Abstract

I. Introduction

II. System Model

In this section, we first introduce the whole wireless system, then, feature processing and data distribution is introduced, followed with the over-the-air computation model, finally, the discriminant analysis is applied to analyze the recieved signal after AirComp.

A. Network Model

Consider a system with one server equipped with a multi-antenna access point (AP) and K multi-antenna sensors, e.g., cameras and radar sensors. The sensors detect a same information source simultaneously, while the server aggerate the real-time sensing results from the sensors to generate a feature vector to make an inference task. Each observation of sensors is a corrupted version of ground-truth data, as each sensor may be influencedb by evironmental factors or its own hardware's factors or the other influences.

The server has N_r receive antennas, while each sensor has $N_t \leq N_r$ transmit antennas. The observed feature of the k-th sensor is denoted as $\mathbf{x}_k = \mathbf{x} + \mathbf{d}_k$, where $\mathbf{x} \in \mathbb{R}^{\mathbf{N}}$ is the ground-true data, $\mathbf{d}_k$ is the observation distortion with the same dimension of the ground-true data, and N is the dimension of the sensing data. The whole system is shown in Fig.1.

B. Feature Processing and Data Distribution

We try to make a real-time inference based on the sensing result. However, the sensing data may be too big, i.e., the dimension N is too large. This may cause a high communication latency, which the users may not want. Thus, it's neccessary to reduce the dimension of the raw sensing data. In this work, principal component analysis (PCA) is used to deal with the sensing data.

With PCA, our work can be regarded as two stages, i.e., the training stage and the inference stage. In the training stage, the server has a fine-tuned offline training dataset, i.e., the ground-true dataset. PCA is first used to deal with this dataset, and a inference model is then trained based on the projected data. In the test stage, PCA model is first broadcasted to each device.

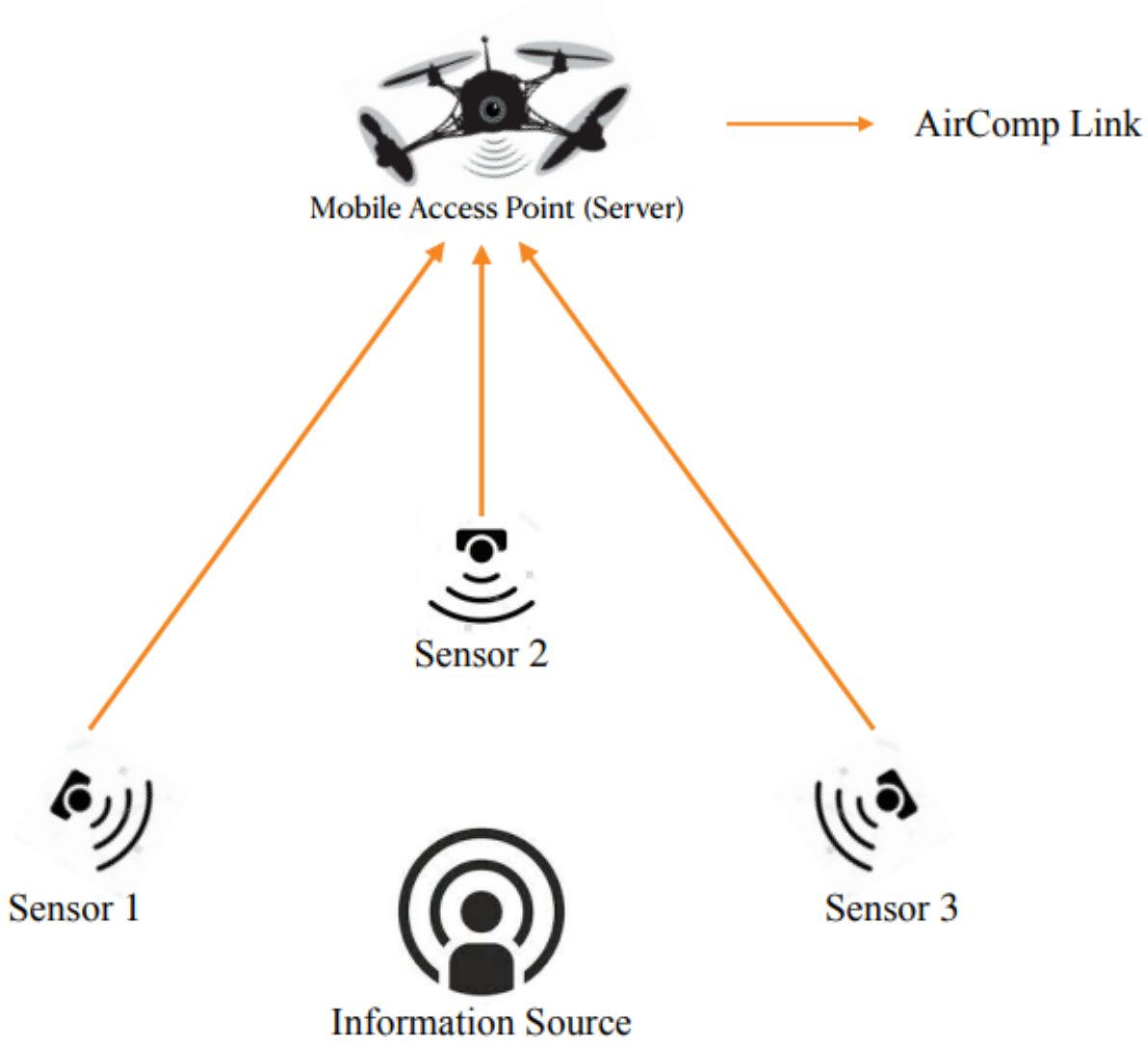

Figure 1: Illustration of the system.

After the devices get the sensing data, they use the PCA model to extract the features, and then transmitted.

Based on the PCA, we can now introduce the data distribution after the PCA. We first assume that after PCA, the data dimension is reduced to M with $M < N$. Recall that the observed feature of the k-th sensor is denoted as $\mathbf{x}_k = \mathbf{x} + \mathbf{d}_k$. After PCA, the feature vector of the k-th sensor can be denoted as

$$\tilde{\mathbf{x}}_k = \tilde{\mathbf{x}} + \tilde{\mathbf{d}}_k, \tag{1}$$

where $\tilde{\cdot}$ means corresponding data after PCA.

The ground-true data $\tilde{\mathbf{x}}$ is assumed to follow a Gaussian mixture given as

$$\tilde{\mathbf{x}} \sim \frac{1}{L}\sum_{l=1}^{L}\mathcal{N}(\mathbf{m}\mu_l, \mathbf{\Sigma}_l), \tag{2}$$

where L denotes the number of total classes of the inference task, and $\mathbf{m}\mu_l = [\mu_{l,1}, \mu_{l,2}, \ldots, \mu_{l,M}]^T$ is the mean vector of the l-th class, $\mathbf{\Sigma}_l = \text{diag}\{\sigma_1^2, \sigma_2^2, \ldots, \sigma_M^2\}$ is the covariance matrix. The projected distortion $\mathbf{d}_k$ is assume to have the following distribution

$$\tilde{\mathbf{d}}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{D}_k), \tag{3}$$

where $\mathbf{D}_k$ is the diagonal covariance matrix with $\mathbf{D}_k = \text{diag}\left\{\delta_{k,1}^2, \delta_{k,2}^2, \ldots, \delta_{k,M}^2\right\}$.

C. Over-the-Air Computation

As diffirent sensor has diffirent physical properties, and observes the information source at diffirent perspective, the distortion of diffirent sensor may be diffirent, too. In a certain sensor, some dimensions may have large distortions while other dimensions may have much smaller distortions. Thus, for more accurate inference, a weighted sum of sensors' data can be considered, i.e.

$$y = \sum_k \mathbf{W}_k\tilde{\mathbf{x}}_k, \tag{4}$$

where $\mathbf{W}_k$ is the weight matrix. The dimensions with small distortions are given a large weight, while dimensions with large distortions are given a small weight.

To this end, the over-the-air computation(AirComp) technique can be used to aggregate the local obeservations $\{\mathbf{x}_k\}$ detected by all devices. The overall model is illustrated in Fig.2.

