{i+n}} \boldsymbol{w} \qquad \text{subject to} \qquad \boldsymbol{w}^H \boldsymbol{a} = 1$$

Optimal weight vector

$$\boldsymbol{w}_{\mathrm{opt}} = \frac{1}{\boldsymbol{a}^H \boldsymbol{R}_{\mathrm{i+n}}^{-1} \boldsymbol{a}} \boldsymbol{R}_{\mathrm{i+n}}^{-1} \boldsymbol{a}$$

The *data covariance matrix* $\boldsymbol{R} = \boldsymbol{R}_{\mathrm{i+n}} + \sigma_{\mathrm{s}}^2 \boldsymbol{a}\boldsymbol{a}^H$ can be used instead of $\boldsymbol{R}_{\mathrm{i+n}}$.

Presence of signal component in $\boldsymbol{R}$ does not affect the *SINR*.

In practice, $\boldsymbol{R}$ is *unavailable* $\Longrightarrow$ *sample estimate* $\hat{\boldsymbol{R}}$ of $\boldsymbol{R}$ is used.

Background

Sample estimate of $\boldsymbol{R}$

$$\hat{\boldsymbol{R}} = \frac{1}{K}\sum_{k=1}^{K} \boldsymbol{x}(k)\boldsymbol{x}^H(k) \quad \longleftarrow \quad \text{contains the signal component!}$$

Practical formulation of the MVDR problem

$$\min_{\boldsymbol{w}} \boldsymbol{w}^H \hat{\boldsymbol{R}} \boldsymbol{w} \qquad \text{subject to} \qquad \boldsymbol{w}^H \boldsymbol{a} = 1$$

Sample matrix inversion (SMI) algorithm [Reed, Mallett, Brennan; 1974]

$$\boldsymbol{w}_{\mathrm{SMI}} = \hat{\boldsymbol{R}}^{-1} \boldsymbol{a} \quad \longleftarrow \quad \text{immaterial constant ommited}$$

Implementation Requirements

- The *steering vector* of the desired signal has to be *known precisely*, when the *signal component is present* in the training data cell.

  Otherwise, the *signal* can be *suppressed* by means of output power (objective) minimization $\longleftarrow$ *signal self-nulling*

- Large number of snapshots (training sample size)

- Stationary training data set

Violation of any of these assumptions is known to lead to a dramatic performance breakdown of adaptive beamforming.

Modelling Error Causes

Typical causes of steering vector mismatch

- Look direction uncertainty and signal pointing errors
- Poor array calibration
- Unknown distortions of array geometry
- Unknown sensor mutual coupling
- Signal local scattering, source spreading, and fading effects
- Near-far wavefront mismodeling
- Unknown wavefront distortions and fluctuations

Drawbacks of SMI beamformer

Why the robustness of SMI beamformer is insufficient?

- The *main cause* of performance degradation is the *presence of the signal component* in the training data cell when the steering vector is mismatched
- Then, the signal can be *misinterpreted* as an interference and *suppressed* by means of adaptive nulling *instead of being protected*
- *Similar effect* when there is *no mismatch* but the *training sample size is small* $\longleftarrow$ *signal cancellation phenomenon*

Diagonal loading SMI method [Cox et al.], [Carlson], [Abramovich]; 80's

$$\boldsymbol{w}_{\mathrm{LSMI}} = \hat{\boldsymbol{R}}_{\mathrm{dl}}^{-1}\boldsymbol{a}, \quad \hat{\boldsymbol{R}}_{\mathrm{dl}} = \hat{\boldsymbol{R}} + \xi \boldsymbol{I}$$

Part II: Deterministic Uncertainty Modeling

Norm-Bounded Additive Mismatch

Actual steering vector

$$\tilde{\boldsymbol{a}} = \boldsymbol{a} + \boldsymbol{\Delta} \neq \boldsymbol{a}$$

Let the *unknown* steering vector mismatch $\boldsymbol{\Delta}$ be *bounded* by some *known* constant $\varepsilon > 0$:

$$\|\boldsymbol{\Delta}\| \leq \varepsilon$$

Our formulation of the robust MVDR beamformer

$$\boxed{\min_{\boldsymbol{w}} \boldsymbol{w}^H \hat{\boldsymbol{R}} \boldsymbol{w} \quad \text{s. t.} \quad |\boldsymbol{w}^H (\boldsymbol{a} + \boldsymbol{\Delta})| \geq 1 \quad \text{for all} \quad \|\boldsymbol{\Delta}\| \leq \varepsilon}$$

Instead of *fixed distortionless response* towards the presumed steering vector $\boldsymbol{a}$, a *distortionless response* is maintained for a continuum of *all possible steering vectors*

Convex Optimization

The constrains guarantee that the distortionless response will be maintained in the *worst case*.

Equivalent problem

$$\min_{\boldsymbol{w}} \boldsymbol{w}^H \hat{\boldsymbol{R}} \boldsymbol{w} \qquad \text{subject to} \qquad \min_{\|\boldsymbol{\Delta}\| \leq \varepsilon} |\boldsymbol{w}^H (\boldsymbol{a} + \boldsymbol{\Delta})| \geq 1$$

Result 1 (if ε is small enough, i.e., $|\boldsymbol{w}^H \boldsymbol{a}| \geq \varepsilon \|\boldsymbol{w}\|$)

$$\min_{\|\boldsymbol{\Delta}\| \leq \varepsilon} |\boldsymbol{w}^H (\boldsymbol{a} + \boldsymbol{\Delta})| = |\boldsymbol{w}^H \boldsymbol{a}| - \varepsilon \|\boldsymbol{w}\|$$

Robust MVDR Beamforming

New problem [Vorobyov, Gershman, Luo; 2003] (using Result 1 and requiring $\mathrm{Im}\{\boldsymbol{w}^H \boldsymbol{a}\} = 0$)

$$\min_{\boldsymbol{w}} \boldsymbol{w}^H \hat{\boldsymbol{R}} \boldsymbol{w} \qquad \text{subject to} \qquad \varepsilon \|\boldsymbol{w}\| \leq \boldsymbol{w}^H \boldsymbol{a} - 1$$

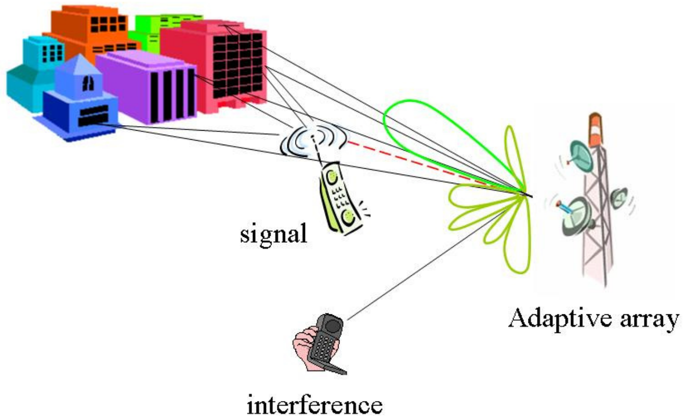
This is *convex second order cone (SOC) programming* problem. It can be easily solved. The *complexity* is the *same as for SMI beamformer*.

Samples of farther extentions of our work

- Ellipsoidal uncertainty set for $\boldsymbol{\Delta}$ [Lorenz, Boyd; 2005]
- Covariance fitting interpretation [Li, Stoica, Wang; 2003]
- Newton-type solutions [Zarifi et al.; 2005], [Lorenz, Boyd]
- Modified uncertainty region [Yu, Er; 2006]
- General rank source [Shahbazpanahi et al.; 2003]

Modeling Error Causes

Coherent signal local scattering / Ricean propagation medium



The *actual* steering vector is formed by $L+1$ *signal paths* as

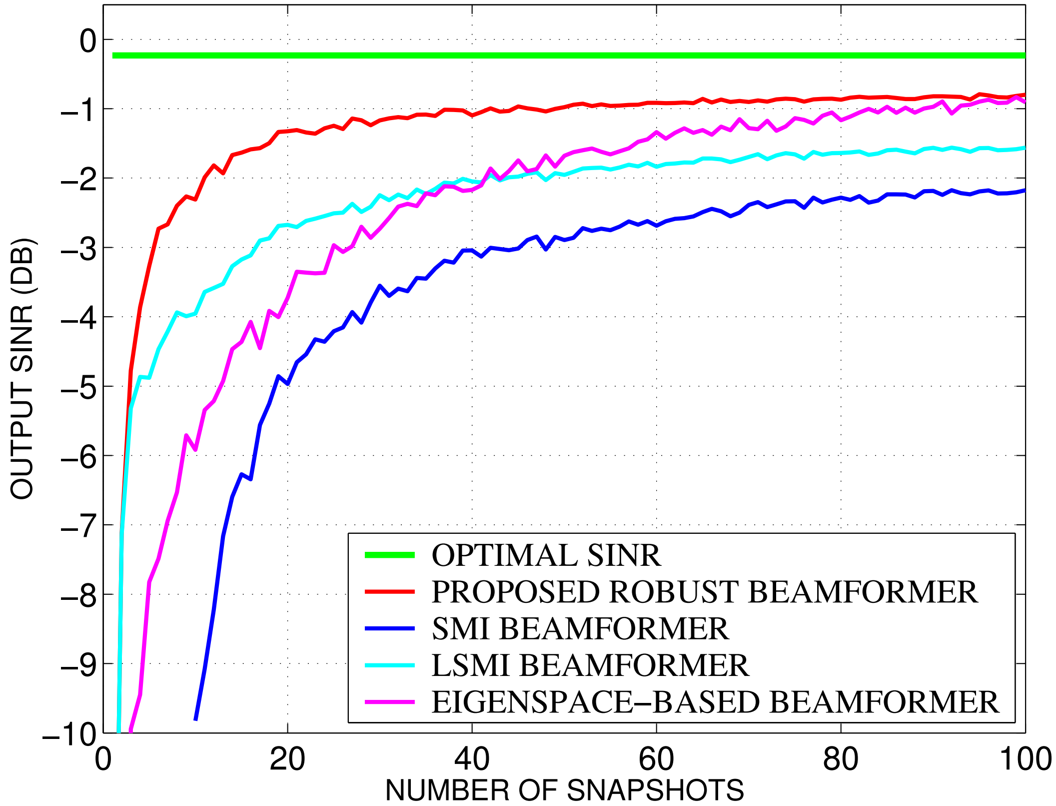
$$\tilde{\boldsymbol{a}} = \boldsymbol{a} + \sum_{i=1}^{L} e^{j\psi_j} \boldsymbol{b}(\theta_i)$$

$\boldsymbol{a}$ is the *direct path / line of side (LOS)*

$\boldsymbol{b}(\theta_i)$ are *scattered paths / non-line of sides (NLOS)*

Simulations

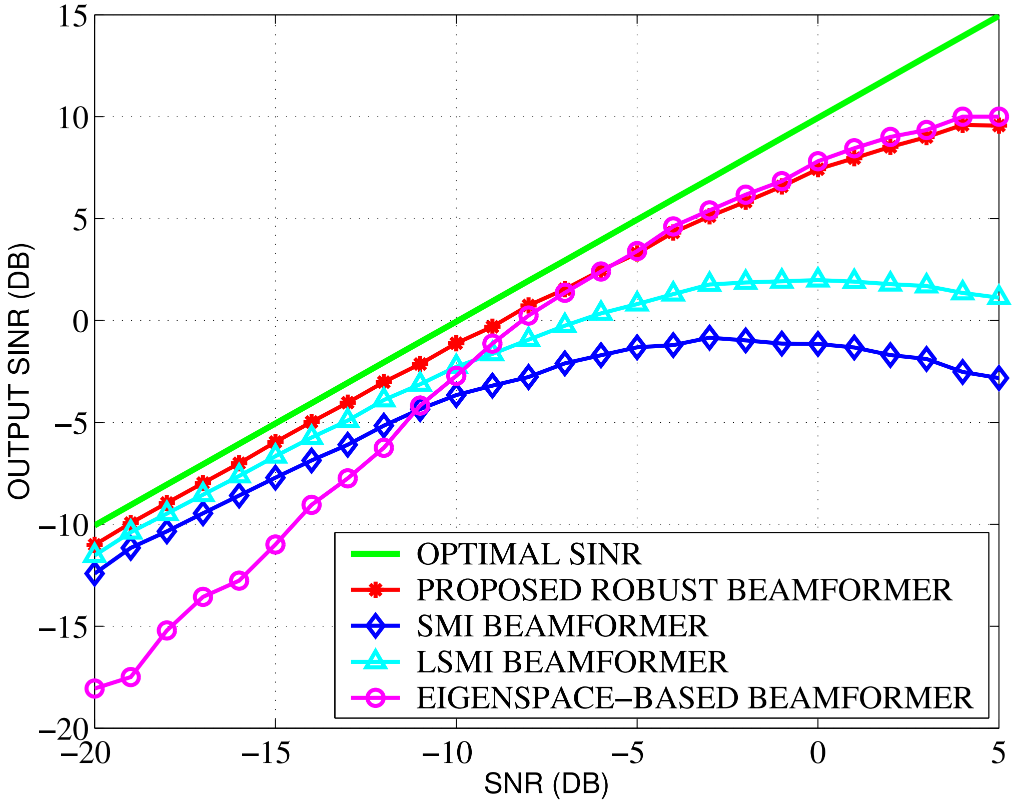
Example 1: Coherent local scattering



ULA, M=10 omnidirectional sensors, half-wavelength spacing; Desired signal DOA is 3°; 2 interferences with DOA's 30° and 50°; INR = -10 dB; signal component is always present; $L = 4$, $\mathrm{std}_{\theta_i} = 2^\circ$; *SNR = -10 dB*

Simulations

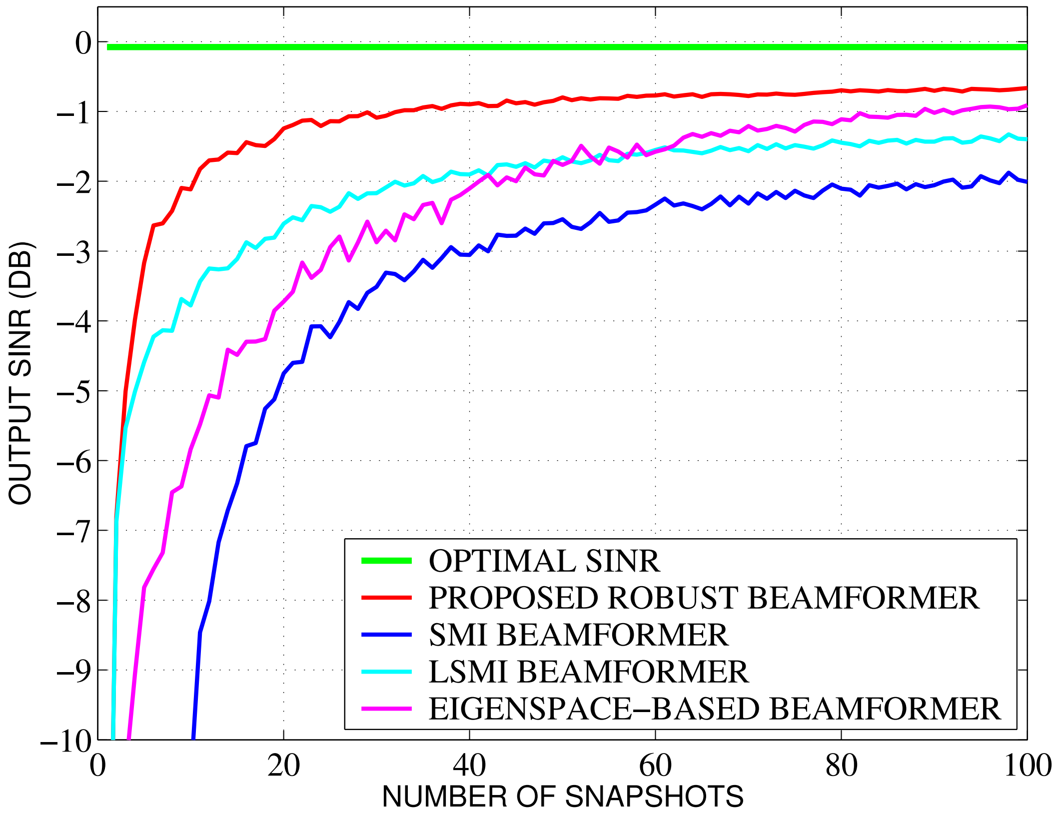
Example 1: Coherent local scattering



ULA, M=10 omnidirectional sensors, half-wavelength spacing; Desired signal DOA is $3°$; 2 interferences with DOA's $30°$ and $50°$; INR = -10 dB; signal component is always present; $L = 4$, $\mathrm{std}_{\theta_i} = 2°$; $N = 30$

Simulations

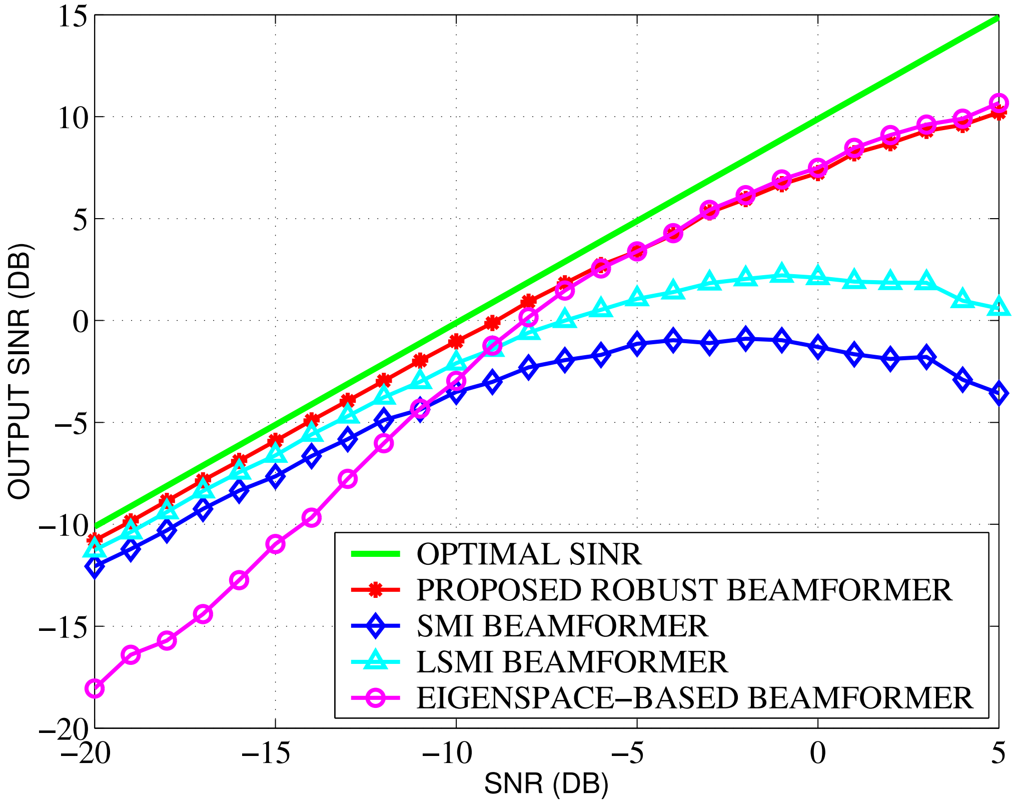
Example 2: Distortion of the signal wavefront



ULA, M=10 omnidirectional sensors, half-wavelength spacing; Desired signal DOA is 3°; 2 interferences with DOA's 30° and 50°; INR = -10 dB; signal component is always present; *independent-increment phase distortions* are Gaussian with the *variance 0.04*; $N = 30$

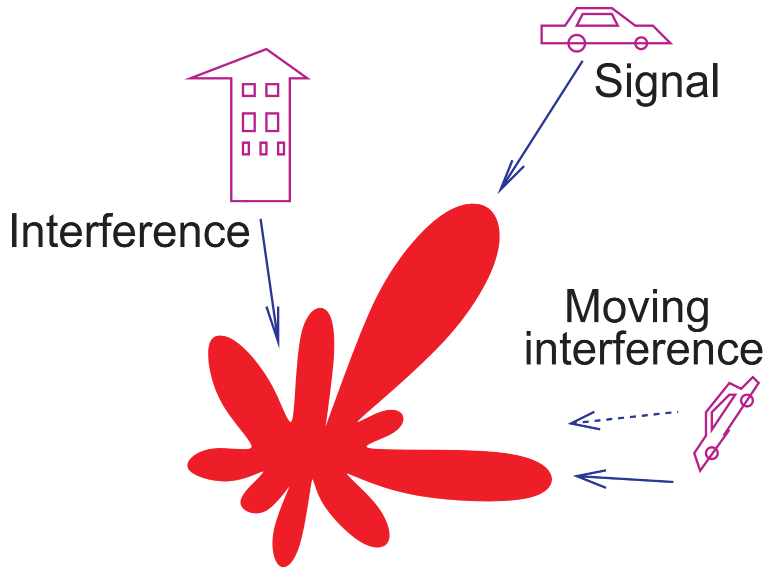
Simulations

Example 2: Distortion of the signal wavefront



ULA, M=10 omnidirectional sensors, half-wavelength spacing; Desired signal DOA is 3°; 2 interferences with DOA's 30° and 50°; INR = -10 dB; signal component is always present; *independent-increment phase distortions* are Gaussian with the *variance 0.04*; $N = 30$

Moving Interferences



Matrix taper [Mailloux], [Zatman], [Guerci]

$$\boldsymbol{w}_{\mathrm{T}} \propto \hat{\boldsymbol{R}}_{\mathrm{T}}^{-1} \boldsymbol{a}\,, \qquad \hat{\boldsymbol{R}}_{\mathrm{T}} = \hat{\boldsymbol{R}} \odot \boldsymbol{T}$$

where $\boldsymbol{T}$ is a specially designed taper matrix: $[\boldsymbol{T}]_{ij} = \mathrm{sinc}(\rho|i-j|)$.

Moving Interferences

Actual data matrix and steering vector

$$\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{Z}, \quad \|\boldsymbol{Z}\|_{\mathrm{F}} \leq \delta$$

$$\tilde{\boldsymbol{a}} = \boldsymbol{a} + \boldsymbol{\Delta}, \quad \|\boldsymbol{\Delta}\| \leq \varepsilon$$

$\boldsymbol{Z}$ takes into account *interference nonstationarity*

Our robust problem formulation

$$\min_{\boldsymbol{w}} \max_{\boldsymbol{Z}} \|\boldsymbol{Y}^H \boldsymbol{w}\|, \quad \|\boldsymbol{Z}\|_{\mathrm{F}} \leq \delta$$

$$\text{subject to} \qquad |\boldsymbol{w}^H \tilde{\boldsymbol{a}}| \geq 1 \quad \text{for all} \quad \|\boldsymbol{\Delta}\| \leq \varepsilon$$

Moving Interferences

Result 2 (for objective)

$$\max_{\|\boldsymbol{Z}\|_{\mathrm{F}} \leq \delta} \|\boldsymbol{Y}^H \boldsymbol{w}\| = \|\boldsymbol{X}^H \boldsymbol{w}\| + \delta \|\boldsymbol{w}\|$$

New problem (using Results 1 and 2)

$$\min_{\boldsymbol{w}} \|\boldsymbol{U}\boldsymbol{w}\| + \delta \|\boldsymbol{w}\|$$
$$\text{subject to} \quad \varepsilon \|\boldsymbol{w}\| \leq \boldsymbol{w}^H \boldsymbol{a} - 1$$

where $\boldsymbol{U} = \boldsymbol{X}^H$ or $\boldsymbol{U} = \sqrt{N}\hat{\boldsymbol{R}}_{\mathrm{T}}^{1/2}$

This is a SOC programming problem!

Nonconvex norm-constrained beamforming

Considers an additional norm constraint for beamformer weights in a more general setting. The problem becomes [Li et. al., 2004]

$$\min_{\tilde{\boldsymbol{a}}} \tilde{\boldsymbol{a}}^H \hat{\boldsymbol{R}}^{-1} \tilde{\boldsymbol{a}} \quad \text{s.t.} \quad \|\tilde{\boldsymbol{a}} - \boldsymbol{a}\|_2 \leq \epsilon_a, \ \|\tilde{\boldsymbol{a}}\|_2^2 = N$$

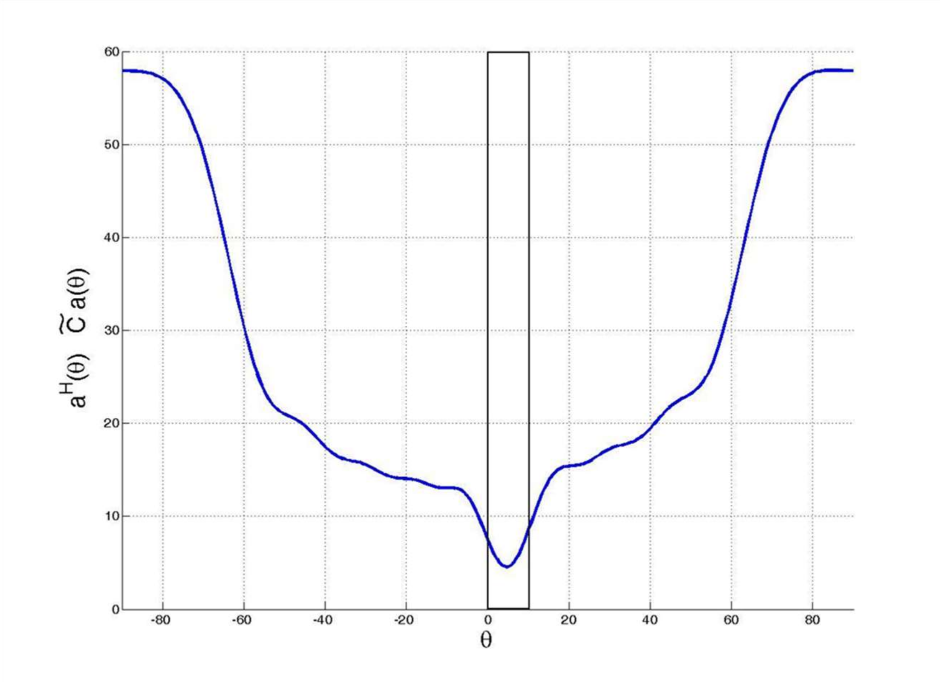
Nonconvex norm constraint $\|\tilde{\boldsymbol{a}}\|_2^2 = N$ is added. *This is a nonconvex problem, but it can be solved via SDP relaxation followed by a rank-reduction postprocessing* [Huang and Palomar, 2010]. *Always has an optimal rank-one solution!*

Better problem [Kabazibasmenj, Vorobyov, Hassanien, 2012]

$$\min_{\tilde{\boldsymbol{a}}} \tilde{\boldsymbol{a}}^H \hat{\boldsymbol{R}}^{-1} \tilde{\boldsymbol{a}} \quad \text{s.t.} \quad \tilde{\boldsymbol{a}}^H \tilde{\boldsymbol{C}} \tilde{\boldsymbol{a}} \leq \Delta_0, \ \|\tilde{\boldsymbol{a}}\|_2^2 = N$$

$\tilde{\boldsymbol{C}} = \int_{\tilde{\Theta}} \boldsymbol{a}(\theta)\boldsymbol{a}^H(\theta)\, d\theta$ and $\tilde{\Theta}$ is the complement of the angular sector $\Theta = [\theta_{\min}, \theta_{\max}]$ where the desired signal is located;
$\Delta_0 \triangleq \max_{\theta\in\Theta} \boldsymbol{a}^H(\theta)\tilde{\boldsymbol{C}}\boldsymbol{a}(\theta)$ represents the boundary to distinguish approximately whether or not the direction of $\boldsymbol{a}$ is in the actual signal angular sector Θ.

