Joint Node Activation, Beamforming and Phase-Shifting Control In IoT Sensor Network Assisted by Reconfigurable Intelligent Surface

Yang Liu, Qingjiang Shi, Qingqing Wu, Jun Zhao, and Ming Li*

***Abstract*—Power saving and battery-life extension have always been a critical concern for IoT network deployment. One effective solution is to switch wireless devices into sleep mode to save power. This paper considers the power control in an IoT network via jointly activating IoT sensors and designing their transmit beamforming. Besides, inspired by the great potential of reconfigurable intelligent surface (RIS) in energy saving, we additionally introduce RIS to further lower the sensors' power consumption. The considered problem is highly challenging due to its combinatorial nature, the highly non-convex quality-of-service (QoS) constraint and the hardware restrictions from the RIS. By exploiting the cutting-the-edge majorization minimization (MM) and the penalty dual decomposition (PDD) frameworks, we have successfully developed highly efficient solutions to tackle this problem. Our proposed solutions can achieve nearly identical performance with that of the exhaustive search but with a much lower complexity. Besides, as revealed by the numerical experiments, our proposed sensor activation scheme can switch off a large portion of sensors under mild QoS requirements, which significantly reduces power expenditure. Moreover, the deployment of RIS can bring an additional $45\% - 70\%$ power saving compared to the no-RIS case.**

***Index terms*— IoT sensor network, reconfigurable intelligent surface (RIS), group sparsity, decentralized algorithm.**

I. Introduction

A. Background

The widespread usage of Internet of Things (IoT) network has become a predominant driving force for the development of 6G technology [1]. Recently, a great number of novel communication protocols have been proposed and are being constructed to embrace the forthcoming vast deployment of IoT networks in extremely diverse applications, including the Bluetooth Low Energy (BLE) [2], Zigbee [3], Z-Wave [4], Low-Power Wi-Fi (also called IEEE 802.11ah) [5], LoRa [6], SigFox [7] and so on.

Beamforming has been shown to be capable of improving system performance significantly for multi-antenna communication networks. One most seminal model of applying beamforming in IoT networks has been introduced in [8], where a center node coordinates multiple sensors to perform data fusion via collaboratively optimizing beamformers. The star-topology deployment of sensors in [8] is explicitly supported by most of the mainstream protocols, including Bluetooth Classic/LE, Zigbee, Z-Wave, Wi-Fi, LoRa and so on. Based on the model established in [8], extensive research has been conducted to study beamforming designs in various settings [9]–[23]. [9] have successfully developed efficient beamforming methods towards optimal linear estimation and power control. Within the same setting, precoding optimization for signal tracking is considered in [10] and [11]. The communication throughput is optimized in [12]. Robust transceiver design counter to channel estimation error is studied in [13]. Recently, IoT sensor networks with new exciting features have also attracted great attentions. For instance, beamforming design of wireless sensors capable of wireless power transferring (WPT) are investigated in [14] and [15]. Sensor networks with a fusion center equipped with large-scale antennas is considered in [16], [17]. Joint precoding design to perform estimation and detection of sparse signals are researched in [18] and [19], respectively. Collaborative beamforming design for data compression-transmission-estimation is proposed in [20]. Besides, transceiver design for a newly emerging paradigm that simultaneously computes and transmits data (namely computation over the air) is investigated in [21]–[23].

Although IoT technology has a wide range of applications, it is still facing a number of difficulties [24], among which, devices' power consumption and battery-life are most concerned. In reality, it is usually inconvenient or even infeasible to replace the IoT devices' batteries. Nowadays, various standards are making research to improve the IoT devices' power saving. For instance, as an evolution of Bluetooth Classic, BLE reduces power consumption via putting devices into sleep for more fractions of time [2]. Similarly, ZigBee manages to maintain as many nodes as possible working in the reduced-function device (RFD) mode, which can greatly lower the power expenditure compared to the full-function device (FFD) mode [3]. Besides, the rising IEEE 802.15.4-based protocols (e.g. 802.15.4k and 802.15.4e [25]) and the emerging

Y. Liu and M. Li are with the School of Information and Communication Engineering, Dalian University of Technology, Dalian, China, email: yangliu_613@dlut.edu.cn, mli@dlut.edu.cn. M. Li's is supported by Grant No. 61971088, Grant No. DUT20GJ214, Grant No. 2020-MS108, and Grant No. 2021D08. Y. Liu is supported by Grant No. DUT20RC(3)029.

Q. Shi is with the School of Software Engineering, Tongji University, Shanghai, China, and also with the Shenzhen Research Institute of Big Data, Shenzhen, China email: shiqj@tongji.edu.cn.

Q. Wu is with Department of Electrical and Computer Engineering, University of Macau, Macau, China, email: qingqingwu@um.edu.mo. Q. Wu is supported by Grant No. 0119/2020/A3, Grant No. 2021A1515011900 and Grant No. 2020B1212030003.

J. Zhao is with the School of Computer Science and Engineering, Nanyang Technological University, Singapore, email: junzhao@ntu.edu.sg. J. Zhao is supported by Nanyang Technological University Startup Grant, NTU-WASP Joint Project, and Singapore Ministry of Education Academic Research Fund Tier 1 RG115/19, RG97/20, RG24/20, Tier 2 MOE2019-T2-1-176.

*M. Li is the corresponding author.

TABLE I: CC2652P [29] Power Consump. ($\mathsf{V}_{\mathsf{DCC}}=3.3$V)

Work Mode	I_{wk} (mA)	P_{wk} (mW)	$\mathsf{P}_{\mathsf{TRP}}^{\max}$ (mW)
TX 0dBm	7.3	24	1
TX 5dBm	9.6	31.7	3.2
TX 10dBm	22	72.6	10
TX 20dBm	85	280.1	100
Standby	$0.94e{-}3$	$3.1e{-}3$	—

Low-Power Wi-Fi [5] (an energy efficient amendment of the classical Wi-Fi) all focus on improving power efficiency to well support the machine-to-machine (M2M) type communications. Their most effective solution is the dynamic power management (DPM) strategy [26]. Its key idea is simple—just deactivate as many devices as possible while maintaining the performance of IoT networks. For instance, [27] and [28] propose to save power by activating only a subset of the sensors while maintaining the network's connectivity/sensing capability. As a matter of fact, the recently thriving standards, e.g. BLE, ZigBee, Low-Power Wi-Fi and IEEE 802.15.4e, have extensively exploited this technique via reducing duty cycles and/or switching off redundant sensor nodes.

B. Motivation

Although several existing works have shown that beamforming can effectively lower sensors' power consumption, e.g. [9], [12], [16], [21], [22], all these works restrict onto the pure transmit radio power (TRP) and hence their power model is incomplete. In fact, IoT devices' non-transmit power consumption, which comes from signal processing, radio frequency (RF) chains, hardware peripherals, can be much larger that the TRP amount. To clarify this fact, we take the IoT development tool—CC2652P device [29] by Texas Instruments as an example. The device CC2652P is powered with working voltage $\mathsf{V}_{\mathsf{DCC}}=3.3$V and can operate in different TX modes, whose working current (I_{wk}), operating power consumption (P_{wk}) and maximal supporting TRP ($\mathsf{P}_{\mathsf{TRP}}^{\max}$) are specified in Table I [29]. In Table I, the working power consumption P_{wk} includes all the consumed energy coming from the micro controlling unit (MCU), RF chains (e.g. power amplifiers, DAC, mixers and phase-shifters), multiple SRAMs as well as hardware peripherals [29]. Table I clearly indicates two facts: i) the non-transmit power is generally much (3-20 times) larger than the TRP; ii) sleep (i.e. Standby) mode can significantly save power. Motived by these observations and inspired by the aforementioned DPM technique [26]–[28], we propose to reduce the total power consumption via sending as many sensors into sleep as possible when designing beamformers. This idea naturally leads to the sensor activation problem.

Besides, to further boost power saving, we propose deploying reconfigurable intelligent surface (RIS) in the IoT sensing network. Very recently, RIS has been cast with great attentions [30], [31]. It is a planar surface comprising an array of reflecting elements, which are built on meta-materials and can reflect the impinging electromagnetic (EM) waveforms in a programmable manner. RIS has been shown to have great potentials in improving system's power saving performance [32]–[34]. For instance, [33] has shown that RIS can achieve an N^2 power scaling law (N is the number of reflecting elements), which exceeds the well known linear scaling law for the classical multi-antenna arrays. Authors in [32] have shown RIS can significantly improve the energy efficiency. Enlightened by the recent exciting findings about RIS [30]–[34], we propose to introduce the RIS to enhance the beamforming gain and deactivate more sensors. This idea has not yet been investigated in the line of wireless sensing research [8]–[22].

C. Contributions

This paper conducts a comprehensive research on the power control problem in a star-topology IoT sensing network studied in [8]–[23]. We propose to perform sensor node selection, i.e. activating the minimal number of sensors to save power. Besides, via deploying RIS device, we jointly design sensor selection, active sensors' beamforming and RIS configuration to minimize the sensors' power consumption. Specifically, the novelty and contribution include

(i) We are the first to take into account the non-transmit power consumption of IoT sensors, which is indeed much larger than TRP (see Table I) and has been ignored in previous literature [9], [12], [16], [21], [22]. This new consideration naturally gives rise to the sensor activation problem. This problem is very challenging since it is non-convex, non-smooth and has a combinatorial nature.

(ii) We propose to exploit RIS to deactivate more sensor nodes. This idea has not been studied in the research on wireless sensing [8]–[23]. The deployment of RIS also brings new challenge—the complicated mean square error (MSE) metric for high-dimensional signal estimation is very difficult to optimize with respect to RIS that is constrained by non-convex constant-modulus. This challenge cannot be solved by any existing methods in RIS relevant literature [30]–[34]. We have developed new consecutive MM procedures to tackle this difficulty and yield closed-form solution.

(iii) To solve the posed non-smooth and non-convex difficult problem, we jointly adopt the seminal sparsity inducing method [35]–[39] and the cutting-the-edge majorization minimization (MM) framework [40]–[42] to come up with an easily-updating and convergence-guaranteed iterative solution. Its sparse-structured output gives rise to efficient sensor activation schemes and can well compete the exhaustive-search-all-pattern solution, which has prohibitively high complexity.

(iv) Furthermore, via utilizing the state-of-the-art penalty dual decomposition (PDD) methodology [43], we proceed to develop a highly efficient solution which a) updates all variables thoroughly analytically and does not depend on any numerical solvers, e.g. CVX [44]; b) decentralizes the computations and significantly decreases the time complexity. This newly proposed closed-form based and decentralized solution is never seen in any relevant literature [8]–[23], [30]–[39].

(v) As a comprehensive research, we also provide algorithmic and theoretical solutions to the feasibility characterization imposed by the quality of service (QoS) requirement. We design an analytically updating and parallelizable algorithm to identify the feasibility of the proposed problem.

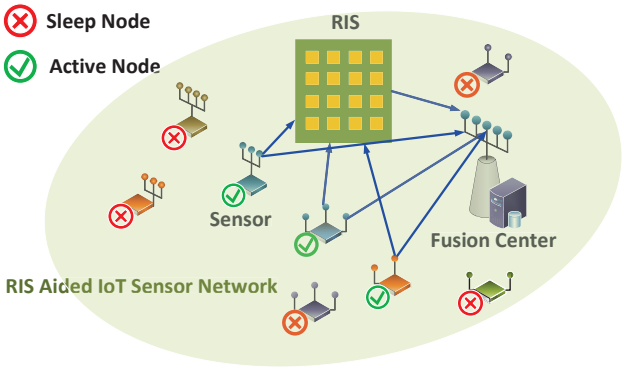

Fig. 1: Model of IoT Wireless Sensor Network Aided by RIS

Moreover, for the special but meaningful case of scalar source signal, our newly derived analytic MSE bound successfully generalizes the known results in the existing literature [9], [15].

(vi) Last but not least, extensive numerical results illustrate that significant benefit in power saving due to sensor activation and RIS configuration.

II. System Model and Problem Formulation

In this section we introduce the system model of the IoT sensor network as shown in Fig.1 and formally pose the joint sensor activation and beamforming design problem.

A. System and Signal Model

We consider the IoT sensor network in [8] composed of L wireless sensors and a fusion center (FC). We label the sensors from 1 to L and denote the entire set of sensors as $\mathcal{L} \triangleq \{1, \cdots, L\}$. The geometrically distributed sensors observe one common physical event, which can be modelled as an unknown parameter vector $\mathbf{x} \in \mathbb{C}^K$. Each sensor $\ell \in \mathcal{L}$ employs a predefined linear filter $\mathbf{A}_\ell \in \mathbb{C}^{K_\ell \times K}$ to process its observation locally [8]. Thus the observed signal at the ℓ-th sensor can be expressed as

$$\mathbf{y}_\ell = \mathbf{A}_\ell \mathbf{x} + \mathbf{n}_\ell, \quad \ell \in \mathcal{L}, \tag{1}$$

where $\mathbf{n}_\ell \in \mathbb{C}^{K_\ell}$ represents the observation noise. Without loss of generality, it is convenient to assume that $\mathbb{E}\{\mathbf{x}\} = \mathbf{0}$, $\mathbb{E}\{\mathbf{x}\mathbf{x}^H\} = \mathbf{I}_K$ and $\mathbf{n}_\ell$ has zero mean and covariance matrix $\mathbb{E}\{\mathbf{n}_\ell \mathbf{n}_\ell^H\} = \mathbf{\Sigma}_\ell$ with $\mathbf{\Sigma}_\ell \succ 0$. The sensors convey their observations to the FC, where the collected data is fused and processed. The ℓ-th sensor and the FC are equipped with N_ℓ and M antennas, respectively. All the sensors employ linear transceivers to manipulate signals. Denote the beamformer at the ℓ-th sensor as $\mathbf{B}_\ell \in \mathbb{C}^{N_\ell \times K_\ell}$ and all the sensors' beamformers $\mathcal{B}$ as $\mathcal{B} \triangleq \{\mathbf{B}_\ell\}_{\ell=1}^L$. Suppose $\mathbf{H}_\ell \in \mathbb{C}^{M \times N_\ell}$ is the wireless channel from the ℓ-th sensor to the FC.