In each slot, each sensor detects the information resorce to obtain the local observation $\{\mathbf{x}_k\}$. Then, they use PCA to extrat the features $\{\tilde{\mathbf{x}}_k\}$ from the observation based on the assumed data distribution. As different sensors may have different obeservations with different observation distortions, their observations may have different importance. Thus, the feature is transformed into transmit form, and then multiplied by the pre-coders, which is learnable and denote the importance of each observation. After that, the server uses the AirComp technique to aggregate the weighted features to generate the overall feature and finish a inference based on the overall feature.

The transmit symbol of device k is denoted as $\mathbf{s}_k \in \mathbb{C}^{N_t}$. Without loss of generality, each element of $\mathbf{s}_k$ is decomposed of two feature elements, given as

$$s_{k,i} = \tilde{x}_{k,(2i-1)} + j\tilde{x}_{k,2i},\ 1 \le i \le N_t,\ 1 \le k \le K, \tag{5}$$

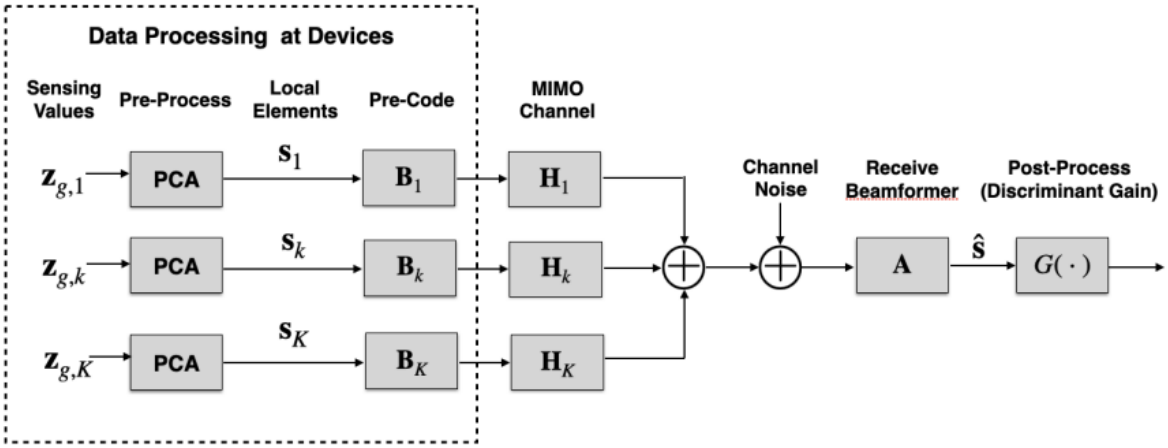

Figure 2: Illustration of the AirComp system.

where $s_{k,i} \in \mathbb{C}$ is the i-th element of the transmit symbol $\mathbf{s}_k$, $x_{k,(2i-1)}$ and $x_{k,2i}$ are the $(2i-1)$-th and $2i$-th observed feature dimensions of device k.

At the server, the received signal can be written as

$$\hat{\mathbf{y}} = \sum_{k=1}^{K} \mathbf{H}_k \mathbf{B}_k \mathbf{s}_k + \mathbf{n}, \tag{6}$$

where the $\mathbf{B}_k \in \mathbb{C}^{N_t \times N_t}$ is the pre-coding matrix of sensor k, denoting the importance of the feature of the k-th sensor, the $\mathbf{H}_k \in \mathbb{C}^{N_r \times N_t}$ is the channel matrix of sensor k and $\mathbf{n} \in \mathbb{C}^{N_r}$ is the noise with $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \delta_0^2 \mathbf{I})$, the δ_0^2 is the noise variance. Then, the server uses a receive beamforming matrix to aggeragate all local transmit symbol $\mathbf{s}_k$ to generate the estimates of the ground-true feature. Specifically, the received symbol after beamforming is given as

$$\hat{\mathbf{s}} = \mathbf{A}\hat{\mathbf{y}} = \mathbf{A} \sum_{k=1}^{K} \mathbf{H}_k \mathbf{B}_k \mathbf{s}_k + \mathbf{A}\mathbf{n}, \tag{7}$$

where $\mathbf{A} \in \mathbb{C}^{N_t \times N_r}$ is the beamforming matrix.

D. Discriminate Gain

One of the most important thing of this system is, how to design the AirComp parameters, so that we can have the most accurate inference result? The authors of [x] propose a metric named discriminant gain to reach this goal. The discriminant gain is derived from a well-known metric named *symmetric Kullback-Leibler (KL) divergence*.

Recall that the sensors' features follow a Gaussian mixture with L classes, given as

$$\tilde{\mathbf{x}} \sim \sum_{l=1}^{L} \mathcal{N}(\mathbf{m}\mu_l, \mathbf{\Sigma}_l). \tag{8}$$

Pairwise discriminant gain measures the discernibility between two classes in the feature space. Consider an arbitrary class pair, say classes l and l', and a data from the Gaussian mixtrue, say $\tilde{\mathbf{x}}$. Based on the distribution of $\mathbf{x}$, the *KL divergence* is used to define the pair-wise discriminant gain:

$$\begin{aligned} G_{l,l'}(\tilde{\mathbf{x}}) &= KL\left[\mathcal{N}(\mathbf{m}\mu_l, \mathbf{\Sigma}) \| \, \mathcal{N}(\mathbf{m}\mu_{l'}, \mathbf{\Sigma})\right] + KL\left[\mathcal{N}(\mathbf{m}\mu_{l'}, \mathbf{\Sigma}) \| \, \mathcal{N}(\mathbf{m}\mu_l, \mathbf{\Sigma})\right], \\ &= (\mathbf{m}\mu_l - \mathbf{m}\mu_{l'})^T \mathbf{\Sigma}^{-1} (\mathbf{m}\mu_l - \mathbf{m}\mu_{l'}), \\ &= \sum_{m-1}^{M} G_{l,l'}(\tilde{x}_m), \end{aligned} \tag{9}$$

where $\tilde{x}_m$ is the m-th element of $\tilde{\mathbf{x}}$ and $G_{l,l'}(\tilde{x}_m)$ is given as

$$G_{l,l'}(\tilde{x}_m) = \frac{(\mu_{l,m} - \mu_{l',m})^2}{\sigma_m^2}, \ 1 \le m \le M. \tag{10}$$

Then, the overall discriminant gain is defined as the average over all pair-wise discriminant gains, given as

$$G(\tilde{\mathbf{x}}) = \frac{2}{L(L-1)} \sum_{l'=1}^{L} \sum_{l<l'} G_{l,l'}(\tilde{\mathbf{x}}) = \sum_{m=1}^{M} G(\tilde{x}_m), \tag{11}$$

where $G(\tilde{x}_m)$ is the discriminant gain of the m-th dimension, given as

$$G(\tilde{x}_m) = \frac{2}{L(L-1)} \sum_{l'=1}^{L} \sum_{l<l'} \frac{(\mu_{l,m} - \mu_{l',m})^2}{\sigma_m^2}, \ 1 \le m \le M. \tag{12}$$

Discriminant gain is a measure of the distance between diffirernt classes and the dispersion degree of every class. When the discriminant gain is larger, the distance between diffirent classes is larger, and the data in every class is more compact [x], which can make the classifier works easier and more accurate. Thus, to reach a more accurate inference result, our goal is to maximize the disciminant gain in the reciever by tuning the AirComp parameters.

E. Post Processing

When the discriminant gain is maiximum, it seems that the accuracy will be high. However, recall that the recived symbol is a linear function of the tranmit signal, i.e.,

$$\hat{\mathbf{s}} = \sum_{k=1}^{K} \mathbf{C}_k \mathbf{s}_k + \mathbf{A}\mathbf{n}, \tag{13}$$

where $\mathbf{A}\mathbf{H}_k\mathbf{B}_k = \mathbf{C}_k, \quad 1 \leq k \leq K$. With this linear function, the feature space may be changed, which can cause that the original trained classifier doesn't work. Hence, we need a post processing at the reciever to keep the feature space unchanged and meanwhile, the post processing will not reduce the discriminant gain.