New Constraint Visualized



Nonconvex double-sided norm constraint

Add double-sided norm constraint to account for gain perturbations in the steering vector. [Huang and Vorobyov, 2019]

$$\min_{\tilde{\boldsymbol{a}}} \tilde{\boldsymbol{a}}^H \hat{\boldsymbol{R}}^{-1} \tilde{\boldsymbol{a}} \text{ s.t. } \tilde{\boldsymbol{a}}^H \boldsymbol{C} \tilde{\boldsymbol{a}} \geq \Delta_1, \; N(1-\eta_1) \leq \|\tilde{\boldsymbol{a}}\|_2^2 \leq N(1+\eta_2), \; \|\boldsymbol{V}^H(\tilde{\boldsymbol{a}} - \boldsymbol{a}_0)\|_2^2 \leq \epsilon_u$$

$\boldsymbol{a}_0 = \boldsymbol{a}(\theta_0)$, $\theta_0 = (\theta_{\max} + \theta_{\min})/2$ is the middle value of the region Θ;
$\boldsymbol{V} \in \mathbb{C}^{N \times N}$ – in the generalized similarity constraint together with $\boldsymbol{a}_0$ and ϵ_u;
$\boldsymbol{C} = \int_{\Theta} \boldsymbol{a}(\theta)\boldsymbol{a}^H(\theta)\, d\theta$; and Δ_1, η_1, and η_2 are selected values.

- The quadratic constraint is inverted
- The norm constrained is relaxed and generalized
- The similarity constraint is generalized

These problem is nonconvex, but can be often exactly or conditionally solved through SDR.

Part III: General Form of Worst-Case SINR Maximization

Data Sample Covariance Matrix Error

- Assume further that the sample covariance matrix $\hat{\boldsymbol{R}}$ and the interference-plus-noise covariance (INC) matrix $\boldsymbol{R}_{\mathrm{i+n}}$ have mismatch.
- Common reasons for the mismatch:
  - Finite sample size (estimation error)
  - Non-stationarity of signals
  - Model mismatch
  - Measurement noise
- Let $\boldsymbol{R} \triangleq \boldsymbol{R}_{\mathrm{i+n}}$ (dropping the subscript for easy notation without causing confuses).
- Suppose that the uncertainty sets $\mathcal{A}$ for $\boldsymbol{R}$, and $\mathcal{B}$ for $\boldsymbol{a}$ are known.

Worst-Case SINR Maximization

- An optimal robust adaptive beamforming (RAB) solution $\boldsymbol{w}^\star$ is obtained by solving the worst-case SINR maximization problem:

$$\underset{\boldsymbol{w}\in\mathcal{W}}{\text{maximize}} \quad \underset{\tilde{\boldsymbol{a}} \in \mathcal{A}, \boldsymbol{R} \in \mathcal{B}}{\text{minimize}} \quad \frac{|\boldsymbol{w}^H \tilde{\boldsymbol{a}}|^2}{\boldsymbol{w}^H \boldsymbol{R} \boldsymbol{w}}.$$

- Here $\mathcal{W}$ is the feasible set of RAB vectors $\boldsymbol{w}$.
- The uncertainty sets $\mathcal{A}$, $\mathcal{B}$ and the feasible set $\mathcal{W}$ are either convex or non-convex. For instance,
  - uncertainty set $\mathcal{A}$:
    - convex $\mathcal{A} = \{\tilde{\boldsymbol{a}} \mid \|\tilde{\boldsymbol{a}} - \boldsymbol{a}\| \leq \epsilon_0\}$;
    - nonconvex $\mathcal{A} = \{\tilde{\boldsymbol{a}} \mid \|\tilde{\boldsymbol{a}} - \boldsymbol{a}\| \leq \epsilon_0,\ N(1-\eta_1) \leq \|\tilde{\boldsymbol{a}}\|^2 \leq N(1+\eta_2)\}$;

Worst-Case SINR Maximization

- uncertainty set $\mathcal{B}$:
  - convex $\mathcal{B} = \{\boldsymbol{R} \mid \|\boldsymbol{R} - \hat{\boldsymbol{R}}\|_2 \leq \gamma_0,\ \boldsymbol{R} \succeq \boldsymbol{0}\}$;
  - nonconvex $\mathcal{B} = \{\boldsymbol{R} \mid \|\boldsymbol{R} - \hat{\boldsymbol{R}}\|_2 \leq \gamma_0,\ \text{Rank}\,(\boldsymbol{R}) = N/2,\ \boldsymbol{R} \succeq \boldsymbol{0}\}$;
- feasible set $\mathcal{W}$:
  - convex $\mathcal{W} = \{\boldsymbol{w} \mid |\boldsymbol{a}(\theta_j)^H \boldsymbol{w}| \leq \sqrt{\eta},\ j = 1, \ldots, J\}$ with the grid $\Theta = \{\theta_1, \ldots, \theta_J\}$ approximating the sidelobe regions (enforcing the sidelobe level control);
  - nonconvex $\mathcal{W} = \{\boldsymbol{w} \mid \|\boldsymbol{w}\|_0 \leq 2N/3\}$.

Minimax SINR

- Another RAB solution $\boldsymbol{w}^\star$ can be defined by estimating the best signal steering vector $\boldsymbol{a}^\star$ and INC matrix $\boldsymbol{R}^\star$ and forming the MVDR RAB solution:

$$\boldsymbol{w}^\star \triangleq \boldsymbol{w}_{\mathrm{MVDR}} = (\boldsymbol{R}^\star)^{-1}\boldsymbol{a}^\star.$$

- This solution can by obtained by solving the minimax SINR problem:

$$\boxed{\underset{\boldsymbol{a}\in\mathcal{A},\boldsymbol{R}\in\mathcal{B}}{\text{minimize}} \quad \underset{\boldsymbol{w}\in\mathcal{W}}{\text{maximize}} \quad \frac{|\boldsymbol{w}^H\boldsymbol{a}|^2}{\boldsymbol{w}^H\boldsymbol{R}\boldsymbol{w}}}$$

with feasible set $\mathcal{W} = \mathbb{C}^N/\{\mathbf{0}\}$.

- In other words, solve the problem which is assumed solvable efficiently:

$$\boxed{\underset{\boldsymbol{a}\in\mathcal{A},\boldsymbol{R}\in\mathcal{B}}{\text{minimize}} \quad \boldsymbol{a}^H\boldsymbol{R}^{-1}\boldsymbol{a}}$$

obtaining $(\boldsymbol{a}^\star, \boldsymbol{R}^\star)$.

Maximin SINR vs. Minimax SINR

- Only for one scenario when $\mathcal{W} = \mathbb{C}^N/\{\mathbf{0}\}$, and $\mathcal{A}$ and $\mathcal{B}$ are convex and compact [Kim et al., 2008], the two RAB problems (via the maximin SINR and minimax SINR criteria) are tantamount:

$$\underset{\boldsymbol{w}\neq\mathbf{0}}{\text{maximize}} \quad \underset{\boldsymbol{a}\in\mathcal{A},\boldsymbol{R}\in\mathcal{B}}{\text{minimize}} \; \frac{|\boldsymbol{w}^H\boldsymbol{a}|^2}{\boldsymbol{w}^H\boldsymbol{R}\boldsymbol{w}} = \underset{\boldsymbol{a}\in\mathcal{A},\boldsymbol{R}\in\mathcal{B}}{\text{minimize}} \quad \underset{\boldsymbol{w}\neq\mathbf{0}}{\text{maximize}} \; \frac{|\boldsymbol{w}^H\boldsymbol{a}|^2}{\boldsymbol{w}^H\boldsymbol{R}\boldsymbol{w}}.$$

- Otherwise, the trivial case holds:

$$\underset{\boldsymbol{a}\in\mathcal{A},\boldsymbol{R}\in\mathcal{B}}{\text{minimize}} \quad \underset{\boldsymbol{w}\in\mathcal{W}}{\text{maximize}} \; \frac{|\boldsymbol{w}^H\boldsymbol{a}|^2}{\boldsymbol{w}^H\boldsymbol{R}\boldsymbol{w}} \geq \underset{\boldsymbol{w}\in\mathcal{W}}{\text{maximize}} \quad \underset{\boldsymbol{a}\in\mathcal{A},\boldsymbol{R}\in\mathcal{B}}{\text{minimize}} \; \frac{|\boldsymbol{w}^H\boldsymbol{a}|^2}{\boldsymbol{w}^H\boldsymbol{R}\boldsymbol{w}}.$$

Maximin SINR vs. Minimax SINR

- Why the worst-case SINR maximization based RAB problem is useful and solved to find an optimal beamforming solution $\boldsymbol{w}^{\star}$, rather than the minimax SINR problem (an upper bound for the maximin SINR problem)?
  - When evaluating $\boldsymbol{w}^{\star}$, the real data is outside the preset uncertainty set very far away.
  - The minimax problem is a hard problem (e.g., a nonconvex optimization problem) while the maximin problem is easy (e.g., a convex or hidden convex problem).
  - In some fast changing environment (e.g., an automotive radar receiver), the beamforming solution obtained from the minimax SINR problem using the out-of-date data (after which the data has been collected for many times) is not guaranteed to have better array performance than the beamforming solution for the maximin problem.

Uncertainty Sets of Steering Vectors

- Let us focus on how the RAB solution $\boldsymbol{w}^\star$ is affected by uncertainty sets of the desired signal steering vectors.
- So, we consider $\mathcal{W} = \{\boldsymbol{w} \in \mathbb{C}^N \mid \boldsymbol{w} \neq \mathbf{0}\}$, and

$$\mathcal{B}_1 = \{\boldsymbol{R} \mid \|\boldsymbol{R} - \hat{\boldsymbol{R}}\|_F^2 \leq \gamma, \, \boldsymbol{R} \succeq \mathbf{0}\}.$$

- Also, a nonconvex uncertainty set of the signal steering vectors

$$\mathcal{A}_1 = \{\tilde{\boldsymbol{a}} \mid \|\tilde{\boldsymbol{a}} - \boldsymbol{a}\|^2 \leq \epsilon, \, N(1-\eta_1) \leq \|\tilde{\boldsymbol{a}}\|^2 \leq N(1+\eta_2)\}$$

  includes a similarity constraint and a norm perturbation constraint of the steering vectors.
- Some other generalized uncertainty sets can be studied (but not herein):
  - $\mathcal{A}_1' = \{\tilde{\boldsymbol{a}} = \boldsymbol{Q}\boldsymbol{u} + \boldsymbol{a} \mid \|\boldsymbol{u}\| \leq 1, \, N(1-\eta_1) \leq \|\tilde{\boldsymbol{a}}\|^2 \leq N(1+\eta_2)\}$, where $\boldsymbol{Q} \in \mathbb{C}^{N \times M}$ [Lorenz and Boyd, 2005].
  - If $\boldsymbol{Q}$ is of full rank, $\mathcal{A}_1'' = \{\tilde{\boldsymbol{a}} \mid \|(\boldsymbol{Q}^H\boldsymbol{Q})^{-1}\boldsymbol{Q}^H(\tilde{\boldsymbol{a}} - \boldsymbol{a})\| \leq 1, \, N(1-\eta_1) \leq \|\tilde{\boldsymbol{a}}\|^2 \leq N(1+\eta_2)\}$.

MVDR RAB Optimization Problem

- The minimax SINR problem is written as, with uncertainty sets $\mathcal{A}_1$ and $\mathcal{B}_1$:

$$\boxed{\underset{\tilde{\boldsymbol{a}} \in \mathcal{A}_1, \boldsymbol{R} \in \mathcal{B}_1}{\text{minimize}} \quad \underset{\boldsymbol{w} \neq \boldsymbol{0}}{\text{maximize}} \quad \frac{|\boldsymbol{w}^H \tilde{\boldsymbol{a}}|^2}{\boldsymbol{w}^H \boldsymbol{R} \boldsymbol{w}}.}$$

- It is further expressed as

$$\underset{\tilde{\boldsymbol{a}} \in \mathcal{A}_1, \boldsymbol{R} \in \mathcal{B}_1}{\text{minimize}} \quad \tilde{\boldsymbol{a}}^H \boldsymbol{R}^{-1} \tilde{\boldsymbol{a}}.$$

- An equivalent nonconvex optimization is derived:

$$\begin{aligned} \underset{t, \tilde{\boldsymbol{a}}, \boldsymbol{R}}{\text{minimize}} \quad & t \\ \text{subject to} \quad & \begin{bmatrix} t & \tilde{\boldsymbol{a}}^H \\ \tilde{\boldsymbol{a}} & \boldsymbol{R} \end{bmatrix} \succeq \boldsymbol{0}, \\ & \|\tilde{\boldsymbol{a}} - \boldsymbol{a}\|^2 \leq \epsilon, \\ & N(1 - \eta_1) \leq \|\tilde{\boldsymbol{a}}\|^2 \leq N(1 + \eta_2), \\ & \|\boldsymbol{R} - \hat{\boldsymbol{R}}\|_F^2 \leq \gamma. \end{aligned}$$

MVDR RAB Optimization Problem

- However, if we assume that the sample covariance matrix is sufficiently good, i.e., $\boldsymbol{R} = \hat{\boldsymbol{R}}$, then the previous problem is simplified into the hidden convex QCQP:

$$\begin{aligned} \underset{\tilde{\boldsymbol{a}}}{\text{minimize}} \quad & \tilde{\boldsymbol{a}}^H \hat{\boldsymbol{R}}^{-1} \tilde{\boldsymbol{a}} \\ \text{subject to} \quad & \|\tilde{\boldsymbol{a}} - \boldsymbol{a}\|^2 \leq \epsilon, \\ & N(1-\eta_1) \leq \|\tilde{\boldsymbol{a}}\|^2 \leq N(1+\eta_2). \end{aligned}$$

- It is known that it is a hidden convex problem, namely, a globally optimal solution $\tilde{\boldsymbol{a}}^\star$ can be obtained within a polynomial-time complexity.
- If $\tilde{\boldsymbol{a}}^\star$ is optimal for the QCQP, then the MVDR RAB vector is defined as:

$$\boldsymbol{w}^\star = \hat{\boldsymbol{R}}^{-1} \tilde{\boldsymbol{a}}^\star.$$

- This RAB solution will be evaluated in our late simulation, with comparision to other proposed RAB vectors.

Worst-Case SINR Maximization

- The worst-case SINR maximization based RAB design is:

$$\underset{\boldsymbol{w}\neq\boldsymbol{0}}{\text{maximize}} \quad \underset{\tilde{\boldsymbol{a}}\in\mathcal{A}_1, \boldsymbol{R}\in\mathcal{B}_1}{\text{minimize}} \quad \frac{|\boldsymbol{w}^H\tilde{\boldsymbol{a}}|^2}{\boldsymbol{w}^H\boldsymbol{R}\boldsymbol{w}}.$$

- Evidently, it can be recast into (since the two parameters are separable):

$$\underset{\boldsymbol{w}\neq\boldsymbol{0}}{\text{maximize}} \quad \frac{\underset{\tilde{\boldsymbol{a}}\in\mathcal{A}_1}{\text{minimize}}\ |\boldsymbol{w}^H\tilde{\boldsymbol{a}}|^2}{\underset{\boldsymbol{R}\in\mathcal{B}_1}{\text{maximize}}\ \boldsymbol{w}^H\boldsymbol{R}\boldsymbol{w}}.$$

- It is not hard to verify that $\boldsymbol{w}^H(\hat{\boldsymbol{R}}+\sqrt{\gamma}\boldsymbol{I})\boldsymbol{w} = \underset{\boldsymbol{R}\in\mathcal{B}_1}{\text{maximize}}\ \boldsymbol{w}^H\boldsymbol{R}\boldsymbol{w}$.
- Letting $\bar{\boldsymbol{R}} = \hat{\boldsymbol{R}}+\sqrt{\gamma}\boldsymbol{I}$, this is further equivalent to the following problem:

$$\begin{aligned} \underset{\boldsymbol{w}\neq\boldsymbol{0}}{\text{minimize}} \quad & \boldsymbol{w}^H\bar{\boldsymbol{R}}\boldsymbol{w} \\ \text{subject to} \quad & \underset{\tilde{\boldsymbol{a}}\in\mathcal{A}_1}{\text{minimize}}\ |\boldsymbol{w}^H\tilde{\boldsymbol{a}}|^2 \geq 1. \end{aligned}$$

Worst-Case SINR Maximization

- Related existing works:
  - The scenario of the uncertainty set $\mathcal{A}_1 = \{\tilde{\boldsymbol{a}} \mid \|\tilde{\boldsymbol{a}} - \boldsymbol{a}\|^2 \leq \epsilon\}$ has been studied in an early paper [Vorobyov, Gershman, Luo, 2003].
  - Therein, the worst-case SINR maximization problem was transformed into an SOCP problem.
  - The case where $\mathcal{W} = \{\boldsymbol{w} \mid |\Re(\hat{\boldsymbol{a}}_0^H \boldsymbol{w}) \geq 1 + \epsilon_0\|\boldsymbol{w}\|, |\boldsymbol{a}(\theta_j)^H \boldsymbol{w}| + \epsilon_j\|\boldsymbol{w}\| \leq \sqrt{\eta}, j = 1, \ldots, J\}$ with the grid $\Theta = \{\theta_1, \ldots, \theta_J\}$ enforcing the robust mainlobe and sidelobe levels control, was consided [Huang, Li, Vorobyov, 2024] to maximize the worst-case INP.
  - It can be recast into an SOCP again, with one half grid size but lower lower sidelobes in the normalized beampattern.
  - Many others....

Maximin & Nonconvex Uncertainty Set

- To solve the worst-case SINR maximization based RAB problem

$$\boxed{\underset{\boldsymbol{w}\neq\boldsymbol{0}}{\text{minimize}}\;\; \boldsymbol{w}^H\bar{\boldsymbol{R}}\boldsymbol{w} \;\;\text{subject to}\;\; \underset{\tilde{\boldsymbol{a}}\in\mathcal{A}_1}{\text{minimize}}\; |\boldsymbol{w}^H\tilde{\boldsymbol{a}}|^2 \geq 1,}$$

  only the QCQP problem in the constraint shall be dealt with.
- By the way, additional constraints on $\boldsymbol{w}$ can be considered if necessary; for example, a sparsity constraint on the beamforming vector; then the RAB problem is formulated into

$$\begin{array}{ll} \underset{\boldsymbol{w}\neq\boldsymbol{0}}{\text{minimize}} & \boldsymbol{w}^H\bar{\boldsymbol{R}}\boldsymbol{w} \\ \text{subject to} & \underset{\tilde{\boldsymbol{a}}\in\mathcal{A}_1}{\text{minimize}}\; |\boldsymbol{w}^H\tilde{\boldsymbol{a}}|^2 \geq 1 \\ & \|\boldsymbol{w}\|_0 \leq 2N/3. \end{array}$$

Hiddenly Convex QCQP and Its Dual

- The inhomogeneous QCQP in the constraint is written as:

$$\begin{array}{ll} \underset{\tilde{\boldsymbol{a}} \in \mathbb{C}^N}{\text{minimize}} & \tilde{\boldsymbol{a}}^H \boldsymbol{w}\boldsymbol{w}^H \tilde{\boldsymbol{a}} \\ \text{subject to} & \|\tilde{\boldsymbol{a}} - \boldsymbol{a}\|^2 \leq \epsilon, \\ & N(1-\eta_1) \leq \|\tilde{\boldsymbol{a}}\|^2 \leq N(1+\eta_2). \end{array}$$

- Observe that $\boldsymbol{a}$ with $\|\boldsymbol{a}\|^2 = N$ is a strictly feasible point of the QCQP.
- Then the SDP relaxation problem of it has an interior point too (in fact, the interior point can be constructed from the strictly feasible point $\boldsymbol{a}$ of the QCQP).
- The dual of the SDP can be shown strictly feasible.
- Therefore, it follows from the linear conic programming strong duality theorem that the QCQP, the SDP relaxation and its dual problems are solvable and share the equal optimal value [Huang and Palomar, 2014].

QMI Reformulation of Maximin SINR

- The dual problem for the SDP or the QCQP is the LMI problem:

$$\begin{aligned} \underset{y_1,y_2,y_3,y_4}{\text{maximize}} \quad & N(1-\eta_1)y_2 + N(1+\eta_2)y_3 + y_4 \\ \text{subject to} \quad & \begin{bmatrix} \boldsymbol{w}\boldsymbol{w}^H - (y_1+y_2+y_3)\boldsymbol{I} & y_1\boldsymbol{a} \\ y_1\boldsymbol{a}^H & -y_4 - y_1(\|\boldsymbol{a}\|^2 - \epsilon) \end{bmatrix} \succeq \boldsymbol{0}, \\ & y_1 \le 0,\ y_2 \ge 0,\ y_3 \le 0,\ y_4 \in \mathbb{R}. \end{aligned}$$

- The worst-case SINR maximization problem RAB design problem:

$$\underset{\boldsymbol{w}\neq\boldsymbol{0}}{\text{minimize}} \ \boldsymbol{w}^H\bar{\boldsymbol{R}}\boldsymbol{w} \quad \text{subject to} \quad \underset{\tilde{\boldsymbol{a}}\in\mathcal{A}_1}{\text{minimize}} \ |\boldsymbol{w}^H\tilde{\boldsymbol{a}}|^2 \ge 1,$$

has an equivalent quadratic matrix inequality (QMI) problem formulation (as shown in the next slide).

QMI Reformulation of Maximin SINR

- The QMI problem w.r.t. beamforming vector $\boldsymbol{w}$ is expressed as:

$$\begin{aligned}
\underset{\boldsymbol{w}, \{y_i\}}{\text{minimize}} \quad & \boldsymbol{w}^H \bar{\boldsymbol{R}} \boldsymbol{w} \\
\text{subject to} \quad & N(1-\eta_1)y_2 + N(1+\eta_2)y_3 + y_4 \geq 1, \\
& \begin{bmatrix} \boldsymbol{w}\boldsymbol{w}^H - (y_1+y_2+y_3)\boldsymbol{I} & y_1\boldsymbol{a} \\ y_1\boldsymbol{a}^H & -y_4 - y_1(\|\boldsymbol{a}\|^2 - \epsilon) \end{bmatrix} \succeq \boldsymbol{0}, \\
& \boldsymbol{w} \in \mathbb{C}^N,\ y_1 \leq 0,\ y_2 \geq 0,\ y_3 \leq 0,\ y_4 \in \mathbb{R}.
\end{aligned}$$

- A QMI problem is hard to solve, in general, with higher computational cost.
- The quadratic matrix inequaliy constraint implies

$$y_1 + y_2 + y_3 \leq 0.$$

LMI Relaxation of QMI Problem

- The LMI relaxation for the QMI problem is formulated as

$$\begin{aligned} \underset{\boldsymbol{W}, \{y_i\}}{\text{minimize}} \quad & \text{tr}\,(\bar{\boldsymbol{R}}\boldsymbol{W}) \\ \text{subject to} \quad & N(1-\eta_1)y_2 + N(1+\eta_2)y_3 + y_4 \geq 1, \\ & \begin{bmatrix} \boldsymbol{W} - (y_1+y_2+y_3)\boldsymbol{I} & y_1\boldsymbol{a} \\ y_1\boldsymbol{a}^H & -y_4 - y_1(\|\boldsymbol{a}\|^2 - \epsilon) \end{bmatrix} \succeq \boldsymbol{0}, \\ & y_1 + y_2 + y_3 \leq 0, \\ & \boldsymbol{W} \succeq \boldsymbol{0},\ y_1 \leq 0,\ y_2 \geq 0,\ y_3 \leq 0,\ y_4 \in \mathbb{R}. \end{aligned}$$

- Solving the SDR may not reach an optimal rank-one solution.
- Thus, how to design an efficient algorithm to return an optimal rank-one solution $\boldsymbol{W}^\star = \boldsymbol{w}^\star \boldsymbol{w}^{\star H}$ for the LMI problem is a key question to the worst-case SINR maximization RAB problem.

A Tightened LMI Relaxation

- Not many approaches to solve the SDP relaxation problem for a QMI problem guaranteeing a rank-one solution, years ago
- Here, we would like to present an efficient method to get a rank-one approximate solution.
- Suppose that any feasible $\boldsymbol{W} \neq \mathbf{0}$ (otherwise, $\boldsymbol{w} = \mathbf{0}$ is optimal for the QMI problem).
- The statement that a nonzero $\boldsymbol{Y} \succeq \mathbf{0}$ is of rank one is equivalent to $\lambda_2(\boldsymbol{Y}) \leq 0$, or

$$\lambda_1(\boldsymbol{Y}) + \lambda_2(\boldsymbol{Y}) \leq \lambda_1(\boldsymbol{Y}),$$

where the eigenvalues $\lambda_1(\boldsymbol{Y}) \geq \lambda_2(\boldsymbol{Y}) \geq \cdots \geq \lambda_N(\boldsymbol{Y}) \geq 0$ are ordered in a descending way.

A Tightened LMI Relaxation

- The rank-one condition $\lambda_1(\boldsymbol{Y}) + \lambda_2(\boldsymbol{Y}) \leq \lambda_1(\boldsymbol{Y})$ is universal, since it can be plugged to the SDP relaxation, such that the tightened SDR is identical to the original QMI or QCQP problem.
- Therefore, the original QMI problem is reformulated as:

$$
\begin{aligned}
\underset{\boldsymbol{W},\{y_i\}}{\text{minimize}} \quad & \operatorname{tr}(\bar{\boldsymbol{R}}\boldsymbol{W}) \\
\text{subject to} \quad & (N-\eta_1)y_2 + (N+\eta_2)y_3 + y_4 = 1, \\
& \begin{bmatrix} \boldsymbol{W} - (y_1+y_2+y_3)\boldsymbol{I} & y_1\boldsymbol{a} \\ y_1\boldsymbol{a}^H & -y_4 - y_1(\|\boldsymbol{a}\|^2 - \epsilon) \end{bmatrix} \succeq \boldsymbol{0}, \\
& y_1 + y_2 + y_3 \leq 0, \\
& \lambda_1(\boldsymbol{W}) + \lambda_2(\boldsymbol{W}) \leq \lambda_1(\boldsymbol{W}), \\
& \boldsymbol{W} \succeq \boldsymbol{0},\ y_1 \leq 0,\ y_2 \geq 0,\ y_3 \leq 0,\ y_4 \in \mathbb{R}.
\end{aligned}
$$

LMI for Sum of K Largest Eigenvalues

- However, $\lambda_1(\boldsymbol{W})$ and $\lambda_1(\boldsymbol{W})+\lambda_2(\boldsymbol{W})$ are convex for $\boldsymbol{W} \succeq \mathbf{0}$.
- Suppose that $S_K(\boldsymbol{W})$ is the sum of K largest eigenvalues of Hermitian matrix $\boldsymbol{W}$ ($K \leq N$). Then the epigraph $\{(\boldsymbol{W}, t) \mid S_K(\boldsymbol{W}) \leq t\}$ of the function admits the LMI representation:

$$\begin{aligned} t - Ks - \operatorname{tr} \boldsymbol{Z} &\geq 0 \\ \boldsymbol{Z} - \boldsymbol{W} + s\boldsymbol{I} &\succeq \mathbf{0} \\ \boldsymbol{Z} &\succeq \mathbf{0} \end{aligned}$$

where $\boldsymbol{Z}$ is an $N \times N$ Hermitian matrix and s is a real number.*

*A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization*, MPS-SIAM Series on Optimization, SIAM, Philadelphia, 2001.