Beyond the setting in [8], we utilize RIS device to improve performance [30], [31]. The RIS has J reflecting elements and each element's functional is modelled as a unit-modulus phase-shifter $\phi_j \triangleq e^{-\jmath\theta_j}$, with $\theta_j \in [0, 2\pi)$, $j \in \{1, \cdots, J\} \triangleq \mathcal{J}$. For ease of discussion, we will alternatively use $\boldsymbol{\phi} \triangleq [\phi_1, \cdots, \phi_J]^T$ and $\mathbf{\Phi} \triangleq \mathsf{Diag}\{\boldsymbol{\phi}\}$ to represent the phase-shift settings of the RIS. Suppose that the ℓ-th sensor-RIS and the RIS-FC channels are $\mathbf{G}_\ell \in \mathbb{C}^{N_\ell \times J}$ and $\mathbf{F} \in \mathbb{C}^{J \times M}$, respectively. Here the highly bandwidth-efficient coherent-sum multiple access channel (MAC) is considered [8]. Besides, as in [8], we assume that the channel state information (CSI) is available, which could be obtained via the state-of-the-art channel estimation techniques [45], [46].

The received signal $\mathbf{r}$ at the FC is expressed as

$$\begin{aligned} \mathbf{r} &= \sum\nolimits_{\ell=1}^L \Big(\mathbf{H}_\ell + \mathbf{F}\mathbf{\Phi}\mathbf{G}_\ell\Big)\mathbf{B}_\ell\big(\mathbf{A}_\ell\mathbf{x} + \mathbf{n}_\ell\big) + \mathbf{n}_0 \\ &= \Big(\sum\nolimits_{\ell=1}^L \widetilde{\mathbf{H}}_\ell(\mathbf{\Phi})\mathbf{B}_\ell\mathbf{A}_\ell\Big)\mathbf{x} + \mathbf{n} = \widehat{\mathbf{H}}(\mathcal{B}, \mathbf{\Phi})\mathbf{x} + \mathbf{n}. \end{aligned} \tag{2}$$

where $\mathbf{n}_0$ is the additive receive noise at the FC, $\mathbf{n}$ is the total noise given as $\mathbf{n} \triangleq \big(\sum_{\ell=1}^L \widetilde{\mathbf{H}}_\ell(\mathbf{\Phi})\mathbf{B}_\ell\mathbf{n}_\ell + \mathbf{n}_0\big)$, $\widetilde{\mathbf{H}}_\ell(\mathbf{\Phi}) \triangleq \big(\mathbf{H}_\ell + \mathbf{F}\mathbf{\Phi}\mathbf{G}_\ell\big)$ and $\widehat{\mathbf{H}}(\mathcal{B}, \mathbf{\Phi}) \triangleq \sum_{\ell=1}^L \widetilde{\mathbf{H}}_\ell(\mathbf{\Phi})\mathbf{B}_\ell\mathbf{A}_\ell$. Without loss of generality, we assume that $\mathbf{n}_0$ has zero mean and white covariance matrix, i.e. $\mathbb{E}\{\mathbf{n}_0\mathbf{n}_0^H\} = \sigma_0^2\mathbf{I}_M$. Besides, in view of the fact that the sensors and the FC are spatially distributed, it is reasonable to assume that $\mathbf{n}_\ell$'s are mutually uncorrelated. It is easily verified that the total noise $\mathbf{n}$ has zero mean and covariance matrix given by

$$\mathbf{\Sigma}_{\mathbf{n}}(\mathcal{B}, \mathbf{\Phi}) \triangleq \mathbb{E}\{\mathbf{n}\mathbf{n}^H\} = \sigma_0^2\mathbf{I}_M + \sum\nolimits_\ell \widetilde{\mathbf{H}}_\ell(\mathbf{\Phi})\mathbf{B}_\ell\mathbf{\Sigma}_\ell\mathbf{B}_\ell^H\widetilde{\mathbf{H}}_\ell(\mathbf{\Phi})^H.$$

The FC collects the noisy observations from all the sensors and processes the obtained signals with the linear filter $\mathbf{W}^H$ to obtain an estimate $\widehat{\mathbf{x}}$ of $\mathbf{x}$:

$$\widehat{\mathbf{x}} = \mathbf{W}^H\mathbf{r} = \mathbf{W}^H\big(\widehat{\mathbf{H}}(\mathcal{B}, \mathbf{\Phi})\mathbf{x} + \mathbf{n}\big). \tag{3}$$

A standard performance metric to evaluate the signal recovery quality is the MSE [47], which is defined as

$$\mathsf{MSE} \triangleq \mathsf{Tr}\Big\{\mathbb{E}\Big[\big(\mathbf{x} - \widehat{\mathbf{x}}\big)\big(\mathbf{x} - \widehat{\mathbf{x}}\big)^H\Big]\Big\}. \tag{4}$$

According to classical estimation theory [47], for given $\mathcal{B}$ and $\mathbf{\Phi}$, the optimal linear receiver $\mathbf{W}^\star = \big[\widehat{\mathbf{H}}(\mathcal{B}, \mathbf{\Phi})\widehat{\mathbf{H}}(\mathcal{B}, \mathbf{\Phi})^H + \mathbf{\Sigma}_{\mathbf{n}}(\mathcal{B}, \mathbf{\Phi})\big]^{-1}\widehat{\mathbf{H}}(\mathcal{B}, \mathbf{\Phi})$ yields the minimal MSE given as

$$\mathsf{MSE}(\mathcal{B}, \mathbf{\Phi}) = \mathsf{Tr}\Big\{\big[\mathbf{I}_K + \widehat{\mathbf{H}}(\mathcal{B}, \mathbf{\Phi})^H\mathbf{\Sigma}_{\mathbf{n}}(\mathcal{B}, \mathbf{\Phi})^{-1}\widehat{\mathbf{H}}(\mathcal{B}, \mathbf{\Phi})\big]^{-1}\Big\}. \tag{5}$$

B. Power Consumption Model

The sensor's consumed power comes from two parts—i) the transmit radio power (TRP) and ii) the power maintaining its proper functionality, including the signal processing by MCU, RF-chain components (amplifiers, ADCs, mixers, etc.), and peripheral hardware circuits. This realistic power model is elaborated in Sec.I-B and illustrated by Table I. To lower sensors' power consumption, inspired by the idea of DPM [26]–[28], we consider switching off as many sensors as possible, provided the data estimation quality can be guaranteed. This strategy is also widely utilized in most of the emerging IoT oriented protocols, e.g. BLE, ZigBee, Z-Wave, 802.15.4e, LoRa and so on. Specifically, the active sensors that transmit signals (i.e. $\|\mathbf{B}_\ell\|_2 > 0$) operate at a high power consumption level $\mathsf{P}_{\mathsf{act}}$. Comparatively, the sleep sensors that do not transmit (i.e. $\|\mathbf{B}_\ell\|_2 = 0$) operate in sleep mode and hence expend much lower power $\mathsf{P}_{\mathsf{slp}}$ (e.g. the Standby Mode in Table I). The average transmission power for the ℓ-th sensor is $\mathbb{E}\big\{\mathsf{Tr}\{\mathbf{B}_\ell(\mathbf{A}_\ell\mathbf{x} + \mathbf{n}_\ell)(\mathbf{A}_\ell\mathbf{x} + \mathbf{n}_\ell)^H\mathbf{B}_\ell^H\}\big\} = \mathsf{Tr}\big\{\mathbf{B}_\ell\big(\mathbf{A}_\ell\mathbf{A}_\ell^H + \mathbf{\Sigma}_\ell\big)\mathbf{B}_\ell^H\big\}$. It is zero for sleep nodes. Denote

the set of active and sleep sensors as $\mathcal{A}$ and $\mathcal{S}$, respectively, i.e. $\mathcal{A}\bigcup\mathcal{S}=\mathcal{L}$ and $\mathcal{A}\bigcap\mathcal{S}=\emptyset$. Then the total power consumption of the IoT sensors can be represented as

$$\begin{aligned}\mathsf{Pwr}(\mathcal{B})&=\mathsf{P}_{\mathsf{act}}|\mathcal{A}|+\mathsf{P}_{\mathsf{slp}}|\mathcal{S}|+\sum\nolimits_{\ell}\mathsf{Tr}\{\mathbf{B}_\ell(\mathbf{A}_\ell\mathbf{A}_\ell^H+\mathbf{\Sigma}_\ell)\mathbf{B}_\ell^H\}\\&=(\mathsf{P}_{\mathsf{act}}-\mathsf{P}_{\mathsf{slp}})|\mathcal{A}|+L\mathsf{P}_{\mathsf{slp}}+\sum\nolimits_{\ell}\mathsf{Tr}\{\mathbf{B}_\ell(\mathbf{A}_\ell\mathbf{A}_\ell^H+\mathbf{\Sigma}_\ell)\mathbf{B}_\ell^H\}\\&=\mathsf{P}_{\mathsf{d}}\Big\|[\|\mathbf{B}_1\|_2,\cdots,\|\mathbf{B}_L\|_2]\Big\|_0+L\mathsf{P}_{\mathsf{slp}}+\sum\nolimits_{\ell}\mathsf{Tr}\{\mathbf{B}_\ell(\mathbf{A}_\ell\mathbf{A}_\ell^H+\mathbf{\Sigma}_\ell)\mathbf{B}_\ell^H\},\end{aligned}\tag{6}$$

where the constant positive $\mathsf{P}_{\mathrm{d}}\triangleq\mathsf{P}_{\mathrm{act}}-\mathsf{P}_{\mathrm{slp}}$ and $|\cdot|$ is the cardinality of a set. Note that $\|\mathbf{a}\|_0$ returns the number of non-zero elements of vector $\mathbf{a}$ and therefore the ℓ_0-norm in (6) indeed counts the number of active sensors.

C. Problem Formulation

To summarize, the minimization of sensors' power consumption via jointly activating sensors, optimizing transmit beamforming and RIS phase-shifting with data fusion quality guaranteed can be formulated as the following problem (P0)

$$\begin{aligned}(\mathsf{P0}):\min_{\mathcal{B},\mathbf{\Phi}}.\ &\mathsf{Pwr}(\mathcal{B}),&(7a)\\\text{s.t. }&\mathsf{MSE}(\mathcal{B},\mathbf{\Phi})\le\epsilon_0,&(7b)\\&\mathsf{Tr}\{\mathbf{B}_\ell(\mathbf{A}_\ell\mathbf{A}_\ell^H+\mathbf{\Sigma}_\ell)\mathbf{B}_\ell^H\}\le\mathsf{P}_\ell,\ \forall\ell\in\mathcal{L},&(7c)\\&|\phi_j|=1,\ \forall j\in\mathcal{J},&(7d)\end{aligned}$$

where ϵ_0 is the predefined MSE target to be achieved and P_ℓ is the maximal TRP of each sensor. Note the (7d) is due to that each RIS element has constant modulus [30], [31].

III. Algorithm Design

In this section we develop algorithms to solve (P0). Here we assume that a feasible solution $(\mathcal{B}_0,\mathbf{\Phi}_0)$ to (P0) is available, which is required by the MM framework. The feasibility characterization will be elaborated in Sec.IV.

A. A Brief Introduction to Multi-Block MM Framework

In this subsection, we briefly review the seminal multi-block MM framework [40]–[42]. Consider the following non-convex problem

$$\begin{aligned}(\mathsf{P}_{\mathsf{Org}}):\ \min_{\mathbf{x}\in\mathcal{X}}.\ &f_0(\mathbf{x}),&(8a)\\\text{s.t. }&f_i(\mathbf{x})\le0,\ i=1,\cdots,m,&(8b)\end{aligned}$$

where the functions $f_i(\bullet)$, $\forall i\in\{0,1,\cdots,m\}$, can be non-convex. The entire variable $\mathbf{x}$ has n blocks, i.e. $\mathbf{x}\triangleq[\mathbf{x}_1^T,\cdots,\mathbf{x}_n^T]^T$. Besides, the feasible set $\mathcal{X}=\prod_{j=1}^n\mathcal{X}_j$, i.e. $\mathbf{x}_j\in\mathcal{X}_j$ with $\mathcal{X}_j$ being feasible region of $\mathbf{x}_j$, $\forall j\in\{1,\cdots,n\}$. Note that $\mathcal{X}_j$ does not need to be convex as long as the problem (9) can be solved. Denote $\mathbf{x}_{-j}=[\mathbf{x}_1^T,\cdots,\mathbf{x}_{j-1}^T,\mathbf{x}_{j+1}^T,\cdots,\mathbf{x}_n^T]^T$. Then the multi-block MM (namely BSUM) framework ([41], [42]) alternatively updates each block in block coordinate descent (BCD) manner on top level. That is for the t-th update, the block $j:=t\ \mathsf{mod}(n)+1$ will be updated by solving the following problem

$$\begin{aligned}\mathbf{x}_j^{(t)}:=\min_{\mathbf{x}_j\in\mathcal{X}_j}.\ &u_0(\mathbf{x}_j|\mathbf{x}_j^{(t-1)};\mathbf{x}_{-j}^{(t-1)}),&(9)\\\text{s.t. }&u_i(\mathbf{x}_j|\mathbf{x}_j^{(t-1)};\mathbf{x}_{-j}^{(t-1)})\le0,\ i=1,\cdots,m,\end{aligned}$$

The functions $u_i(\bullet|\mathbf{x}_j^{(t-1)};\mathbf{x}_{-j}^{(t-1)})$, $\forall,i,j$, are convex. In (9), $\mathbf{x}_{-j}^{(t-1)}$ are fixed as constants. The construction of $u_i(\bullet)$ relies on $\mathbf{x}_j^{(t-1)}$ (i.e. in the proximity of the latest value $\mathbf{x}_j^{(t-1)}$ of $\mathbf{x}_j$). The convex approximations $u_i(\bullet|\mathbf{x}_j^{(t-1)};\mathbf{x}_{-j}^{(t-1)})$ should satisfy the following MM upper-bounding conditions, $\forall i\in\{0,1,\cdots,m\}$ and $\forall j\in\{1,\cdots,n\}$

$$\begin{aligned}&\text{C1):}\quad u_i(\mathbf{x}_j^{(t-1)}|\mathbf{x}_j^{(t-1)};\mathbf{x}_{-j}^{(t-1)})=f_i(\mathbf{x}^{(t-1)}),&(10)\\&\text{C2):}\quad u_i(\mathbf{x}_j|\mathbf{x}_j^{(t-1)};\mathbf{x}_{-j}^{(t-1)})\ge f_i(\mathbf{x}_j,\mathbf{x}_{-j}^{(t-1)}),\ \forall\mathbf{x}_j\in\mathcal{X}_j;\\&\text{C3):}\quad \nabla_{\mathbf{x}_j}u_i(\mathbf{x}_j^{(t-1)}|\mathbf{x}_j^{(t-1)};\mathbf{x}_{-j}^{(t-1)})=\nabla_{\mathbf{x}_j}f_i(\mathbf{x}_j^{(t-1)},\mathbf{x}_{-j}^{(t-1)}).\end{aligned}$$

In a nutshell, the key idea of multi-block MM framework is to convexify the objective and constraints of each block via upper-bounds that approximates the original ones very well (see (10)).