Recall that the transmit feature of the k-th device can be write as $\tilde{\mathbf{x}}_k = \tilde{\mathbf{x}} + \tilde{\mathbf{d}}_k$. Similarly, the transmit symbol can be write as $\mathbf{s}_k = \mathbf{s} + \mathbf{e}_k$, where the $\mathbf{s}_k$ is the transmit symbol of the k-th device, $\mathbf{s}$ is corresponding ground-true feature of the transmit form, and $\mathbf{e}_k$ is the distortion of the k-th device of the transmit form. Thus, (13) can be rewrited as

$$\hat{\mathbf{s}} = \sum_{k=1}^{K} \mathbf{C}_k\mathbf{s} + \sum_{k=1}^{K} \mathbf{C}_k\mathbf{e}_k + \mathbf{A}\mathbf{n}. \tag{14}$$

It should be notice that, the original classifier is trained based on the fine-tuned feature, i.e., the ground-true feature. Hence, we can use following operator to post process the $\hat{\mathbf{s}}$,

$$\tilde{\mathbf{s}} = (\sum_{k=1}^{K} \mathbf{C}_k)^{-1}\hat{\mathbf{s}} = \mathbf{s} + (\sum_{k=1}^{K} \mathbf{C}_k)^{-1} \sum_{k=1}^{K} \mathbf{C}_k\mathbf{e}_k + (\sum_{k=1}^{K} \mathbf{C}_k)^{-1}\mathbf{A}\mathbf{n}. \tag{15}$$

With this post processing, the mean of every class is the same with the ground-truth, i.e., $\mathbb{E}[\tilde{\mathbf{s}}] = \mathbb{E}[\mathbf{s}]$, as $\tilde{\mathbf{d}}_k$ and $\mathbf{n}$'s mean is 0. Simultaneously, with the discriminant gain becoming larger, though the centre of each class remains unchanged, every class becomes more compact, i.e., the variance of each class is smaller.

And it's should be noticed that the following theorem is true:

Theorem 1. *Assume* $\mathbf{x}$ *has Gaussian mixture distribution, and let* $\mathbf{y} = \mathbf{C}\mathbf{x}$. *When* $\mathbf{C}$ *is a diagonal real matrix, the discriminant gain of* $\mathbf{y}$ *and* $\mathbf{x}$ *is the same.*

Proof: xxx □

With this theorem, if $(\sum_{k=1}^{K} \mathbf{C}_k)^{-1}$ is a diagonal real matrix, then our post processing will not change the discriminant gain that has been maximized. And this can be achieved by a polular technicue named *zero-forcing* (ZF) pre-coding which will be introduced later.

III. Problem Formulation

In this section, the received feature vector is first derived, followed by formulating the problem of discriminant gain maximization under the constraints of transmit power.

First, the estimated feature vector, which is denoted as $\hat{\mathbf{x}}$, is derived from the received symbol after beamforming, say $\hat{\mathbf{s}}$. Specifically, the feature elements are recovered from the received

symbol follows the same way in (5). That's to say, the real part and imaginary part of the i-th element of $\hat{\mathbf{s}}$ are the $(2i-1)$-th and the $2i$-th element of $\hat{\mathbf{x}}$, respectively:

$$\begin{cases} \hat{x}_{2i-1} = \mathrm{Re}\left(\hat{s}_i\right), \\ \hat{x}_{2i} = \mathrm{Im}\left(\hat{s}_i\right), \end{cases} \quad 1 \le i \le N_t, \tag{16}$$

where $\mathrm{Re}(\cdot)$ and $\mathrm{Im}(\cdot)$ are the functions to take the real part and imaginary part of a complex number, $\hat{x}_{2i-1}$ and $\hat{x}_{2i}$ are the $(2i-1)$-th and the $2i$-th element of $\hat{\mathbf{x}}$, and $\hat{s}_i$ is the i-th element of $\hat{\mathbf{s}}$ defined in (7).

Then, the objective of this work is to maximize the discriminant gain of the received estimated feature vector, given as

$$\max \;\; G(\hat{\mathbf{x}}), \tag{17}$$

where $G(\cdot)$ represents the discriminant gain and $\hat{\mathbf{x}}$ is the estimated feature vector defined in (16). Besides, there is one constraint on the transmit power of each device, given by

$$\mathrm{tr}(\mathbf{B}_k \mathbf{B}_k^H) \le P_k, \quad 1 \le k \le K, \tag{18}$$

where $\mathbf{B}_k$ is the transmit beamforming matrix of device k and P_k is device k's residual transmit power getting rid of the transmit symbol variance.

Overall, the problem can be formulated as maximizing the discriminant gain of the received estimated feature vector by jointly designing the transmit pre-coders and the receiving beamformer, given by

$$\begin{aligned} (\mathrm{P1}) \quad & \max_{\mathbf{A}, \{\mathbf{B}_k\}} \;\; G(\hat{\mathbf{x}}), \\ & \text{s.t.} \;\; \mathrm{tr}(\mathbf{B}_k \mathbf{B}_k^H) \le P_k, \; 1 \le k \le K. \end{aligned} \tag{19}$$

Note that in (P1), different elements of the received estimated feature vector $\hat{\mathbf{x}}$ are coupled due to the transmission via a MIMO channel, making its the objective function, say the discriminant gain $G(\hat{\mathbf{x}})$, non-convex and difficult to address. In the sequel, (P1) is first simplified and then solved.

IV. Problem Simplification

In this section, the popular *zero-forcing* (ZF) pre-coders (see e.g., [X-X]) are first used to simplify the objective function of (P1). Then, a simplified problem is derived.

A. ZF pre-coders

Following [X-X], the ZF pre-coding design is given as

$$\mathbf{A}\mathbf{H}_k\mathbf{B}_k = \mathbf{C}_k, \quad 1 \leq k \leq K, \tag{20}$$

where

$$\mathbf{C}_k = \mathrm{diag}\{c_{k,1}, c_{k,2}, \ldots, c_{k,N_t}\}, \tag{21}$$

and $c_{k,i} \geq 0$ representing the receive signal strength of the i-th element of device k. It follows that the ZF pre-coders can be derived as

$$\mathbf{B}_k = (\mathbf{A}\mathbf{H}_k)^{\dagger}\mathbf{C}_k, \tag{22}$$

where † denotes the pseudo-inverse. Thus, the constraints become

$$\mathrm{tr}\left(\mathbf{C}_k^2(\mathbf{A}\mathbf{H}_k)^{\dagger}(\mathbf{A}\mathbf{H}_k)^{\dagger H}\right) \leq P_k, \ \forall k. \tag{23}$$

B. Received Estimated Feature Vector

In this part, the received feature vector is obtained by analyzing the received symbols with ZF pre-coders, as described in the sequel.

First, by substituting the ZF pre-coding design in (20) into the received symbol in (7), it can be derived as

$$\hat{\mathbf{s}} = \sum_{k=1}^{K} \mathbf{C}_k\mathbf{s}_k + \mathbf{A}\mathbf{n}, \tag{24}$$

where $\mathbf{C}_k$ is the received signal strength diagonal matrix defind in (21), $\hat{\mathbf{s}}_k$ defined in (5) is the transmit symbol of device k, $\mathbf{A}$ is the receive beamforming matrix, and $\mathbf{n}$ is the channel noise.

Then, according to (24), an arbitrary element of $\hat{\mathbf{s}}$, say $\hat{s}_i$, can be written as

$$\hat{s}_i = \sum_{k=1}^{K} c_{k,i}s_{k,i} + a_i, \quad 1 \leq i \leq N_t, \tag{25}$$

where $c_{k,i}$ defined in (21) is the receive signal strength of the i-th element of device k and a_i is the i-th element of $\mathbf{An}$.

Besides, according to (16), the feature elements associated to an arbitrary element of $\hat{\mathbf{s}}$, say the i-th can be written as

$$\begin{cases} \hat{x}_{2i-1} = \Re\left(\hat{s}_i\right) = \Re\left(\sum_{k=1}^{K} c_{k,i}s_{k,i} + a_i\right), \\ \hat{x}_{2i} = \Im\left(\hat{s}_i\right) = \Im\left(\sum_{k=1}^{K} c_{k,i}s_{k,i} + a_i\right), \end{cases} \forall i, \tag{26}$$

where the notations follow that in (24) and (16). Overall, the estimated feature vector at the server can be written as

$$\hat{\mathbf{x}} = [..., \hat{x}_{2i-1}, \hat{x}_{2i}...,]^T, \tag{27}$$

where $\hat{x}_{2i-1}$ and $\hat{x}_{2i}$ defined in (26) are the $(2i-1)$-th and the $2i$-th featrue elements of $\hat{\mathbf{x}}$.