Equivalent Description of Rank-One Matrix

- Thus, the rank-one condition for $\boldsymbol{W} \succeq \mathbf{0}$, $\lambda_1(\boldsymbol{W}) + \lambda_2(\boldsymbol{W}) \leq \lambda_1(\boldsymbol{W})$, amounts to

$$\lambda_1(\boldsymbol{W}) - 2s - \operatorname{tr} \boldsymbol{Z} \geq 0, \ \boldsymbol{Z} - \boldsymbol{W} + s\boldsymbol{I} \succeq \mathbf{0}, \boldsymbol{Z} \succeq \mathbf{0},$$

- Here, the variables are $\boldsymbol{W}$, $\boldsymbol{Z}$ and s, with two LMIs and one inequality constraints.
- However, the inequality condition is nonconvex (since $\lambda_1(\boldsymbol{W})$ is convex).
- Nevertheless, it is known that

$$\boxed{\lambda_1(\boldsymbol{W}) = \text{maximize } \operatorname{tr}(\boldsymbol{W}\boldsymbol{X}) \ \text{ subject to } \operatorname{tr} \boldsymbol{X} = 1, \ \boldsymbol{X} \succeq \mathbf{0}.}$$

Bilinear Reformulation of QMI Problem

- A bilinear matrix inequality (BLMI) form for the QMI problem is built: [Huang, Fu, Vorobyov, Luo, 2023]:

$$
\begin{aligned}
\text{minimize} \quad & \operatorname{tr}(\bar{\boldsymbol{R}}\boldsymbol{W}) \\
\text{subject to} \quad & N(1-\eta_1)y_2 + N(1+\eta_2)y_3 + y_4 = 1, \\
& \begin{bmatrix} \boldsymbol{W} - (y_1+y_2+y_3)\boldsymbol{I} & y_1\boldsymbol{a} \\ y_1\boldsymbol{a}^H & -y_4 - y_1(\|\boldsymbol{a}\|^2 - \epsilon) \end{bmatrix} \succeq \boldsymbol{0}, \\
& y_1 + y_2 + y_3 \le 0, \\
& \operatorname{tr}(\boldsymbol{W}\boldsymbol{X}) - 2s - \operatorname{tr}\boldsymbol{Z} \ge 0, \\
& \boldsymbol{Z} - \boldsymbol{W} + s\boldsymbol{I} \succeq \boldsymbol{0}, \\
& \operatorname{tr}\boldsymbol{X} = 1, \\
& \boldsymbol{W} \succeq \boldsymbol{0},\ \boldsymbol{Z} \succeq \boldsymbol{0}, \boldsymbol{X} \succeq \boldsymbol{0}, \\
& y_1 \le 0,\ y_2 \ge 0,\ y_3 \le 0,\ y_4 \in \mathbb{R},\ s \in \mathbb{R}.
\end{aligned}
$$

Bilinear Reformulation of QMI Problem

- The optimization variables are $\boldsymbol{W}$, $\boldsymbol{Z}$, $\boldsymbol{X}$, $\{y_i\}$ and s.
- The only bilinar term $\text{tr}(\boldsymbol{W}\boldsymbol{X})$ in the fourth constraint is non-convex (w.r.t. $\boldsymbol{W}$ and $\boldsymbol{X}$ jointly), while all others are convex.
- The red and blue constraints are tantamount to the rank-one condition for $\boldsymbol{W}$.
- In words, the bilinear optimization problem is equivalent to the original worst-case SINR maximization based RAB problem.
- So, we only need to deal with the bilinear term $\text{tr}(\boldsymbol{W}\boldsymbol{X})$.
- If a feasible solution $\boldsymbol{W}$ for the BLMI problem is computed efficiently, then the solution $\boldsymbol{W}$ must be rank-one.
- Note that the bigger $\text{tr}(\boldsymbol{W}\boldsymbol{X})$ is, the larger the feasible set of the BLMI problem is.

Approximation for the Bilinear Problem

- Solve LMI relaxation problem of the QMI reformulation for the maximin SINR, obtaining $\boldsymbol{W}^\star$ as an initial point, namely, $\boldsymbol{W}_k \triangleq \boldsymbol{W}^\star$, $k = 0$;
- Update the bilinear inequality constraint to

$$\operatorname{tr}\left(\boldsymbol{W}_k \boldsymbol{X}\right) - 2s - \operatorname{tr}\left(\boldsymbol{Z}\right) \geq 0,$$

and solve the LMI problem:

Approximation for the Bilinear Problem

$$
\begin{array}{ll}
\text{minimize} & \operatorname{tr}\left(\bar{\boldsymbol{R}}\boldsymbol{W}\right) \\
\text{subject to} & (N-\eta_1)y_2 + (N+\eta_2)y_3 + y_4 = 1, \\
& \begin{bmatrix} \boldsymbol{W} - (y_1+y_2+y_3)\boldsymbol{I} & y_1\boldsymbol{a} \\ y_1\boldsymbol{a}^H & -y_4 - y_1(\|\boldsymbol{a}\|^2 - \epsilon) \end{bmatrix} \succeq \boldsymbol{0}, \\
& y_1 + y_2 + y_3 \le 0, \\
& \operatorname{tr}\left(\boldsymbol{W}_k\boldsymbol{X}\right) - 2s - \operatorname{tr}\left(\boldsymbol{Z}\right) \ge 0, \\
& \boldsymbol{Z} - \boldsymbol{W} + s\boldsymbol{I} \succeq \boldsymbol{0}, \\
& \operatorname{tr}\boldsymbol{X} = 1, \\
& \boldsymbol{W} \succeq \boldsymbol{0},\ \boldsymbol{Z} \succeq \boldsymbol{0},\ \boldsymbol{X} \succeq \boldsymbol{0}, \\
& y_1 \le 0,\ y_2 \ge 0,\ y_3 \le 0,\ y_4 \in \mathbb{R},\ s \in \mathbb{R}.
\end{array}
$$

– Once a new solution $\boldsymbol{W}^\star$ is obtained, set $\boldsymbol{W}_{k+1} \triangleq \boldsymbol{W}^\star$ and $k = k+1$.

Approximation for the Bilinear Problem

- In this way, an iterative procedure is established until a stopping criterion (for instance, $\|\boldsymbol{W}_k - \boldsymbol{W}_{k-1}\|_2 \leq \xi$) is fulfilled.

- The approximation algorithm is summarized as follows.

1. Let $k = 0$, and $\boldsymbol{W}_0 = \boldsymbol{W}^\star$ (optimal for the LMI relaxation problem for the QMI problem);
2. Do
3. Solve the LMI problem (previous slide), obtaining solution $\boldsymbol{W}_{k+1}$;
4. $k = k + 1$;
5. Until $\|\boldsymbol{W}_k - \boldsymbol{W}_{k-1}\|_2 \leq \xi$;
6. Output $\boldsymbol{w}^\star = \sqrt{\lambda_1}\boldsymbol{w}_1$ (where $\boldsymbol{W}_k = \lambda_1 \boldsymbol{w}_1 \boldsymbol{w}_1^H$ indeed is rank-1).

Convergence Properties

- It is verified that $\mathrm{tr}\,(\bar{\boldsymbol{R}}\boldsymbol{W}_{k-1}) \geq \mathrm{tr}\,(\bar{\boldsymbol{R}}\boldsymbol{W}_{k})$ for $k = 1, 2, \ldots$, since $\boldsymbol{W}_{k-1}$ is feasible for the LMI in the k-th step (previous slide).
- Since the finally output solution must be feasible for the BLMI problem, it is of rank one.

Rank-Reduction for LMI Relaxation

- Claim: For $\boldsymbol{W} \succeq \mathbf{0}$, $\text{tr}\,(\boldsymbol{W}) = \|\boldsymbol{W}\|_F = \text{tr}\,(\boldsymbol{W}\boldsymbol{W})/\|\boldsymbol{W}\|_F$ $\iff$ $\boldsymbol{W}$ is of rank one. [Huang, Yang, Vorobyov, 2022]
- The LMI relaxation can be solved iteratively by solving

$$\begin{aligned}
\underset{\boldsymbol{W},\{y_i\}}{\text{minimize}} \quad & \text{tr}\,(\bar{\boldsymbol{R}}\boldsymbol{W}) + \rho\Big(\text{tr}\,(\boldsymbol{W}) - \frac{\text{tr}\,(\boldsymbol{W}\boldsymbol{W}_k)}{\|\boldsymbol{W}_k\|_F}\Big) \\
\text{subject to} \quad & N(1-\eta_1)y_2 + N(1+\eta_2)y_3 + y_4 \geq 1, \\
& \begin{bmatrix} \boldsymbol{W} - (y_1+y_2+y_3)\boldsymbol{I} & y_1\boldsymbol{a} \\ y_1\boldsymbol{a}^H & -y_4 - y_1(\|\boldsymbol{a}\|^2 - \epsilon) \end{bmatrix} \succeq \mathbf{0}, \\
& y_1 + y_2 + y_3 \leq 0, \\
& \boldsymbol{W} \succeq \mathbf{0},\ y_1 \leq 0,\ y_2 \geq 0,\ y_3 \leq 0,\ y_4 \in \mathbb{R}.
\end{aligned}$$

- This rank-one PSD condition is weaker than $\|\boldsymbol{W}\|_2 = \|\boldsymbol{W}\|_F$, or $\|\boldsymbol{W}\|_2 = \|\boldsymbol{W}\|_*$.

Simulation Setup

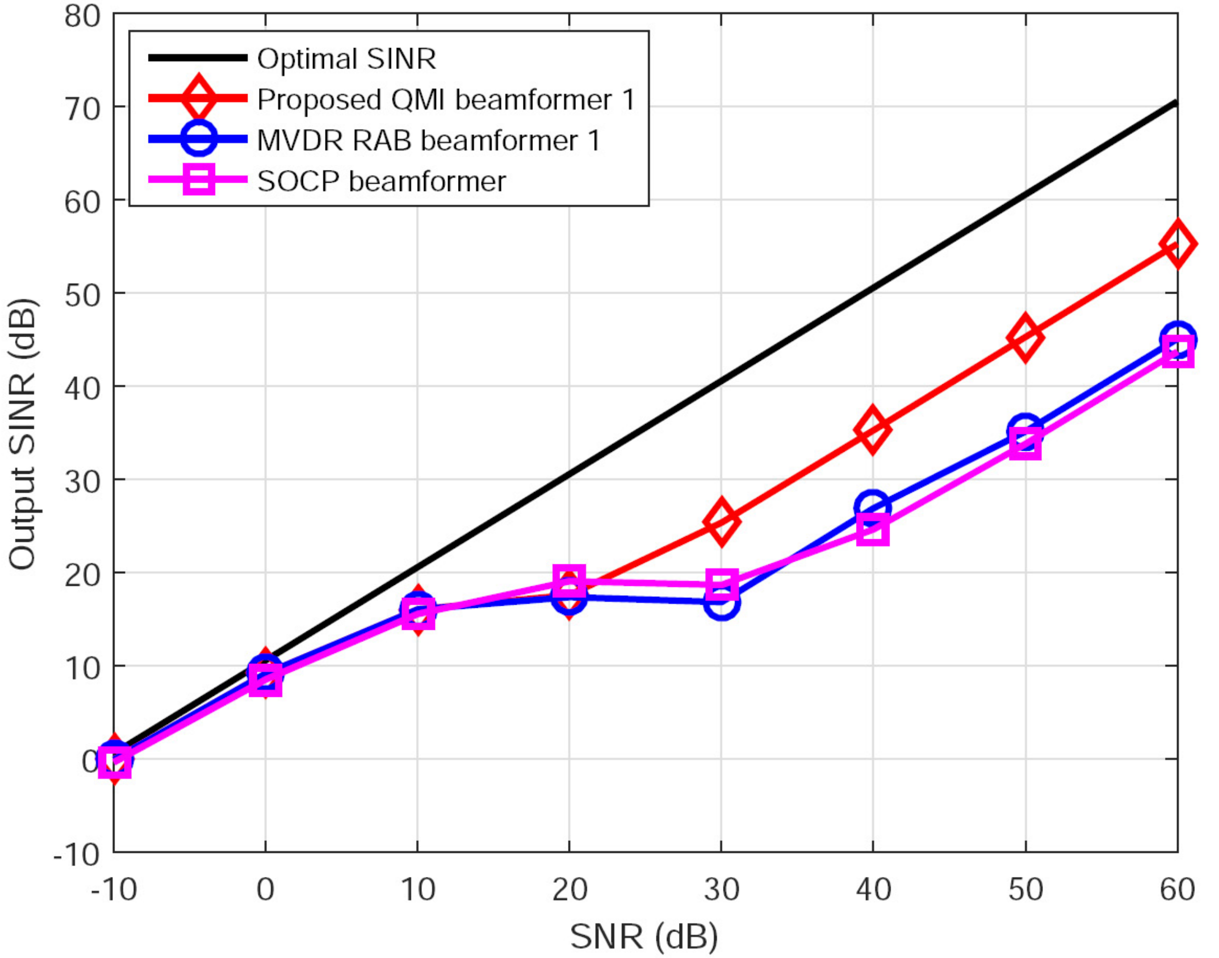
- Scenario settings:
  - A uniformly linear array with $N = 12$ sensors spaced half a wavelength;
  - the actual signal direction $\theta = 7^\circ$;
  - the presumed direction $\theta_0 = 5^\circ$;
  - two interferers located in the directions of $\theta_{1,2} = \pm 15^\circ$ with the same interference-to-noise ratio (INR) of 30 dB;
  - the array noise: a spatially and temporally white Gaussian vector with zero mean and covariance $\boldsymbol{I}$;
  - wavefront distortion: The phase increments are independent Gaussian variables each with zero mean and standard deviation 0.02, with the zero initial phase distortion;

Simulation Setup

  - parameters $\eta_1 = \eta_2 = 0.2$, and $\epsilon = 0.3N$;
  - the radius $\sqrt{\gamma} = 0.1\lambda_N(\hat{\boldsymbol{R}})$.
- All results are averaged over 200 simulation runs.
- Three robust beamformers are compared: The proposed one, SOCP beamformer, and MVDR RAB beamformer.
- The SOCP beamformer is a solution for the RAB design problem in [Vorobyov, Gershman, Luo, 2003].
- The MVDR RAB beamformer is the optimal beamformer for the minimax SINR problem with $\boldsymbol{R} = \hat{\boldsymbol{R}}$.

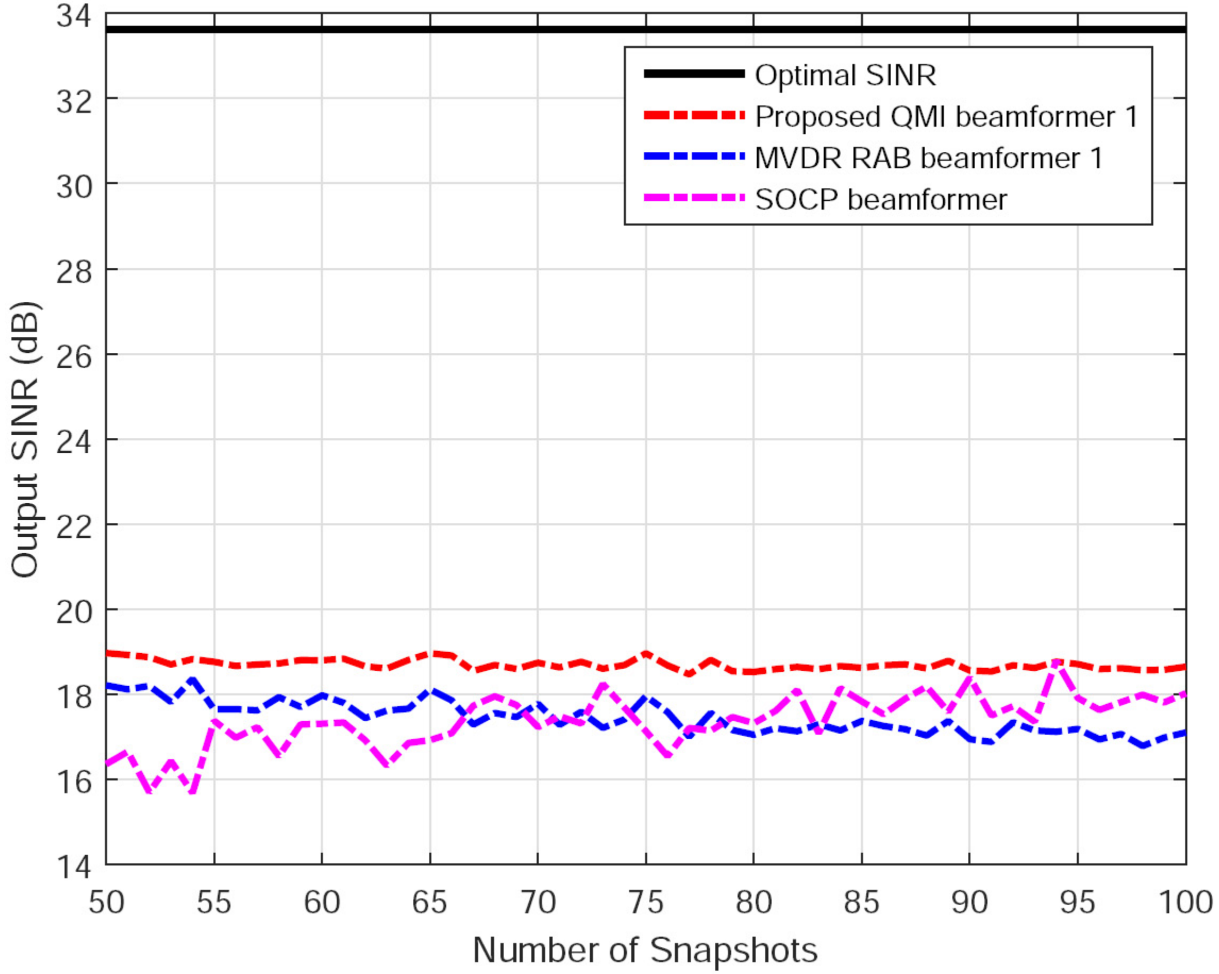
Beamformer Output SINR vs SNR

The average beamformer output SINR versus the SNR, with number of snapshots $T = 100$:



Array Output SINR vs Snapshot Numbers

The average array output SINR versus the number of snapshots, with SNR=23 dB (due to $200 = 10^{23/10}$):



Lessons

- The RAB designs can be implemented via the worst-case SINR maximization (maximin SINR) and the minimax SINR problems.
- MVDR RAB beamforming optimization problem is identical to the minimax SINR problem, which could be a hard problem (it depends on the predefined uncertainty sets).
- The worst-case SINR maximization based RAB problem can be reformulated into a QMI problem and further reexpressed as a bi-linear optimization problem, via which an approximation algorithm can be designed.
- Alternatively, the QMI is solved by using a certain rank-reduction process to obtain a rank-one solution for the LMI relaxation problem.

Part IV: Stochastic Uncertainty Modeling

Probabilistically-constrained Beamformer

Our formulation of the robust probabilistically-constrained beamformer

$$\min_{\boldsymbol{w}} \boldsymbol{w}^H \hat{\boldsymbol{R}} \boldsymbol{w} \qquad \text{s. t.} \qquad \Pr\{|\boldsymbol{w}^H(\boldsymbol{a} + \boldsymbol{\Delta})| \geq 1\} \geq p$$

$$\boldsymbol{\Delta} \sim \mathcal{N}_{\mathcal{C}}(\boldsymbol{0}, \boldsymbol{C}_{\boldsymbol{\Delta}})$$

- instead of maintaining a distortionless response for a continuum of *all possible steering vectors* $\tilde{\boldsymbol{a}}$, our constrain guarantees that the distortionless response will be maintained *with a certain "sufficient" probability*
- $|\boldsymbol{w}^H(\boldsymbol{a} + \boldsymbol{\Delta})|^2$ is *Ricean distributed* – how to solve?

Well approximated by SOCP for both Gaussian and worst-case distributed mismatch.

Probabilistically-constrained Beamformer

Approximation of the probability in the constraint

$$\begin{aligned}\Pr\{|\boldsymbol{w}^H\tilde{\boldsymbol{a}}| \geq 1\} &\geq \Pr\{|\mathrm{Re}\{\boldsymbol{w}^H\tilde{\boldsymbol{a}}\}| \geq 1 \cap |\mathrm{Im}\{\boldsymbol{w}^H\tilde{\boldsymbol{a}}\}| \geq 1\} \\ &= \Pr\{|\mathrm{Re}\{\boldsymbol{w}^H\tilde{\boldsymbol{a}}\}| \geq 1\}\ \Pr\{|\mathrm{Im}\{\boldsymbol{w}^H\tilde{\boldsymbol{a}}\}| \geq 1\} \\ &= \left(\Pr\{|\mathrm{Re}\{\boldsymbol{w}^H\tilde{\boldsymbol{a}}\}| \geq 1\}\right)^2 = \left(\Pr\{|\mathrm{Im}\{\boldsymbol{w}^H\tilde{\boldsymbol{a}}\}| \geq 1\}\right)^2\end{aligned}$$

Approximate problem

$$\min_{\boldsymbol{w}} \boldsymbol{w}^H(\hat{\boldsymbol{R}} + \xi \boldsymbol{I})\boldsymbol{w} \qquad \text{s. t.} \qquad \Pr\{|\mathrm{Re}\{\boldsymbol{w}^H\tilde{\boldsymbol{a}}\}| \geq 1\} \geq \sqrt{p}$$

$$\mathrm{Re}\{\boldsymbol{w}^H\tilde{\boldsymbol{a}}\} \sim \mathcal{N}_{\mathcal{R}}\left(\mathrm{Re}\{\boldsymbol{w}^H\boldsymbol{a}\},\ \|\boldsymbol{C}_{\boldsymbol{\Delta}}^{1/2}\boldsymbol{w}\|^2/2\right)$$

Probabilistically constrained Beamforming

Using the standard error function for Gaussian distribution

$$\begin{aligned}\Pr\{|\mathrm{Re}\{\boldsymbol{w}^H\tilde{\boldsymbol{a}}\}| \geq 1\} &= \Pr\{\mathrm{Re}\{\boldsymbol{w}^H\tilde{\boldsymbol{a}}\} \geq 1\} \\ &- \Pr\{\mathrm{Re}\{\boldsymbol{w}^H\tilde{\boldsymbol{a}}\} \geq -1\} \\ &= \frac{1}{2}\left[\mathrm{erf}\left(\frac{1-\mathrm{Re}\{\boldsymbol{w}^H\boldsymbol{a}\}}{\|\boldsymbol{C}_{\boldsymbol{\Delta}}^{1/2}\boldsymbol{w}\|}\right)\right. \\ &- \left.\mathrm{erf}\left(\frac{-1-\mathrm{Re}\{\boldsymbol{w}^H\boldsymbol{a}\}}{\|\boldsymbol{C}_{\boldsymbol{\Delta}}^{1/2}\boldsymbol{w}\|}\right)\right]\end{aligned}$$

If the mismatch's distribution is worst-case, we derive a generalization of Chebyshov inequality (has to hold for all distributions). The solution has the same SOCP structure!

Part V: Distributional Uncertainty Modeling

Distributionally Robust Optimization (DRO)

Consider both $\boldsymbol{R}_{\mathrm{i+n}} \in \mathcal{H}^N$ and $\boldsymbol{a} \in \mathbb{C}^N$ as random variables.

The RAB problem can be formulated as a stochastic programming problem:

$$\begin{aligned} \underset{\boldsymbol{w}}{\text{minimize}} \quad & \mathbb{E}[\boldsymbol{w}^H \boldsymbol{R}_{\mathrm{i+n}} \boldsymbol{w}] \\ \text{subject to} \quad & \mathbb{E}[\boldsymbol{w}^H \boldsymbol{a}\boldsymbol{a}^H \boldsymbol{w}] \geq 1, \end{aligned}$$

However, the exact knowledge of the underlying distribution might not always be available.

- distributionally robust optimization (DRO) is employed
  - considers uncertainty sets that capture the uncertainty about the true distribution and aims to maximize the worst-case SINR over a range of possible distributions.

$$\boxed{\begin{aligned} \underset{\boldsymbol{w}}{\text{minimize}} \quad & \max_{G_1 \in \mathcal{D}_1} \mathbb{E}_{G_1}[\boldsymbol{w}^H \boldsymbol{R}_{\mathrm{i+n}} \boldsymbol{w}] \\ \text{subject to} \quad & \min_{G_2 \in \mathcal{D}_2} \mathbb{E}_{G_2}[\boldsymbol{w}^H \boldsymbol{a}\boldsymbol{a}^H \boldsymbol{w}] \geq 1, \end{aligned}}$$

Distributionally Robust Beamforming

$$
\begin{aligned}
\underset{\boldsymbol{w}}{\text{minimize}} \quad & \max_{G_1 \in \mathcal{D}_1} \mathbb{E}_{G_1}[\boldsymbol{w}^H \boldsymbol{R}_{\mathrm{i+n}} \boldsymbol{w}] \\
\text{subject to} \quad & \min_{G_2 \in \mathcal{D}_2} \mathbb{E}_{G_2}[\boldsymbol{w}^H \boldsymbol{a}\boldsymbol{a}^H \boldsymbol{w}] \geq 1,
\end{aligned}
$$

Distributional uncertainty sets $\mathcal{D}_1$ and $\mathcal{D}_2$ account for uncertainty in probability distributions of steering vector and INC matrix.

Two main approaches:

- **Moment-based DRO:**

  first-/second-order moment constraints
- **Wasserstein DRO:**

  uncertainty sets based on probability distribution geometry

Moment-based DRO-based RAB

$$\mathcal{D}_1 = \left\{ G_1 \in \mathcal{M}_1 \;\middle|\; \begin{array}{l} \mathbb{P}_{G_1}(\boldsymbol{R}_{\mathrm{i+n}} \in \mathcal{Z}_1) = 1 \\ \mathbb{E}_{G_1}[\boldsymbol{R}_{\mathrm{i+n}}] \succeq \mathbf{0} \\ \|\mathbb{E}_{G_1}[\boldsymbol{R}_{\mathrm{i+n}}] - \boldsymbol{S}_0\|_F \leq \rho_1 \end{array} \right\},$$

- **Support Constraint**:
  - Ensures $\boldsymbol{R}_{\mathrm{i+n}}$ remains within physically meaningful support set $\mathcal{Z}_1$.
  - Encodes prior knowledge (e.g., limits on interference/noise power).
- **PSD Constraint:**
  - Guarantees positive semidefiniteness, consistent with covariance matrix properties.
- **Similarity Constraint:**
  - Keeps expected INC matrix close to empirical estimate $\boldsymbol{S}_0$.
  - $\boldsymbol{S}_0$ can be approximated by sample data covariance matrix $\widehat{\boldsymbol{R}} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{y}(t)\boldsymbol{y}^H(t)$ or improved reconstruction.
  - ρ_1 acts as a trust level: Small ρ_1: strong reliance on data. Large ρ_1: more flexibility to handle mismodeling.