B. SOCP-Based Solution

One prominent difficulty of (P0) comes from the non-convex and non-smooth ℓ_0-norm (activation pattern) in its objective. One straightforward method to tackle it is to exhaustively search all possible activation patterns. With each specific activation pattern determined, (P0) reduces to a power minimization problem [12]. However this brute-force method is obviously too cumbersome due to its prohibitively high complexity (totally 2^L-1 activation patterns). Here we adopt the group sparsity inducing method [38], which can effectively construct block-sparse-structured variables and has been extensively used in diverse research areas including machine learning, signal processing and wireless communications [35]–[37]. Its key idea is simple—approximate the non-smooth ℓ_0-norm with a smooth function $f_{\mathsf{smth}}(\cdot)$ and then optimize it. A bunch of well-behaved candidates of f_{smth} have already been found in the literature, including parameterized logarithm/exponential/arc-tangent functions and the ℓ_p-norms $(0<p<1)$ [36], [37]. Here we adopt the ℓ_p-norm and approximate $\mathsf{Pwr}(\mathcal{B})$ by the following function (with the constant $L\mathsf{P}_{\mathsf{slp}}$ omitted)

$$\widetilde{\mathsf{Pwr}}(\mathcal{B})=\sum\nolimits_{\ell}\mathsf{P}_{\mathsf{d}}\big(\|\mathbf{B}_\ell\|_2^2+\delta_0^2\big)^{p/2}+\sum\nolimits_{\ell}\mathsf{Tr}\{\mathbf{B}_\ell(\mathbf{A}_\ell\mathbf{A}_\ell^H+\mathbf{\Sigma}_\ell)\mathbf{B}_\ell^H\},$$

where p chosen in $(0,1)$ and δ_0 is a tiny positive constant to avoid ill-conditioned behaviours [38]. Therefore, we turn to optimize the following problem

$$\begin{aligned}(\mathsf{P0}'):\min_{\mathcal{B},\mathbf{\Phi}}.\ &\widetilde{\mathsf{Pwr}}(\mathcal{B}),&(11a)\\\text{s.t. }&\text{(7b), (7c), (7d),}&(11b)\end{aligned}$$

The problem $(\mathsf{P0}')$, although smooth now, is still challenging due to the non-convex ℓ_p-norm in the objective and its non-convex MSE and constant modulus constraints. In the following, we adopt the aforementioned MM methodology. To this end, we first present the following useful result that is proved in Appendix A.

Lemma 1. *Assume the matrix-valued functions* $\widehat{\mathbf{H}}(\mathbf{Z})$ *and* $\mathbf{\Sigma_n}(\mathbf{Z})$ *with respect to the complex matrix variable* $\mathbf{Z}$ *are continuously differentiable and define*

$$\mathbf{Q}(\mathbf{Z}) \triangleq \mathbf{\Sigma_n}(\mathbf{Z})+\widehat{\mathbf{H}}(\mathbf{Z})\widehat{\mathbf{H}}(\mathbf{Z})^H, \tag{12a}$$

$$\mathsf{MSE}(\mathbf{Z}) \triangleq \mathsf{Tr}\big\{\big[\mathbf{I}_K+\widehat{\mathbf{H}}(\mathbf{Z})^H\mathbf{\Sigma_n}(\mathbf{Z})^{-1}\widehat{\mathbf{H}}(\mathbf{Z})\big]^{-1}\big\}. \tag{12b}$$

Besides, assume $\mathbf{Z}_0$ *is a given value of* $\mathbf{Z}$ *and define*

$$\widehat{\mathbf{H}}_0 \triangleq \widehat{\mathbf{H}}(\mathbf{Z}_0), \mathbf{\Sigma}_{\mathbf{n},0} \triangleq \mathbf{\Sigma_n}(\mathbf{Z}_0), \mathbf{Q}_0 \triangleq \mathbf{\Sigma}_{\mathbf{n},0}+\widehat{\mathbf{H}}_0\widehat{\mathbf{H}}_0^H, \tag{13}$$

Then an upper-bound $\mathsf{U}\big(\mathbf{Z}|\mathbf{Z}_0\big)$ *of* $\mathsf{MSE}\big(\mathbf{Z}\big)$ *is given as follows*

$$\mathsf{U}\big(\mathbf{Z}|\mathbf{Z}_0\big)\triangleq\mathsf{Tr}\big\{\mathbf{Q}_0^{-1}\widehat{\mathbf{H}}_0\widehat{\mathbf{H}}_0^H\mathbf{Q}_0^{-1}\mathbf{Q}(\mathbf{Z})\big\}-2\mathsf{Re}\big\{\widehat{\mathbf{H}}_0^H\mathbf{Q}_0^{-1}\mathbf{H}\big(\mathbf{Z}\big)\big\}+K.$$

Furthermore, the upper-bound $\mathsf{U}\big(\mathbf{Z}|\mathbf{Z}_0\big)$ *satisfies the MM conditions C1)-C3) in (10).*

In the following, we solve the problem (P0) in a block coordinate descent (BCD) manner [41]. Firstly we optimize $\mathbf{\Phi}$ with $\mathcal{B}$ being fixed at $\mathcal{B}_0$. The update of $\mathbf{\Phi}$ is to solve the following problem.

$$(\mathsf{P1}): \mathsf{Find}\ \mathbf{\Phi}, \tag{14a}$$

$$\mathsf{s.t.}\ \mathsf{MSE}(\mathbf{\Phi};\mathcal{B}_0)-\epsilon_0 \le 0, \tag{14b}$$

$$|\phi_j| = 1,\ \forall j \in \mathcal{J}. \tag{14c}$$

The problem $(\mathsf{P1})$ is kind of weird due to its missing objective. To overcome this difficulty, since $\mathbf{\Phi}$ does not affect the objective of (P0), we turn to minimize the MSE metric, i.e.

$$(\mathsf{P2}): \min_{\mathbf{\Phi}}.\ \mathsf{MSE}(\mathbf{\Phi};\mathcal{B}_0) \tag{15a}$$

$$\mathsf{s.t.}\ |\phi_j| = 1,\ \forall j \in \mathcal{J}. \tag{15b}$$

The underlying rationale to solve (P2) is to obtain a "more" feasible $\mathbf{\Phi}$, which benefits the subsequent optimization of $\mathcal{B}$ via enlarging its associated feasible region. To tackle the intractable $\mathsf{MSE}(\mathbf{\Phi};\mathcal{B}_0)$, we adopt the MM method. Assume that $\mathbf{\Phi}_0$ is a feasible point of $\mathbf{\Phi}$ and denote

$$\widehat{\mathbf{H}}_0 \triangleq \widehat{\mathbf{H}}(\mathcal{B}_0,\mathbf{\Phi}_0), \mathbf{\Sigma}_{\mathbf{n},0} \triangleq \mathbf{\Sigma_n}(\mathcal{B}_0,\mathbf{\Phi}_0), \mathbf{Q}_0 \triangleq \mathbf{\Sigma}_{\mathbf{n},0}+\widehat{\mathbf{H}}_0\widehat{\mathbf{H}}_0^H,$$
$$\mathbf{S}_0 \triangleq \widehat{\mathbf{H}}_0^H\mathbf{Q}_0^{-1}, \mathbf{T}_0 \triangleq \mathbf{Q}_0^{-1}\mathbf{H}_0\mathbf{H}_0^H\mathbf{Q}_0^{-1}. \tag{16}$$

Invoking Lemma 1 and substituting $\mathbf{Z}$ therein with $\mathbf{\Phi}$, $\mathsf{MSE}\big(\mathbf{\Phi};\mathcal{B}_0\big)$ can be upper-bounded as follows

$$\begin{aligned}\mathsf{MSE}\big(\mathbf{\Phi};\mathcal{B}_0\big) &\le \mathsf{U}_{\mathbf{\Phi}}\big(\mathbf{\Phi}|\mathbf{\Phi}_0;\mathcal{B}_0\big) = \\ &\sum\nolimits_{\ell=1}^{L}\mathsf{Tr}\big\{\mathbf{F}^H\mathbf{T}_0\mathbf{F}\mathbf{\Phi}\mathbf{G}_\ell\mathbf{B}_{0,\ell}\mathbf{\Sigma}_\ell\mathbf{B}_{0,\ell}^H\mathbf{G}_\ell^H\mathbf{\Phi}^H\big\} \\ &+2\sum\nolimits_{\ell=1}^{L}\mathsf{Re}\big\{\mathsf{Tr}\big\{\mathbf{G}_\ell\mathbf{B}_{0,\ell}\mathbf{\Sigma}_\ell\mathbf{B}_{0,\ell}^H\mathbf{H}_\ell^H\mathbf{T}_0\mathbf{F}\mathbf{\Phi}\big\}\big\} \\ &+\sum\nolimits_{i,j=1}^{L}\mathsf{Tr}\big\{\mathbf{F}^H\mathbf{T}_0\mathbf{F}\mathbf{\Phi}\mathbf{G}_i\mathbf{B}_{0,i}\mathbf{A}_i\mathbf{A}_j^H\mathbf{B}_{0,j}^H\mathbf{G}_j^H\mathbf{\Phi}^H\big\} \\ &+2\sum\nolimits_{i,j=1}^{L}\mathsf{Re}\big\{\mathsf{Tr}\big\{\mathbf{G}_i\mathbf{B}_{0,i}\mathbf{A}_i\mathbf{A}_j^H\mathbf{B}_{0,j}^H\mathbf{H}_j^H\mathbf{T}_0\mathbf{F}\mathbf{\Phi}\big\}\big\} \\ &-2\sum\nolimits_{\ell=1}^{L}\mathsf{Re}\{\mathbf{G}_\ell\mathbf{B}_{0,\ell}\mathbf{A}_\ell\mathbf{S}_0\mathbf{F}\mathbf{\Phi}\}+c_1,\end{aligned} \tag{17}$$

where c_1 is some constant independent of $\mathbf{\Phi}$. Adopting the identities $\mathsf{Tr}\big\{\mathbf{A}\mathbf{\Phi}\mathbf{B}\mathbf{\Phi}^H\big\} = \phi^H\big[\mathbf{A}\odot\mathbf{B}^T\big]\phi$ and $\mathsf{Tr}\big\{\mathbf{C}\mathbf{\Phi}\big\} = \mathsf{Re}\{\mathsf{diag}(\mathbf{C})^H\phi\}$, we can rewrite the upper-bound of MSE w.r.t. $\mathbf{\Phi}$ in (17) as follows

$$\mathsf{MSE}\big(\mathbf{\Phi};\mathcal{B}_0\big)\le\mathsf{U}_{\mathbf{\Phi}}\big(\mathbf{\Phi}|\mathbf{\Phi}_0;\mathcal{B}_0\big)=\phi^H\mathbf{S}\phi+2\mathsf{Re}\{\mathbf{s}^H\phi\}+c_1, \tag{18}$$

where the parameters in (18) are defined as follows

$$\begin{aligned}\mathbf{S}\triangleq&\big(\mathbf{F}^H\mathbf{T}_0\mathbf{F}\big)\odot\Big(\sum\nolimits_\ell\mathbf{G}_\ell\mathbf{B}_{0,\ell}\mathbf{\Sigma}_\ell^*\mathbf{B}_{0,\ell}^T\mathbf{G}_\ell^T+\sum\nolimits_{i,j}\mathbf{G}_j^*\mathbf{B}_{0,j}^*\mathbf{A}_j^*\mathbf{A}_i^T\mathbf{B}_{0,i}^T\mathbf{G}_i^T\Big), \\ \mathbf{s}\triangleq&-\mathsf{diag}\Big(\sum\nolimits_\ell\mathbf{G}_\ell^*\mathbf{B}_{0,\ell}^*\mathbf{A}_\ell^*\mathbf{S}_0^*\mathbf{F}^*\Big)+\mathsf{diag}\Big(\sum\nolimits_\ell\mathbf{G}_\ell^*\mathbf{B}_{0,\ell}^*\mathbf{\Sigma}_\ell^*\mathbf{B}_{0,\ell}^T\mathbf{H}_\ell^T\mathbf{T}_0^*\mathbf{F}^*\Big) \\ &+\sum\nolimits_{i,j}\mathsf{diag}\big(\mathbf{G}_i^*\mathbf{B}_{0,i}^*\mathbf{A}_i^*\mathbf{A}_j^T\mathbf{B}_{0,j}^T\mathbf{H}_j^T\mathbf{T}_0^*\mathbf{F}^*\big).\end{aligned} \tag{19}$$

Therefore, after the MM transformation, solving (P2) turns into solving the following problem

$$(\mathsf{P3}): \min_{\mathbf{\Phi}}.\ \phi^H\mathbf{S}\phi+2\mathsf{Re}\{\mathbf{s}^H\phi\}, \tag{20a}$$

$$\mathsf{s.t.}\ |\phi_j| = 1,\ \forall j \in \mathcal{J}, \tag{20b}$$

The above problem is still difficult due to the non-convex constraints (20b). Fortunately, this difficulty can be resolved, again, via MM idea. Recall that ϕ_0 (i.e. $\mathbf{\Phi}_0$) is a feasible solution to (P2). The following fact stands

$$\begin{aligned}&\phi^H\mathbf{S}\phi+2\mathsf{Re}\big\{\mathbf{s}^H\phi\big\} \\ \overset{(a)}{=}&\big(\phi-\phi_0\big)^H\mathbf{S}\big(\phi-\phi_0\big)+2\mathsf{Re}\{(\mathbf{S}\phi_0+\mathbf{s})^H(\phi-\phi_0)\}+c_2 \\ \overset{(b)}{\le}&\lambda_{\max}(\mathbf{S})\big\|\phi-\phi_0\big\|_2^2+2\mathsf{Re}\{(\mathbf{S}\phi_0+\mathbf{s})^H(\phi-\phi_0)\}+c_2 \\ =&\lambda_{\max}(\mathbf{S})\|\phi\|_2^2+2\mathsf{Re}\{(\mathbf{S}\phi_0+\mathbf{s}-\lambda_{\max}(\mathbf{S})\phi_0)^H\phi\}+c_3,\end{aligned} \tag{21}$$

where (a) performs the second-order Taylor expansion at ϕ_0, (b) is due to $\lambda_{\max}(\mathbf{S})\mathbf{I}_J \succeq \mathbf{S}$ with $\lambda_{\max}(\mathbf{S})$ representing the maximal eigenvalue of $\mathbf{S}$, and c_2 and c_3 are constants independent of ϕ. Since we obtain the above upper-bound of (20a) via modifying the 2-nd order term of its Taylor expansion, clearly (21) has identical the 0-th and 1-st order values with those of (20a) at ϕ_0. This implies (21) satisfies C1)-C3) in (10). Therefore, instead of optimizing the objective in (P3) directly, we turn to optimize its upper-bound in (21). The key point of (21) lies in the fact that for any feasible ϕ satisfying (20b), the quadratic term in (21) is indeed a constant. Hence the update of ϕ reduces to solving the following problem

$$(\mathsf{P4}): \min_{\mathbf{\Phi}}.\ \mathsf{Re}\{(\mathbf{S}\phi_0+\mathbf{s}-\lambda_{\max}(\mathbf{S})\phi_0)^H\phi\}, \tag{22a}$$

$$\mathsf{s.t.}\ |\phi_j| = 1,\ \forall j \in \mathcal{J}, \tag{22b}$$

whose optimal solution can be immediately obtained as

$$\phi^\star = \exp\big(\jmath\angle\big(\lambda_{\max}(\mathbf{S})\phi_0-\mathbf{S}\phi_0-\mathbf{s}\big)\big), \tag{23}$$

with both $\exp(\cdot)$ and $\angle(\cdot)$ being element-wise operations.