C. Discriminant Gain

To further derive the discriminant gain, the distrubution of the estimated elements, (i.e., $\hat{x}_i, \hat{x}_k$ which are the i-th and the k-th elements of $\hat{\mathbf{x}}$ with $i \neq k, 1 \leq i, k \leq M$), is first investigated.

Proposition 1. *$\forall 1 \leq i, j \leq M$, $\hat{x}_i$ and $\hat{x}_j$ are independent.*

Proof: According to (7) and (20), the following equations can be derived

$$\hat{x}_i = \sum_{k=1}^{K} c_{k,i} x_{k,i} + a_n,$$
$$\hat{x}_j = \sum_{k=1}^{K} c_{k,j} x_{k,j} + a_{n'},$$

where a_n and $a_{n'}$ are the corresponding elements of the real part or the imaginary part of $\mathbf{An}$ according to (20). As $x_{k,i}$ and $x_{k,j}$ are different elements of the observation of the k-th sensor, which are thus independent, and the observations of differnet sensors are also independent, the first parts of $\hat{x}_i$ and $\hat{x}_j$, i.e., $\sum_{k=1}^{K} c_{k,i} x_{k,i}$ and $\sum_{k=1}^{K} c_{k,j} x_{k,j}$ are independent.

It should be noticed that there are four situations of a_n and $a_{n'}$:

- the real part and the imaginary part of the same elements of $\mathbf{An}$,
- the real part of two different elements of $\mathbf{An}$,
- the real part and the imaginary part of two different elements of $\mathbf{An}$,
- the imaginary part of two different elements of $\mathbf{An}$,

which can be formulized as

$$a_n = \Re(\mathbf{A}_l \mathbf{n}), a_{n'} = \Im(\mathbf{A}_l \mathbf{n}),$$
$$or \quad a_n = \Re(\mathbf{A}_l \mathbf{n}), a_{n'} = \Re(\mathbf{A}_{l'} \mathbf{n}),$$
$$or \quad a_n = \Re(\mathbf{A}_l \mathbf{n}), a_{n'} = \Im(\mathbf{A}_{l'} \mathbf{n}),$$
$$or \quad a_n = \Im(\mathbf{A}_l \mathbf{n}), a_{n'} = \Im(\mathbf{A}_{l'} \mathbf{n}),$$

where $\mathbf{A}_l$ and $\mathbf{A}'_l$ are the corresponding rows of $\mathbf{A}$ which lead to a_n or a'_n.

It should be noticed that for any situation above, a_n and a'_n are independent. However, we only prove the first situation here, as the rest proofs are similar.

Assume the subscript $_{\Re}$ and $_{\Im}$ denote the real part and the imaginary part respectively, e.g., $\mathbf{A}_{\Re,l} = \Re(\mathbf{A}_l)$. Thus, the following equation can be derived

$$\begin{aligned}\mathbf{A}_l\mathbf{n} &= (\mathbf{A}_{\Re,l} + j\mathbf{A}_{\Im,l})(\mathbf{n}_{\Re} + j\mathbf{n}_{\Im}),\\ &= (\mathbf{A}_{\Re,l}\mathbf{n}_{\Re} - \mathbf{A}_{\Im,l}\mathbf{n}_{\Im}) + j(\mathbf{A}_{\Im,l}\mathbf{n}_{\Re} + \mathbf{A}_{\Re,l}\mathbf{n}_{\Im}),\end{aligned}$$

which leads to

$$\begin{aligned}a_n &= \mathbf{A}_{\Re,l}\mathbf{n}_{\Re} - \mathbf{A}_{\Im,l}\mathbf{n}_{\Im},\\ a_{n'} &= \mathbf{A}_{\Im,l}\mathbf{n}_{\Re} + \mathbf{A}_{\Re,l}\mathbf{n}_{\Im}.\end{aligned}$$

Therefore, the covariance of a_n and $a_{n'}$ can be derived as

$$\begin{aligned}COV(a_n, a_{n'}) &= \mathbb{E}[a_n a_{n'}] - \mathbb{E}[a_n]\mathbb{E}[a_{n'}],\\ &= \mathbb{E}[(\mathbf{A}_{\Re,l}\mathbf{n}_{\Re} - \mathbf{A}_{\Im,l}\mathbf{n}_{\Im})(\mathbf{A}_{\Im,l}\mathbf{n}_{\Re} + \mathbf{A}_{\Re,l}\mathbf{n}_{\Im})] - \mathbb{E}[\mathbf{A}_{\Re,l}\mathbf{n}_{\Re} - \mathbf{A}_{\Im,l}\mathbf{n}_{\Im}]\mathbb{E}[\mathbf{A}_{\Im,l}\mathbf{n}_{\Re} + \mathbf{A}_{\Re,l}\mathbf{n}_{\Im}].\end{aligned}$$

As $\mathbf{n}_{\Re}$ and $\mathbf{n}_{\Im}$ are independent, it's obvious that

$$\begin{aligned}&\mathbb{E}[(\mathbf{A}_{\Re,l}\mathbf{n}_{\Re} - \mathbf{A}_{\Im,l}\mathbf{n}_{\Im})(\mathbf{A}_{\Im,l}\mathbf{n}_{\Re} + \mathbf{A}_{\Re,l}\mathbf{n}_{\Im})]\\ &= \mathbf{A}_{\Re,l}\mathbb{E}[\mathbf{n}_{\Re}\mathbf{n}_{\Re}^T]\mathbf{A}_{\Im,l}^T - \mathbf{A}_{\Im,l}\mathbb{E}[\mathbf{n}_{\Im}\mathbf{n}_{\Im}^T]\mathbf{A}_{\Re,l}^T + \mathbf{A}_{\Re,l}\mathbb{E}[\mathbf{n}_{\Re}]\mathbf{A}_{\Re,l}\mathbb{E}[\mathbf{n}_{\Im}] - \mathbf{A}_{\Im,l}\mathbb{E}[\mathbf{n}_{\Im}]\mathbf{A}_{\Im,l}\mathbb{E}[\mathbf{n}_{\Re}],\end{aligned}$$

and

$$\begin{aligned}&\mathbb{E}[\mathbf{A}_{\Re,l}\mathbf{n}_{\Re} - \mathbf{A}_{\Im,l}\mathbf{n}_{\Im}]\mathbb{E}[\mathbf{A}_{\Im,l}\mathbf{n}_{\Re} + \mathbf{A}_{\Re,l}\mathbf{n}_{\Im}]\\ &= \mathbf{A}_{\Re,l}\mathbb{E}[\mathbf{n}_{\Re}]\mathbb{E}[\mathbf{n}_{\Re}^T]\mathbf{A}_{\Im,l}^T - \mathbf{A}_{\Im,l}\mathbb{E}[\mathbf{n}_{\Im}]\mathbb{E}[\mathbf{n}_{\Im}^T]\mathbf{A}_{\Re,l}^T + \mathbf{A}_{\Re,l}\mathbb{E}[\mathbf{n}_{\Re}]\mathbf{A}_{\Re,l}\mathbb{E}[\mathbf{n}_{\Im}] - \mathbf{A}_{\Im,l}\mathbb{E}[\mathbf{n}_{\Im}]\mathbf{A}_{\Im,l}\mathbb{E}[\mathbf{n}_{\Re}].\end{aligned}$$

Due to $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \delta_0^2\mathbf{I})$, the covariance can be furthe formulized as

$$\begin{aligned}COV(a_n, a_{n'}) &= \mathbf{A}_{\Re,l}\mathbb{E}[\mathbf{n}_{\Re}\mathbf{n}_{\Re}^T]\mathbf{A}_{\Im,l}^T - \mathbf{A}_{\Im,l}\mathbb{E}[\mathbf{n}_{\Im}\mathbf{n}_{\Im}^T]\mathbf{A}_{\Re,l}^T - \mathbf{A}_{\Re,l}\mathbb{E}[\mathbf{n}_{\Re}]\mathbb{E}[\mathbf{n}_{\Re}^T]\mathbf{A}_{\Im,l}^T + \mathbf{A}_{\Im,l}\mathbb{E}[\mathbf{n}_{\Im}]\mathbb{E}[\mathbf{n}_{\Im}^T]\mathbf{A}_{\Re,l}^T,\\ &= \delta_0^2(\mathbf{A}_{\Re,l}\mathbf{A}_{\Im,l}^T - \mathbf{A}_{\Im,l}\mathbf{A}_{\Re,l}^T),\\ &= 0,\end{aligned}$$

which means a_n and $a_{n'}$ are independent.