Moment-based DRO-based RAB

$$\mathcal{D}_2 = \left\{ G_2 \in \mathcal{M}_2 \;\middle|\; \begin{array}{l} \mathbb{P}_{G_2}(\boldsymbol{a} \in \mathcal{Z}_2) = 1 \\ \|\mathbb{E}_{G_2}[\boldsymbol{a}] - \boldsymbol{a}_0\| \leq \gamma_1 \\ \mathbb{E}_{G_2}[\boldsymbol{a}\boldsymbol{a}^H] \preceq (1+\gamma_2)\boldsymbol{\Sigma} + \boldsymbol{a}_0\boldsymbol{a}_0^H \end{array} \right\},$$

- **Support Constraint:**
  - Restricts $\boldsymbol{a}$ to feasible region $\mathcal{Z}_2$ (e.g., angular sector, element magnitude/phase limits).
  - Encodes prior knowledge, excludes unrealistic directions.
- **First-Order Moment Constraint:**
  - Ensures expected steering vector remains close to nominal mean $\boldsymbol{a}_0$
  - Parameter γ_1 controls tolerance: Small γ_1: tight trust in $\boldsymbol{a}_0$. Large γ_1: allows flexibility for mismodeling.
- **Second-Order Moment Constraint:**
  - Bounds variability of $\boldsymbol{a}$.
  - Parameter γ_2 controls robustness against uncertainty in covariance.

The Maximization Problem in the Objective

- The inner maximization problem in the objective of the DRO-based RAB problem is rewritten as (the subscript of $\boldsymbol{R}_{\mathrm{i+n}}$ is dropped for notational simplicity.)

$$
\begin{aligned}
\underset{G_1 \in \mathcal{M}_1}{\text{maximize}} \quad & \int_{\mathcal{Z}_1} \boldsymbol{w}^H \boldsymbol{R} \boldsymbol{w} \, dG_1(\boldsymbol{R}) \\
\text{subject to} \quad & \int_{\mathcal{Z}_1} dG_1(\boldsymbol{R}) = 1, \\
& \int_{\mathcal{Z}_1} \boldsymbol{R} \, dG_1(\boldsymbol{R}) \succeq \mathbf{0}, \\
& \left\| \int_{\mathcal{Z}_1} \boldsymbol{R} \, dG_1(\boldsymbol{R}) - \boldsymbol{S}_0 \right\|_F \leq \rho_1.
\end{aligned}
$$

- The dual problem is cast as

$$
\begin{aligned}
\underset{\boldsymbol{X}, \boldsymbol{Y}}{\text{minimize}} \; & \rho_1 \|\boldsymbol{X}\|_F + \delta_{\mathcal{Z}_1}(\boldsymbol{w}\boldsymbol{w}^H + \boldsymbol{X} + \boldsymbol{Y}) - \operatorname{tr}(\boldsymbol{X}\boldsymbol{S}_0) \\
\text{subject to} \; & \boldsymbol{X} \in \mathcal{H}^N, \boldsymbol{Y} \succeq \mathbf{0} \; (\in \mathcal{H}_+^N),
\end{aligned}
$$

  where $\delta_{\mathcal{Z}_1}(\cdot)$ is defined as the support function of $\mathcal{Z}_1$, and it is convex.
- Clearly, the dual problem is a finite-dimension convex optimization problem.
- It is verified that the strong duality holds between them.

- Assuming prior knowledge of an upper bound on the total interference-plus-noise power, the support set $\mathcal{Z}_1$ is defined as

$$\mathcal{Z}_1 = \{\boldsymbol{R} \in \mathcal{H}^N \mid \text{tr}(\boldsymbol{R}) \leq \rho_2, \boldsymbol{R} \succeq \boldsymbol{0}\}$$

  Then, the corresponding support function is expressed as (note that in this case, the second constraint becomes redundant and can be excluded; therefore, dual variable $\boldsymbol{Y}$ is omitted):

$$\delta_{\mathcal{Z}_1}(\boldsymbol{w}\boldsymbol{w}^H + \boldsymbol{X}) = \rho_2 \lambda_{\max}(\boldsymbol{w}\boldsymbol{w}^H + \boldsymbol{X}).$$

- In the case where the Frobenius norm of the INC matrix is assumed to be upper-bounded, the set $\mathcal{Z}_1$ can be expressed as

$$\mathcal{Z}_1 = \{\boldsymbol{R} \in \mathcal{H}^N \mid \|\boldsymbol{R}\|_F \leq \rho_2, \boldsymbol{R} \succeq \boldsymbol{0}\}$$

  Then, the following identity holds for the support function:

$$\delta_{\mathcal{Z}_1}(\boldsymbol{w}\boldsymbol{w}^H + \boldsymbol{X}) = \rho_2 \left\| \boldsymbol{w}\boldsymbol{w}^H + \boldsymbol{X} \right\|_F.$$

- Consequently, the dual problem is obtained by replacing the support function.

The Minimization Problem in the Constraint

- The minimization problem in the constraint of the DRO-based RAB problem can be expressed as

$$
\begin{aligned}
\underset{G_2 \in \mathcal{M}_2}{\text{minimize}} \quad & \int_{\mathcal{Z}_2} \boldsymbol{a}^H \boldsymbol{w}\boldsymbol{w}^H \boldsymbol{a}\, dG_2(\boldsymbol{a}) \\
\text{subject to} \quad & \int_{\mathcal{Z}_2} dG_2(\boldsymbol{a}) = 1, \\
& \left\| \int_{\mathcal{Z}_2} \boldsymbol{a}\, dG_2(\boldsymbol{a}) - \boldsymbol{a}_0 \right\| \le \gamma_1, \\
& \int_{\mathcal{Z}_2} \boldsymbol{a}\boldsymbol{a}^H\, dG_2(\boldsymbol{a}) \preceq (1+\gamma_2)\boldsymbol{\Sigma} + \boldsymbol{a}_0\boldsymbol{a}_0^H.
\end{aligned}
$$

- The dual problem can be derived as follows

$$
\begin{aligned}
\underset{x,\boldsymbol{x},\boldsymbol{Z}}{\text{maximize}} \quad & x + \Re(\boldsymbol{a}_0^H \boldsymbol{x}) - \gamma_1 \|\boldsymbol{x}\| - \text{tr}(\boldsymbol{Z}((1+\gamma_2)\boldsymbol{\Sigma} + \boldsymbol{a}_0\boldsymbol{a}_0^H)) \\
\text{subject to} \quad & \boldsymbol{a}^H(\boldsymbol{w}\boldsymbol{w}^H + \boldsymbol{Z})\boldsymbol{a} - \Re(\boldsymbol{a}^H \boldsymbol{x}) - x \ge 0, \ \forall \boldsymbol{a} \in \mathcal{Z}_2, \\
& \boldsymbol{Z} \succeq \boldsymbol{0}, \boldsymbol{x} \in \mathbb{C}^N, x \in \mathbb{R}.
\end{aligned}
$$

- The strong duality between the two problems holds.

- Suppose that $\mathcal{Z}_2 = \mathbb{C}^N$. The semi-infinite inequality constraint in the dual is equivalent to the following quadratic matrix inequality (QMI):

$$\begin{bmatrix} \boldsymbol{w}\boldsymbol{w}^H + \boldsymbol{Z} & -\frac{\boldsymbol{x}}{2} \\ -\frac{\boldsymbol{x}^H}{2} & -x \end{bmatrix} \succeq \mathbf{0}.$$

- Additionally, suppose that $\mathcal{Z}_2$ is defined by at most two quadratic constraints. Specifically,

$$\boldsymbol{a}^H \boldsymbol{Q}_i \boldsymbol{a} + 2\Re(\boldsymbol{q}_i^H \boldsymbol{a}) + q_i \leq 0, \quad i = 1, \ldots, k,$$

where $k \in \{1, 2\}$. According to the S-procedure, if there exists a complex vector $\overline{\boldsymbol{a}}$ in the interior of the set $\mathcal{Z}_2$, the semi-infinite inequality constraint can be reformulated as the following QMI:

$$\begin{bmatrix} \boldsymbol{w}\boldsymbol{w}^H + \boldsymbol{Z} & -\frac{\boldsymbol{x}}{2} \\ -\frac{\boldsymbol{x}^H}{2} & -x \end{bmatrix} \succeq \sum_{i=1}^{k} \tau_i \begin{bmatrix} \boldsymbol{Q}_i & \boldsymbol{q}_i \\ \boldsymbol{q}_i^H & q_i \end{bmatrix}, \; \tau_i \leq 0.$$

- In the ideal scenario where there is no uncertainty in the steering vector $\boldsymbol{a}$, its norm satisfies $\|\boldsymbol{a}\|^2 = N$.
- Consequently, a reasonable choice for the support set $\mathcal{Z}_2$ is:

$$\mathcal{Z}_2 = \{\boldsymbol{a} \in \mathbb{C}^N \mid (1-\Delta)N \leq \|\boldsymbol{a}\|^2 \leq (1+\Delta)N\},$$

  where $0 < \Delta < 1$ quantifies the allowable deviation from the ideal norm.
- Therefore, the QMI constraint can be rewritten as:

$$\begin{bmatrix} \boldsymbol{w}\boldsymbol{w}^H + \boldsymbol{Z} & -\frac{\boldsymbol{x}}{2} \\ -\frac{\boldsymbol{x}^H}{2} & -x \end{bmatrix} \succeq \tau_1 \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & -(1+\Delta)N \end{bmatrix} + \tau_2 \begin{bmatrix} -\boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & (1-\Delta)N \end{bmatrix},$$

  $\tau_1, \tau_2 \leq 0.$

Equivalent QMI Reformulation

- The original DRO-based RAB problem can be transformed into

$$
\begin{aligned}
\text{minimize} \quad & \rho_1 \|\boldsymbol{X}\|_F + \rho_2 \left\| \boldsymbol{w}\boldsymbol{w}^H + \boldsymbol{X} \right\|_F - \text{tr}(\boldsymbol{X}\boldsymbol{S}_0) \\
\text{subject to} \quad & x + \Re(\boldsymbol{a}_0^H \boldsymbol{x}) - \gamma_1 \|\boldsymbol{x}\| - \text{tr}(\boldsymbol{Z}((1+\gamma_2)\boldsymbol{\Sigma} + \boldsymbol{a}_0 \boldsymbol{a}_0^H)) \geq 1, \\
& \begin{bmatrix} \boldsymbol{w}\boldsymbol{w}^H + \boldsymbol{Z} & -\frac{\boldsymbol{x}}{2} \\ -\frac{\boldsymbol{x}^H}{2} & -x \end{bmatrix} \succeq \\
& \tau_1 \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & -(1+\Delta)N \end{bmatrix} + \tau_2 \begin{bmatrix} -\boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & (1-\Delta)N \end{bmatrix}, \\
& \tau_1, \tau_2 \leq 0, \\
& \boldsymbol{w}, \boldsymbol{x} \in \mathbb{C}^N, \ \boldsymbol{X} \in \mathcal{H}^N, \ \boldsymbol{Z} \succeq \boldsymbol{0}, \ x, \tau_1, \tau_2 \in \mathbb{R}.
\end{aligned}
$$

- This is a nonconvex QMI problem (w.r.t. $\boldsymbol{w}$).

- The conventional LMI relaxation technique can be applied, namely, the following LMI problem is solved:

$$
\begin{aligned}
\text{minimize} \quad & \rho_1 \|\boldsymbol{X}\|_F + \rho_2 \|\boldsymbol{W} + \boldsymbol{X}\|_F - \operatorname{tr}(\boldsymbol{X}\boldsymbol{S}_0) \\
\text{subject to} \quad & x + \Re(\boldsymbol{a}_0^H \boldsymbol{x}) - \gamma_1 \|\boldsymbol{x}\| - \operatorname{tr}(\boldsymbol{Z}((1+\gamma_2)\boldsymbol{\Sigma} + \boldsymbol{a}_0 \boldsymbol{a}_0^H)) \geq 1, \\
& \begin{bmatrix} \boldsymbol{W} + \boldsymbol{Z} & -\frac{\boldsymbol{x}}{2} \\ -\frac{\boldsymbol{x}^H}{2} & -x \end{bmatrix} \succeq \tau_1 \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & -(1+\Delta)N \end{bmatrix} + \tau_2 \begin{bmatrix} -\boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & (1-\Delta)N \end{bmatrix}, \\
& \tau_1, \tau_2 \leq 0, \\
& \boldsymbol{x} \in \mathbb{C}^N, \ \boldsymbol{X} \in \mathcal{H}^N, \ \boldsymbol{W}, \boldsymbol{Z} \succeq \boldsymbol{0}, \ x, \tau_1, \tau_2 \in \mathbb{R}.
\end{aligned}
$$

- If the LMI problem has a rank-one optimal solution $\boldsymbol{W}^\star = \boldsymbol{w}^\star \boldsymbol{w}^{\star H}$, then $\boldsymbol{w}^\star$ is optimal for the QMI problem.
- A rank-one solution procedure is desired when the LMI relaxation problem admits an optimal solution $\boldsymbol{W}^\star$ of rank more than one.

Rank-One Solution Procedure for the LMI Relaxation

- Observe that if nonzero $\boldsymbol{W} \succeq \mathbf{0}$ is of rank one, then $\text{tr}(\boldsymbol{W}) = \|\boldsymbol{W}\|_F$, and vice versa.
- The previous condition can also take the form: $\text{tr}(\boldsymbol{W}) - \frac{\text{tr}(\boldsymbol{W}\boldsymbol{W})}{\|\boldsymbol{W}\|_F} = 0$.
- Therefore, at iteration k of a procedure, the following LMI problem with a penalty term on the rank-one solution constraint is solved:

$$
\begin{aligned}
\text{minimize} \quad & \rho_1 \|\boldsymbol{X}\|_F + \rho_2 \|\boldsymbol{W} + \boldsymbol{X}\|_F - \text{tr}(\boldsymbol{X}\boldsymbol{S}_0) + \alpha \left(\text{tr}(\boldsymbol{W}) - \frac{\text{tr}(\boldsymbol{W}\boldsymbol{W}_k)}{\|\boldsymbol{W}_k\|_F} \right) \\
\text{subject to} \quad & x + \Re(\boldsymbol{a}_0^H \boldsymbol{x}) - \gamma_1 \|\boldsymbol{x}\| - \text{tr}(\boldsymbol{Z}((1+\gamma_2)\boldsymbol{\Sigma} + \boldsymbol{a}_0 \boldsymbol{a}_0^H)) \geq 1, \\
& \begin{bmatrix} \boldsymbol{W} + \boldsymbol{Z} & -\frac{\boldsymbol{x}}{2} \\ -\frac{\boldsymbol{x}^H}{2} & -x \end{bmatrix} \succeq \tau_1 \begin{bmatrix} \boldsymbol{I} & \mathbf{0} \\ \mathbf{0} & -(1+\Delta)N \end{bmatrix} + \tau_2 \begin{bmatrix} -\boldsymbol{I} & \mathbf{0} \\ \mathbf{0} & (1-\Delta)N \end{bmatrix}, \\
& \tau_1, \tau_2 \leq 0, \\
& \boldsymbol{x} \in \mathbb{C}^N,\ \boldsymbol{X} \in \mathcal{H}^N,\ \boldsymbol{W}, \boldsymbol{Z} \succeq \mathbf{0},\ x, \tau_1, \tau_2 \in \mathbb{R}.
\end{aligned}
$$

Finding a Rank-One Solution for the LMI Relaxation

Algorithm 1 Finding a Rank-One Solution for LMI Relaxation Problem

Input: $\boldsymbol{S}_0, \boldsymbol{\Sigma}, \boldsymbol{a}_0, \gamma_1, \gamma_2, \rho_1, \rho_2, \alpha, \eta$;

Output: A rank-one solution $\boldsymbol{w}^\star \boldsymbol{w}^{\star H}$ for problem;

1: Solve LMI problem, returning $\boldsymbol{W}^\star$
2: `if` $\boldsymbol{W}^\star$ is of rank one `then`
3: Output $\boldsymbol{w}^\star$ with $\boldsymbol{W}^\star = \boldsymbol{w}^\star \boldsymbol{w}^{\star H}$, and terminate;
4: `end`
5: $k \leftarrow 0$; Let $\boldsymbol{W}_k$ be the optimal (high-rank) solution $\boldsymbol{W}^\star$ for LMI problem;
6: `do`
7: Solve the LMI problem with penalty, obtaining solution $\boldsymbol{W}_{k+1}$;
8: $k \leftarrow k+1$;
9: **until** $\text{tr}(\boldsymbol{W}_k) - \frac{\text{tr}(\boldsymbol{W}_k \boldsymbol{W}_{k-1})}{\|\boldsymbol{W}_{k-1}\|_F} \leq \eta$
10: Output $\boldsymbol{w}^\star$ with $\boldsymbol{W}_k = \boldsymbol{w}^\star \boldsymbol{w}^{\star H}$.

Finding a Rank-One Solution for the LMI Relaxation

- It can be shown that the sequence of the optimal values is descent, and the penalty term converges to zero.
- The terminating condition implies that the output solution $\boldsymbol{W}_k$ is a rank-one solution, since $\mathrm{tr}(\boldsymbol{W}_k) \approx \frac{\mathrm{tr}(\boldsymbol{W}_k \boldsymbol{W}_{k-1})}{\|\boldsymbol{W}_{k-1}\|_F} \approx \|\boldsymbol{W}_k\|_F$
- The computational complexity is dominated by solving the LMI problem (12) in each iteration, which is manageable since the problem has only one inequality constraint and one LMI constraint with size N + 1.

Alternative Uncertainty Sets for Steering Vector

- $\mathcal{D}_2' = \left\{ G_2 \in \mathcal{M}_2 \;\middle|\; \begin{array}{l} \mathbb{P}_{G_2}(\boldsymbol{a} \in \mathcal{Z}_2) = 1 \\ (\mathbb{E}_{G_2}[\boldsymbol{a}] - \bar{\boldsymbol{a}})^H \boldsymbol{Q}^{-1} (\mathbb{E}_{G_2}[\boldsymbol{a}] - \bar{\boldsymbol{a}}) \leq \gamma_1 \\ \|\mathbb{E}_{G_2}[(\boldsymbol{a} - \bar{\boldsymbol{a}})(\boldsymbol{a} - \bar{\boldsymbol{a}})^H] - \bar{\boldsymbol{\Sigma}}\|_F \leq \gamma_2 \end{array} \right\},$
  - the second constraint ensures that the mean of the random vector $\boldsymbol{a}$ resides within an ellipsoid of size γ_1, centered at the empirical mean $\bar{\boldsymbol{a}}$, with the ellipsoid's shape defined by the positive definite matrix $\boldsymbol{Q} \succ \boldsymbol{0}$.
  - The third constraint in forces that $\mathbb{E}_{G_2}[(\boldsymbol{a} - \bar{\boldsymbol{a}})(\boldsymbol{a} - \bar{\boldsymbol{a}})^H]$, which is called the centered second-order moment matrix of $\boldsymbol{a}$ and is different from the covariance matrix unless $\mathbb{E}_{G_2}[\boldsymbol{a}] = \bar{\boldsymbol{a}}$ (e.g., when $\gamma_1 = 0$), is located in a ball of radius γ_2, centered at the empirical covariance matrix $\bar{\boldsymbol{\Sigma}}$.
  - Accordingly, the DRO-based RAB problem can be recast into

$$\begin{array}{ll} \text{minimize} & \rho_1 \|\boldsymbol{X}\|_F + \rho_2 \left\| \boldsymbol{w}\boldsymbol{w}^H + \boldsymbol{X} \right\|_F - \text{tr}(\boldsymbol{X}\boldsymbol{S}_0) \\ \text{subject to} & -x + 2\Re(\bar{\boldsymbol{a}}^H \boldsymbol{x}) - \text{tr}(\boldsymbol{Z}(\bar{\boldsymbol{\Sigma}} - \bar{\boldsymbol{a}}\bar{\boldsymbol{a}}^H)) \geq 1 + 2\sqrt{\gamma_1} \|\boldsymbol{Q}^{\frac{1}{2}} \boldsymbol{x}\| + \gamma_2 \|\boldsymbol{Z}\|_F, \\ & \begin{bmatrix} \boldsymbol{w}\boldsymbol{w}^H + \boldsymbol{Z} & -\boldsymbol{Z}\bar{\boldsymbol{a}} - \boldsymbol{x} \\ -\bar{\boldsymbol{a}}^H \boldsymbol{Z} - \boldsymbol{x}^H & x \end{bmatrix} \succeq \tau_1 \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & -(1+\Delta)N \end{bmatrix} + \tau_2 \begin{bmatrix} -\boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & (1-\Delta)N \end{bmatrix}, \\ & \tau_1, \tau_2 \leq 0. \\ & \boldsymbol{w}, \boldsymbol{x} \in \mathbb{C}^N, \; \boldsymbol{X}, \boldsymbol{Z} \in \mathcal{H}^N, \; x \in \mathbb{R}. \end{array}$$

Alternative Uncertainty Sets for Steering Vector

- $\mathcal{D}_2'' = \left\{ G_2 \in \mathcal{M}_2 \;\middle|\; \begin{array}{l} \mathbb{P}_{G_2}(\boldsymbol{a} \in \mathcal{Z}_2) = 1 \\ (\mathbb{E}_{G_2}[\boldsymbol{a}] - \bar{\boldsymbol{a}})^H \boldsymbol{Q}^{-1}(\mathbb{E}_{G_2}[\boldsymbol{a}] - \bar{\boldsymbol{a}}) \leq \gamma_1 \\ \mathbb{E}_{G_2}[(\boldsymbol{a} - \bar{\boldsymbol{a}})(\boldsymbol{a} - \bar{\boldsymbol{a}})^H] \preceq (1+\gamma_2)\bar{\boldsymbol{\Sigma}} \end{array} \right\},$
  - The third constraint allows us to control how far the realization might be from $\bar{\boldsymbol{a}}$ on average.
  - Accordingly, the DRO-based RAB problem can be recast into

$$
\begin{array}{ll}
\text{minimize} & \rho_1 \|\boldsymbol{X}\|_F + \rho_2 \|\boldsymbol{w}\boldsymbol{w}^H + \boldsymbol{X}\|_F - \text{tr}(\boldsymbol{X}\boldsymbol{S}_0) \\
\text{subject to} & -x + 2\Re(\bar{\boldsymbol{a}}^H \boldsymbol{x}) - \text{tr}(\boldsymbol{Z}(\bar{\boldsymbol{\Sigma}} - \bar{\boldsymbol{a}}\bar{\boldsymbol{a}}^H)) \geq 1 + 2\sqrt{\gamma_1}\|\boldsymbol{Q}^{\frac{1}{2}}\boldsymbol{x}\| + \gamma_2 \text{tr}(\bar{\boldsymbol{\Sigma}}\boldsymbol{Z}), \\
& \left[\begin{array}{cc} \boldsymbol{w}\boldsymbol{w}^H + \boldsymbol{Z} & -\boldsymbol{Z}\bar{\boldsymbol{a}} - \boldsymbol{x} \\ -\bar{\boldsymbol{a}}^H \boldsymbol{Z} - \boldsymbol{x}^H & x \end{array} \right] \succeq \\
& \tau_1 \left[\begin{array}{cc} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & -(1+\Delta)N \end{array} \right] + \tau_2 \left[\begin{array}{cc} -\boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & (1-\Delta)N \end{array} \right], \\
& \tau_1, \tau_2 \leq 0. \\
& \boldsymbol{w}, \boldsymbol{x} \in \mathbb{C}^N,\ \boldsymbol{X} \in \mathcal{H}^N,\ \boldsymbol{Z} \succeq \boldsymbol{0},\ x \in \mathbb{R}.
\end{array}
$$

Alternative Uncertainty Set for INC Matrix

- $\mathcal{D}_1' = \left\{ G_1 \in \mathcal{M}_1 \;\middle|\; \begin{array}{l} \mathbb{P}_{G_1}(\boldsymbol{R}_{\mathrm{i+n}} \in \mathcal{Z}_1) = 1 \\ \mathbb{E}_{G_1}[\boldsymbol{R}_{\mathrm{i+n}}] \succeq (1-\rho_1)(\boldsymbol{S}_0 + \epsilon \boldsymbol{I}) \\ \mathbb{E}_{G_1}[\boldsymbol{R}_{\mathrm{i+n}}] \preceq (1+\rho_1)(\boldsymbol{S}_0 + \epsilon \boldsymbol{I}) \end{array} \right\}$,
  - the second and third constraints ensure that the expected value of the INC matrix remains in close proximity to a diagonally loaded version of its empirical mean, with the allowable deviation regulated by ρ_1.
  - It enhances robustness against the influence of the desired signal present in the training data.
  - Accordingly, the DRO-based RAB problem can be recast into

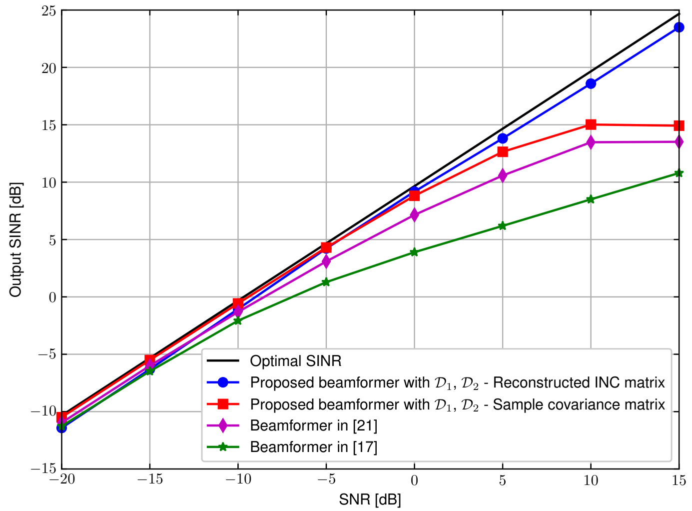
$$
\begin{aligned}
\text{minimize} \quad & \rho_2 \|\boldsymbol{w}\boldsymbol{w}^H + \boldsymbol{X} - \boldsymbol{X}'\|_F + \mathrm{tr}(((1+\rho_1)\boldsymbol{X}' - (1-\rho_1)\boldsymbol{X})(\boldsymbol{S}_0 + \epsilon \boldsymbol{I})) \\
\text{subject to} \quad & x + \Re(\boldsymbol{a}_0^H \boldsymbol{x}) - \gamma_1 \|\boldsymbol{x}\| - \mathrm{tr}(\boldsymbol{Z}((1+\gamma_2)\boldsymbol{\Sigma} + \boldsymbol{a}_0 \boldsymbol{a}_0^H)) \geq 1, \\
& \begin{bmatrix} \boldsymbol{w}\boldsymbol{w}^H + \boldsymbol{Z} & -\frac{\boldsymbol{x}}{2} \\ -\frac{\boldsymbol{x}^H}{2} & -x \end{bmatrix} \succeq \tau_1 \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & -(1+\Delta)N \end{bmatrix} + \tau_2 \begin{bmatrix} -\boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & (1-\Delta)N \end{bmatrix}, \\
& \tau_1, \tau_2 \leq 0, \\
& \boldsymbol{w}, \boldsymbol{x} \in \mathbb{C}^N, \;, \; \boldsymbol{X}, \boldsymbol{X}', \boldsymbol{Z} \succeq \boldsymbol{0}, \; x, \tau_1, \tau_2 \in \mathbb{R}.
\end{aligned}
$$

Simulation Setup

- Scenario: A uniform linear array with N = 10 sensors spaced half a wavelength;
  - The angular sector of interest $\Theta = [0^\circ, 10^\circ]$;
  - The actual signal direction $\theta = 5^\circ$;
  - The presumed direction $\theta_0 = 1^\circ$;
  - Two interferers located in the directions of $\theta_1 = -5^\circ$ and $\theta_2 = 15^\circ$ with the same interference-to-noise ratio (INR) of 30 dB;
  - The array noise: a spatially and temporally white Gaussian vector with zero mean and covariance $\boldsymbol{I}$;
  - Wavefront distortion: The phase increments are independent Gaussian variables each with zero mean and standard deviation 0.02;
  - $\boldsymbol{S}_0$ is the sampling covariance matrix, unless otherwise stated;
  - $\boldsymbol{a}_0 = \frac{1}{L}\sum_{l=1}^{L} \boldsymbol{d}(\theta_l)$ and $\Sigma = \frac{1}{L}\sum_{l=1}^{L}(\boldsymbol{d}(\theta_l) - \boldsymbol{a}_0)(\boldsymbol{d}(\theta_l) - \boldsymbol{a}_0)^H$;
  - Parameters $\rho_1 = 0.001\|\boldsymbol{S}_0\|_F$, $\rho_2 = 1.1\text{tr}(\boldsymbol{S}_0)$, and $\alpha = 10^3$.
- All results are averaged over 200 simulation runs.