Next we study the optimization of $\mathcal{B}$ when $\mathbf{\Phi}$ are fixed at the value of $\mathbf{\Phi}_0$. We assume that $\mathcal{B}_0$ is a feasible value of $\mathcal{B}$ and follow the notations in (13). Invoking Lemma 1 again and substituting $\mathbf{Z}$ therein with $\mathcal{B}$, an upper-bound of MSE w.r.t. $\mathcal{B}$ can be obtained as follows (after some manipulations)

$$\begin{aligned}&\mathsf{MSE}\big(\mathcal{B};\mathbf{\Phi}_0\big) \le \mathsf{U}_{\mathcal{B}}\big(\mathcal{B}|\mathcal{B}_0;\mathbf{\Phi}_0\big) = \\ &-\sum\nolimits_{\ell=1}^{L}2\mathsf{Re}\{\mathsf{Tr}\{\mathbf{A}_\ell\mathbf{S}_0\widetilde{\mathbf{H}}_{0,\ell}\mathbf{B}_\ell\}\}+\sum_\ell\mathsf{Tr}\big\{\mathbf{T}_0\widetilde{\mathbf{H}}_{0,\ell}\mathbf{B}_\ell\mathbf{\Sigma}_\ell\mathbf{B}_\ell^H\widetilde{\mathbf{H}}_{0,\ell}^H\big\} \\ &+\sum\nolimits_{i,j}\mathsf{Tr}\big\{\mathbf{T}_0\widetilde{\mathbf{H}}_{0,i}\mathbf{B}_i\mathbf{A}_i\mathbf{A}_j^H\mathbf{B}_j^H\widetilde{\mathbf{H}}_{0,j}^H\big\}+\sigma_0^2\mathsf{Tr}\big\{\mathbf{T}_0\big\},\end{aligned} \tag{24}$$

where $\widetilde{\mathbf{H}}_{0,\ell} \triangleq \widetilde{\mathbf{H}}_\ell(\mathbf{\Phi}_0)$, $\forall \ell$, is used for short. To simplify the following exposition, we introduce the following notations:

$$\begin{aligned}
&\mathbf{b}_\ell \triangleq \mathsf{vec}\big(\mathbf{B}_\ell\big),\ \mathbf{Q}_{i,j} \triangleq \big(\mathbf{A}_i^* \mathbf{A}_j^T\big) \otimes \big(\widetilde{\mathbf{H}}_{0,i}^H \mathbf{T}_0 \widetilde{\mathbf{H}}_{0,j}\big), \qquad (25)\\
&\mathbf{b} \triangleq [\mathbf{b}_1^T, \cdots, \mathbf{b}_L^T]^T,\ \mathbf{P}_\ell \triangleq \mathbf{\Sigma}_\ell^* \otimes \big(\widetilde{\mathbf{H}}_{0,\ell}^H \mathbf{T}_0 \widetilde{\mathbf{H}}_{0,\ell}\big),\\
&\mathbf{P} \triangleq \mathsf{Diag}\big(\big[\mathbf{P}_1, \cdots, \mathbf{P}_L\big]\big),\ \mathbf{Q} \triangleq [\mathbf{Q}]_{i,j},\ c_5 \triangleq \sigma_0^2 \mathsf{Tr}\{\mathbf{T}_0\},\\
&\mathbf{p}_\ell \triangleq \mathsf{vec}\big(\widetilde{\mathbf{H}}_{0,\ell}^H \mathbf{S}_0^H \mathbf{A}_\ell^H\big),\ \mathbf{p} \triangleq [\mathbf{p}_1^T, \cdots, \mathbf{p}_L^T]^T,\ \forall i,j,\ell \in \mathcal{L}.
\end{aligned}$$

Then the upper-bound in (24) can be written compactly as

$$\mathsf{MSE}\big(\mathcal{B}; \mathbf{\Phi}_0\big) \le \mathsf{U}_{\mathcal{B}}\big(\mathcal{B}|\mathcal{B}_0; \mathbf{\Phi}_0\big) = \mathbf{b}^H\big(\mathbf{P}+\mathbf{Q}\big)\mathbf{b} - 2\mathsf{Re}\{\mathbf{p}^H\mathbf{b}\} + c_5. \quad (26)$$

Besides the originally non-convex MSE constraint resolved above, another difficulty in optimizing $\mathcal{B}$ is its objective $\widetilde{\mathsf{Pwr}}(\mathcal{B})$, which reads

$$\sum\nolimits_\ell \mathsf{P_d}\big(\|\mathbf{b}_\ell\|_2^2 + \delta_0^2\big)^{\frac{p}{2}} + \sum\nolimits_\ell \mathbf{b}_\ell^H \mathbf{D}_\ell \mathbf{b}_\ell, \qquad (27)$$

where the second term representing the rewritten transmission power with $\mathbf{D}_\ell \triangleq \big(\mathbf{A}_\ell \mathbf{A}_\ell^H + \mathbf{\Sigma}_\ell\big)^T \otimes \mathbf{I}_{N_\ell}$, $\forall \ell \in \mathcal{L}$. The ℓ_p-norm, although smooth, is still non-convex. Fortunately, the seminal iterative re-weighted least square (IRWLS) method [38], [39], has adopted MM methodology to tackle it. In fact, since the function $h(x) = (x+\delta_0^2)^{\frac{p}{2}}$ with $0 < p < 1$ is concave in x, therefore

$$\big(x+\delta_0^2\big)^{\frac{p}{2}} \overset{(a)}{\le} \big(x_0+\delta_0^2\big)^{\frac{p}{2}} + \frac{p}{2}\big(x_0+\delta_0^2\big)^{\frac{p}{2}-1}\big(x-x_0\big), \qquad (28)$$

where x_0 is any feasible value and (a) utilizes the fact that the first order Taylor expansion at any point of a concave function forms an upper-bound of the function. Clearly, the MM conditions C1)-C3) in (10) are satisfied by the upper-bound in (28). By substituting the x and x_0 in (28) with $\mathbf{B}_\ell$ and $\mathbf{B}_{0,\ell}$, respectively, into the upper-bound in (28) and applying it to every ℓ_p-norm term in (27), we turn to solve the following problem

$$\sum\nolimits_\ell w_\ell \|\mathbf{b}_\ell\|_2^2 + \sum\nolimits_\ell \mathbf{b}_\ell^H \mathbf{D}_\ell \mathbf{b}_\ell + c_8, \qquad (29)$$

where $w_\ell \triangleq \frac{p}{2}\big(\|\mathbf{b}_{0,\ell}\|_2^2 + \delta_0^2\big)^{\frac{p}{2}-1}$ and c_8 is a constant. Therefore, the update of $\mathcal{B}$ can be performed by solving the subsequent problem

$$\begin{aligned}
(\mathsf{P5}): \min_{\mathbf{b}}\ & \sum\nolimits_\ell w_\ell \|\mathbf{b}_\ell\|_2^2 + \sum\nolimits_\ell \mathbf{b}_\ell^H \mathbf{D}_\ell \mathbf{b}_\ell, && (30a)\\
\text{s.t. } & \mathbf{b}^H\big(\mathbf{P}+\mathbf{Q}\big)\mathbf{b} - 2\mathsf{Re}\{\mathbf{p}^H\mathbf{b}\} + c_5 \le \epsilon_0, && (30b)\\
& \mathbf{b}_\ell^H \mathbf{D}_l \mathbf{b}_\ell \le \mathsf{P}_\ell,\ \forall \ell \in \mathcal{L}, && (30c)
\end{aligned}$$

The problem $(\mathsf{P5})$ is a second order cone programming (SOCP) problem and can be solved via off-the-shelf numerical solvers, e.g. CVX [44]. The SOCP based solution to solve the problem $(\mathsf{P0}')$ is summarized in Alg.1.

Proposition 1. *Assume that Alg.1 starts from a feasible point $(\mathcal{B}^{(0)}, \mathbf{\Phi}^{(0)})$. Then*

1) Alg.1 falls in the multi-block MM framework;

2) the iterates $\{\mathcal{B}^{(t)}, \mathbf{\Phi}^{(t)}\}$ generated by Alg.1 are always feasible to $(\mathsf{P0}')$, and hence $(\mathsf{P0})$;

3) The sequence $\{\widetilde{\mathsf{Pwr}}(\mathcal{B}^{(t)})\}$ monotonically decreases.

Algorithm 1: Solve $(\mathsf{P0}')$

Initialization: Generate feasible $\mathcal{B}^{(0)}$ and $\mathbf{\Phi}^{(0)}$, $t = 0$;
repeat
  update $\mathcal{B}^{(t+1)}$ by solving (P5) (by SOCP or Alg.2);
  update $\mathbf{\Phi}^{(t+1)}$ by (23) ;
  $t++$;
until *convergence*;

Proof. To see 1), the update of $\mathcal{B}$ follows the MM framework since i) $\mathsf{U}_{\mathcal{B}}(\cdot|\mathcal{B}^{(t+1)}; \mathbf{\Phi}^{(t+1)})$ follows C1)-C3), (proved in Lemma 1) and so does (30a) (recall the discussion below (28)). For $\mathbf{\Phi}$, we notice the following preliminary fact: if $u_1(\mathbf{x}|\mathbf{x}_0)$ is an upper-bound of $f(\mathbf{x})$ satisfying C1)-C3) at $\mathbf{x}_0$ and $u_2(\mathbf{x}|\mathbf{x}_0)$ is an upper-bound of $u_1(\mathbf{x}|\mathbf{x}_0)$ satisfying C1)-C3) at $\mathbf{x}_0$, then $u_2(\mathbf{x}|\mathbf{x}_0)$ is also an upper-bound of $f(\mathbf{x})$ satisfying C1)-C3) at $\mathbf{x}_0$. Indeed, $u_2(\mathbf{x}_0|\mathbf{x}_0) = u_1(\mathbf{x}_0|\mathbf{x}_0) = f(\mathbf{x}_0)$, $u_2(\mathbf{x}|\mathbf{x}_0) \ge u_1(\mathbf{x}|\mathbf{x}_0) \ge f(\mathbf{x})$ and $\nabla u_2(\mathbf{x}_0) = \nabla u_1(\mathbf{x}_0) = \nabla f(\mathbf{x}_0)$. Taking the $u_2(\cdot|\mathbf{x}_0)$ and $u_1(\cdot|\mathbf{x}_0)$ as the functions in (21) and (18), respectively, we can see that the update of $\mathbf{\Phi}$ also complies with MM framework.

For 2) and 3), assume that $(\mathcal{B}^{(t)}, \mathbf{\Phi}^{(t)})$ is feasible. Define (21) as $u_2(\cdot|\mathbf{\Phi}^{(t)}; \mathcal{B}^{(t)})$, which is indeed equivalent to (22a). For the update of $\mathcal{B}^{(t+1)}$ (when $\mathbf{\Phi}^{(t)}$ is fixed), we have

$$\mathsf{MSE}(\mathcal{B}^{(t+1)}, \mathbf{\Phi}^{(t)}) \le \mathsf{U}_{\mathcal{B}}(\mathcal{B}^{(t+1)}|\mathcal{B}^{(t)}; \mathbf{\Phi}^{(t)}) \le \epsilon_0; \qquad (31)$$

where the first inequality is due to C2) and the second inequality holds since $\mathcal{B}^{(t+1)}$ is obtained by solving (P5), which implies (30b) naturally stands. For the update of $\mathbf{\Phi}^{(t+1)}$ (when $\mathcal{B}^{(t+1)}$ is fixed), we have

$$\begin{aligned}
\mathsf{MSE}(\mathcal{B}^{(t+1)}, \mathbf{\Phi}^{(t+1)}) &\overset{(a)}{\le} u_2(\mathbf{\Phi}^{(t+1)}|\mathbf{\Phi}^{(t)}; \mathcal{B}^{(t+1)}) \qquad (32)\\
&\overset{(b)}{\le} u_2(\mathbf{\Phi}^{(t)}|\mathbf{\Phi}^{(t)}; \mathcal{B}^{(t+1)}) \overset{(c)}{=} \mathsf{MSE}(\mathcal{B}^{(t+1)}, \mathbf{\Phi}^{(t)}) \overset{(d)}{\le} \epsilon_0.
\end{aligned}$$

where (a) is due to C2), (b) holds since $\mathbf{\Phi}^{(t+1)}$ is optimal to (P4), (c) is due to C1) and (d) holds because of (31). Therefore $(\mathcal{B}^{(t+1)}, \mathbf{\Phi}^{(t+1)})$ is still feasible, i.e. (P5) is still feasible. Lastly, denote the function in (29) as $u_3(\cdot|\mathcal{B}^{(t)}, \mathbf{\Phi}^{(t+1)})$, then by similar arguments as above,

$$\begin{aligned}
\widetilde{\mathsf{Pwr}}(\mathcal{B}^{(t+1)}) &\le u_3(\mathcal{B}^{(t+1)}|\mathcal{B}^{(t)}, \mathbf{\Phi}^{(t)}) \qquad (33)\\
&\le u_3(\mathcal{B}^{(t)}|\mathcal{B}^{(t)}; \mathbf{\Phi}^{(t)}) = \widetilde{\mathsf{Pwr}}(\mathcal{B}^{(t)}).
\end{aligned}$$

Therefore, the sequence $\{\widetilde{\mathsf{Pwr}}(\mathcal{B}^{(t)})\}$ is decreasing. □

C. PDD-Based Solution

The SOCP-based solution in Alg.1 updates $\mathcal{B}$ via numerically solving the SOCP problem $(\mathsf{P5})$. Its shortcoming is twofold: i) it depends on numerical solvers; ii) its complexity increases extensively when the number of sensors grows (as will be clear in Sec.V). Therefore, we are motivated to further explore low-complexity solutions.