This ends the proof. □

With Proposition 1, the discriminant gains of different estimated elements can be separately analyzed.

To further investigate the discriminant gain, the distributions of the estimated elements should be derived. First, reformula (20) with (7) as

$$\begin{cases} \hat{x}_{2i-1} = \sum_{k=1}^{K} c_{k,i}\tilde{x}_{k,2i-1} + \Re\left(a_i\right), \\ \hat{x}_{2i} = \sum_{k=1}^{K} c_{k,i}\tilde{x}_{k,2i} + \Im\left(a_i\right), \end{cases} \quad \forall i. \tag{28}$$

Recall that the local elements $x_{k,2i-1}$ and $x_{k,2i}$ are given as

$$\begin{aligned} \tilde{x}_{k,2i-1} &= x_{2i-1} + d_{k,2i-1}, \\ \tilde{x}_{k,2i} &= x_{2i} + d_{k,2i}, \end{aligned} \quad \forall k, \tag{29}$$

where x_{2i-1} and x_{2i} are elements of the ground true vector with

$$\begin{aligned} x_{2i-1} &\sim \frac{1}{L}\sum_{l=1}^{L}\mathcal{N}(\mu_{l,2i-1}, \sigma_{2i-1}^2), \\ x_{2i} &\sim \frac{1}{L}\sum_{l=1}^{L}\mathcal{N}(\mu_{l,2i}, \sigma_{2i}^2), \end{aligned} \tag{30}$$

and $d_{k,2i-1}, d_{k,2i}$ are elements of the observation distortion with

$$\begin{aligned} d_{k,2i-1} &\sim \mathcal{N}(0, \delta_{k,2i-1}^2), \\ d_{k,2i} &\sim \mathcal{N}(0, \delta_{k,2i}^2). \end{aligned} \tag{31}$$

Thus, we can further obtain the following lemma

Lemma 1. *The distrubution of the estimated elements $\hat{x}_{2i-1}, \hat{x}_{2i}$ shown as*

$$\begin{aligned} \hat{x}_{2i-1} &\sim \frac{1}{L}\sum_{l=1}^{L}\mathcal{N}(\hat{\mu}_{l,2i-1}, \hat{\sigma}_{2i-1}^2), \\ \hat{x}_{2i-1} &\sim \frac{1}{L}\sum_{l=1}^{L}\mathcal{N}(\hat{\mu}_{l,2i}, \hat{\sigma}_{2i}^2), \end{aligned} \tag{32}$$

where the means $\{\hat{\mu}_{l,2i-1}, \hat{\mu}_{l,2i}\}$ and the variance $\{\hat{\sigma}_{2i-1}^2, \hat{\sigma}_{2i}^2\}$ are given as

$$\begin{cases} \hat{\mu}_{l,2i-1} = \mu_{l,2i-1}\sum_{k=1}^{K} c_{k,i}, \\ \hat{\mu}_{l,2i} = \mu_{l,2i}\sum_{k=1}^{K} c_{k,i}, \\ \hat{\sigma}_{2i-1}^2 = \sigma_{2i-1}^2(\sum_{k=1}^{K} c_{k,i})^2 + \sum_{k=1}^{K}\delta_{k,2i-1}^2 c_{k,i}^2 + \frac{\delta_0^2}{2}(\mathbf{A}_{\Re,i}\mathbf{A}_{\Re,i}^T + \mathbf{A}_{\Im,i}\mathbf{A}_{\Im,i}^T), \\ \hat{\sigma}_{2i}^2 = \sigma_{2i}^2(\sum_{k=1}^{K} c_{k,i})^2 + \sum_{k=1}^{K}\delta_{k,2i}^2 c_{k,i}^2 + \frac{\delta_0^2}{2}(\mathbf{A}_{\Re,i}\mathbf{A}_{\Re,i}^T + \mathbf{A}_{\Im,i}\mathbf{A}_{\Im,i}^T). \end{cases} \tag{33}$$

Proof: Take subscript $2i$ as example, as the proof of the subscript $2i-1$ is similar.

The mean $\hat{\mu}_{l,2i}$ is obtained as follow

$$\begin{aligned}\hat{\mu}_{l,2i} &= \mathbb{E}_l\left[\sum_{k=1}^{K} c_{k,i}\tilde{x}_{k,2i} + \Im(a_i)\right] = \mathbb{E}_l\left[\sum_{k=1}^{K} c_{k,i}x_{2i} + \sum_{k=1}^{K} c_{k,i}d_{k,2i} + \Im(a_i)\right] \\ &= \sum_{k=1}^{k} \mu_{l,2i}c_{k,i} + \mathbb{E}\left[\Im(\mathbf{A}_i\mathbf{n})\right], \\ &= \sum_{k=1}^{k} \mu_{l,2i}c_{k,i} + \mathbb{E}\left[\mathbf{A}_{\Im,i}\mathbf{n}_{\Re} + \mathbf{A}_{\Re,i}\mathbf{n}_{\Im}\right], \\ &= \sum_{k=1}^{k} \mu_{l,2i}c_{k,i}.\end{aligned}$$

The variance $\hat{\sigma}_{2i}^2$ is obtained as follow

$$\begin{aligned}\hat{\sigma}_{2i}^2 &= \mathbb{E}_l\left[(\sum_{k=1}^{K} c_{k,i}\tilde{x}_{k,i} + \Im(a_i))^2\right] - \mathbb{E}_l\left[\sum_{k=1}^{K} c_{k,i}\tilde{x}_{k,i} + \Im(a_i)\right]^2, \\ &= \mathbb{E}_l\left[(x_{2i}\sum_{k=1}^{K} c_{k,i} + \sum_{k=1}^{K} c_{k,i}d_{k,2i} + \mathbf{A}_{\Im,i}\mathbf{n}_{\Re} + \mathbf{A}_{\Re,i}\mathbf{n}_{\Im})^2\right] - (\sum_{k=1}^{k} \mu_{l,2i}c_{k,i})^2, \\ &= \mathbb{E}_l\left[(x_{2i}\sum_{k=1}^{K} c_{k,i})^2\right] + \mathbb{E}_l\left[(\sum_{k=1}^{K} c_{k,i}d_{k,2i})^2\right] + \mathbb{E}_l\left[(\mathbf{A}_{\Im,i}\mathbf{n}_{\Re})^2\right] + \mathbb{E}_l\left[(\mathbf{A}_{\Re,i}\mathbf{n}_{\Im})^2\right] - (\sum_{k=1}^{k} \mu_{l,2i}c_{k,i})^2, \\ &= (\sigma_{2i}^2 + \mu_{l,2i}^2)(\sum_{k=1}^{K} c_{k,i})^2 + \sum_{k=1}^{K} \delta_{k,2i}^2 c_{k,i}^2 + \frac{\delta_0^2}{2}(\mathbf{A}_{\Re,i}\mathbf{A}_{\Re,i}^T + \mathbf{A}_{\Im,i}\mathbf{A}_{\Im,i}^T) - (\sum_{k=1}^{k} \mu_{l,2i}c_{k,i})^2, \\ &= \sigma_{2i}^2(\sum_{k=1}^{K} c_{k,i})^2 + \sum_{k=1}^{K} \delta_{k,2i}^2 c_{k,i}^2 + \frac{\delta_0^2}{2}(\mathbf{A}_{\Re,i}\mathbf{A}_{\Re,i}^T + \mathbf{A}_{\Im,i}\mathbf{A}_{\Im,i}^T).\end{aligned}$$