Simulation Result: Average Array Output SINR versus SNR

- Number of snapshots $T = 100$.
- Proposed DRO-based beamformer with $\mathcal{D}_1$ and $\mathcal{D}_2$ outperforms [Li'18]*, [Zhang'15]†.
- Performance drops at high SNR due to poor INC matrix approximation.
- Reconstructed INC matrix ([Gu'12]‡) improves robustness.



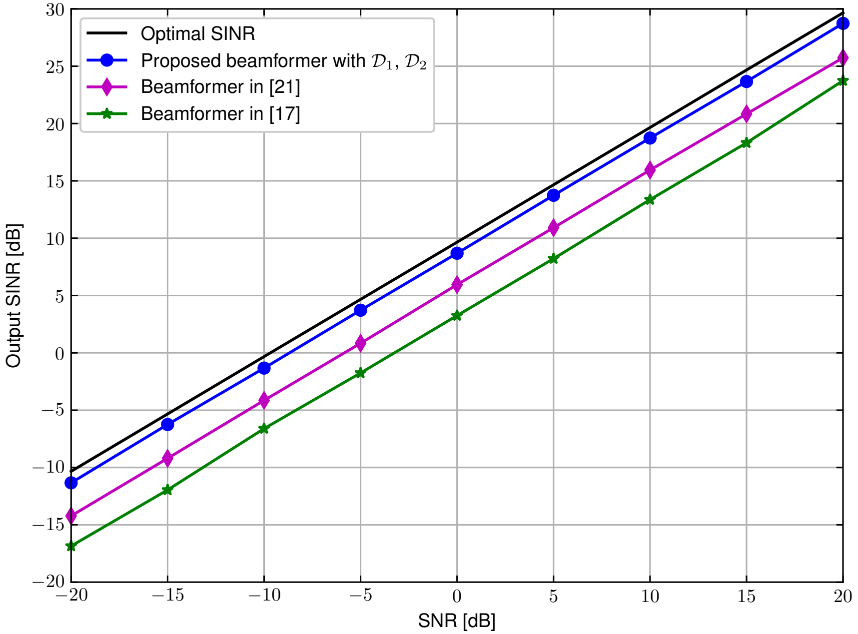
* B. Li, Y. Rong, J. Sun and K. L. Teo, "A distributionally robust minimum variance beamformer design," *IEEE Signal Processing Letters*, vol. 25, no. 1, pp. 105–109, Jan. 2018.

† X. Zhang, Y. Li, N. Ge and J. Lu, "Robust minimum variance beamforming under distributional uncertainty," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Brisbane, Australia, Apr. 2015, pp. 2514–2518.

‡ Y. Gu and A. Leshem, "Robust adaptive beamforming based on interference covariance matrix reconstruction and steering vector estimation," *IEEE Transactions on Signal Processing*, vol. 60, no. 7, pp. 3881–3885, July 2012.

Simulation Result: Average Array Output SINR versus SNR

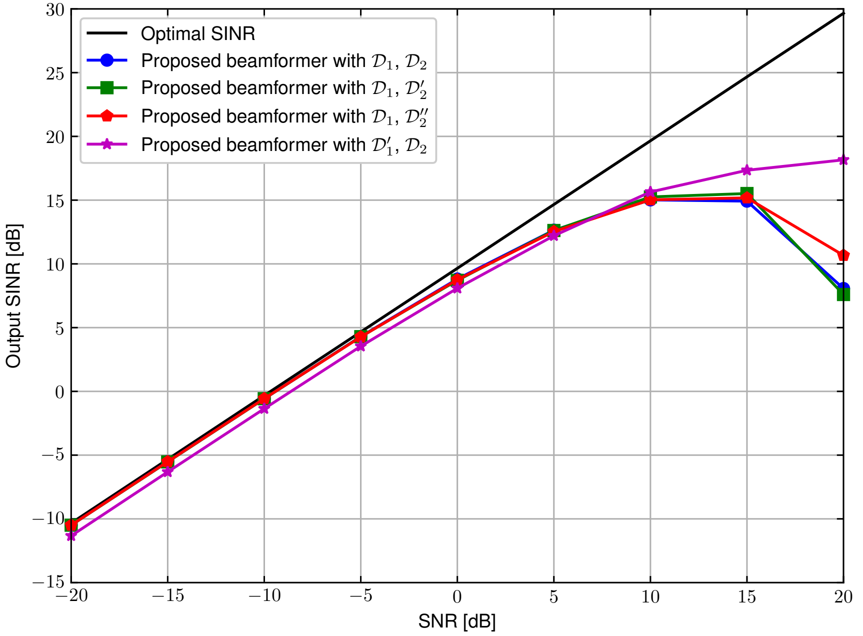
- Number of snapshots $T = 10$.
- Exact knowledge of the actual INC matrix
  - No distortion from desired signal in covariance.
  - Avoids performance degradation at high SNR.
- Proposed DRO-based beamformer
  - Consistently superior SINR across the full SNR range.



Perfect INC knowledge eliminates high-SNR performance loss.

Simulation Result: Effect of Different Uncertainty Sets

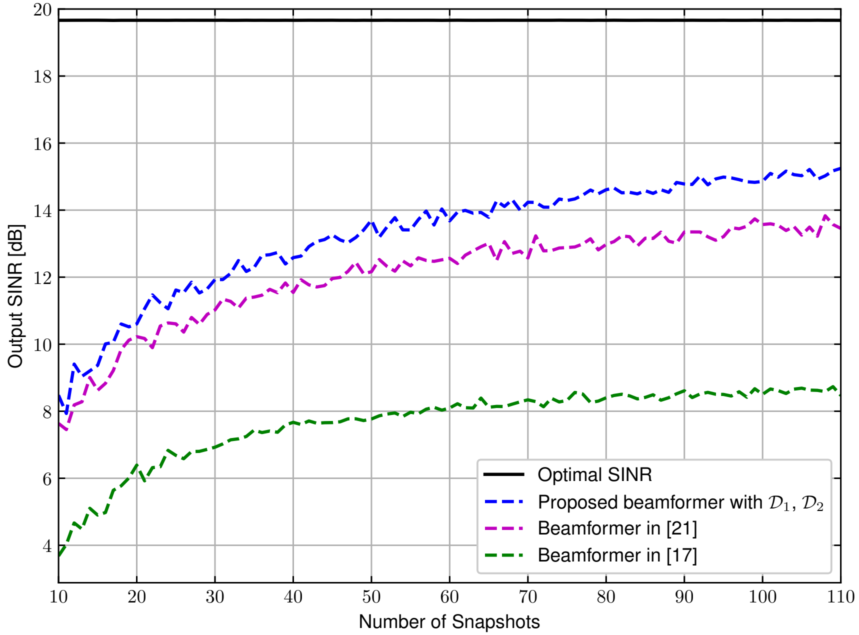
- Average Array Output SINR versus SNR. Number of snapshots $T = 100$.
- High-SNR performance
  - $\mathcal{D}_1'$ improves SINR at high SNR.
  - Incorporates diagonal loading constraints. Keeps expected INC matrix close to a diagonally loaded empirical mean.
  - Effectively mitigates influence of desired signal in training samples.



Smartly designed uncertainty sets enhance robustness against signal contamination.

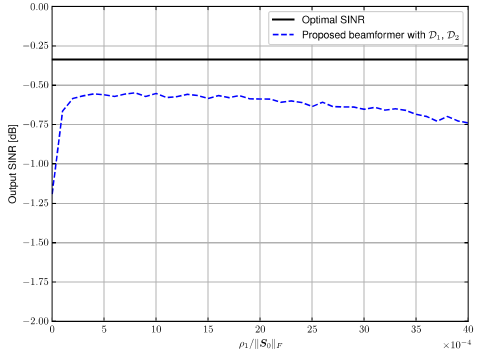
Simulation Result: SINR vs. Number of Snapshots

- Fixed SNR at 10 dB. Varying number of snapshots.
- Proposed beamformer outperforms beamformer in [Zhang'15] and [Li'18] across all snapshot counts.

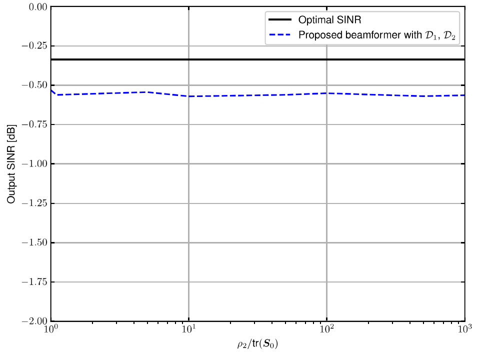


Simulation Result: Parameter Sensitivity

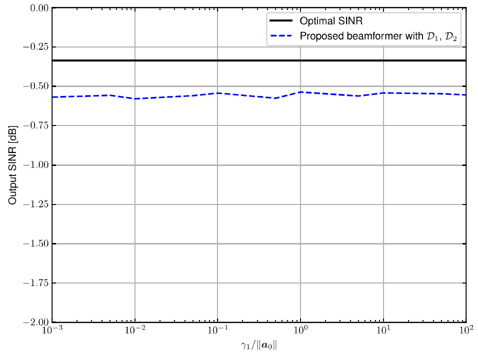
- Beamformer maintains consistent SINR across a wide range of values for all four parameters.
- Demonstrates robustness and parameter insensitivity.
- Exact parameter tuning is not critical.



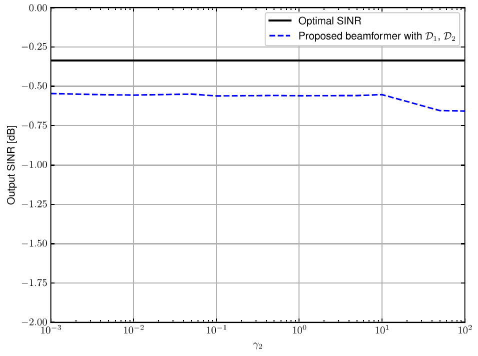
(a)



(b)



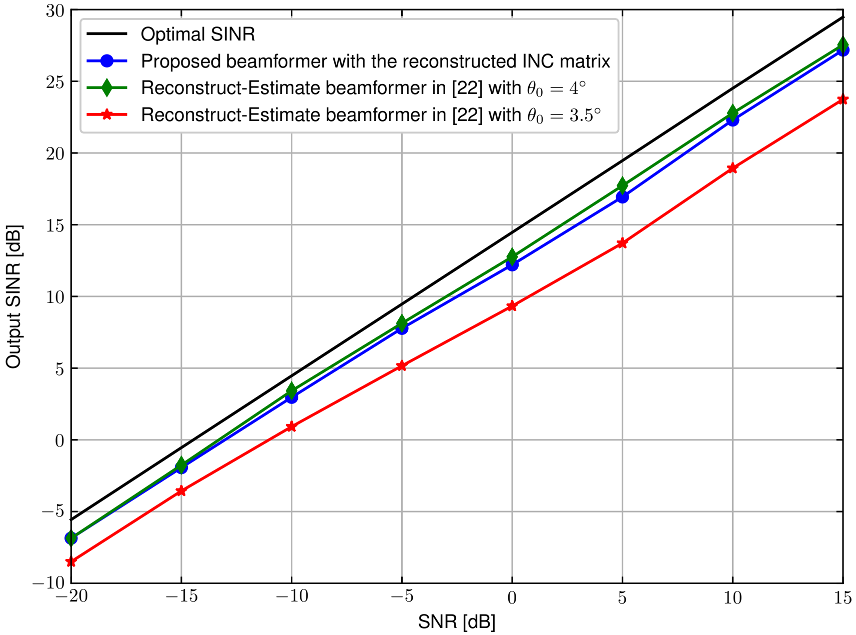
(c)



(d)

Simulation Result: High-Resolution Scenario ($N = 32$ sensors)

- Scenario setup
  - Desired signal direction: 5°, presumed signal direction: $\theta_0 = 4^\circ$, desired sector $[\theta_0 - 2^\circ, \theta_0 + 2^\circ]$. Interferers at 0° and 8°.
  - DRO-based beamformer uses reconstructed INC matrix [Gu'12].
- DRO-based beamformer performs comparable to beamformer in [Gu'12].
- Beamformer in [Gu'12] is sensitive to the presumed direction.
- Proposed DRO-based beamformer does not rely on the presumed direction. ($\boldsymbol{a}_0$ is estimated as the mean of steering vectors generated from random directions within the desired sector)



Summary of Moment-based DRO-based RAB

- We have studied the DRO-based RAB problem of maximizing the worst-case SINR over the distributional sets for random INC matrix and desired signal steering vector.
- The RAB problem is transformed into a nonconvex QMI problem via the strong duality theorem of linear conic programming.
- The QMI problem is tackled by iteratively solving a sequence of LMI relaxation problems with a penalty term on the rank-one constraint.
- The sequence of the optimal values for the LMI relaxation problems is descent, which means that the algorithm always outputs a rank-one solution when the penalty weight is large enough.
- Numerical results show that the proposed beamformer outperforms the other two existing beamformers in terms of the array output SINR.

Wasserstein DRO

- Drawbacks of moment-based DRO
  - ignores higher-order distributional structure
  - Cannot adapt to geometry of actual data $\rightarrow$ robustness is limited
- Wasserstein DRO
  - Uncertainty sets defined via Wasserstein distance from empirical distribution
  - Geometrically interpretable, data-driven formulation
  - Avoids restrictive moment assumptions
  - Define uncertainty set of distributions as a Wasserstein ball:

$$\mathcal{B}_\epsilon(\widehat{P}) = \left\{ P \in \mathcal{M}(\mathcal{X}) \ : \ W_\alpha(P, \widehat{P}) \leq \epsilon \right\}.$$

Wasserstein Ball

$$\mathcal{B}_\epsilon(\widehat{P}) = \left\{ P \in \mathcal{M}(\mathcal{X}) \ : \ W_\alpha(P, \widehat{P}) \leq \epsilon \right\},$$

where W_α = Wasserstein distance, $\widehat{P}$ = empirical distribution.

- Wasserstein metric measures minimal transport cost to transform one distribution into another

$$W_\alpha(P, Q) = \left(\inf_{\gamma \in \Gamma(P,Q)} \int_{\mathcal{X} \times \mathcal{X}} d^\alpha(\boldsymbol{x}, \boldsymbol{y}) \, d\gamma(\boldsymbol{x}, \boldsymbol{y}) \right)^{1/\alpha},$$

- Advantages:
  - Captures full distributional shape, not only moments
  - Naturally data-driven (uncertainty ball around empirical distribution)
  - Radius ϵ controls conservativeness

Wasserstein DRO-based RAB

- Assuming distributional uncertainty for the steering vector, Wasserstein DRO-based RAB problem in real domain is represented by

$$\min_{\boldsymbol{w}_{\mathrm{r}}} \ \boldsymbol{w}_{\mathrm{r}}^T \hat{\boldsymbol{R}}_{\mathrm{r}} \boldsymbol{w}_{\mathrm{r}} \quad \text{s.t.} \quad \inf_{P \in \mathcal{B}_\epsilon(\widehat{P})} \ \mathbb{E}_{\boldsymbol{a}_{\mathrm{r}} \sim P}[\boldsymbol{w}_{\mathrm{r}}^T \boldsymbol{a}_{\mathrm{r}}] \geq 1,$$

- Using the strong duality property in Wasserstein DRO, the dual formulation corresponding to the constraint can be expressed as

$$\min_{\lambda \geq 0} \left\{ \lambda \epsilon^\alpha + \mathbb{E}_{\hat{\boldsymbol{a}}_{\mathrm{r}} \sim \widehat{P}_M} \left[\sup_{\boldsymbol{a}_{\mathrm{r}} \in \mathbb{R}^{2N}} \left\{ -\boldsymbol{w}_{\mathrm{r}}^T \boldsymbol{a}_{\mathrm{r}} - \lambda d^\alpha (\hat{\boldsymbol{a}}_{\mathrm{r}}, \boldsymbol{a}_{\mathrm{r}}) \right\} \right] \right\}.$$

- Let $\alpha = 1$ and the metric d be the norm distance function. The dual formulation can be simplified as

$$\min_{\lambda \geq \|\boldsymbol{w}_{\mathrm{r}}\|_*} \left\{ \lambda \epsilon + \mathbb{E}_{\hat{\boldsymbol{a}}_{\mathrm{r}} \sim \widehat{P}_M} \left[-\boldsymbol{w}_{\mathrm{r}}^T \hat{\boldsymbol{a}}_{\mathrm{r}} \right] \right\} = \epsilon \left\| \boldsymbol{w}_{\mathrm{r}} \right\|_* - \boldsymbol{w}_{\mathrm{r}}^T \bar{\boldsymbol{a}}_{\mathrm{r}},$$

Main Results (Norm-based Cost)

- If Wasserstein distance uses $\ell 2$ **norm** as transport cost $\rightarrow$
  - DRO formulation reduces to second-order cone program (SOCP)
  - Equivalent to worst-case norm-bounded RAB

$$\min_{\boldsymbol{w}_{\mathrm{r}}} \ \boldsymbol{w}_{\mathrm{r}}^T \hat{\boldsymbol{R}}_{\mathrm{r}} \boldsymbol{w}_{\mathrm{r}} \quad \text{s.t.} \ \ \epsilon \left\| \boldsymbol{w}_{\mathrm{r}} \right\|_2 \leq \boldsymbol{w}_{\mathrm{r}}^T \bar{\boldsymbol{a}}_{\mathrm{r}} - 1.$$

- Interpretation:
  - Wasserstein radius $\equiv$ bound on possible steering vector error norm
  - Provides geometric interpretation of norm-constrained uncertainty sets

Established worst-case norm-bounded RAB solution justified from DRO.

Main Results (Quadratic Cost)

- If Wasserstein metric uses quadratic cost (Mahalanobis-like metric) $\rightarrow$

$$d(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{2}(\boldsymbol{x} - \boldsymbol{y})^T \boldsymbol{\Lambda}(\boldsymbol{x} - \boldsymbol{y}),$$

  - DRO formulation recovers ellipsoidal uncertainty sets
  - Equivalent to ellipsoidal RAB in prior literature

$$\min_{\boldsymbol{w}_\mathrm{r}} \ \boldsymbol{w}_\mathrm{r}^T \hat{\boldsymbol{R}}_\mathrm{r} \boldsymbol{w}_\mathrm{r} \quad \text{s.t.} \quad \sqrt{2\epsilon} \left\| \boldsymbol{\Lambda}^{\frac{-1}{2}} \boldsymbol{w}_\mathrm{r} \right\|_2 \le \boldsymbol{w}_\mathrm{r}^T \bar{\boldsymbol{a}}_\mathrm{r} - 1,$$

  - Specifically, the ellipsoidal uncertainty set takes the form

$$\mathcal{E}(\bar{\boldsymbol{a}}_\mathrm{r}, 2\epsilon \boldsymbol{\Lambda}^{-1}) = \left\{ \boldsymbol{a}_\mathrm{r} \,\middle|\, (\boldsymbol{a}_\mathrm{r} - \bar{\boldsymbol{a}}_\mathrm{r})^T \boldsymbol{\Lambda} (\boldsymbol{a}_\mathrm{r} - \bar{\boldsymbol{a}}_\mathrm{r}) \le 2\epsilon \right\},$$

  - Ellipsoidal sets $\leftrightarrow$ Chance-constrained formulations (under Gaussianity)

This observation underscores a fundamental connection between the metric employed in the Wasserstein DRO-based RAB formulation and the structure of the uncertainty set in deterministic robust models.

Wasserstein DRO for INC Matrix Uncertainty

- The Wasserstein DRO-based formulation corresponding to the objective is given by

$$\min_{\boldsymbol{w}_\mathrm{r}} \sup_{Q \in \mathcal{B}_\rho(\widehat{Q})} \mathbb{E}_{\boldsymbol{R}_\mathrm{r} \sim Q}[\boldsymbol{w}_\mathrm{r}^T \boldsymbol{R}_\mathrm{r} \boldsymbol{w}_\mathrm{r}],$$

- we adopt the Frobenius norm as the underlying metric, leading to the following dual formulation for the inner supremum:

$$\min_{\lambda \geq 0} \left\{ \lambda\rho + \mathbb{E}_{\check{\boldsymbol{R}}_\mathrm{r} \sim \widehat{Q}} \left[\sup_{\boldsymbol{R}_\mathrm{r}} \left\{ \boldsymbol{w}_\mathrm{r}^T \boldsymbol{R}_\mathrm{r} \boldsymbol{w}_\mathrm{r} - \lambda \|\boldsymbol{R}_\mathrm{r} - \check{\boldsymbol{R}}_\mathrm{r}\|_F \right\} \right] \right\}.$$

- The dual problem simplifies to

$$\min_{\lambda \geq \|\boldsymbol{w}_\mathrm{r}\|_2^2} \left\{ \lambda\rho + \mathbb{E}_{\check{\boldsymbol{R}}_\mathrm{r} \sim \widehat{Q}} \left[\boldsymbol{w}_\mathrm{r}^T \check{\boldsymbol{R}}_\mathrm{r} \boldsymbol{w}_\mathrm{r} \right] \right\} = \boldsymbol{w}_\mathrm{r}^T \hat{\boldsymbol{R}}_\mathrm{r} \boldsymbol{w}_\mathrm{r} + \rho \|\boldsymbol{w}_\mathrm{r}\|_2^2.$$

- Therefore, the optimization problem is equivalent to

$$\min_{\boldsymbol{w}_\mathrm{r}} \left\{ \boldsymbol{w}_\mathrm{r}^T \hat{\boldsymbol{R}}_\mathrm{r} \boldsymbol{w}_\mathrm{r} + \rho \|\boldsymbol{w}_\mathrm{r}\|_2^2 = \boldsymbol{w}_\mathrm{r}^T (\hat{\boldsymbol{R}}_\mathrm{r} + \rho \boldsymbol{I}) \boldsymbol{w}_\mathrm{r} \right\}.$$

This result suggests that introducing the Wasserstein uncertainty in the INC matrix induces a regularization effect, effectively approximating the INC matrix as a diagonally loaded version of the sample data covariance matrix.

Summary of Wasserstein DRO-based RAB

- Introduced a novel Wasserstein DRO-based RAB formulation ensuring robustness against distributional uncertainties in beamforming
- Unlike conventional DRO-RAB methods (moment-based), our approach uses the 1-Wasserstein metric to construct uncertainty sets
- Provides a comprehensive representation of uncertainty beyond first- and second-order moments
- Demonstrated that the choice of Wasserstein cost metric fundamentally shapes the resulting RAB formulation
- Revealed a deep connection between distributional and deterministic robustness
- Established Wasserstein DRO as a unifying framework, where deterministic robust formulations emerge as special cases under appropriate cost functions

Part VI: Learning to Optimize (L2O) Beamforming and Other Applications

Classic Optimizer *vs* **Learned Optimizer**

- Classic Optimizer
  - manually designed
  - few or no tuning parameters

- Learned Optimizer
  - train over a set of similar optimizees
  - solve unseen optimizees from the same distribution

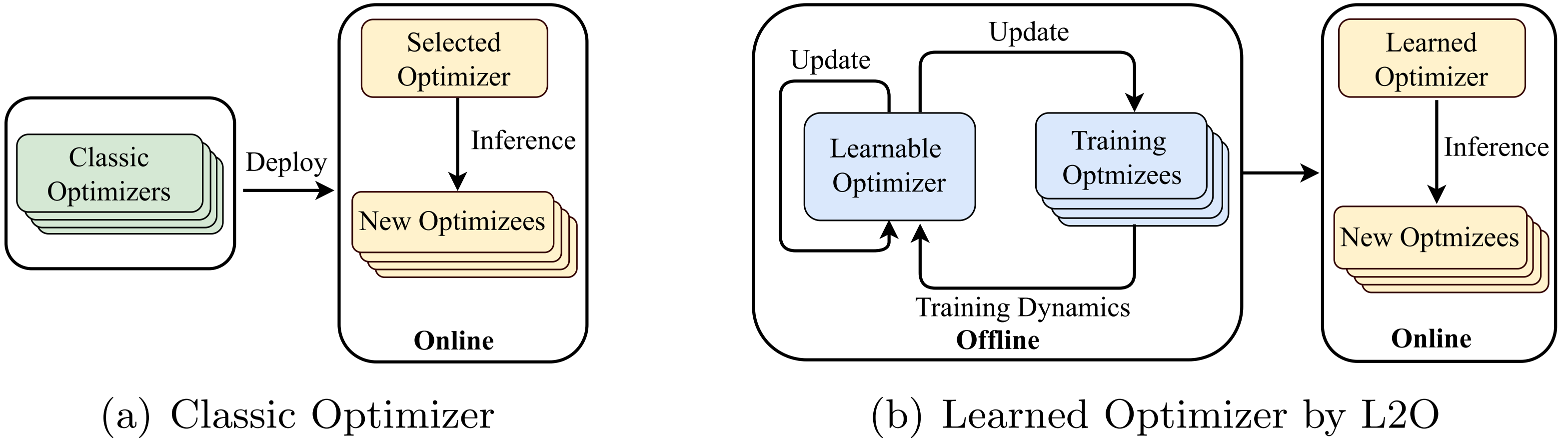


Fig. Deployment procedures of classic or learned optimizers [CHL22].

[CHL22] Chen T, Chen X, Chen W, Heaton H, Liu J, Wang Z, Yin W. Learning to optimize: A primer and a benchmark. *Journal of Machine Learning Research*. 2022;23(189):1-59.

Model-free Neural Networks (NNs)

- fully data driven
- fast inference
- slow training

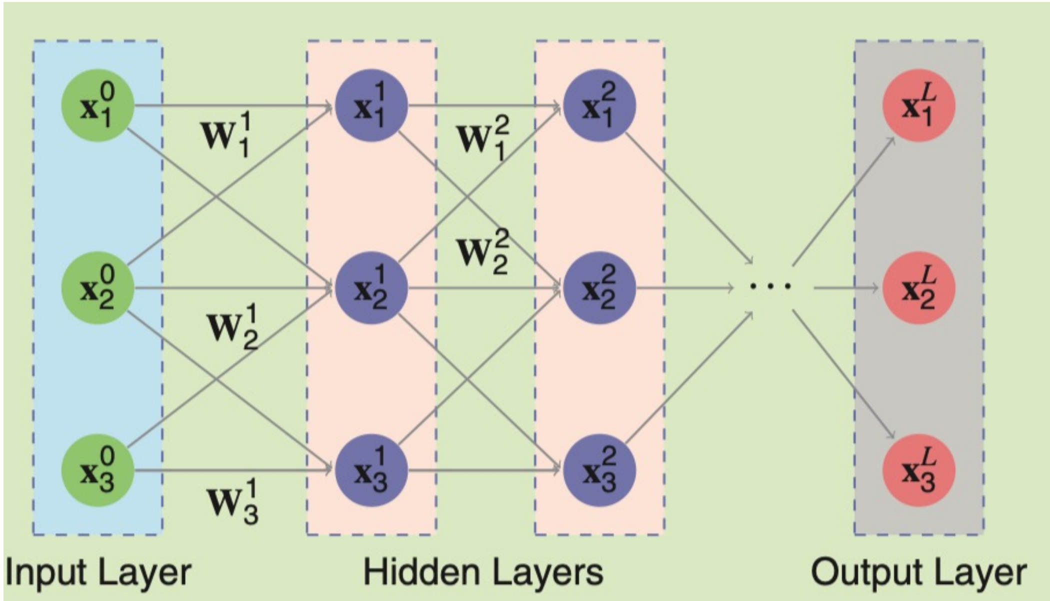


Fig. A conventional neural network architecture [MLE21].