Firstly, we introduce an useful conclusion as stated in the following lemma

Lemma 2. *[12, Thm.1] (Convex Trust Region Problem) Consider the following convex trust region problem:*

$$(\mathsf{P_{TR}}) \min_{\boldsymbol{x}\in\mathbb{C}^n}. \ \boldsymbol{x}^H\boldsymbol{A}\boldsymbol{x} - 2\mathsf{Re}\{\boldsymbol{a}^H\boldsymbol{x}\}, \tag{34a}$$

$$\text{s.t. } \|\boldsymbol{x}\|_2^2 \le \tau^2, \tag{34b}$$

where $\boldsymbol{A} \succcurlyeq 0$ and has eigenvalue decomposition $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^H$. Then the optimal solution to ($\mathsf{P_{TR}}$) is given as

$$\boldsymbol{x}^\star = \boldsymbol{U}\big(\boldsymbol{\Lambda}+\mu^\star\boldsymbol{I}_n\big)^{-1}\boldsymbol{U}^H\boldsymbol{a}, \tag{35}$$

where the scalar $\mu^\star$ is non-negative and its value can be determined highly efficiently (e.g. bisection search).

Lemma 2 indicates a critical fact—a convex quadratic problem with a bounded-ball constraint can be efficiently solved analytically. This conclusion will be extensively exploited in the following algorithm design.

Now by introducing the following notations

$$\widetilde{\mathbf{P}} \triangleq \mathbf{Q}+\mathbf{P}, \ \mathbf{Y} \triangleq \mathsf{Diag}\big(\{w_1\mathbf{I}_{K_1N_1},\cdots,w_L\mathbf{I}_{K_LN_L}]\}\big), \tag{36a}$$

$$\mathbf{D} \triangleq \mathsf{Diag}\big([\{\mathbf{D}_1,\cdots,\mathbf{D}_L\}\big), \ \mathbf{R} \triangleq \mathbf{Y}+\mathbf{D}, \ c_6 \triangleq c_5-\epsilon_0, \tag{36b}$$

we rewrite the problem $(\mathsf{P5})$ in a compact form as follows

$$(\mathsf{P6}): \min_{\mathbf{b}}. \ \mathbf{b}^H\mathbf{R}\mathbf{b}, \tag{37a}$$

$$\text{s.t. } \mathbf{b}^H\widetilde{\mathbf{P}}\mathbf{b} - 2\mathsf{Re}\{\mathbf{p}^H\mathbf{b}\} + c_6 \le 0, \tag{37b}$$

$$\mathbf{b}_\ell^H\mathbf{D}_\ell\mathbf{b}_\ell \le \mathsf{P}_\ell, \ \forall \ell \in \mathcal{L}. \tag{37c}$$

Note that $\mathbf{R}$ has a block diagonal structure. Therefore if the constraint (37b) could be "eliminated" in some way, the different block variables $\{\mathbf{B}_\ell\}$ in $(\mathsf{P6})$ can be decoupled and consequently efficiently optimized via Lemma 2. Inspired by this observation, we introduce a copy $\mathbf{y}$ of $\mathbf{b}$ to decouple the troublesome constraint (37b) from others, which results in the following equivalent form of $(\mathsf{P6})$

$$(\mathsf{P7}): \min_{\mathbf{b},\mathbf{y}}. \ \mathbf{b}^H\mathbf{R}\mathbf{b}, \tag{38a}$$

$$\text{s.t. } \mathbf{y}^H\widetilde{\mathbf{P}}\mathbf{y} - 2\mathsf{Re}\{\mathbf{p}^H\mathbf{y}\} + c_6 \le 0, \tag{38b}$$

$$\mathbf{b}_\ell^H\mathbf{D}_\ell\mathbf{b}_\ell \le \mathsf{P}_\ell, \ \forall \ell \in \mathcal{L}, \tag{38c}$$

$$\mathbf{b} = \mathbf{y}. \tag{38d}$$

Following the cutting-the-edge PDD framework [43], the augmented Lagrangian function $\mathsf{L}_\rho(\mathbf{b},\mathbf{y};\boldsymbol{\lambda})$ of $(\mathsf{P7})$ is given as

$$\mathsf{L}_\rho(\mathbf{b},\mathbf{y};\boldsymbol{\lambda}) \triangleq \mathbf{b}^H\mathbf{R}\mathbf{b} + \mathsf{Re}\{\boldsymbol{\lambda}(\mathbf{b}-\mathbf{y})\} + \frac{1}{2\rho}\|\mathbf{b}-\mathbf{y}\|_2^2. \tag{39}$$

The PDD method is a two-layer iterative procedure [43], with its inner layer updating the primal variables by way of any schemes complying with the generic block successive upper-bound minimization (BSUM) framework while the outer layer selectively updating the Lagrangian multiplier $\boldsymbol{\lambda}$ or the penalty coefficient ρ. Specifically, applying the PDD method to (P7), the inner layer alternatively optimizes $\mathbf{b}$ and $\mathbf{y}$ towards minimizing $\mathsf{L}_\rho(\mathbf{b},\mathbf{y};\boldsymbol{\lambda})$ with $\boldsymbol{\lambda}$ and ρ being fixed. This procedure is specified in the following.

Define $\mathbf{q} \triangleq \rho\boldsymbol{\lambda} + \mathbf{b}$. The update of $\mathbf{y}$ is meant to solve

$$(\mathsf{P8}): \min_{\mathbf{y}}. \ \|\mathbf{y}\|_2^2 - 2\mathsf{Re}\{\mathbf{q}^H\mathbf{y}\}, \tag{40a}$$

$$\text{s.t. } \mathbf{y}^H\widetilde{\mathbf{P}}\mathbf{y} - 2\mathsf{Re}\{\mathbf{p}^H\mathbf{y}\} + c_6 \le 0. \tag{40b}$$

In order to exploit Lemma 2 to solve $(\mathsf{P8})$, we need firstly transfer it into the standard form of $(\mathsf{P_{TR}})$, which has a bounded norm constraint (34b). To this end, two different cases, according to the singularity of matrix $\widetilde{\mathbf{P}}$, should be considered and will be elaborated in the following.

CASE-1: **$\widetilde{\mathbf{P}}$ is invertible**. By introducing the following notations

$$\mathbf{s} \triangleq \widetilde{\mathbf{P}}^{\frac{1}{2}}\mathbf{y} - \widetilde{\mathbf{P}}^{-\frac{1}{2}}\mathbf{p}, \ \mathbf{t} \triangleq \widetilde{\mathbf{P}}^{-\frac{1}{2}}\mathbf{q} - \widetilde{\mathbf{P}}^{-\frac{3}{2}}\mathbf{p}, \tag{41}$$

the problem $(\mathsf{P8})$ can be equivalently written as

$$(\mathsf{P9}): \min_{\mathbf{s}}. \ \mathbf{s}^H\widetilde{\mathbf{P}}^{-1}\mathbf{s} - 2\mathsf{Re}\{\mathbf{t}^H\mathbf{s}\}, \tag{42a}$$

$$\text{s.t. } \|\mathbf{s}\|_2^2 \le \mathbf{p}^H\widetilde{\mathbf{P}}^{-1}\mathbf{p} - c_6, \tag{42b}$$

Under the assumption that a feasible solution exists, it can be readily verified that $\mathbf{p}^H\widetilde{\mathbf{P}}^{-1}\mathbf{p} - c_6 \ge 0$. By virtue of Lemma 2, $\mathbf{s}^\star$ can be obtained analytically. By (41), the update of $\mathbf{y}$ is immediately obtained as

$$\mathbf{y}^\star = \widetilde{\mathbf{P}}^{-\frac{1}{2}}\mathbf{s}^\star + \widetilde{\mathbf{P}}^{-1}\mathbf{p}. \tag{43}$$

CASE-2: **$\widetilde{\mathbf{P}}$ is singular**. Note that this case is indeed not uncommon and will definitely occur under specific system settings. For instance, it can be proved that when $M < K$, $\widetilde{\mathbf{P}}$ will always be rank-deficient. To invoke Lemma 2, we will first modify the constraint (40b) via leveraging the MM methodology. Assume that $\mathbf{y}_0$ is any feasible solution to $(\mathsf{P8})$ obtained previously. The following identities stand

$$\begin{aligned}
&\mathbf{y}^H\widetilde{\mathbf{P}}\mathbf{y} - 2\mathsf{Re}\{\mathbf{p}^H\mathbf{y}\} + c_6 \\
=&(\mathbf{y}-\mathbf{y}_0)^H\widetilde{\mathbf{P}}(\mathbf{y}-\mathbf{y}_0) + 2\mathsf{Re}\big\{\big(\widetilde{\mathbf{P}}\mathbf{y}_0-\mathbf{p}\big)^H\big(\mathbf{y}-\mathbf{y}_0\big)\big\} \\
&+\big(\mathbf{y}_0^H\widetilde{\mathbf{P}}\mathbf{y}_0 - 2\mathsf{Re}\{\mathbf{p}^H\mathbf{y}_0\} + c_6\big) &\text{(44a)}\\
\le&(\mathbf{y}-\mathbf{y}_0)^H\big(\widetilde{\mathbf{P}}+\delta\mathbf{I}\big)(\mathbf{y}-\mathbf{y}_0) + 2\mathsf{Re}\big\{\big(\widetilde{\mathbf{P}}\mathbf{y}_0-\mathbf{p}\big)^H\big(\mathbf{y}-\mathbf{y}_0\big)\big\} \\
&+\big(\mathbf{y}_0^H\widetilde{\mathbf{P}}\mathbf{y}_0 - 2\mathsf{Re}\{\mathbf{p}^H\mathbf{y}_0\} + c_6\big) &\text{(44b)}\\
=&\mathbf{y}^H\big(\widetilde{\mathbf{P}}+\delta\mathbf{I}\big)\mathbf{y} - 2\mathsf{Re}\big\{\big(\mathbf{p}+\delta\mathbf{y}_0\big)^H\mathbf{y}\big\} + \big(c_6+\delta\|\mathbf{y}_0\|_2^2\big), &\text{(44c)}
\end{aligned}$$

where δ is a small positive constant. Following similar arguments used for (21), the function in (44c) actually forms an upper-bound of (40b) and satisfies the conditions C1)-C3) in (10). Therefore, by MM method, we turn to replace the constraint (40b) by (44c). For ease of notations, we denote

$$\widehat{\mathbf{P}} \triangleq \widetilde{\mathbf{P}}+\delta\mathbf{I}, \ \widehat{\mathbf{p}} \triangleq \mathbf{p}+\delta\mathbf{y}_0, \ p_0 \triangleq c_6+\delta\|\mathbf{y}_0\|_2^2. \tag{45}$$

We replace the original problem $(\mathsf{P8})$ with its conservative version given as

$$(\mathsf{P10}): \min_{\mathbf{y}}. \ \|\mathbf{y}\|_2^2 - 2\mathsf{Re}\{\mathbf{q}^H\mathbf{y}\}, \tag{46a}$$

$$\text{s.t. } \mathbf{y}^H\widehat{\mathbf{P}}\mathbf{y} - 2\mathsf{Re}\{\widehat{\mathbf{p}}^H\mathbf{y}\} + p_0 \le 0, \tag{46b}$$

which now turns the "$\mathbf{P}$" matrix invertible at the expense of shrinking the feasible region slightly. Up to here, it is immediately clear that we can just play the tricks in (41-43) again followed by invoking Lemma 2 to update $\mathbf{y}$ analytically.

After discussing the update of $\mathbf{y}$, in the following, we proceed to elaborate the minimization of $\mathsf{L}_\rho\big(\mathbf{b},\mathbf{y};\boldsymbol{\lambda}\big)$ w.r.t. $\mathbf{b}$. The update of $\mathbf{b}$ is to solve the following problem

$$(\mathsf{P11}):\min_{\mathbf{b}}\ \mathbf{b}^H\mathbf{R}\mathbf{b}+\mathsf{Re}\big\{\boldsymbol{\lambda}\big(\mathbf{b}-\mathbf{y}\big)\big\}+\frac{1}{2\rho}\big\|\mathbf{b}-\mathbf{y}\big\|_2^2, \quad (47a)$$

$$\text{s.t. } \mathbf{b}_\ell^H\mathbf{D}_\ell\mathbf{b}_\ell\le\mathsf{P}_\ell,\forall\ell\in\mathcal{L}, \quad (47b)$$

Recall that $\mathbf{R}$ is block diagonal. Hence the above problem is actually decomposed into L independent sub-problems with each being given as

$$(\mathsf{P12}_\ell):\min_{\mathbf{b}_\ell}\ \mathbf{b}_\ell^H\big[(w_\ell+\frac{1}{2\rho})\mathbf{I}+\mathbf{D}_\ell\big]\mathbf{b}_\ell-\mathsf{Re}\big\{\big(\rho^{-1}\mathbf{y}_\ell-\boldsymbol{\lambda}_\ell\big)^H\mathbf{b}_\ell\big\},$$
$$\text{s.t. } \mathbf{b}_\ell^H\mathbf{D}_\ell\mathbf{b}_\ell\le\mathsf{P}_\ell, \quad (48)$$

where $\mathbf{y}_\ell$ and $\boldsymbol{\lambda}_\ell$ are segmental sub-vector of $\mathbf{y}$ and $\boldsymbol{\lambda}$, respectively, that are aligned with the $\mathbf{b}_\ell$, i.e. $\mathbf{y}\triangleq[\mathbf{y}_\ell^T,\cdots,\mathbf{y}_L^T]^T$ and $\boldsymbol{\lambda}\triangleq[\boldsymbol{\lambda}_1^T,\cdots,\boldsymbol{\lambda}_L^T]^T$. These L independent sub-problems $\{(\mathsf{P12}_\ell)\}_{\ell=1}^L$ can be solved by invoking Lemma 2 in parallel.

The inner layer of PDD procedure alternatively updates $\mathbf{y}$ and $\mathbf{b}$ until convergence. For each outer layer iteration, exactly one of the following two actions will be taken [43]

A1: when the equality $\mathbf{b}=\boldsymbol{y}$ is nearly satisfied, the Lagrangian multiplier $\boldsymbol{\lambda}$ will be updated in a gradient ascent manner, which is given by

$$\boldsymbol{\lambda}:=\boldsymbol{\lambda}+\rho^{-1}\big(\mathbf{b}-\mathbf{y}\big); \quad (49)$$

A2: when the difference between $\mathbf{b}$ and $\mathbf{y}$ exceeds a predefined threshold, the penalty parameter ρ^{-1} will be augmented, forcing the equality constraint $\mathbf{b}=\mathbf{y}$ to be approached in the subsequent iterations.