This ends the proof. □

Thus, for any element pair, say, $(\hat{x}_{2i-1}, \hat{x}_{2i})$, the discriminant gain can be derived as

$$\begin{aligned}G(\hat{x}_{2i-1}) &= \frac{2}{L(L-1)}\sum_{l'=1}^{L}\sum_{l<l'}\frac{(\hat{\mu}_{l,2i-1} - \hat{\mu}_{l',2i-1})^2}{\hat{\sigma}_{2i-1}^2}, \\ G(\hat{x}_{2i}) &= \frac{2}{L(L-1)}\sum_{l'=1}^{L}\sum_{l<l'}\frac{(\hat{\mu}_{l,2i} - \hat{\mu}_{l',2i})^2}{\hat{\sigma}_{2i}^2}.\end{aligned} \tag{34}$$

As the total discriminant gain is the sum of the discriminant gain of every single element, our prroblem can be derived as

$$\text{P2} \quad \begin{aligned} \max_{\mathbf{A},\{\mathbf{C}_k\}} \ & G = \frac{2}{L(L-1)} \sum_{i=1}^{N_t} (G(\hat{x}_{2i-1}) + G(\hat{x}_{2i})), \\ \text{s.t.} \ & \text{tr}(\mathbf{C}_k^2 (\mathbf{A}\mathbf{H}_k)^{\dagger} (\mathbf{A}\mathbf{H}_k)^{\dagger H}) \leq P_k, \ \forall k, \\ & c_{k,i} \geq 0, \ \forall i, k. \end{aligned} \tag{35}$$

V. Problem Solution

In this section, the Problem P2 is solved. The semidefininte relaxation (SDR) is first used, and then Successive Convex Approximation (SCA) is used to solved the final problem.

A. SDR Simplification

Assume that $\mathbf{A}\mathbf{H}_k$ is full rank matrixs, i.e., $(\mathbf{A}\mathbf{H}_k)^{\dagger} = (\mathbf{A}\mathbf{H}_k)^{-1}$. Then, the first constrain becomes

$$\text{tr}(\mathbf{C}_k^2 (\mathbf{H}_k^H \mathbf{A}^H \mathbf{A} \mathbf{H}_k)^{-1}) \leq P_k, \ \forall k. \tag{36}$$

Denote $\tilde{\mathbf{A}}_{\Re,i}, \tilde{\mathbf{A}}_{\Im,i}$ to be the transposation of $\mathbf{A}_{\Re,i}, \mathbf{A}_{\Im,i}$. Thus, we have

$$\begin{aligned} \mathbf{A}^H \mathbf{A} &= \mathbf{A}_{\Re}^H \mathbf{A}_{\Re} + \mathbf{A}_{\Im}^H \mathbf{A}_{\Im}, \\ &= \sum_{i=1}^{N_t} (\tilde{\mathbf{A}}_{\Re,i} \tilde{\mathbf{A}}_{\Re,i}^T - \tilde{\mathbf{A}}_{\Im,i} \tilde{\mathbf{A}}_{\Im,i}^T). \end{aligned} \tag{37}$$

Let $\mathbf{F}_{\Re,i} = \tilde{\mathbf{A}}_{\Re,i} \tilde{\mathbf{A}}_{\Re,i}^T$ and $\mathbf{F}_{\Im,i} = \tilde{\mathbf{A}}_{\Im,i} \tilde{\mathbf{A}}_{\Im,i}^T$. Thus, $\{\mathbf{F}_{\Re,i}\}, \{\mathbf{F}_{\Im,i}\}$ are all symmetric matrixs, and problem P2 becomes

$$\text{P3} \quad \begin{aligned} \max_{\{\mathbf{F}_{\Re,i}\},\{\mathbf{F}_{\Im,i}\},\{\mathbf{C}_k\}} \ & G = \frac{2}{L(L-1)} \sum_{i=1}^{N_t} (G(\hat{x}_{2i-1}) + G(\hat{x}_{2i})), \\ \text{s.t.} \ & \text{tr}(\mathbf{C}_k^2 (\mathbf{H}_k^H (\sum_{i=1}^{N_t} \mathbf{F}_{\Re,i} + \mathbf{F}_{\Im,i}) \mathbf{H}_k)^{-1}) \leq P_k, \ \forall k, \\ & c_{k,i} \geq 0, \ \forall i, k, \\ & \text{tr}(\mathbf{F}_{\Re,i}) \geq 0, \ \forall i \\ & \text{tr}(\mathbf{F}_{\Im,i}) \geq 0, \ \forall i \\ & \text{Rank}(\mathbf{F}_{\Re,i}) = \text{Rank}(\mathbf{F}_{\Im,i}) = 1. \end{aligned} \tag{38}$$

As the objective function is the sum of following term

$$\frac{(\hat{\mu}_{l,2i-1}-\hat{\mu}_{l',2i-1})^2}{\hat{\sigma}^2_{2i-1}}+\frac{(\hat{\mu}_{l,2i}-\hat{\mu}_{l',2i})^2}{\hat{\sigma}^2_{2i}} \tag{39}$$

We only need to analyze one of the two terms. Take the term with subscript $2i$ as example, it can be expanded as follow

$$\frac{(\hat{\mu}_{l,2i}-\hat{\mu}_{l',2i})^2}{\hat{\sigma}^2_{2i}}=\frac{(\mu_{l,2i}-\mu_{l',2i})^2(\sum_{k=1}^K c_{k,i})^2}{\sigma^2_{2i}(\sum_{k=1}^K c_{k,i})^2+\sum_{k=1}^K \delta^2_{k,2i}c^2_{k,i}+\frac{\delta_0^2}{2}(\text{tr}(\mathbf{F}_{\Re,i})+\text{tr}(\mathbf{F}_{\Im,i}))} \tag{40}$$

Using SDR, the low rank constrains are simply dropped. To further simplify the problem, a series of auxiliary variable $\{\alpha_{l,l',i}\}$ is introduced such that the following equations holds.

$$\begin{aligned}\alpha_{l,l',2i-1}&=\frac{(\mu_{l,2i-1}-\mu_{l',2i-1})^2(\sum_{k=1}^K c_{k,i})^2}{\sigma^2_{2i-1}(\sum_{k=1}^K c_{k,i})^2+\sum_{k=1}^K \delta^2_{k,2i-1}c^2_{k,i}+\frac{\delta_0^2}{2}(\text{tr}(\mathbf{F}_{\Re,i})+\text{tr}(\mathbf{F}_{\Im,i}))},\\ \alpha_{l,l',2i}&=\frac{(\mu_{l,2i}-\mu_{l',2i})^2(\sum_{k=1}^K c_{k,i})^2}{\sigma^2_{2i}(\sum_{k=1}^K c_{k,i})^2+\sum_{k=1}^K \delta^2_{k,2i}c^2_{k,i}+\frac{\delta_0^2}{2}(\text{tr}(\mathbf{F}_{\Re,i})+\text{tr}(\mathbf{F}_{\Im,i}))},\end{aligned} \quad \forall l,l',i. \tag{41}$$

Thus our problem becomes

$$\text{P4}\quad \begin{aligned}\max_{\{\mathbf{F}_{\Re,i}\},\{\mathbf{F}_{\Im,i}\},\{\mathbf{C}_k\},\{\alpha_{l,l',i}\}}\ &\frac{2}{L(L-1)}\sum_{l'=1}^{L}\sum_{l<l'}\sum_{i=1}^{N_t}\alpha_{l,l',2i-1}+\alpha_{l,l',2i},\\ \text{s.t.}\ \ &\text{tr}(\mathbf{C}_k^2(\mathbf{H}_k^H(\sum_{i=1}^{N_t}\mathbf{F}_{\Re,i}+\mathbf{F}_{\Im,i})\mathbf{H}_k)^{-1})\le P_k,\ \forall k,\\ &c_{k,i}\ge 0,\ \forall i,k,\\ &\text{tr}(\mathbf{F}_{\Re,i})\ge 0,\ \forall i\\ &\text{tr}(\mathbf{F}_{\Im,i})\ge 0,\ \forall i\\ &(\alpha_{l,l',2i-1}\sigma^2_{2i-1}-\beta_{l,l',2i-1})(\sum_{k=1}^K c_{k,i})^2+\alpha_{l,l',2i-1}(\sum_{k=1}^K \delta^2_{k,2i-1}c^2_{k,i}+\frac{\delta_0^2}{2}(\text{tr}(\mathbf{F}_{\Re,i})+\text{tr}(\mathbf{F}_{\Im,i})))=0,\ \forall l,l',i\\ &(\alpha_{l,l',2i}\sigma^2_{2i}-\beta_{l,l',2i})(\sum_{k=1}^K c_{k,i})^2+\alpha_{l,l',2i}(\sum_{k=1}^K \delta^2_{k,2i}c^2_{k,i}+\frac{\delta_0^2}{2}(\text{tr}(\mathbf{F}_{\Re,i})+\text{tr}(\mathbf{F}_{\Im,i})))=0,\ \forall l,l',i,\end{aligned} \tag{42}$$

where $\beta_{l,l',2i-1}=(\mu_{l,2i-1}-\mu_{l',2i-1})^2$ and the same with subscript $2i$.