[MLE21] V. Monga, Y. Li and Y. C. Eldar, "Algorithm Unrolling: Interpretable, Efficient Deep Learning for Signal and Image Processing," *IEEE Signal Processing Magazine*, vol. 38, no. 2, pp. 18-44, 2021.

Model-based NNs -- Plug-and-Play (PnP)

- modify existing optimization algorithms
- examples:
  - ✓ plug-and-play
  - ✓ algorithm unrolling

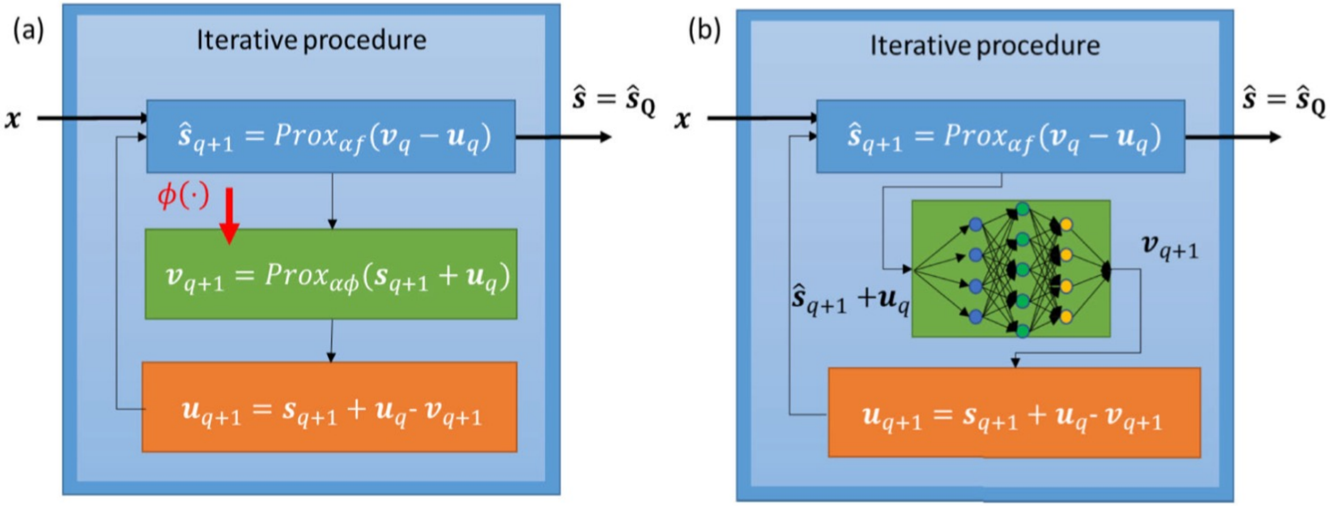


Fig. Illustration of (a) ADMM algorithm; (b) PnP ADMM network [SWC23].

[SWC23] N. Shlezinger, J. Whang, Y. C. Eldar and A. G. Dimakis, "Model-Based Deep Learning," *Proceedings of the IEEE*, vol. 111, no. 5, pp. 465-499, 2023.

Model-based NNs -- Algorithm Unrolling (AU)

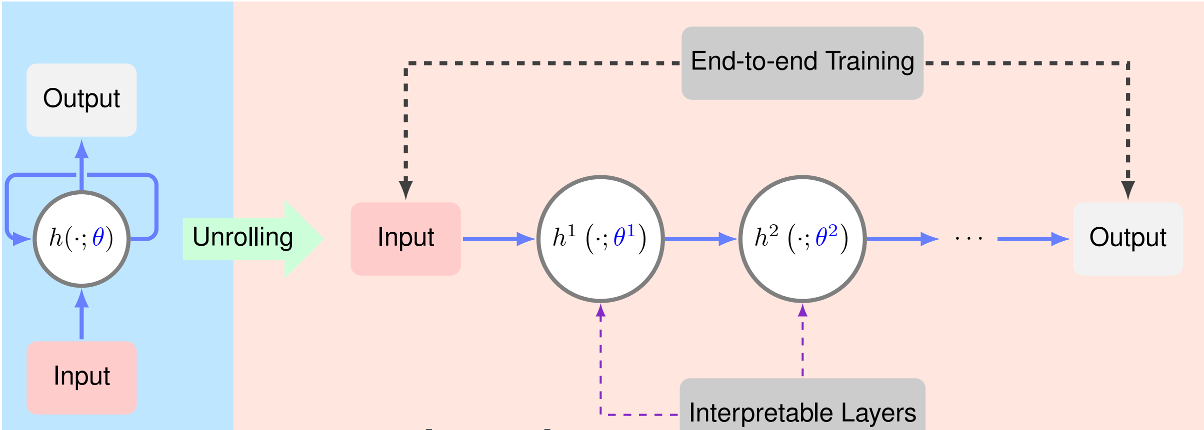


Figure from [MLE21]

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

[MLE21] V. Monga, Y. Li and Y. C. Eldar, "Algorithm Unrolling: Interpretable, Efficient Deep Learning for Signal and Image Processing," *IEEE Signal Processing Magazine*, vol. 38, no. 2, pp. 18-44, 2021.

System model

- **System composition**
  - A dual-function radar-communication base station (N_t transmit antennas and N_r receive antennas)
  - M single-antenna communication users
  - T point-like targets
  - J interferences
- **Transmit signal**

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

$\mathbf{F}_{\mathrm{a}} \in \mathbb{C}^{N_t \times L}$: analog beamforming matrix

$\mathbf{F}_{\mathrm{d}}^{\mathrm{s}} \in \mathbb{C}^{L \times T}$: 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 $\mathbb{E}\{\mathbf{s}\mathbf{s}^{\mathrm{H}}\} = \mathbf{I}_T$

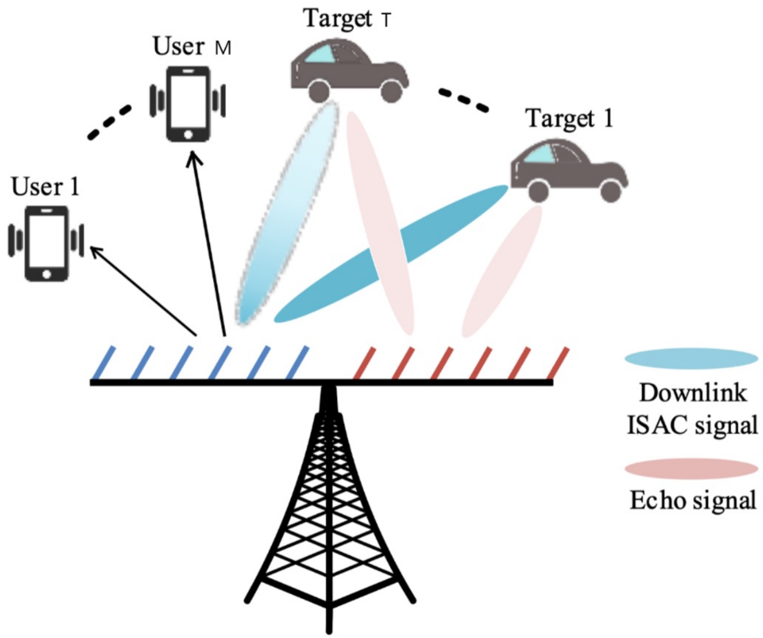


Fig. Illustration of the ISAC system [MWD24] .

[MWD24] C. Meng, Z. Wei, D. Ma, W. Ni, L. Su and Z. Feng, "*Multiobjective-Optimization-Based Transmit Beamforming for Multitarget and Multiuser MIMO-ISAC Systems*," IEEE Internet of Things Journal, vol. 11, no. 18, pp. 29260-29274, 2024.

System model

- **Communication performance metric**

Achievable sum-rate (SR) for M users:

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

where the **communication SINR** for the m th user is:

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

σ_m^2 : noise power at mth user equipment

$\mathbf{h}_m$: communication channel from BS to the mth user

System model

- **Sensing performance metric**

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

Mutual information (MI) for T targets:

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

where the **sensing SINR** for the t th target is:

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

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

θ_l : direction of the target

α_l : propagation reflection coefficient

System model

- **Uncertainty model**

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

$\hat{\mathbf{a}}_{\mathrm{r}}$: presumed receive steering vector

$\boldsymbol{\delta}_{\mathrm{r}}$: steering vector error, bounded by ε_r, i.e., $\boldsymbol{\delta}_{\mathrm{r}} \in \mathcal{A}_{\mathrm{r}} \triangleq \{\boldsymbol{\delta}_{\mathrm{r}} | \frac{1}{N_r} \| \boldsymbol{\delta}_{\mathrm{r}} \|_2 \le \varepsilon_r\}$.

- **Problem formulation (**maximize the communication SR and worst-case sensing MI**)**

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

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

Projected Gradient Descent and Ascent (PGDA)

- **PGD-based minimization over the uncertainty set $\mathcal{A}_{\mathrm{r}}$**
  - **Gradient step**

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

where μ_{r} is the step size for $\boldsymbol{\delta}_{\mathrm{r}}$.

  - **Projection step**

$$\boldsymbol{\delta}_{\mathrm{r}}^{(i+1)} = \min\left\{\frac{\varepsilon_{r} N_{r}}{\left\|\tilde{\boldsymbol{\delta}}_{\mathrm{r}}^{(i+1)}\right\|_{2}}, 1\right\}\tilde{\boldsymbol{\delta}}_{\mathrm{r}}^{(i+1)}$$

Projected Gradient Descent and Ascent (PGDA)

- **PGA-based maximization over $\mathbf{F}_{\rm a}$, $\mathbf{F}_{\rm d}$, and $\mathbf{w}_t$**
  - **Gradient step**

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

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

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

where $\mu_{\rm a}$, $\mu_{\rm d}$, and $\mu_{\rm w}$ are the step sizes for $\mathbf{F}_{\rm a}$, $\mathbf{F}_{\rm d}$, and $\mathbf{w}_t$.

  - **Projection step**

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

Unrolled PGDA

- **Aim**

  Use **algorithm unrolling** to tune step sizes in PGDA algorithm based on data

- **Trainable parameters**

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

  with $\boldsymbol{\mu}_i \triangleq [\mu_{\mathrm{a}}^{(i)}, \mu_{\mathrm{d}}^{(i)}, \mu_{\mathrm{w}}^{(i)}, \mu_{\mathrm{r}}^{(i)}]^{\mathrm{T}}$ representing the step size vector for the ith iteration

- **Loss function** [LS23] $(\gamma = \rho\gamma_{\mathrm{c}} + (1-\rho)\gamma_{\mathrm{s}})$

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

  - weighted sum of losses
  - $\mathcal{D}$: data set containing communication and sensing channel realizations

- **Tune the hyperparameter**

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

Numerical Evaluation

- **System Parameters**
  - $M=4,\ T=8,\ J=4,\ N_t=6,\ N_r=8,\ L=4$
  - Noise variance $\sigma_m^2=\sigma_t^2=\sigma^2=1$, power budget $10\log_{10}\left(\frac{P_t}{\sigma^2}\right)=5\,\mathrm{dB}$
  - weight factor $\rho=0.8$ (i.e., $\gamma=0.8\gamma_{\mathrm{c}}+0.2\gamma_{\mathrm{s}}$)

- **Training settings**
  - Each element of channel $\mathbf{h}$, reflection coefficients α_t, and α_j are drawn from zero-mean complex Gaussian distribution with variances of 0 dB, 2 dB, and 3dB, respectively.
  - Dataset sizes for training and testing: 1000 and 100
  - Stochastic gradient descent (SGD) optimizer with learning rate of 0.01
  - Fixed step size for PGDA and initial step size for unrolled PGDA: 0.1

Numerical Evaluation

▪ Performance evaluation

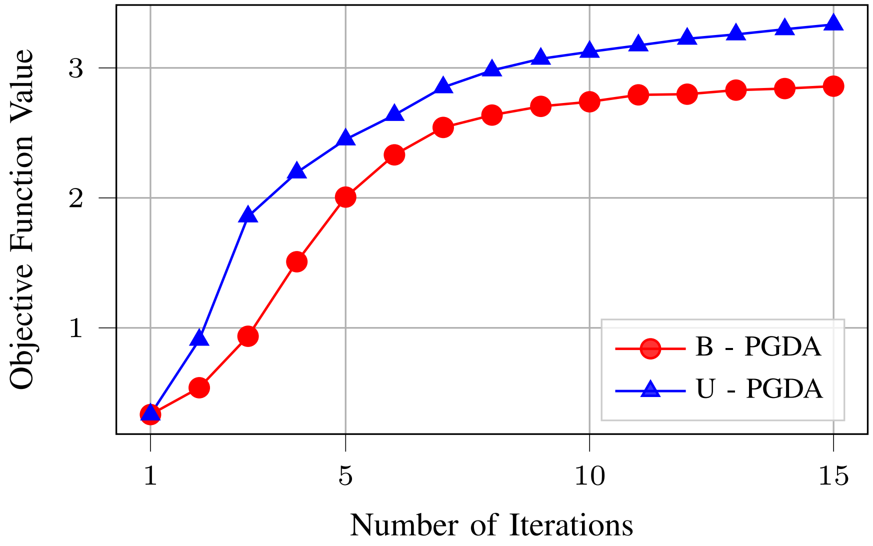


Fig. The objective value versus the number of iterations.

- 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.

Numerical Evaluation

- **Robustness evaluation**

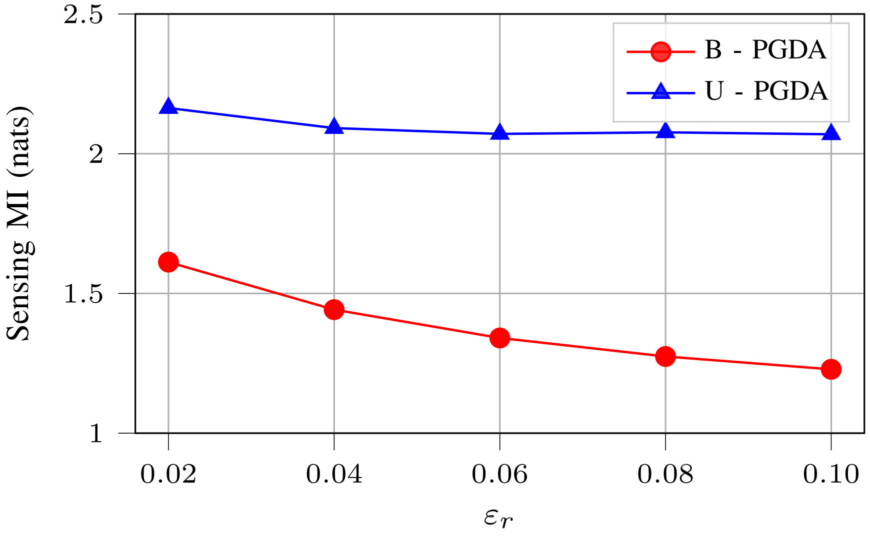


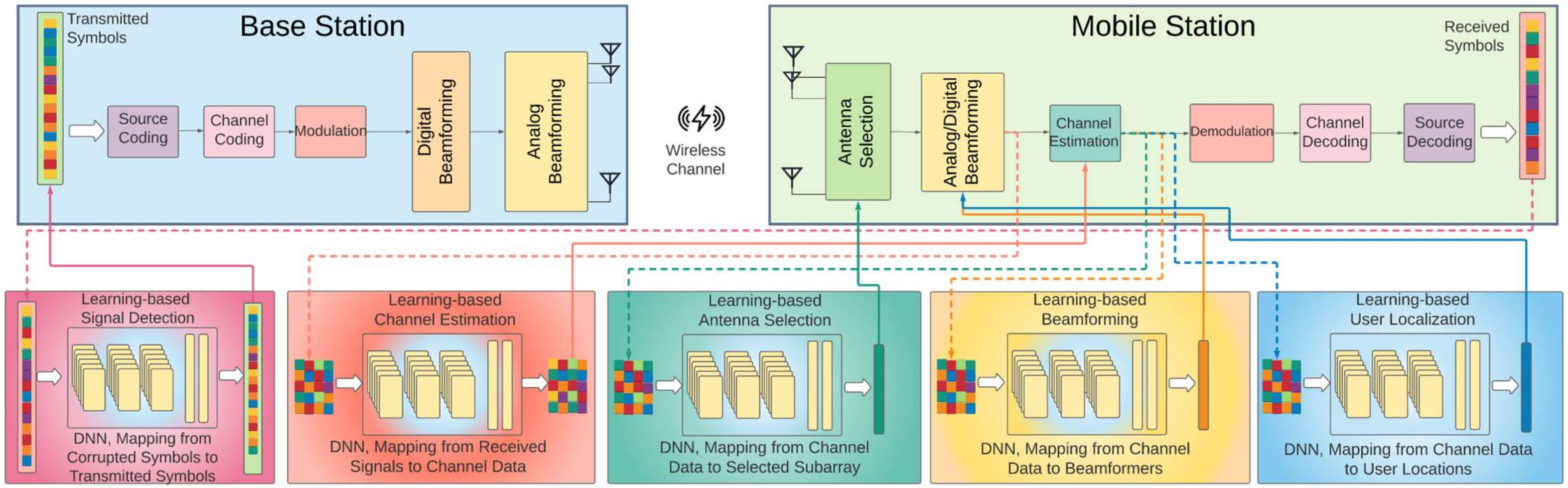
Fig. The sensing MI versus ε_r.

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

Part VII: General Learning-Based Beamforming

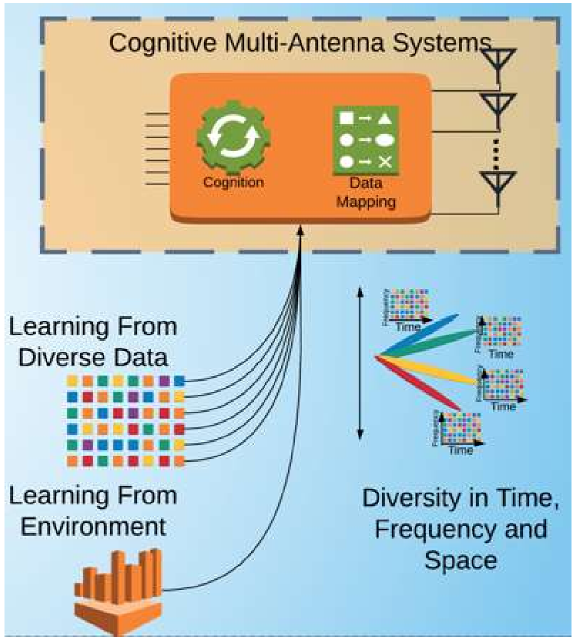
Deep Learning Applications in the Physical Layer

DL can be used for various PHY applications.

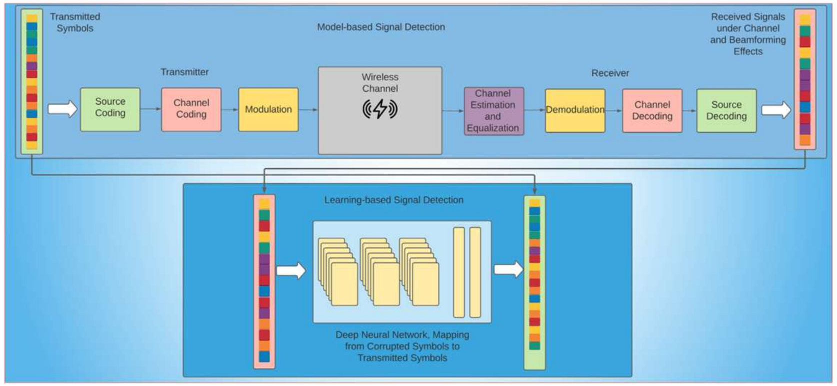


Elbir, Ahmet M. and Kumar Vijay Mishra. "Cognitive Learning-Aided Multi-Antenna Communications.'', IEEE Wireless Communications.
Elbir, Ahmet M., et al. "Federated Learning for Physical Layer Design.", IEEE Communications Magazine, vol. 59, no. 11, pp. 81-87, 2021

Motivation for Learning in Beamforming

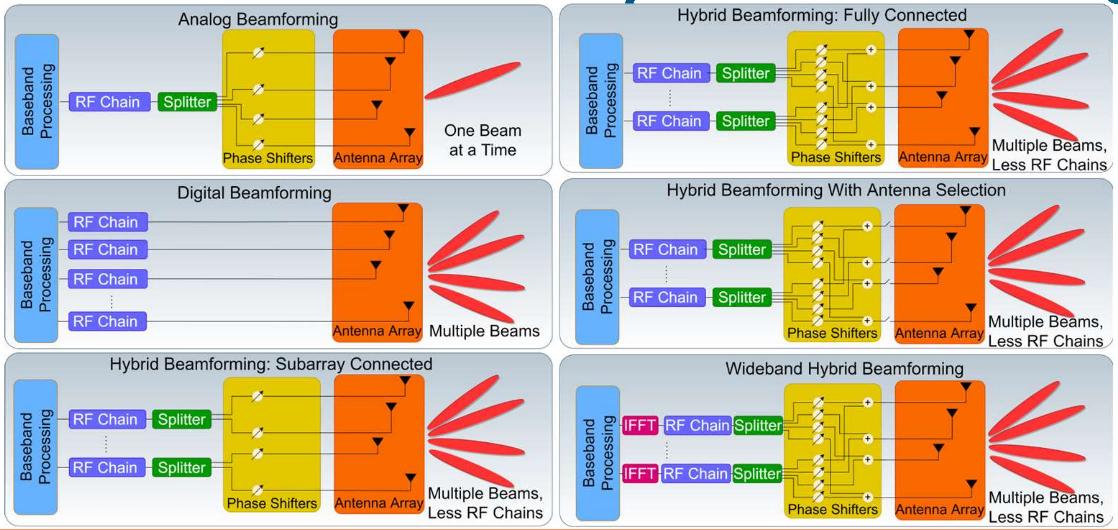


- **Robust** performance against the corruptions/imperfections
- **Easily** updated for incoming future data, **adapt** in response to the environmental changes
- **Lower** post-training **computational complexity**.



Elbir, Ahmet M. and Kumar Vijay Mishra. "Joint Antenna Selection and Hybrid Beamformer Design Using Unquantized and Quantized Deep Learning Networks." IEEE Trans. Wireless Commun., vol. 19, no. 3, 5 Dec. 2019
Elbir, Ahmet M., et al. "Federated Learning for Physical Layer Design.", IEEE Communications Magazine, vol. 59, no. 11, pp. 81-87, 2021
Heath, Robert W., et al. "An Overview of Signal Processing Techniques for Millimeter Wave MIMO Systems." IEEE J. Sel. Top. Signal Process., vol. 10, no. 3, 8 Feb. 2016

Motivation for DL in Hybrid Beamforming



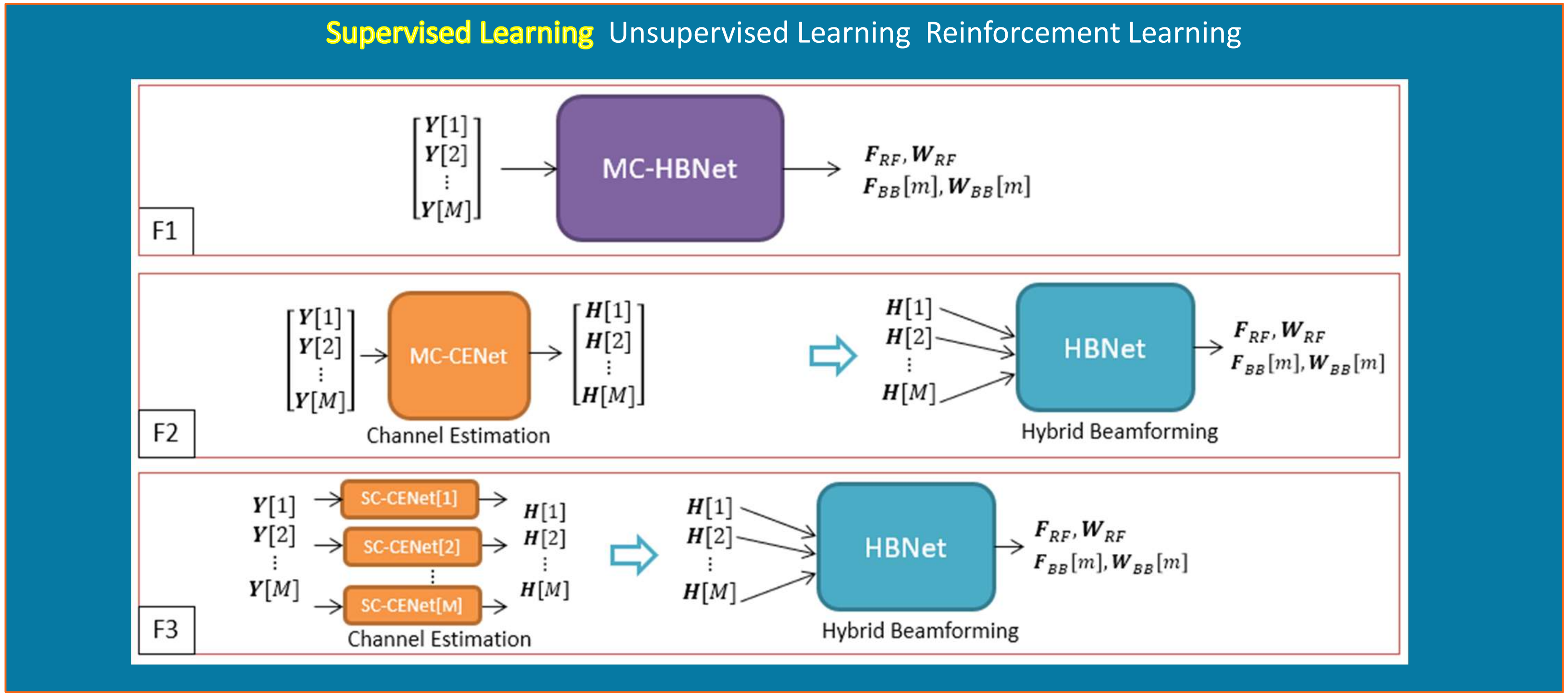
- Therefore, to **reduce** the number of digital RF components, **hybrid analog and baseband beamforming** architectures have been introduced, wherein a **small number of digital beamformers** are employed.
- Hybrid beamforming (HB) is an essential tool for **inexpensive** and yet **effective** transmission for mm-Wave massive MIMO.