The previously developed solution built on the combination of MM and PDD frameworks to solve $(\mathsf{P5})$ is summarized in Alg.2. Note that the $\{\eta^{(k)}\}$ (step-10 in Alg.2) is a parameter sequence set converging to 0 [43] and c is a positive constant smaller than 1. It can be proved that when Slater's condition holds, limit points of the solution iterates yielded by Alg.2 are KKT points of $(\mathsf{P6})$, i.e. optimal solutions since $(\mathsf{P6})$ is convex. Details are omitted for space limitation.

Algorithm 2: PDD Method to Solve $(\mathsf{P5})$

1: initialize $\mathbf{b}^{(0)}$, $\mathbf{y}^{(0)}$, $\boldsymbol{\lambda}^{(0)}$ and $\rho^{(0)}$ and $k=1$;
2: **repeat**
3: set $\mathbf{b}^{(k-1,0)}:=\mathbf{b}^{(k-1)}$, $\mathbf{y}^{(k-1,0)}:=\mathbf{y}(k-1)$, $t=0$;
4: **repeat**
5: update $\mathbf{y}^{(k-1,t+1)}$ by solving $(\mathsf{P8})$ or $(\mathsf{P10})$;
6: update $\mathbf{b}^{(k-1,t+1)}$ by solving $(\mathsf{P12}_\ell)$'s in parallel;
7: $t++$;
8: **until** *convergence*
9: set $\mathbf{y}^{(k)}:=\mathbf{y}^{(k-1,\infty)}$ and $\mathbf{b}^{(k)}:=\mathbf{b}^{(k-1,\infty)}$;
10: **if** $\|\mathbf{b}^{(k)}-\mathbf{y}^{(k)}\|_\infty\le\eta^{(k)}$ **then**
11: $\boldsymbol{\lambda}^{(k+1)}:=\boldsymbol{\lambda}^{(k)}+\frac{1}{\rho^{(k)}}\big(\mathbf{b}^{(k)}-\mathbf{y}^{(k)}\big)$, $\rho^{(k+1)}:=\rho^{(k)}$;
12: **else**
13: $\boldsymbol{\lambda}^{(k+1)}:=\boldsymbol{\lambda}^{(k)}$, $\rho^{(k+1)}:=c\cdot\rho^{(k)}$;
14: **end if**
15: $k++$;
16: **until** $\|\mathbf{b}^{(k)}-\mathbf{y}^{(k)}\|_2$ is sufficiently small

D. Active Sensors' Selection and Beamforming

So far, we can solve $(\mathsf{P0}')$ iteratively, by either the SOCP-based or PDD-based algorithm, and obtain a solution $(\boldsymbol{\mathcal{B}}^\star,\boldsymbol{\Phi}^\star)$. Recall that our original goal is to resolve $(\mathsf{P0})$, while $(\mathsf{P0}')$ is indeed an approximation of it. Therefore, we need to construct a solution to $(\mathsf{P0})$ based on the obtained $(\boldsymbol{\mathcal{B}}^\star,\boldsymbol{\Phi}^\star)$. One notable characteristic of $\boldsymbol{\mathcal{B}}^\star$ is its group (block) sparsity structure—a subset of $\{\mathbf{B}_\ell^\star\}$ have vanishing values (their norms are typically many orders of magnitude smaller than others). In fact, this structure is built up during the process of the iterative reweighting procedure (i.e. the MM iterations), which is indeed the origin of its name "group sparsity inducing". More deep insights can be obtained in [38], [39]. In the following we exploit the group sparsity in $\boldsymbol{\mathcal{B}}^\star$ to obtain a solution to $(\mathsf{P0})$. Intuitively, the smaller $\mathbf{B}_\ell$ is, the less the ℓ-th sensor contributes to data fusion, which implies a higher priority to be switched off. Therefore, we can obtain a "switch-off" ranking from $\boldsymbol{\mathcal{B}}^\star$. Note that such a ranking procedure via exploiting the group sparsity is also adopted in [35]–[37]. Here we propose two ranking criteria in the following.

1) **TX Power** By evaluating the transmit powers $\{\mathsf{P}_{\mathsf{TX},\ell}\}$

$$\mathsf{P}_{\mathsf{TX},\ell}:=\mathsf{Tr}\big\{\mathbf{B}_\ell^\star\big(\mathbf{A}_\ell\mathbf{A}_\ell^H+\boldsymbol{\Sigma}_\ell\big)\mathbf{B}_\ell^{\star H}\big\},\quad\forall\ell\in\mathcal{L}, \quad (50)$$

we can obtain a ranking $\boldsymbol{\pi}^{\mathsf{TX}}\triangleq[\pi_1^{\mathsf{TX}},\cdots,\pi_L^{\mathsf{TX}}]^T$ such that $\mathsf{P}_{\mathsf{TX},\pi_1^{\mathsf{TX}}}\le\mathsf{P}_{\mathsf{TX},\pi_2^{\mathsf{TX}}}\le\cdots\le\mathsf{P}_{\mathsf{TX},\pi_L^{\mathsf{TX}}}$.

2) **MSE** Via calculating the quantities MSE_ℓ in the following

$$\mathsf{MSE}_\ell:=\mathsf{Tr}\big\{\big[\mathbf{I}_K+\mathbf{A}_\ell^H\mathbf{B}_\ell^{\star H}\overline{\mathbf{H}}_\ell^H\overline{\boldsymbol{\Sigma}}_\ell^{-1}\overline{\mathbf{H}}_\ell\mathbf{B}_\ell^\star\mathbf{A}_\ell\big]^{-1}\big\},\forall\ell,$$

where $\overline{\mathbf{H}}_\ell:=\mathbf{H}_\ell+\mathbf{F}\boldsymbol{\Phi}^\star\mathbf{G}_\ell$ and $\overline{\boldsymbol{\Sigma}}_\ell:=\sigma_0^2\mathbf{I}_M+\overline{\mathbf{H}}_\ell\mathbf{B}_\ell^\star\boldsymbol{\Sigma}_\ell\mathbf{B}_\ell^{\star H}\overline{\mathbf{H}}_\ell^H$, we can obtain a ranking $\boldsymbol{\pi}^{\mathsf{MSE}}\triangleq[\pi_1^{\mathsf{MSE}},\cdots,\pi_L^{\mathsf{MSE}}]^T$ such that $\mathsf{MSE}_{\pi_1^{\mathsf{MSE}}}\ge\mathsf{MSE}_{\pi_2^{\mathsf{MSE}}}\ge\cdots\ge\mathsf{MSE}_{\pi_L^{\mathsf{MSE}}}$. Note MSE_ℓ represents MSE performance when the ℓ-th sensor solely transmits the data to the FC. Intuitively, the sensor with a higher MSE_ℓ contributes less to signal recovery.

For the ranking $\boldsymbol{\pi}^i$ determined above, $i\in\{\mathsf{MSE},\mathsf{TX}\}$, the π_ℓ^i with a larger ℓ indicates a higher priority to be activated. Therefore we propose to always activate the sensors with the lowest "swith-off" priorities, i.e. the sensors $\{\pi_Q^i,\pi_{Q+1}^i,\cdots,\pi_L^i\}$. To switch off as many sensors as possible, we seek to determine the maximal integer $Q_{\max}$ such that the active sensors can achieve an $\mathsf{MSE}\le\epsilon_0$. Hence we can conduct a bisection search over $\boldsymbol{\pi}^i$ to identify $Q_{\max}$. Specifically, for a given Q, define the active set $\mathcal{A}(Q)\triangleq\{\pi_Q,\pi_{Q+1},\cdots,\pi_L\}$ and the associated active beamformers $\boldsymbol{\mathcal{B}}(Q)\triangleq\{\mathbf{B}_\ell,\ell\in\mathcal{A}(Q)\}$. Then identifying the feasibility of the active sensors $\mathcal{A}(Q)$ means to solve

$$(\mathsf{P13}_{\mathcal{A}(Q)}):\min_{\boldsymbol{\mathcal{B}}(Q),\boldsymbol{\Phi}}\ \mathsf{MSE}\big(\boldsymbol{\mathcal{B}}(Q),\boldsymbol{\Phi}\big) \quad (51)$$
$$\text{s.t. } \mathsf{Tr}\big\{\mathbf{B}_\ell\big(\mathbf{A}_\ell\mathbf{A}_\ell^H+\boldsymbol{\Sigma}_\ell\big)\mathbf{B}_\ell^H\big\}\le\mathsf{P}_\ell,\ \forall\ell\in\mathcal{A}(Q),$$
$$|\phi_j|=1,\ \forall j\in\mathcal{J},$$

which can be solved highly efficiently by Alg.4 that will be elaborated in Sec.IV shortly. If the obtained MSE by solving

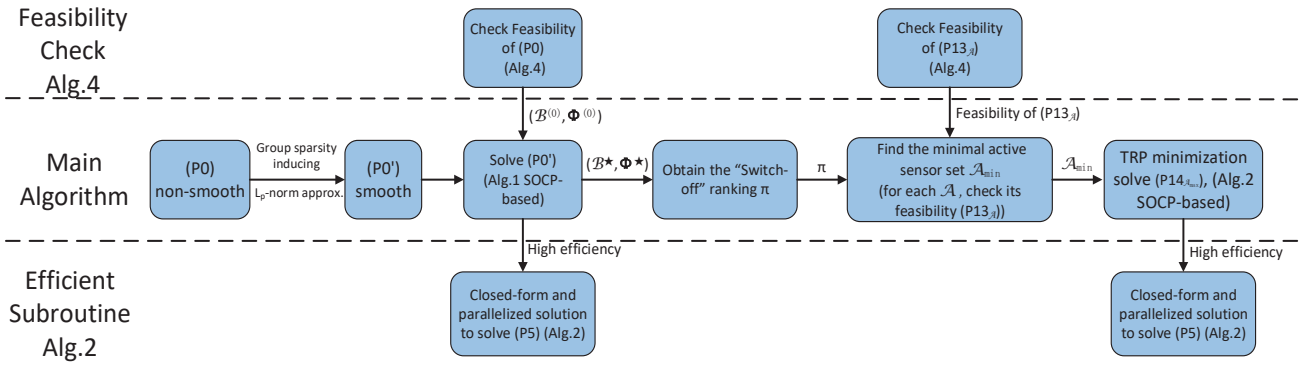

Fig. 2: Main flow and connections between different algorithms

Algorithm 3: Joint. Snsr. Actv. & BF. (**Main Alg.**)

solve (P0′) via Alg.1 to obtain $(\mathcal{B}^{\star}, \mathbf{\Phi}^{\star})$;
obtain a ranking $\boldsymbol{\pi}^i$, $i \in \{\mathsf{MSE}, \mathsf{TX}\}$, from $(\mathcal{B}^{\star}, \mathbf{\Phi}^{\star})$;
set $Q_{\mathsf{low}} := 1$ and $Q_{\mathsf{up}} := L$;
repeat
  $Q := \lceil \frac{Q_{\mathsf{low}}+Q_{\mathsf{up}}}{2} \rceil$; solve $(\mathsf{P13}_{\mathcal{A}(Q)})$ via Alg.4 to obtain $(\mathcal{B}^{\ddagger}(\mathcal{A}(Q)), \mathbf{\Phi}^{\ddagger})$;
  if $\mathsf{MSE}(\mathcal{B}^{\ddagger}(\mathcal{A}(Q)), \mathbf{\Phi}^{\ddagger}) \le \epsilon_0$ **then**
    $Q_{\mathsf{low}} := Q$;
  else
    $Q_{\mathsf{up}} := Q$;
  end if
until $Q_{\mathsf{up}} - Q_{\mathsf{low}} = 1$
set $Q_{\max} := Q_{\mathsf{low}}$; solve $(\mathsf{P14}_{\mathcal{A}(Q_{\max})})$ (SOCP/PDD);

$(\mathsf{P13}_{\mathcal{A}(Q)})$ is smaller than ϵ_0, then we could turn off more sensors (i.e. increase Q). Otherwise more sensors need be activated.

Once the $Q_{\max}$ is determined, we need finally determine the beamforming of the active sensors by solving

$$\begin{aligned}(\mathsf{P14}_{\mathcal{A}(Q_{\max})}): \min_{\mathcal{B}(Q_{\max}), \mathbf{\Phi}} & \sum_{\ell \in \mathcal{A}(Q_{\max})} \mathsf{Tr}\big\{\mathbf{B}_\ell\big(\mathbf{A}_\ell \mathbf{A}_\ell^H + \mathbf{\Sigma}_\ell\big)\mathbf{B}_\ell^H\big\}, \\ \text{s.t. } & \mathsf{MSE}\big(\mathcal{B}_{\mathcal{A}(Q_{\max})}, \mathbf{\Phi}\big) \le \epsilon_0, \\ & \mathsf{Tr}\big\{\mathbf{B}_\ell\big(\mathbf{A}_\ell \mathbf{A}_\ell^H + \mathbf{\Sigma}_\ell\big)\mathbf{B}_\ell^H\big\} \le \mathsf{P}_\ell, \ \forall \ell \in \mathcal{A}(Q_{\max}), \\ & |\phi_j| = 1, \ \forall j \in \mathcal{J}. \end{aligned} \tag{52}$$

To tackle the above problem, we can still use the SOCP-based or PDD-based solutions that is developed to solve (P0′). Details are omitted to avoid repetition. The overall procedure of joint sensor activation, active beamforming and RIS phase configuration is summarized in Alg.3. Besides, we also illustrate the main work flow and the connections between various algorithms in Fig.2.