With Karush–Kuhn–Tucker (KKT) conditions, the following lemma can be derived

Lemma 2. *Problem P4 is equivalent to the following problem*

$$
\begin{aligned}
\min_{\{\mathbf{F}_{\Re,i}\},\{\mathbf{F}_{\Im,i}\},\{\mathbf{C}_k\},\{\alpha_{l,l',i}\}} \quad & -\frac{2}{L(L-1)}\sum_{l'=1}^{L}\sum_{l<l'}\sum_{i=1}^{N_t}(\alpha_{l,l',2i-1}+\alpha_{l,l',2i}),\\
\textit{s.t.} \quad & \textit{tr}(\mathbf{C}_k^2(\mathbf{H}_k^H(\sum_{i=1}^{N_t}\mathbf{F}_{\Re,i}+\mathbf{F}_{\Im,i})\mathbf{H}_k)^{-1}) \le P_k,\ \forall k,\\
& c_{k,i}\ge 0,\ \forall i,k,\\
\textit{P5} \quad & \textit{tr}(\mathbf{F}_{\Re,i})\ge 0,\ \forall i\\
& \textit{tr}(\mathbf{F}_{\Im,i})\ge 0,\ \forall i\\
& (\alpha_{l,l',2i-1}\sigma_{2i-1}^2-\beta_{l,l',2i-1})(\sum_{k=1}^{K}c_{k,i})^2+\alpha_{l,l',2i-1}(\sum_{k=1}^{K}\delta_{k,2i-1}^2c_{k,i}^2+\frac{\delta_0^2}{2}(\textit{tr}(\mathbf{F}_{\Re,i})+\textit{tr}(\mathbf{F}_{\Im,i})))\le 0,\ \forall l,l',i\\
& (\alpha_{l,l',2i}\sigma_{2i}^2-\beta_{l,l',2i})(\sum_{k=1}^{K}c_{k,i})^2+\alpha_{l,l',2i}(\sum_{k=1}^{K}\delta_{k,2i}^2c_{k,i}^2+\frac{\delta_0^2}{2}(\textit{tr}(\mathbf{F}_{\Re,i})+\textit{tr}(\mathbf{F}_{\Im,i})))\le 0,\ \forall l,l',i.
\end{aligned}
\tag{43}
$$

Proof: It's obvious that problem P4 is euivalent to the folloing problem

$$
\begin{aligned}
\min_{\{\mathbf{F}_{\Re,i}\},\{\mathbf{F}_{\Im,i}\},\{\mathbf{C}_k\},\{\alpha_{l,l',i}\}} \quad & -\frac{2}{L(L-1)}\sum_{l'=1}^{L}\sum_{l<l'}\sum_{i=1}^{N_t}(\alpha_{l,l',2i-1}+\alpha_{l,l',2i}),\\
\text{s.t.} \quad & \text{tr}(\mathbf{C}_k^2(\mathbf{H}_k^H(\sum_{i=1}^{N_t}\mathbf{F}_{\Re,i}+\mathbf{F}_{\Im,i})\mathbf{H}_k)^{-1}) \le P_k,\ \forall k,\\
& c_{k,i}\ge 0,\ \forall i,k,\\
& \text{tr}(\mathbf{F}_{\Re,i})\ge 0,\ \forall i\\
& \text{tr}(\mathbf{F}_{\Im,i})\ge 0,\ \forall i\\
& (\alpha_{l,l',2i-1}\sigma_{2i-1}^2-\beta_{l,l',2i-1})(\sum_{k=1}^{K}c_{k,i})^2+\alpha_{l,l',2i-1}(\sum_{k=1}^{K}\delta_{k,2i-1}^2c_{k,i}^2+\frac{\delta_0^2}{2}(\text{tr}(\mathbf{F}_{\Re,i})+\text{tr}(\mathbf{F}_{\Im,i})))=0,\ \forall l,l',i\\
& (\alpha_{l,l',2i}\sigma_{2i}^2-\beta_{l,l',2i})(\sum_{k=1}^{K}c_{k,i})^2+\alpha_{l,l',2i}(\sum_{k=1}^{K}\delta_{k,2i}^2c_{k,i}^2+\frac{\delta_0^2}{2}(\text{tr}(\mathbf{F}_{\Re,i})+\text{tr}(\mathbf{F}_{\Im,i})))=0,\ \forall l,l',i.
\end{aligned}
\tag{44}
$$

Take subscript $2i$ as example, for any $l,l',2i$, let $\tilde{\alpha}_{l,l',2i}$ be the optimal variable of problem P5. The gradient of the Lagarange function of P5 respecte to $\tilde{\alpha}_{l,l',2i}$ needs to be 0, i.e.,

$$
-\frac{2}{L(L-1)}+\lambda_{l,l',2i}(\sigma_{2i-1}^2(\sum_{k=1}^{K}c_{k,i})^2+\sum_{k=1}^{K}\delta_{k,2i}^2c_{k,i}^2+\frac{\delta_0^2}{2}(\text{tr}(\mathbf{F}_{\Re,i})+\text{tr}(\mathbf{F}_{\Im,i})))=0, \tag{45}
$$

where $\lambda_{l,l',2i} \geq 0$ is the Lagrange multiplier of coresponding constrant of P5, i.e.,

$$(\alpha_{l,l',2i}\sigma_{2i}^2 - \beta_{l,l',2i})(\sum_{k=1}^{K} c_{k,i})^2 + \alpha_{l,l',2i}(\sum_{k=1}^{K} \delta_{k,2i}^2 c_{k,i}^2 + \frac{\delta_0^2}{2}(\text{tr}(\mathbf{F}_{\Re,i}) + \text{tr}(\mathbf{F}_{\Im,i}))) \leq 0. \tag{46}$$

As $-\frac{2}{L(L-1)} < 0$, if $\lambda_{l,l',2i} = 0$, then (39) is invalid. Thus, $\lambda_{l,l',2i}$ must be greater than 0, which means the inequality (40) must equal to 0, due to the complementary slackness. Therefore problem P5 is equavalent to (38), and naturally equavalent to P4.

This ends the proof. □

Howerver, problem P5 is still a non-convex problem, due to the first and the last two constraints. Hence, SCA is used to .convert these constraints into convex sets.