Elbir, Ahmet M. and Kumar Vijay Mishra. "Joint Antenna Selection and Hybrid Beamformer Design Using Unquantized and Quantized Deep Learning Networks." IEEE Trans. Wireless Commun., vol. 19, no. 3, 5 Dec. 2019

Elbir, Ahmet M., et al. "Federated Learning for Physical Layer Design.", IEEE Communications Magazine, vol. 59, no. 11, pp. 81-87, 2021

Elbir, Ahmet M., et al. "Twenty-Five Years of Advances in Beamforming: From convex and nonconvex optimization to learning techniques." IEEE Signal Process. Mag., vol. 40, no. 4, 6 June 2023

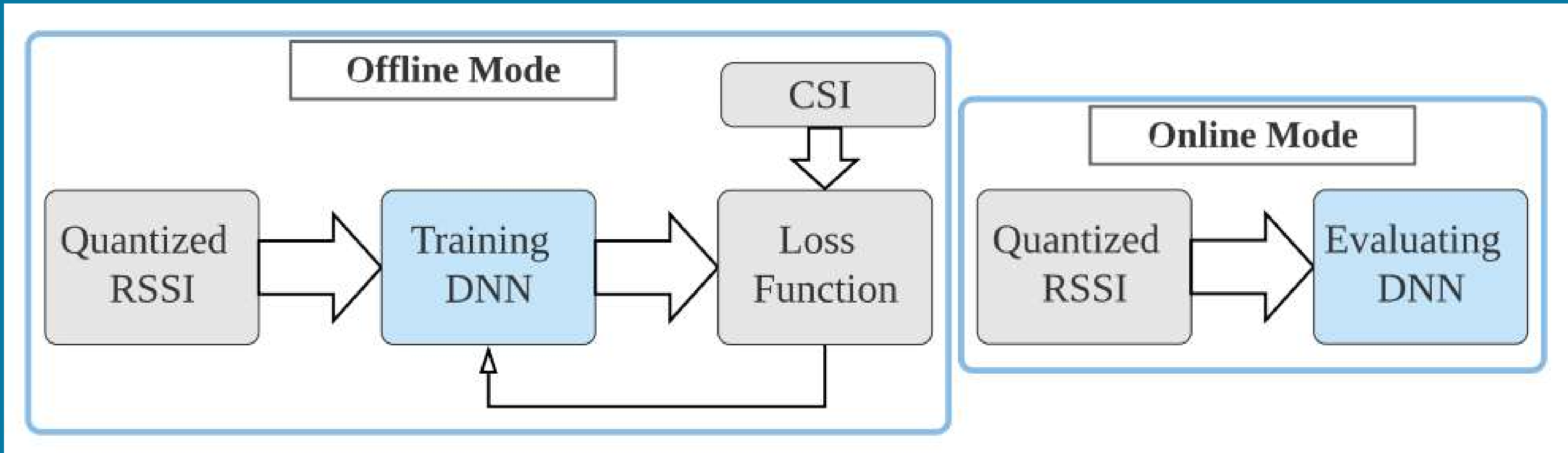
Learning Schemes



Elbir, Ahmet M., et al. "A Family of Deep Learning Architectures for Channel Estimation and Hybrid Beamforming in Multi-Carrier mm-Wave Massive MIMO'', IEEE Transactions on Cognitive Communications and Networking, in press

Learning Schemes

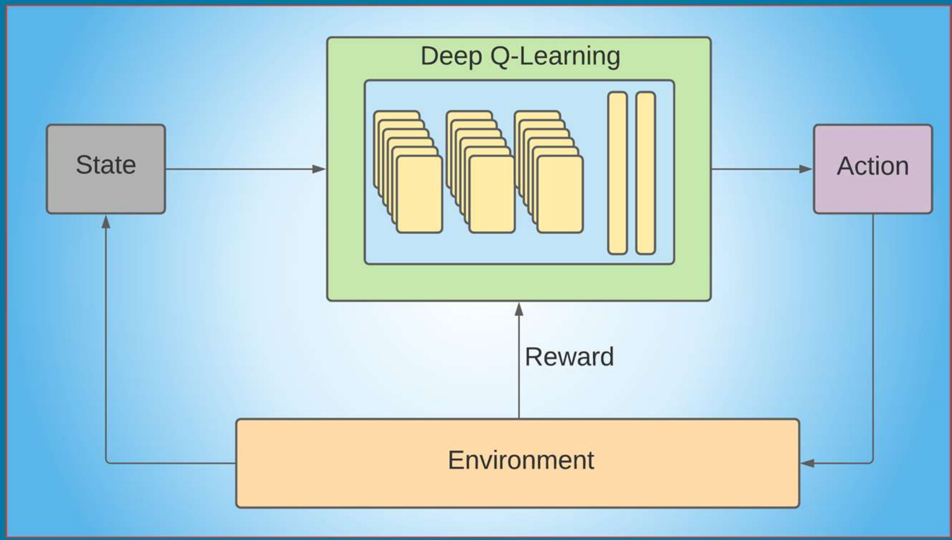
Supervised Learning **Unsupervised Learning** Reinforcement Learning



Hojatian, Hamed, et al. "Unsupervised Deep Learning for Massive MIMO Hybrid Beamforming." arXiv, 30 June 2020
Gao, Jiabao, et al. "Unsupervised Learning for Passive Beamforming." IEEE Commun. Lett., vol. 24, no. 5, 10 Jan. 2020

Learning Schemes

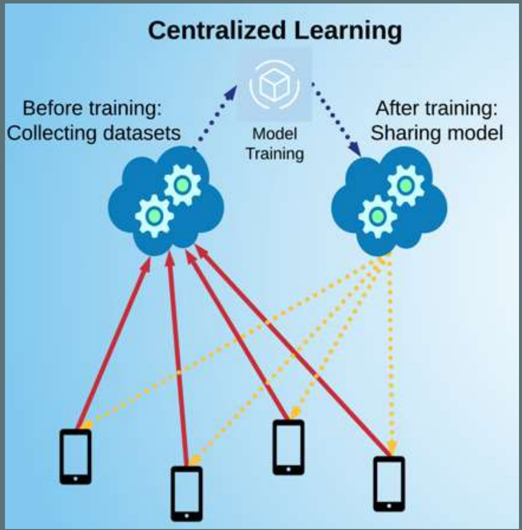
Supervised Learning Unsupervised Learning Reinforcement Learning

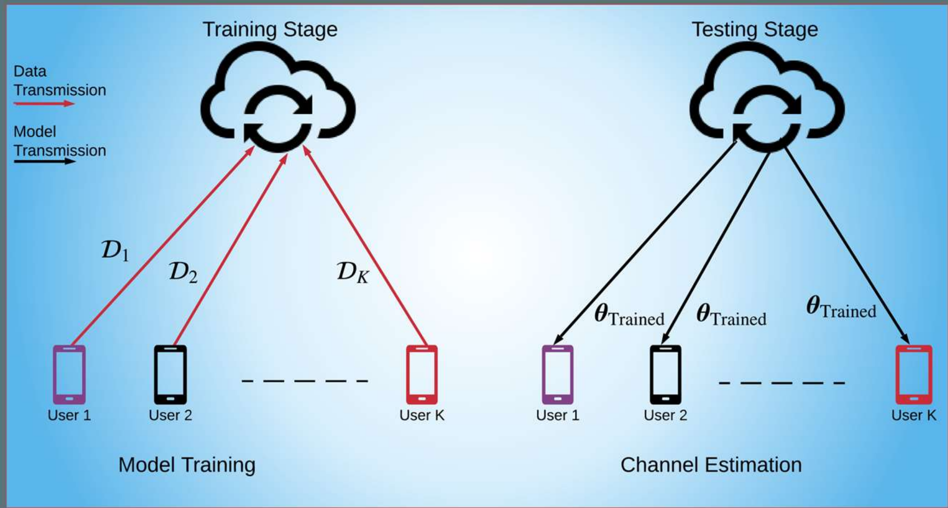


Huang, Chongwen, et al. "Reconfigurable Intelligent Surface Assisted Multiuser MISO Systems Exploiting Deep Reinforcement Learning." arXiv, 24 Feb. 2020

Learning Schemes

Centralized Learning Federated Learning Online Learning Transfer Learning



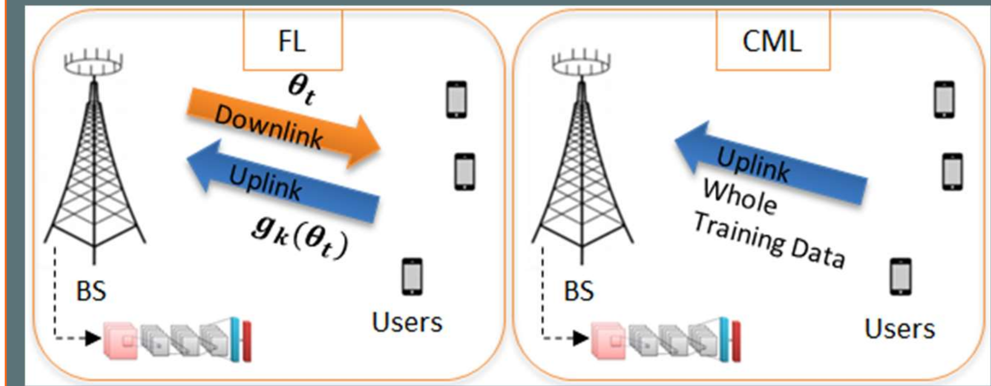


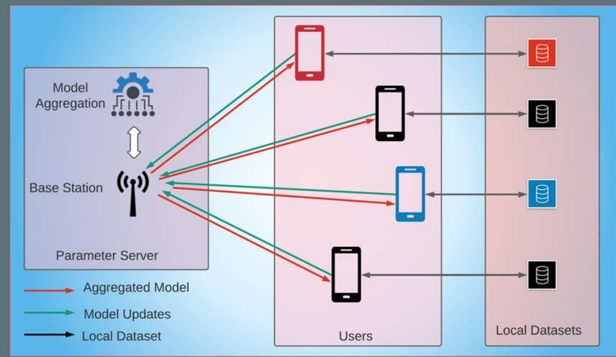
Elbir, Ahmet M. and Sinem Coleri. "Federated Learning for Hybrid Beamforming in mm-Wave Massive MIMO." IEEE Commun. Lett., vol. 24, no. 12, 25 Aug. 2020

Elbir, Ahmet M., et al. Federated Dropout Learning for Hybrid Beamforming with Spatial Path Index Modulation in Multi-User Mmwave-Mimo Systems. ICASSP, 6 June 2021

Learning Schemes

Centralized Learning **Federated Learning** Online Learning Transfer Learning





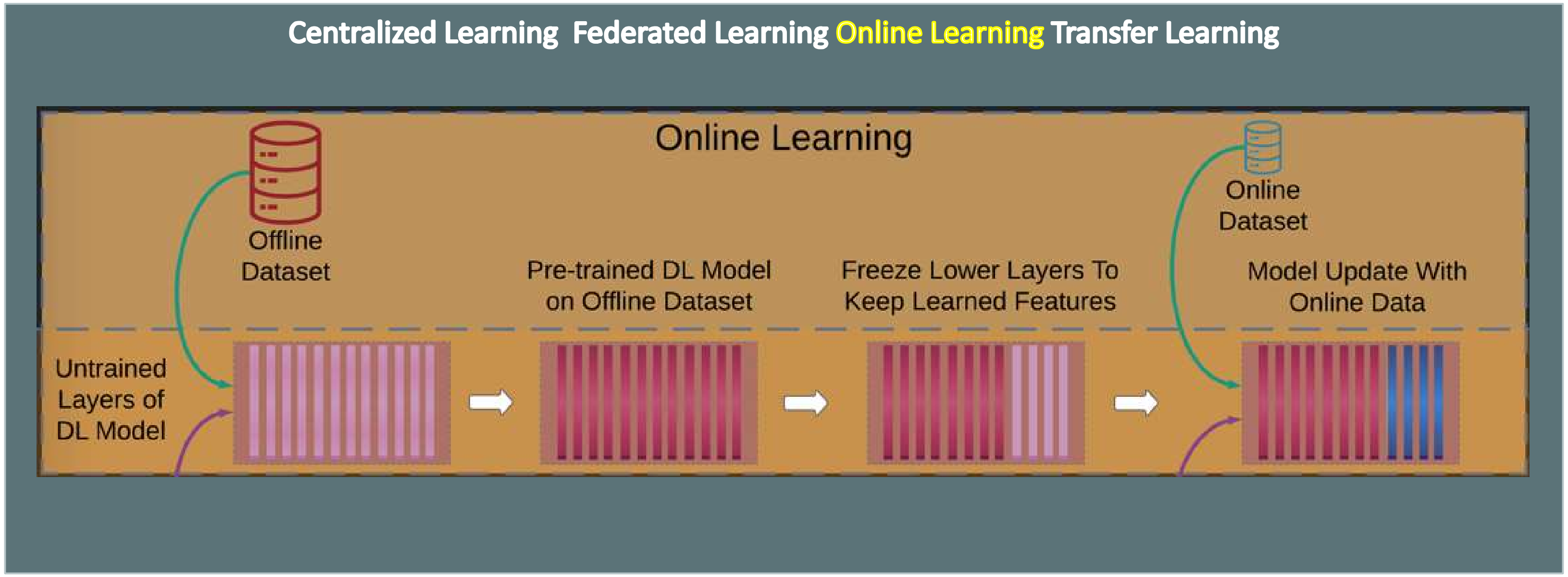
Elbir, Ahmet M. and Sinem Coleri. "Federated Learning for Hybrid Beamforming in mm-Wave Massive MIMO." IEEE Commun. Lett., vol. 24, no. 12, 25 Aug. 2020

Elbir, Ahmet M., et al. Federated Dropout Learning for Hybrid Beamforming with Spatial Path Index Modulation in Multi-User Mmwave-Mimo Systems. ICASSP, 6 June 2021

Elbir, Ahmet M. and Sinem Coleri. "Federated Learning for Channel Estimation in Conventional and IRS-Assisted Massive MIMO.'', IEEE Transactions on Wireless Communications, in press

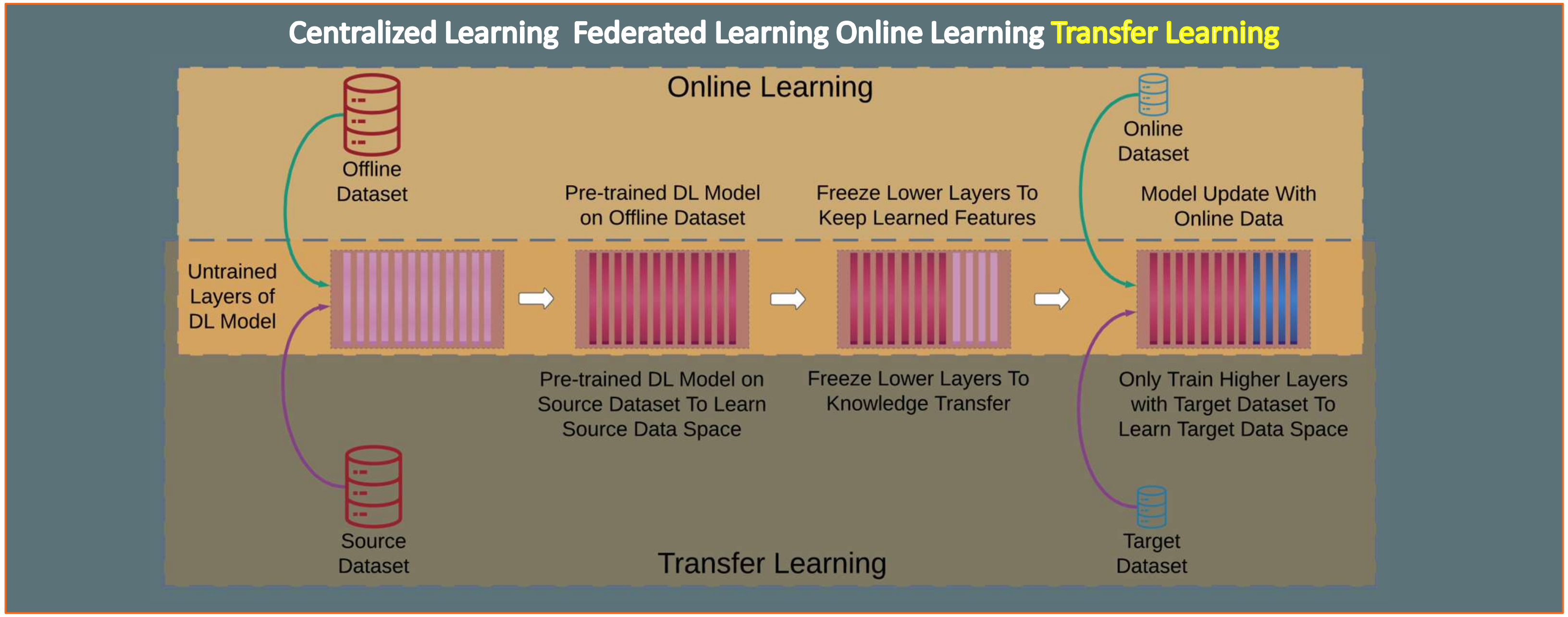
23

Learning Schemes



Elbir, Ahmet M., et al. "A Family of Deep Learning Architectures for Channel Estimation and Hybrid Beamforming in Multi-Carrier mm-Wave Massive MIMO.'' IEEE Transactions on Cognitive Communications and Networking, in press
Elbir, Ahmet M. "A Deep Learning Framework for Hybrid Beamforming Without Instantaneous CSI Feedback." IEEE Trans. Veh. Technol., vol. 69, no. 10, 18 Aug. 2020

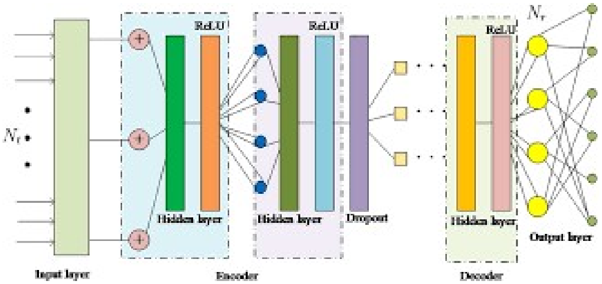
Learning Schemes



Elbir, Ahmet M. and Kumar Vijay Mishra. "Sparse Array Selection Across Arbitrary Sensor Geometries With Deep Transfer Learning." IEEE Trans. Cognit. Commun. Networking, vol. 7, no. 1, 4 June 2020

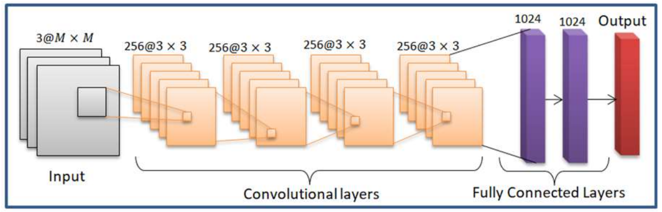
Learning Model Types

- Multi Layer Perceptron (MLP)



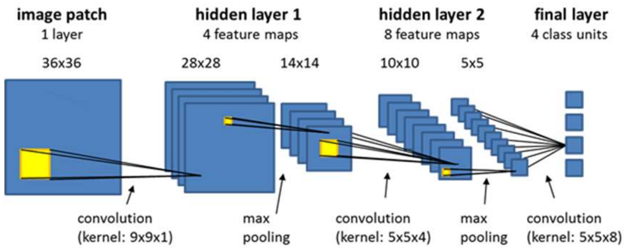
Huang, Hongji, et al. "Deep Learning for Super-Resolution Channel Estimation and DOA Estimation Based Massive MIMO System." IEEE Trans. Veh. Technol., vol. 67, no. 9, 29 June 2018

- Convolutional Neural Network (CNN)



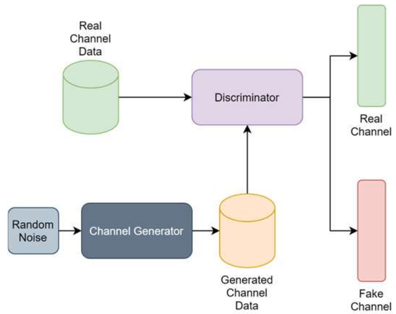
Elbir, Ahmet M. and Kumar Vijay Mishra. "Joint Antenna Selection and Hybrid Beamformer Design Using Unquantized and Quantized Deep Learning Networks." IEEE Trans. Wireless Commun., vol. 19, no. 3, 5 Dec. 2019

- Convolutional-only Neural Network (CoNN)



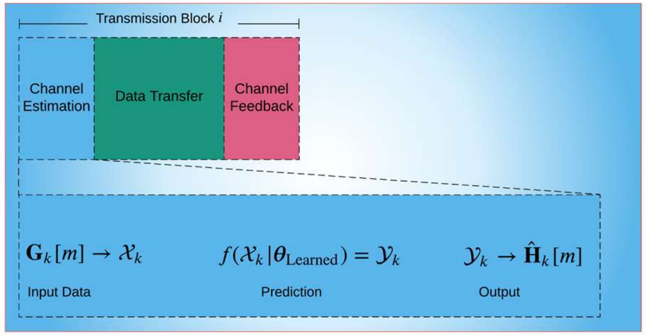
Dong, Peihao, et al. "Deep CNN-Based Channel Estimation for mmWave Massive MIMO Systems." IEEE J. Sel. Top. Signal Process., vol. 13, no. 5, 1 July 2019

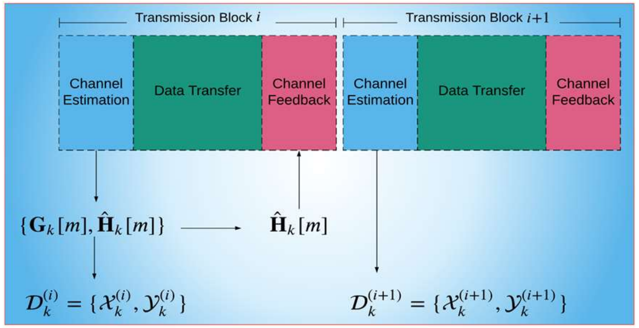
- Generative Adversarial Network (GAN)



Balevi, Eren and Jeffrey G. Andrews. "Unfolded Hybrid Beamforming with GAN Compressed Ultra-Low Feedback Overhead." IEEE Trans. Wireless Commun., 2 July 2021

Data Collection/Generation and Training

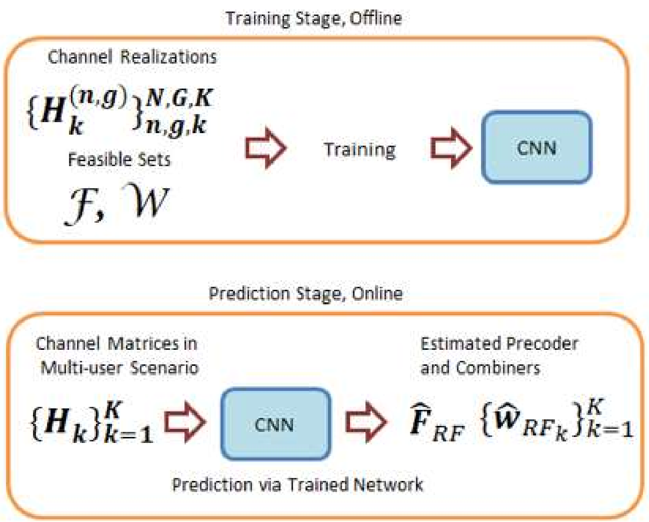




$$\underset{\mathbf{F}_{\text{RF}}, \mathbf{F}_{\text{BB}}}{\text{minimize}} \ ||\mathbf{F}^{\text{opt}} - \mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}||_{\mathcal{F}}^2$$

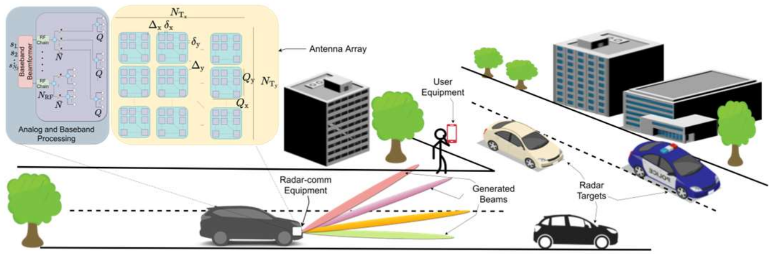
$$\text{subject to: } \mathbf{F}_{\text{RF}} \in \mathcal{F}_{\text{RF}},$$

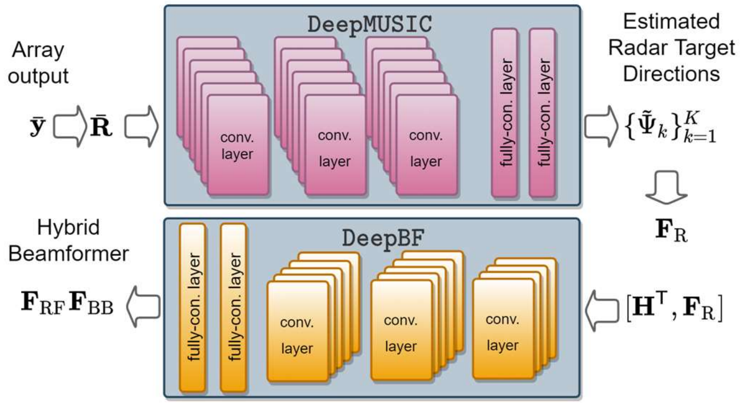
$$||\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}||_{\mathcal{F}}^2 = N_{\text{S}}.$$



Elbir, Ahmet M. and Sinem Coleri. "Federated Learning for Channel Estimation in Conventional and IRS-Assisted Massive MIMO.'', IEEE Transactions on Wireless Communications, in press
Elbir, Ahmet M. and Anastasios K. Papazafeiropoulos. "Hybrid Precoding for Multiuser Millimeter Wave Massive MIMO Systems: A Deep Learning Approach." IEEE Trans. Veh. Technol., vol. 69, no. 1, 4 Nov. 2019

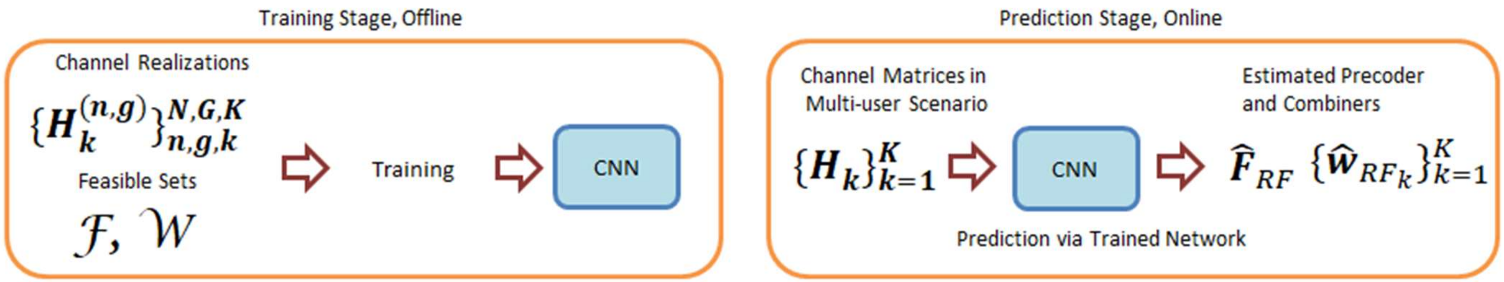
Data Collection/Generation and Training





Elbir, Ahmet M., et al. "Terahertz-Band Joint Ultra-Massive MIMO Radar-Communications: Model-Based and Model-Free Hybrid Beamforming." , IEEE Journal of Selected Topics in Signal Processing, vol. 15, no. 6, pp. 1468-1483, Nov. 2021