E. Complexity

Here we briefly discuss the complexity of our proposed algorithms. For simplicity, we assume that $N_\ell = N_0$ and $K_\ell = K_0$, $\forall \ell \in \mathcal{L}$. According to the complexity analysis in [48], the SOCP method to solve (P5) has a complexity of $\mathcal{O}(C_1 L^{3.5} N_0^2 K_0^2 (L + N_0 K_0))$. For the PDD algorithm, its has two-layer loops. For the inner loop, the update of $\mathbf{y}$ and $\mathbf{b}_\ell$ has a complexity of $\mathcal{O}(L^3 N_0^3 K_0^3)$ and $\mathcal{O}(N_0^3 K_0^3)$, respectively. Therefore the total complexity of the PDD algorithm Alg.2 has a complexity of $\mathcal{O}(C_2 L^3 N_0^3 K_0^3)$. Note that $C_2 = C_{\mathsf{out}} C_{\mathsf{in}}$ with C_{out} and C_{in} denoting the executed times of the outer and inner PDD loops, respectively, and generally taking value from 10 to 20 (refer to Sec.V). For the sensor activation in Alg.3, we need to perform $\log_2(L)$ times of feasibility check, (i.e. solving $(\mathsf{P13}_{\mathcal{A}(Q)})$) with a complexity of $\mathcal{O}(C_3 N_0^3 K_0^3)$ for each check. Besides, the $(\mathsf{P14}_{\mathcal{A}(Q_{\max})})$ can be solved via SOCP or PDD solutions like $(\mathsf{P}(0'))$, with a complexity of $\mathcal{O}(C_1 L_{\min}^{3.5} N_0^2 K_0^2 (L_{\min} + N_0 K_0))$ or $\mathcal{O}(C_2 L_{\min}^3 N_0^3 K_0^3)$, respectively, with $L_{\min} = L - Q_{\max}$.

IV. Feasibility Characterization

In this section, we investigate the feasibility characterization of (P0), which is equivalent to $(\mathsf{P13}_{\mathcal{A}(Q)})$. Note the feasibility check is critical since i) it is required by Alg.1 and Alg.2 to star off; ii) for sensor activation, we need check the feasibility of a given active set $\mathcal{A}(Q)$, i.e. $(\mathsf{P13}_{\mathcal{A}(Q)})$.

A. Algorithmic Feasibility Characterization

First we discuss an algorithmic method that can identify the feasibility of (P0) (or $(\mathsf{P13}_{\mathcal{A}(Q)})$). This is meant to find the possible minimal MSE and compare it with the QoS requirement ϵ_0 while assuring the constraints (7c-7d) are satisfied. That is we need to solve the following problem

$$(\mathsf{P15}): \min_{\mathcal{B}, \mathbf{\Phi}} \ \mathsf{MSE}\big(\mathcal{B}, \mathbf{\Phi}\big) \tag{53a}$$

$$\text{s.t. } \mathsf{Tr}\big\{\mathbf{B}_\ell\big(\mathbf{A}_\ell \mathbf{A}_\ell^H + \mathbf{\Sigma}_\ell\big)\mathbf{B}_\ell^H\big\} \le \mathsf{P}_\ell, \ \forall \ell \in \mathcal{L}, \tag{53b}$$

$$|\phi_j| = 1, \ \forall j \in \mathcal{J}. \tag{53c}$$

For (P15), we still solve it by way of BCD. When $\mathcal{B}$ is given, following identical arguments as previously, $\mathbf{\Phi}$ can be updated via optimizing the upper-bound $\mathsf{U}_{\mathbf{\Phi}}(\cdot)$ developed in (17-18), which is indeed the problem (P3) and has an analytic solution given by (23). Note that (P0′) updates $\mathbf{\Phi}$ by (23) to maintain feasible. In comparison, (P15) updates $\mathbf{\Phi}$ to decrease its objective. When $\mathbf{\Phi}$ is fixed, the update of $\mathcal{B}$ can be executed

via optimizing the upper-bound $\mathsf{U}_{\mathcal{B}}(\cdot)$ given in (26), which is given as (omitting independent constants)

$$(\mathsf{P16}): \min_{\mathbf{b}} . \ \mathbf{b}^H\widetilde{\mathbf{P}}\mathbf{b}-2\mathsf{Re}\{\mathbf{p}^H\mathbf{b}\} \tag{54a}$$

$$\text{s.t. } \mathbf{b}_\ell^H\mathbf{D}_\ell\mathbf{b}_\ell \le \mathsf{P}_\ell, \ \forall \ell \in \mathcal{L}. \tag{54b}$$

The above problem is SOCP and can be numerically solved. Fortunately, efficient solution independent of numerical solvers can still be obtained. Suppose that $\mathbf{b}_0$ is any feasible solution of $\mathbf{b}_0$. Following similar derivations in (44a), we can invoke the MM trick again and upper-bound the objective (54a) by

$$\mathbf{b}^H\widetilde{\mathbf{P}}\mathbf{b} - 2\mathsf{Re}\{\mathbf{p}^H\mathbf{b}\} \tag{55a}$$

$$=(\mathbf{b}-\mathbf{b}_0)^H\widetilde{\mathbf{P}}(\mathbf{b}-\mathbf{b}_0)+2\mathsf{Re}\big\{\big(\widetilde{\mathbf{P}}\mathbf{b}_0-\mathbf{p}\big)^H\mathbf{b}\big\}-\mathbf{b}_0^H\widetilde{\mathbf{P}}\mathbf{b}_0 \tag{55b}$$

$$\le\lambda_{\mathsf{max}}\big(\widetilde{\mathbf{P}}\big)\big\|\mathbf{b}-\mathbf{b}_0\big\|_2^2+2\mathsf{Re}\big\{\big(\widetilde{\mathbf{P}}\mathbf{b}_0-\mathbf{p}\big)^H\mathbf{b}\big\}-\mathbf{b}_0^H\widetilde{\mathbf{P}}\mathbf{b}_0, \tag{55c}$$

which decouples all sensors' beamformers and divide the objective into L separable components. By writing $\mathbf{b}_0$ and $\big(\widetilde{\mathbf{P}}\mathbf{b}_0-\mathbf{p}\big)$ in a form of L segments explicitly, i.e., $\mathbf{b}_0 \triangleq [\mathbf{b}_{0,1}^T,\cdots,\mathbf{b}_{0,L}^T]^T$ and $\widetilde{\mathbf{P}}\mathbf{b}_0-\mathbf{p} \triangleq \mathbf{u} \triangleq [\mathbf{u}_1^T,\cdots,\mathbf{u}_L^T]^T$, the optimization problem to update $\mathbf{b}$, which reads

$$(\mathsf{P17}): \min_{\mathbf{b}} . \ \lambda_{\mathsf{max}}\big(\widetilde{\mathbf{P}}\big)\big\|\mathbf{b}-\mathbf{b}_0\big\|_2^2+2\mathsf{Re}\big\{\mathbf{u}^H\mathbf{b}\big\} \tag{56a}$$

$$\text{s.t. } \mathbf{b}_\ell^H\mathbf{D}_\ell\mathbf{b}_\ell \le \mathsf{P}_\ell, \ \forall \ell \in \mathcal{L}, \tag{56b}$$

can be decomposed into L independent smaller problems

$$(\mathsf{P17}_\ell): \min_{\mathbf{b}_\ell} . \ \lambda_{\mathsf{max}}\big(\widetilde{\mathbf{P}}\big)\big\|\mathbf{b}_\ell-\mathbf{b}_{0,\ell}\big\|_2^2+2\mathsf{Re}\big\{\mathbf{u}_\ell^H\mathbf{b}_\ell\big\} \tag{57a}$$

$$\text{s.t. } \mathbf{b}_\ell^H\mathbf{D}_\ell\mathbf{b}_\ell \le \mathsf{P}_\ell. \tag{57b}$$

Obviously, the above L sub-problems $(\mathsf{P17}_\ell)$ can be solved in parallel with each being figured out analytically via invoking Lemma 2, again. The proposed algorithm to solve the feasible problem $(\mathsf{P15})$ is summarized in Alg.4. The Alg.4 has a complexity of $\mathcal{O}(C_4 \max_{\ell\in\mathcal{L}}\{N_\ell^3K_\ell^3\})$. Besides, following identical arguments as in Prop.1, we can show that Alg.4 falls in the multi-block MM framework and guarantees monotonic improvement in objective and feasibility of solutions. Besides, we can also prove that Alg.4 converge to KKT points of $(\mathsf{P15})$. Due to limit of space, we omit these details.

Remark IV.1. *Note that the problem* $(\mathsf{P15})$ *is non-convex, which implies that its truly optimal value is generally unknown. MM method only guarantees converging to stationary points of non-convex problems, not globally optimal solutions [41]. Therefore, the genuine global optimal value of* $(\mathsf{P15})$ *is indeed unknown. So Alg.4 actually serves as a sufficient condition. If we failed to find a feasible solution via invoking Alg.4, there still exists a possibility that the truly optimal value of* $(\mathsf{P15})$ *is no larger than* ϵ_0 *(i.e.* $(\mathsf{P0})$ *is feasible). Unfortunately, however, since the globally optimal solution cannot be found, we have to claim that* $(\mathsf{P0})$ *is infeasible in this case.*

B. Achievable Analytic MSE Bound for Scalar Signal

In this subsection, we identify the globally optimal value of $(\mathsf{P15})$ analytically for a special but very meaningful setting.

Algorithm 4: Feasibility Characterization Algorithm

1: randomly generate feasible $(\boldsymbol{\mathcal{B}}^{(0)}, \boldsymbol{\Phi}^{(0)})$; $t := 0$;
2: **repeat**
3: **for** $\ell = 1 : L$ In Parallel **do**
4: solve $(\mathsf{P17}_\ell)$ to obtain $\mathbf{b}_\ell^{(t+1)}$ by Lem. 2;
5: **end for**
6: update $\boldsymbol{\Phi}^{(t+1)}$ by (23); $t++$;
7: **until** convergence

This result is presented in Prop. 2 in the following and is proved in Appendix B.

Proposition 2. *When* $K = 1$*, i.e. the target signal* $\mathbf{x}$ *is a scalar, the receiving filter* $\mathbf{W}$ *now becomes* $\mathbf{w}$ *and* $\mathbf{A}_\ell$ *becomes* $\mathbf{a}_\ell$*,* $\forall \ell \in \mathcal{L}$*. In this case, provided the power supplies* $\{\mathsf{P}_\ell\}$ *are sufficient, the infimum of achievable MSE is given as*

$$\mathsf{MSE}_{\mathsf{inf}}=\Big(1+\sum\nolimits_{\ell=1}^{L}\mathbf{a}_\ell^H\boldsymbol{\Sigma}_\ell^{-1}\mathbf{a}_\ell\Big)^{-1}, \tag{58}$$

regardless of the RIS phase-shift settings.

To gain further insight into the achievable MSE bound obtained in Prop.2, we now consider the centralized sensing scenario, where all the observation samples are assumed to be obtained locally at the fusion center. Specifically the overall observation samples $\mathbf{z}$ can be expressed compactly as follows

$$\mathbf{z} = \mathbf{d}x + \mathbf{n}, \tag{59}$$

where $\mathbf{d} \triangleq [\mathbf{a}_1^T,\cdots,\mathbf{a}_L^T]^T$ and $\mathbf{v} \triangleq [\mathbf{n}_1^T,\cdots,\mathbf{n}_L^T]^T$. Clearly the entire observation noise $\mathbf{v}$ has the covariance matrix $\mathbf{R}_{\mathbf{v}} = \mathsf{Diag}\{\boldsymbol{\Sigma}_1,\cdots,\boldsymbol{\Sigma}_L\}$. To obtain an optimal linear estimate of x for the centralized sensing in (59), we apply linear MMSE estimator $\mathbf{w}_{\mathsf{cntr}}^\star = (\mathbf{R}_{\mathbf{v}}+\mathbf{d}\mathbf{d}^H)^{-1}\mathbf{d}$ to filter $\mathbf{z}$ [47]. It can be verified that the corresponding minimum MSE for the centralized sensing scenario is given as [47]

$$\mathsf{MSE}_{\mathsf{cntr}}^\star=\Big(1+\mathbf{d}^H\mathbf{R}_{\mathbf{v}}^{-1}\mathbf{d}\Big)^{-1}=\Big(1+\sum\nolimits_{\ell=1}^{L}\mathbf{a}_\ell^H\boldsymbol{\Sigma}_\ell^{-1}\mathbf{a}_\ell\Big)^{-1}, \tag{60}$$

which coincides with the result in Prop.2. We can therefore conclude that the decentralized sensing scheme can achieve as good MSE performance as that obtained in the centralized sensing scheme as long as the power supplies are sufficient.

Remark IV.2. *Our result generalizes the MSE bound in [9] for single-antenna case. Besides, our result also coincides with the asymptotic MSE lower bound derived in [15] for large scale antennas. Combining our Prop.2 and the result in [15], it is now clear that the performance of distributive sensing can be compensated by transmission power or the number of antennas at sensors. Prop.2 also reveals a fact—fuelled by sufficient power, the receiving filter* $\mathbf{w}$ *and RIS phase-shift* $\boldsymbol{\Phi}$ *will not affect the MSE. At this time the beamformers* $\boldsymbol{\mathcal{B}}$ *will perfectly nullify the impact of the wireless channels and the receive noise.*

V. Numerical Results

In this section, numerical results are presented. Our simulation considers an IoT network deployed in an indoor