Lemma 3. *With SCA, problem P5 can be convert to the following convex problem.*

$$\begin{aligned}
\min_{\{\mathbf{F}_{\Re,i}\},\{\mathbf{F}_{\Im,i}\},\{\mathbf{C}_k\},\{\alpha_{l,l',i}\},\{\mathbf{Y}_k\}} \; & -\frac{2}{L(L-1)} \sum_{l'=1}^{L} \sum_{l<l'} \sum_{i=1}^{N_t} (\alpha_{l,l',2i-1} + \alpha_{l,l',2i}), \\
\textit{s.t.} \; & \textit{tr}(\mathbf{C}_k^2) + \textit{tr}(\mathbf{Y}_k^{-1}) \leq \sqrt{P_k}, \; \forall k, \\
& \mathbf{H}_k^H (\sum_{i=1}^{N_t} \mathbf{F}_{\Re,i} + \mathbf{F}_{\Im,i}) \mathbf{H}_k = \mathbf{Y}_k, \; \forall k, \\
& c_{k,i} \geq 0, \; \forall i, k, \\
& \textit{tr}(\mathbf{F}_{\Re,i}) \geq 0, \; \forall i \\
& \textit{tr}(\mathbf{F}_{\Im,i}) \geq 0, \; \forall i \\
& \sigma_{2i-1}^2 (\sum_{k=1}^{K} c_{k,i})^2 + \sum_{k=1}^{K} \delta_{k,2i-1}^2 c_{k,i}^2 + \frac{\delta_0^2}{2}(\textit{tr}(\mathbf{F}_{\Re,i}) + \textit{tr}(\mathbf{F}_{\Im,i})) \leq \\
& \beta_{l,l',2i-1} \frac{(\sum_{k=1}^{K} c_{k,i}^{(t)})^2}{\alpha_{l,l',2i-1}^{(t)}} + \sum_{k=1}^{K} \frac{2(\sum_{k=1}^{K} c_{k,i}^{(t)})}{\alpha_{l,l',2i-1}^{(t)}} (c_{k,i} - c_{k,i}^{(t)}) \\
& - \frac{(\sum_{k=1}^{K} c_{k,i}^{(t)})^2}{\alpha_{l,l',2i-1}^{(t)2}} (\alpha_{l,l',2i-1} - \alpha_{l,l',2i-1}^{(t)}), \; \forall l, l', i \\
& \sigma_{2i}^2 (\sum_{k=1}^{K} c_{k,i})^2 + \sum_{k=1}^{K} \delta_{k,2i}^2 c_{k,i}^2 + \frac{\delta_0^2}{2}(\textit{tr}(\mathbf{F}_{\Re,i}) + \textit{tr}(\mathbf{F}_{\Im,i})) \leq \\
& \beta_{l,l',2i} \frac{(\sum_{k=1}^{K} c_{k,i}^{(t)})^2}{\alpha_{l,l',2i}^{(t)}} + \sum_{k=1}^{K} \frac{2(\sum_{k=1}^{K} c_{k,i}^{(t)})}{\alpha_{l,l',2i}^{(t)}} (c_{k,i} - c_{k,i}^{(t)}) \\
& - \frac{(\sum_{k=1}^{K} c_{k,i}^{(t)})^2}{\alpha_{l,l',2i}^{(t)2}} (\alpha_{l,l',2i} - \alpha_{l,l',2i}^{(t)}), \; \forall l, l', i
\end{aligned} \tag{47}$$

Proof: Here, we only pay attention to the last two constraints, as the rest of the problem is the same. And take subscript $2i$ as an example.

Notice that the last constraints of P5 can be rewrited as

$$\sigma_{2i}^2(\sum_{k=1}^{K} c_{k,i})^2 + \sum_{k=1}^{K} \delta_{k,2i}^2 c_{k,i}^2 + \frac{\delta_0^2}{2}(\text{tr}(\mathbf{F}_{\Re,i}) + \text{tr}(\mathbf{F}_{\Im,i})) \leq \beta_{l,l',2i} \frac{(\sum_{k=1}^{K} c_{k,i})^2}{\alpha_{l,l',2i}}.\ \forall l, l', i$$

Both of the two sides of this inequation are convex functions. Thus, using the first order Taylor expension of the right side instead reults in a convex constraint.

$$\sigma_{2i}^2(\sum_{k=1}^{K} c_{k,i})^2 + \sum_{k=1}^{K} \delta_{k,2i}^2 c_{k,i}^2 + \frac{\delta_0^2}{2}(\text{tr}(\mathbf{F}_{\Re,i}) + \text{tr}(\mathbf{F}_{\Im,i})) \leq$$

$$\beta_{l,l',2i} \frac{(\sum_{k=1}^{K} c_{k,i}^{(t)})^2}{\alpha_{l,l',2i}^{(t)}} + \sum_{k=1}^{K} \frac{2(\sum_{k=1}^{K} c_{k,i}^{(t)})}{\alpha_{l,l',2i}^{(t)}}(c_{k,i} - c_{k,i}^{(t)}) - \frac{(\sum_{k=1}^{K} c_{k,i}^{(t)})^2}{\alpha_{l,l',2i}^{(t)2}}(\alpha_{l,l',2i} - \alpha_{l,l',2i}^{(t)}),\ \forall l, l', i$$

This ends the proof. □

VI. Simulation Result

In this section, some theoretical analysis is shown, and several simulation results are demonstrated.

A. Simulation Setting

1) Dataset: The wireless sensing simulator in [xx] is used to simulate various high-fidelity human motions and generate the dataset. Four classes data is generated, including child walking, child pacing, adult walking and adult pacing. Our setup is similar to [xx], the heights of children are assumed to be uniformly distributed in $[0.9, 1.2]$ meters, while the interval of the heights of adults is assumed to be $[1.6, 1.9]$ meters. The spped of walking and pacing are assumed to be $0.5H$ and $0.25H$ m/s, where H is the height. The heading of the moving human is set to be uniformly distributed in $[-180°, 180°]$.

2) Data Distribution: We assume that our system has 3 sensors. Recall that every sensor has its own distortion, i.e., $\tilde{\mathbf{d}}_k \in \mathcal{N}(\mathbf{0}, \mathbf{D}_k)$, where $\mathbf{D}_k$ is the diagonal covariance matrix with $\mathbf{D}_k = \text{diag}\left\{\delta_{k,1}^2, \delta_{k,2}^2, \ldots, \delta_{k,M}^2\right\}$. We set $\delta_{k,i}^2, \forall k, i$ uniformly distributed in $[0, 10]$, and use the above distribution to generate the distortion of each sensor.

3) Classifier: Here, Surpport Vector Machine (SVM) and Multiple Layer Perceptron (MLP) are used to be the classifiers. A python package named *sklearn* [xx] is used to implement SVM and MLP. Spacifically, C-Support Vector Classification is used, with default settings of sklearn package. Besides, an MLP with hidden layer size of 24 is used, and the rest of the parameters is default settings of sklearn package.

4) AirComp: A tree-dimensional setting is used. The server is set to be located at (0, 0, 20), while all sensors are uniformly located within a circular region centered at (120, 20, 0) with radius 20 meters. Compute the distance dependent large-scale fading as $T_0(d/d_0)^{-\alpha}$, where T_0 is the path loss at the reference distance $d_0 = 1$ meter, d denotes the distance between the transmitter and receiver, and α is the path loss exponent. Besides, the small-scale fading is set as Rician fading with rician factor β. The final result is obtained by averaging the result over 500 channel realization. We set $\alpha = 3$, $T_0 = -30dB$, $\beta = 3$. For simplification, all power constraints is set to be the same, i.e., $P_k = P_0, \forall k$, where P_0 is the maximum transmit power of all sensors. The variance of channel noise is set to be $\delta_0^2 = -110dBm$.

B. Relationship between Discriminant Gain and Accuracy

Recall that, with the discriminant gain being maximized, the centre of each class remains unchanged after post-processing, every calss becomes more compact, i.e., the variance of each class is smaller. Meanwhile, it's can be shown that the reduction of discriminant gain mainly comes from the distortion and the channal noise. Thus, it's obviously that when the discriminant gain becomes larger, the impact of the distortion and channel noise will be smaller, which leads to the increasing of the accuracy of the classifier which has been trained based on the fine-tuned data, i.e., the ground-true data.

To verify the above theory, we set the number of antennas to be 2, and tune the range of convariance matrix of the distortion $\mathbf{D}_k$ to illustrate the relationship between the inference accuracy and discriminant gain. The result is shown as follow.

[Picture ACC vs DG]

The result shows that the accuracy is positive correlated to the discriminant gain, as we have mentioned above. The accuracy of MLP increases faster than the SVM when the discriminant gain is small, and the SVM finnally outperform the MLP before convergence. This is because when the discriminant is large, i.e., the distortion is small, the MLP is so complicated that it may overfit the taining data, i.e., the ground-true data. When the distortion is small, the data

is similar to the ground-true data, which leads to a lower accuracy because of overfitting. And when the distortion is large, the MLP is more robust than the SVM.

C. Accuracy with Different Numbers of Antenna

We also investigate the relationship between the accuracy and the number of antennas, result is shown as follow.

[picture]

To better illustrate the benifit of our algorithm, we compare our work with Guangxu Zhu [xxx] and benchmark. The benchmark is set to be randomly pick a point in the feasible set of problem P2. It can be seen that our work outperform the rest two lines.

D. Accuracy with Different Tranmission Power

To better understand our algorithm, we show that how the accuracy change when the transmission power change. We also compare our algorithm with above two methods, the result is shown as follow. Our work still ourperform the two methods, and the transmission power becomes larger, the algorithm becames more accurate.

[picturre]

VII. Conclusion