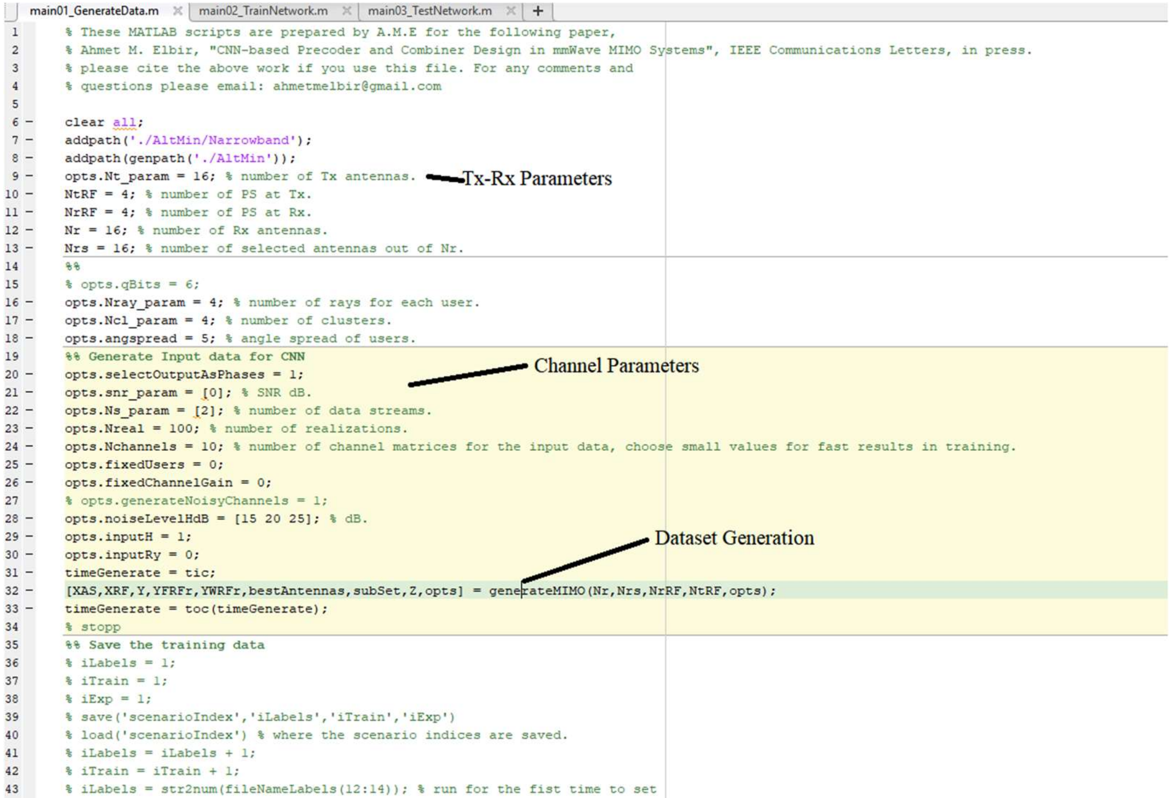
Implementation in MATLAB



Elbir, Ahmet M. and Sinem Coleri. "Federated Learning for Channel Estimation in Conventional and IRS-Assisted Massive MIMO." , IEEE Transactions on Wireless Communications, in press

Elbir, Ahmet M. and Anastasios K. Papazafeiropoulos. "Hybrid Precoding for Multiuser Millimeter Wave Massive MIMO Systems: A Deep Learning Approach." IEEE Trans. Veh. Technol., vol. 69, no. 1, 4 Nov. 2019

Implementation in MATLAB




```matlab
% These MATLAB scripts are prepared by A.M.E for the following paper,
% Ahmet M. Elbir, "CNN-based Precoder and Combiner Design in mmWave MIMO Systems", IEEE Communications Letters, in press.
% please cite the above work if you use this file. For any comments and
% questions please email: ahmetmelbir@gmail.com

clear all;
addpath('./AltMin/Narrowband');
addpath(genpath('./AltMin'));
opts.Nt_param = 16; % number of Tx antennas.
NtRF = 4; % number of PS at Tx.
NrRF = 4; % number of PS at Rx.
Nr = 16; % number of Rx antennas.
Nrs = 16; % number of selected antennas out of Nr.
%%
% opts.qBits = 6;
opts.Nray_param = 4; % number of rays for each user.
opts.Ncl_param = 4; % number of clusters.
opts.angspread = 5; % angle spread of users.
%% Generate Input data for CNN
opts.selectOutputAsPhases = 1;
opts.snr_param = [0]; % SNR dB.
opts.Ns_param = [2]; % number of data streams.
opts.Nreal = 100; % number of realizations.
opts.Nchannels = 10; % number of channel matrices for the input data, choose small values for fast results in training.
opts.fixedUsers = 0;
opts.fixedChannelGain = 0;
% opts.generateNoisyChannels = 1;
opts.noiseLevelHdB = [15 20 25]; % dB.
opts.inputH = 1;
opts.inputRy = 0;
timeGenerate = tic;
[XAS,XRF,Y,YFRFr,YWRFr,bestAntennas,subSet,Z,opts] = generateMIMO(Nr,Nrs,NrRF,NtRF,opts);
timeGenerate = toc(timeGenerate);
% stopp
%% Save the training data
% iLabels = 1;
% iTrain = 1;
% iExp = 1;
% save('scenarioIndex','iLabels','iTrain','iExp')
% load('scenarioIndex') % where the scenario indices are saved.
% iLabels = iLabels + 1;
% iTrain = iTrain + 1;
% iLabels = str2num(fileNameLabels(12:14)); % run for the fist time to set
```

Elbir, Ahmet M. "CNN-Based Precoder and Combiner Design in mmWave MIMO Systems." IEEE Commun. Lett., vol. 23, no. 7, 9 May. 2019

Implementation in MATLAB

```
                else %% Select DOA/DOD for output
                    DOASelected = txang(:,subSetF(qFb,:));
                    DODSelected = rxang(:,subSetF(qWb,:));
                    Z(1,nn).rxang = rxang;
                    Z(1,nn).txang = txang;
                    Z(1,nn).DOASelected = DOASelected;
                    Z(1,nn).DODSelected = DODSelected;
                end
            end
            %% output of the network. classification
            Y(1,j) = (qAb);
%             YRF(1,j) = qFb;
            Yb(1,j) = Y(1,jFirst);
            Z(1,nn).Y = Yb(1,j);
%             Z(1,nn).YRF = [qFb qWb];
            if opts.selectOutputAsPhases == 1
                YFRFr(j,:) = angle(vec(Z(1,nn).FrfSelected));
                YWRFr(j,:) = angle(vec(Z(1,nn).WrfSelected));
%                   if isnan(YFRFr(j,:)) == 1
%                       pause
%                   end
            else % DOA and DOD of users.
                YFRFr(j,:) = pi/180*reshape(DOASelected,[2*NtRF,1]).';
                YWRFr(j,:) = pi/180*reshape(DODSelected,[2*NrRF,1]).';
            end
            XAS(:,:,1,j) = abs(H);
            XAS(:,:,2,j) = real(H);
            XAS(:,:,3,j) = imag(H);
            XRF(:,:,1,j) = abs(H(bestAntennas,:));
            XRF(:,:,2,j) = real(H(bestAntennas,:));
            XRF(:,:,3,j) = imag(H(bestAntennas,:));


            j = j + 1;
            %                 nch
        end
    end
  end
  nch
end
```

Labels

Input Data

Elbir, Ahmet M. "CNN-Based Precoder and Combiner Design in mmWave MIMO Systems." IEEE Commun. Lett., vol. 23, no. 7, 9 May. 2019

Implementation in MATLAB

```
%% DNN Layers.
layers = [imageInputLayer([sizeInputAntennaSelection(1:3)],'Normalization', 'zerocenter');
    convolution2dLayer(2,2^6);
    batchNormalizationLayer
    reluLayer();
    convolution2dLayer(2,2^6);
    batchNormalizationLayer
    reluLayer();
    convolution2dLayer(2,2^6);
    batchNormalizationLayer
    reluLayer();
    fullyConnectedLayer(2^7);
    reluLayer();
    dropoutLayer();
    fullyConnectedLayer(2^7);
    reluLayer();
    dropoutLayer();
    fullyConnectedLayer(Qf);
    softmaxLayer();
    classificationLayer();
    ]
    optsAntennaSelection = trainingOptions('sgdm',...
    'Momentum', 0.9,...
    'InitialLearnRate',0.01,... % The default value is 0.01.
    'MaxEpochs',5000,...
    'MiniBatchSize',miniBatchSize,... % The default is 128.
    'LearnRateSchedule','piecewise',...
    'LearnRateDropFactor',0.9,...
    'LearnRateDropPeriod',30,...
    'L2Regularization',0.0001,... % The default is 0.0001.
    'ExecutionEnvironment', 'auto',...
    'ValidationData',{valDataAntennaSelection,valLabelsAntennaSelection},...
    'ValidationFrequency',validationFrequency,...
    'ValidationPatience', 200,...
    'Plots','none',...
    'Shuffle','every-epoch',...
    'OutputFcn',@(info)stopIfAccuracyNotImproving(info,3));
convnet = trainNetwork(dataAntennaSelection, labelsAntennaSelection, layers, optsAntennaSelection);
```

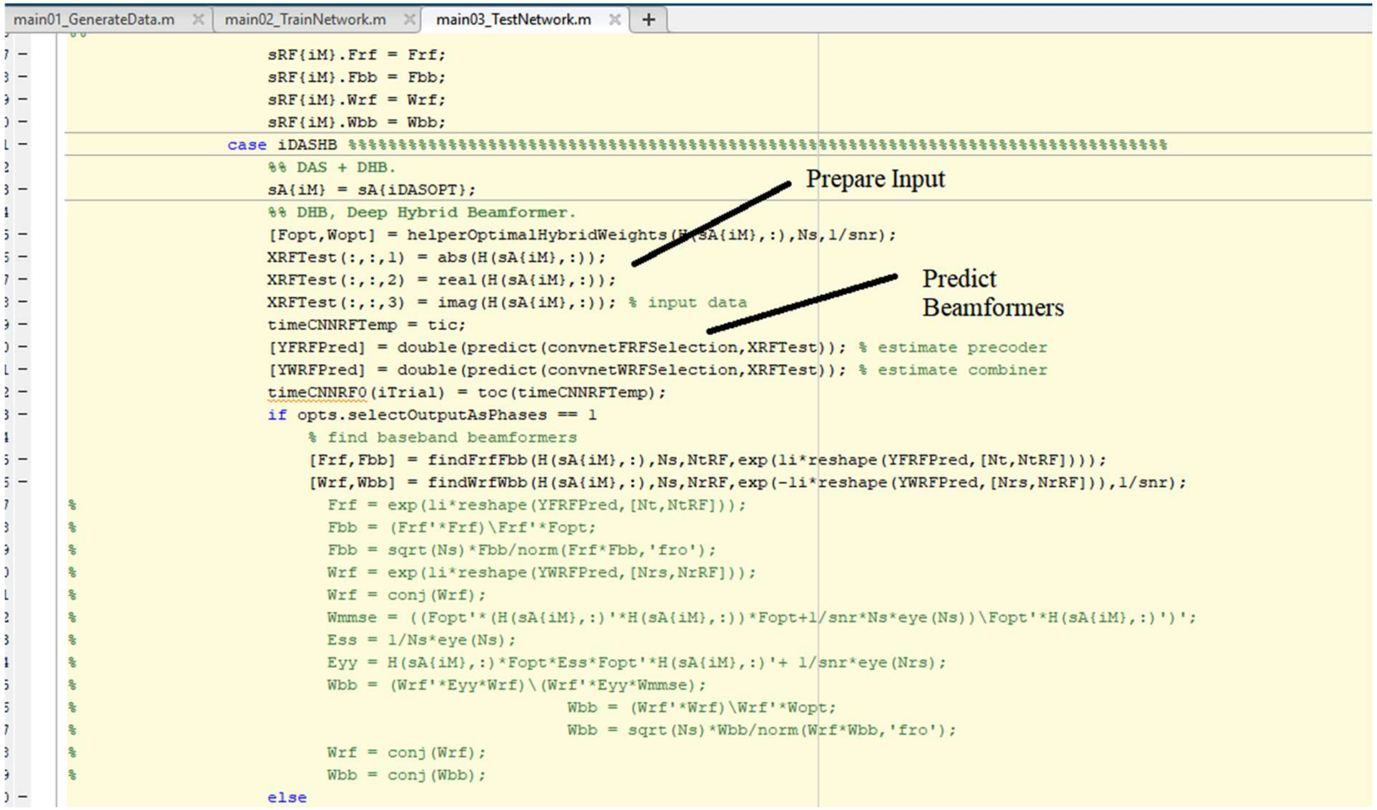
Design Layers

Settings for Training

Train the Model

Elbir, Ahmet M. "CNN-Based Precoder and Combiner Design in mmWave MIMO Systems." IEEE Commun. Lett., vol. 23, no. 7, 9 May. 2019

Implementation in MATLAB



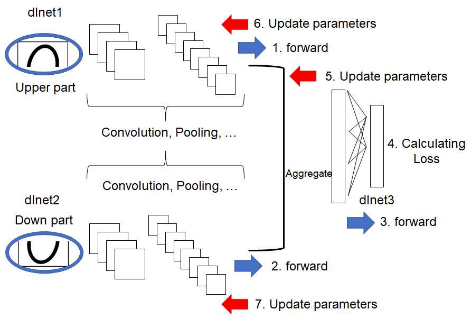

```matlab
            sRF{iM}.Frf = Frf;
            sRF{iM}.Fbb = Fbb;
            sRF{iM}.Wrf = Wrf;
            sRF{iM}.Wbb = Wbb;
        case iDASHB %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
            %% DAS + DHB.
            sA{iM} = sA{iDASOPT};
            %% DHB, Deep Hybrid Beamformer.
            [Fopt,Wopt] = helperOptimalHybridWeights(H(sA{iM},:),Ns,1/snr);
            XRFTest(:,:,1) = abs(H(sA{iM},:));
            XRFTest(:,:,2) = real(H(sA{iM},:));
            XRFTest(:,:,3) = imag(H(sA{iM},:)); % input data
            timeCNNRFTemp = tic;
            [YFRFPred] = double(predict(convnetFRFSelection,XRFTest)); % estimate precoder
            [YWRFPred] = double(predict(convnetWRFSelection,XRFTest)); % estimate combiner
            timeCNNRF0(iTrial) = toc(timeCNNRFTemp);
            if opts.selectOutputAsPhases == 1
                % find baseband beamformers
                [Frf,Fbb] = findFrfFbb(H(sA{iM},:),Ns,NtRF,exp(1i*reshape(YFRFPred,[Nt,NtRF])));
                [Wrf,Wbb] = findWrfWbb(H(sA{iM},:),Ns,NrRF,exp(-1i*reshape(YWRFPred,[Nrs,NrRF])),1/snr);
%                 Frf = exp(1i*reshape(YFRFPred,[Nt,NtRF]));
%                 Fbb = (Frf'*Frf)\Frf'*Fopt;
%                 Fbb = sqrt(Ns)*Fbb/norm(Frf*Fbb,'fro');
%                 Wrf = exp(1i*reshape(YWRFPred,[Nrs,NrRF]));
%                 Wrf = conj(Wrf);
%                 Wmmse = ((Fopt'*(H(sA{iM},:)'*H(sA{iM},:))*Fopt+1/snr*Ns*eye(Ns))\Fopt'*H(sA{iM},:)')';
%                 Ess = 1/Ns*eye(Ns);
%                 Eyy = H(sA{iM},:)*Fopt*Ess*Fopt'*H(sA{iM},:)'+ 1/snr*eye(Nrs);
%                 Wbb = (Wrf'*Eyy*Wrf)\(Wrf'*Eyy*Wmmse);
%                                         Wbb = (Wrf'*Wrf)\Wrf'*Wopt;
%                                         Wbb = sqrt(Ns)*Wbb/norm(Wrf*Wbb,'fro');
%                 Wrf = conj(Wrf);
%                 Wbb = conj(Wbb);
            else
```

Elbir, Ahmet M. "CNN-Based Precoder and Combiner Design in mmWave MIMO Systems." IEEE Commun. Lett., vol. 23, no. 7, 9 May. 2019

Implementation in MATLAB

Useful Links:

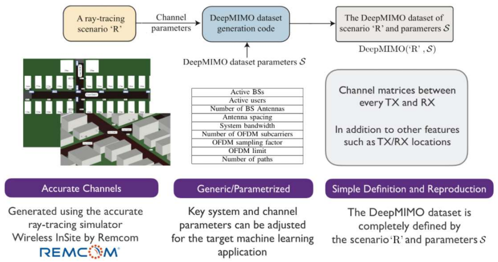
- LTE Downlink Channel Estimation and Equalization: https://www.mathworks.com/help/lte/ug/lte-downlink-channel-estimation-and-equalization.html
- 802.11p Packet Error Rate Simulation for a Vehicular Channel: https://www.mathworks.com/help/wlan/ug/802-11p-packet-error-rate-simulation-for-a-vehicular-channel.html
- Train CNN with Two inputs: https://www.mathworks.com/matlabcentral/fileexchange/74760-image-classification-using-cnn-with-multi-input-cnn
- Reinforcement Learning in MATLAB: https://www.mathworks.com/help/reinforcement-learning/index.html?searchHighlight=reinforcement%20learning&s_tid=srchtitle_reinforcement%20learning_1
- Use parfeval to Train Multiple Deep Learning Networks: https://www.mathworks.com/help/parallel-computing/use-parfeval-to-train-multiple-deep-learning-networks.html

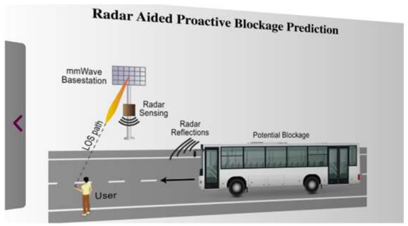


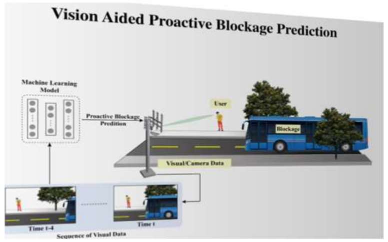
Implementation in MATLAB

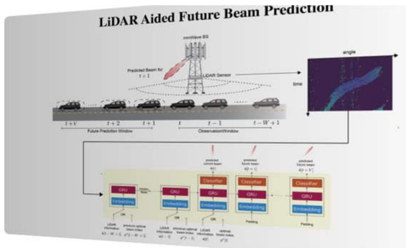
Useful Links:

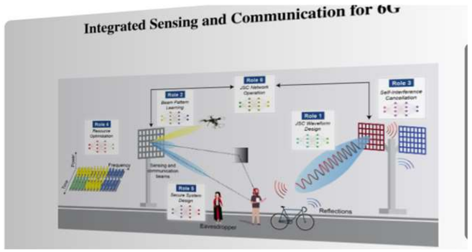
- DeepMIMO: A Generic Deep Learning Dataset: https://www.deepmimo.net/index.html
- DeepSense6g: A Large-Scale Real-World Multi-Modal Sensing and Communication Dataset for 6G Deep Learning Research: https://deepsense6g.net/







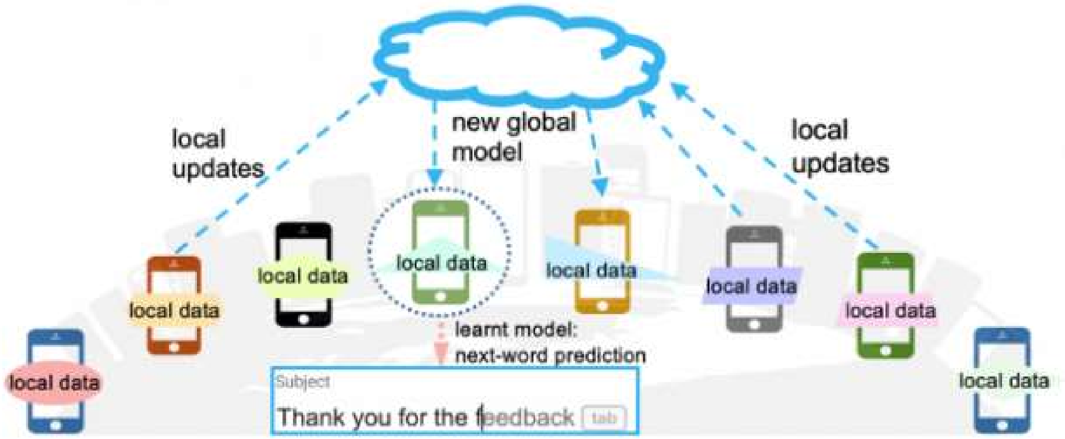




Implementation in MATLAB

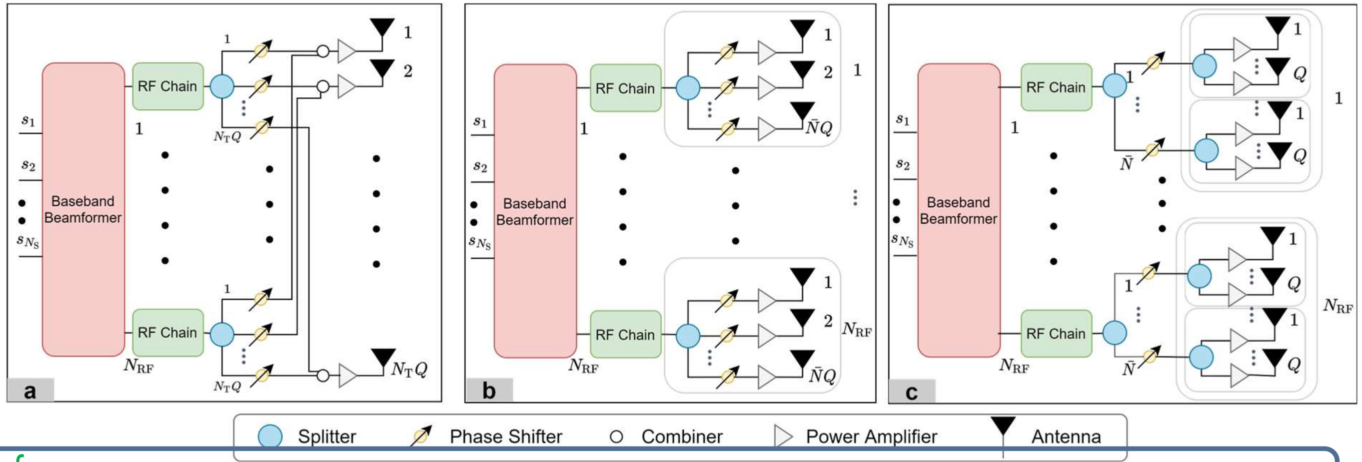
Useful Links:

- Federated Learning in Python: https://towardsdatascience.com/federated-learning-a-step-by-step-implementation-in-tensorflow-aac568283399
- Federated Learning in MATLAB: https://www.mathworks.com/help/deeplearning/ug/train-network-using-federated-learning.html



Hybrid Beamforming via Deep Learning
(Joint Radar-Communications)

Fully-connected | Array-of-subarrays (AoSA) | Group-of-subarrays (GoSA)



Number of phase shifters:	$N_T Q N_{RF}$	$\bar{N}Q \quad (\bar{N} = \frac{N_T}{N_{RF}})$	N_T

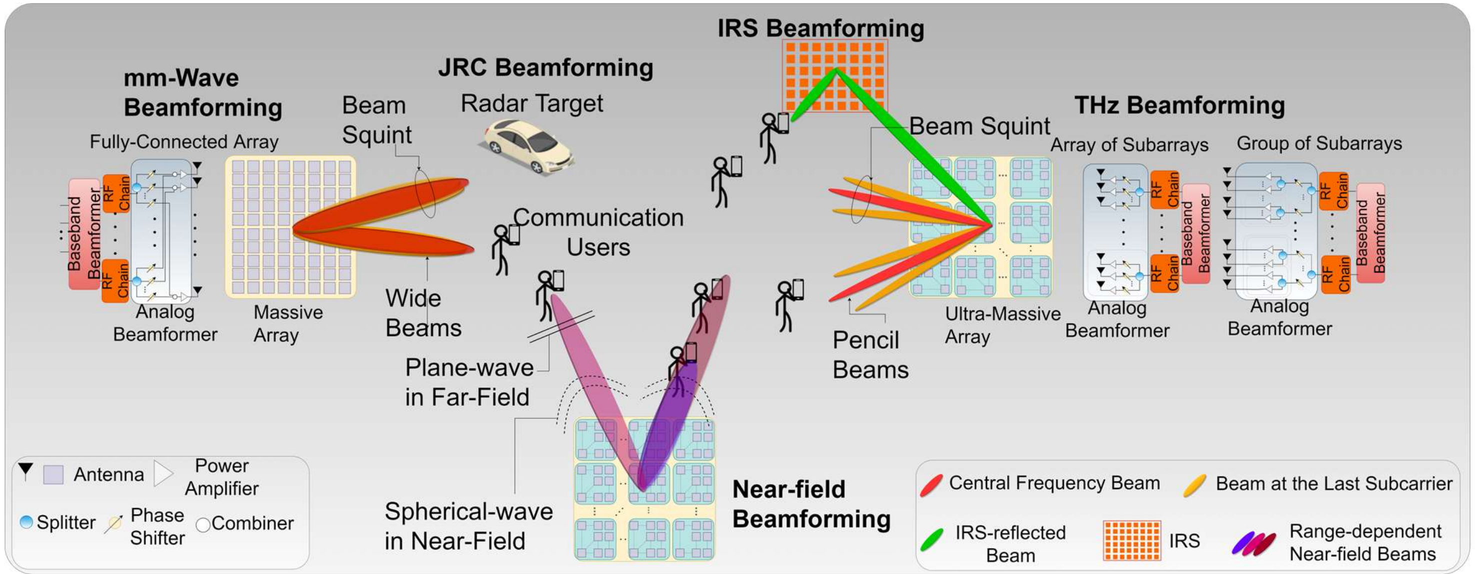
Major Challanges in THz Hybrid Beamforming :

- High path loss: LoS-dominant with multiple NLoS channel
- **Ultra-massive number of antennas: Group-of-subarrays (GoSA)**
- Complexity: Deep-learning-based solutions

Elbir, Ahmet M., et al. "Terahertz-Band Joint Ultra-Massive MIMO Radar-Communications: Model-Based and Model-Free Hybrid Beamforming.'', IEEE Journal of Selected Topics in Signal Processing, vol. 15, no. 6, pp. 1468-1483, Nov. 2021

Summary and Emerging Techniques

- Beamforming has a rich research heritage spanning several decades
- Beamforming in other applications: ultrasound, quantum, acoustics, synthetic apertures
- mmWave vs THz, intelligent surfaces, near-field, joint radar-comms
- Learning techniques offer a very interesting future outlook



Lessons and Future

- Advances in robust optimization are quite important! So, advancing math for robust optimization is of high significance.
- Optimization algorithms can be mapped relatively straightforwardly to network architectures that can be learned (L2O). Powerful tool to find better solutions, but we can do better in terms of computational complexity.
- Solutions for complex beamforming problems follow the same basic principles, but face challenges that can be efficiently addressed only by learning
- Robust optimization is still instrumental in the design of network architectures, learning based solutions; deriving guarantees, which are useful to know in practice


40000
THANKS!