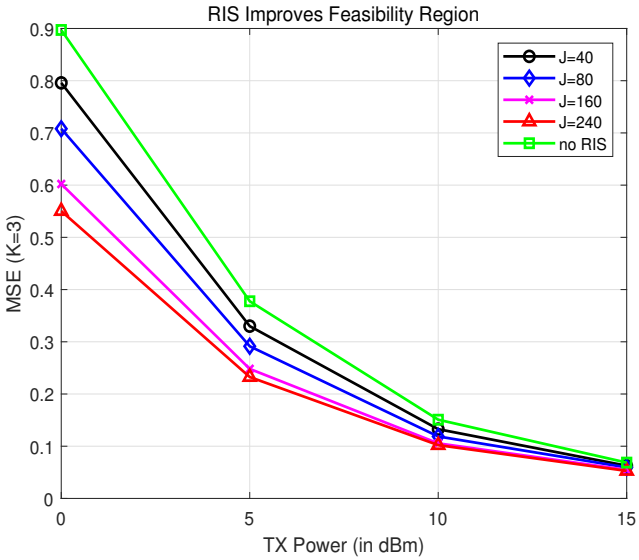

Fig. 3: Achievable MSE with Different RIS Settings

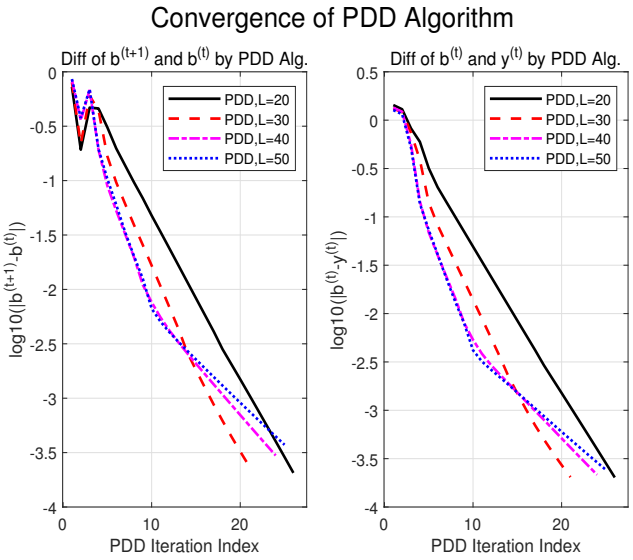

Fig. 4: Convergence of PDD Algorithm

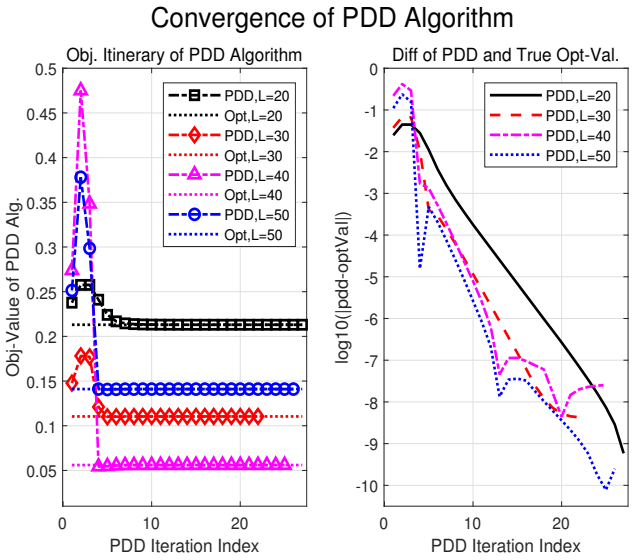

Fig. 5: Convergence of PDD Algorithm

environment, where all the wireless sensors developed by the CC2652P devices. A fusion center is located at a height of 1.2 meters (m) above the floor. An RIS is deployed on the 4.0m high ceiling right above the FC. The IoT network has $L = 20$ wireless sensors that are uniformly distributed around the FC on the floor with a distance no larger than 20m. We test the TX 5dBm and TX 10dBm modes of the CC2652P, whose active sensor power $\mathsf{P}_{\mathsf{act}}$ is 31.7mW and 72.6mW, respectively (see Table I). By contrast, the Standby mode $\mathsf{P}_{\mathsf{slp}}$ of the CC2652P's is negligible, i.e. $\mathsf{P}_{\mathsf{slp}} \approx 0$. We set $K = 3$ and $M = 4$, respectively, if not explicitly specified. The RIS-FC channel is light-of-sight (LoS) path and has a fading exponent of -3.0. The pathloss of the sensor-RIS and sensor-FC links have fading exponents of -3.5 and -4.0, respectively. The noise spectrum density is -100dBm/Hz and the working bandwidth of the sensor network is 10MHz. We take $p = 0.5$ for ℓ_p-norm and $\delta_0 = 10^{-3}$ in (27).

Firstly, in Fig.3, we examine the impact of RIS on improving feasible MSE performance. In the experiment, $L = 20$ wireless sensors are randomly deployed. The horizontal axis represents the maximal TRP of each sensor. As shown in the figure, we obtain the achievable MSE via Alg.4 associated with different number of RIS elements. As a benchmark we also present the achievable MSE performance corresponding to the no-RIS case. Clearly, more reflecting elements can improve the MSE performance, especially in the scenario where transmission powers are low.

Secondly, as shown in Fig.4 and Fig.5, we testify the PDD algorithm's performance for solving (P5). In Fig.4, we fix the number of RIS reflecting elements at $J = 40$ and vary the number of wireless sensors L from 20 to 50. For the PDD algorithm implementation, the penalty adjustment parameter $c = 0.75$. The inner layer of PDD algorithm will terminate the BCD updating once the relative improvement of the objective is below 0.005. $\eta^{(k)}$ is set as the 0.9 times the maximal element-wise difference of $\mathbf{b}^{(k-1)}$ and $\mathbf{y}^{(k-1)}$. The left part of Fig.4 represents the difference of $\mathbf{b}$ associated with two consecutive PDD outer iterations. The right half of Fig.4 plots the difference of $\mathbf{b}$ and $\mathbf{y}$ along with the progress of PDD outer loops. Besides, Fig.5 illustrates the objective itinerary obtained by the PDD algorithm. The left half of Fig.5 compares the objective value itinerary of the PDD iterates when solving (P5) as well as its true optimal value obtained via CVX. The right half of Fig.5 illustrates the convergence behaviour of objective value of PDD solution more precisely in log domain. Combining the results in Fig.4 and Fig.5, we conclude that generally the PDD algorithm converges well enough within 20 iterations.

Next, in Fig.6, we examine the PDD-based solution's performance when solving the sparsity-inducing power minimization problem (P0′). In the simulation, we test a bunch of various

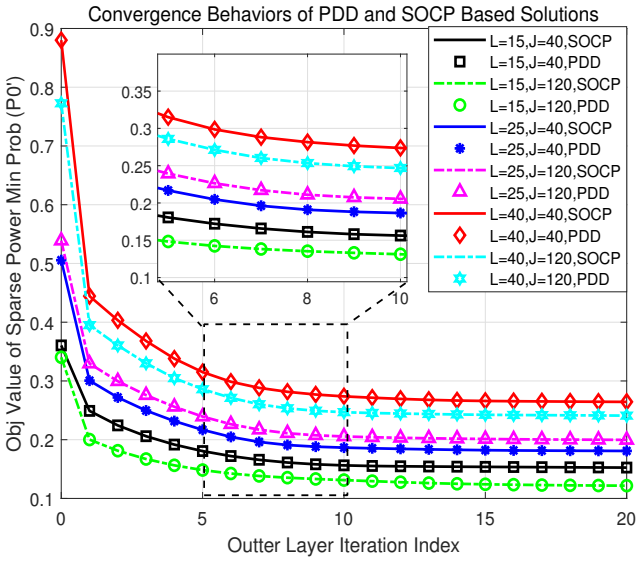

Fig. 6: Convergence of PDD-Based Solution

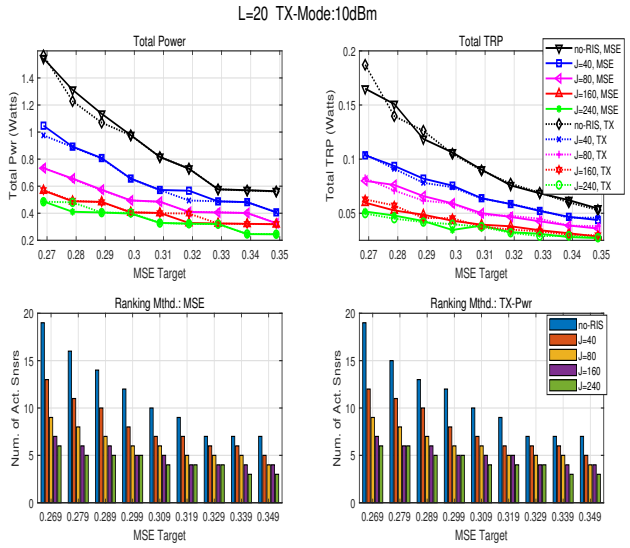

Fig. 7: Joint Sensor Activation and Beamforming (TX 10dBm)

TABLE II: MATLAB Running Time Per Iteration (in sec.)

Alg.	$L=20$	$L=50$	$L=100$	$L=150$	$L=200$
SOCP	3.017	39.21	346.7	1274.5	2922.0
PDD	1.461	14.50	157.5	627.4	1649.5

system settings where L range from 15 to 40 and the RIS has 40 or 120 reflecting elements. For each specific system setting, whose channel coefficients are randomly generated, the SOCP-based and PDD-based solutions are executed from a common feasible starting point. As shown in Fig.6, not very surprisingly, the two peers yield almost identical performance. The results presented in Fig.6 together with those in Fig.4 and Fig.5 convince us that our proposed PDD-based solution can obtain as good performance as that of the SOCP-based counterpart, which relies on numerical solvers.

In the following, in Table II, we compare the time complexity of the PDD-based and SOCP-based solutions. Considering the fact that the PDD-based solution yields almost identical objective iterates with that of the SOCP-based rival, as suggested in Fig.6, it suffices to compare the time complexity of each outer layer iteration, which corresponds to solving the problem (P5). By fixing $J = 40$, we test the Matlab run time consumed for solving (P5) associated with different number of sensors L. Note that for PDD implementation, the update of L sensors are carried out in parallel and consequently the corresponding run time is tantamount to updating one $\mathbf{b}_\ell$. As shown in Table II, for sensor network with moderate or large scale, the run time of PDD-based solution is clearly advantageous over its SOCP-based counterpart.

In Fig.7 and Fig.8, we examine the effectiveness of our proposed joint sensor activation, active and passive beamforming design, where $L = 20$ sensors implemented by CC2652P device work at TX 10dBm and TX 5dBm modes [29], respectively. Different numbers of RIS reflecting elements (i.e. J) are tested, including the no-RIS case as a benchmark. The proposed switch-off ranking criteria based on TX power and MSE for sensor activation are tested in the experiment, which are labelled as "TX" and "MSE" in the legend, respectively. In each plot, the upper-left figure shows the optimized total power consumptions with respect to different MSE requirement, which comprises both $\mathsf{P}_{\mathsf{act}}$ and TRP. The upper-right figure illustrates the total TRP of all activated sensors. The activated number of sensors obtained by the "TX" and "MSE" ranking are presented in the two lower figures. As can be seen, the two proposed ranking criteria have nearly identical performance. More importantly, the employment of RIS can effectively decrease the number of activated sensors. Take $J = 160$ RIS elements as an example, the number of activated sensors can be generally reduced by 50% compared to the no-RIS scenario, which yields a power saving by 45%-70%. Moreover, by comparing Fig.7 with Fig.8, more benefit in power saving is observed in TX 5dBm mode than in the 10dBm case. This is because that TX 5dBm mode has relative higher $\mathsf{P}_{\mathsf{act}}$ compared to its supporting TRP $\mathsf{P}_{\mathsf{TRP}}^{\max}$, as reflected in Table.I.

In Fig.9, to testify the optimality of our proposed sensor activation schemes, we compare them with the exhaustive search method in both TX 10dBm and 5dBm working modes. Note that in our test, there are totally $L = 15$ sensors. In the exhaustive search, we just examine all possible $2^L-1=32767$ sensor activation patterns and, for each specific activation pattern $\mathcal{A}$, we check its feasibility (i.e. solve (P13$_\mathcal{A}$)) and then perform TX beamforming design (i.e. solve (P14$_\mathcal{A}$))) if (P13$_\mathcal{A}$) is feasible. As illustrated in Fig.9, our proposed activation schemes, whether the TX or MSE ranking being used, yield almost identical number of activated sensors as that of the exhaustive search.

At last, in Fig.10, we examine the achievable MSE bounds for $K = 1$ obtained in Prop.2. Experiments are conducted for $L = 20$ and $L = 30$ combined with various RIS settings, with the MSE bound in Prop.2 also presented in the plot. Note that the MSE curves in Fig.10 are obtained via invoking Alg.4. Clearly, when the transmission power is sufficient (above 40dBm), regardless of the RIS settings, the analytic

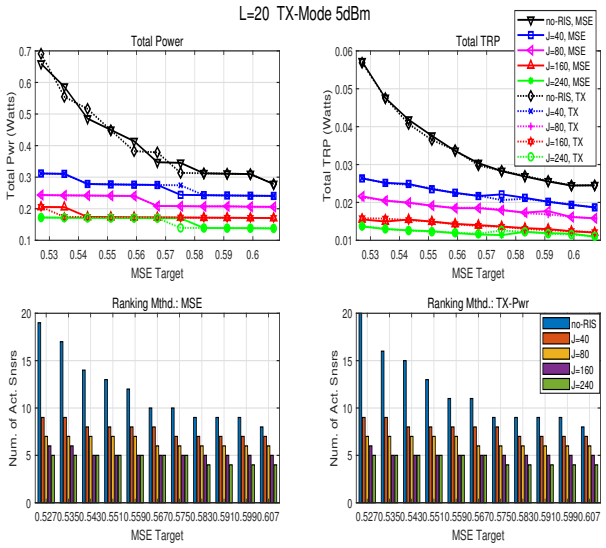

Fig. 8: Joint Sensor Selection and Beamforming (TX 5dBm)

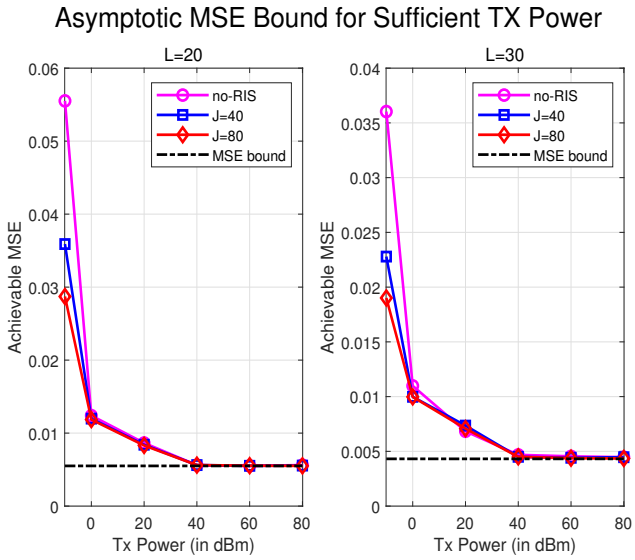

Fig. 10: Asymptotic Achievable MSE Bound

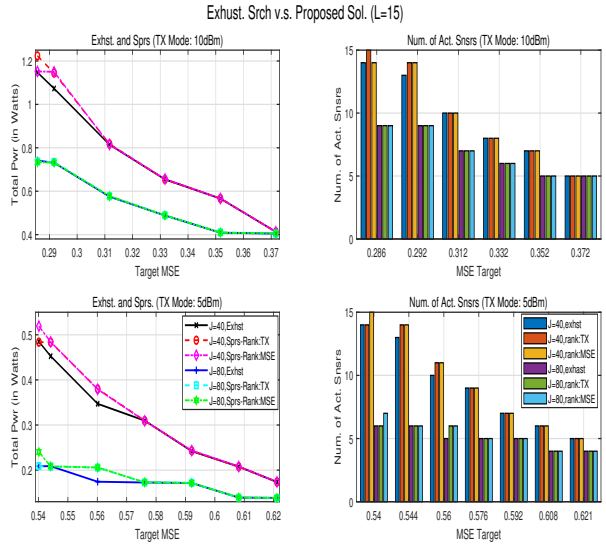

Fig. 9: Sparse Inducing Solution and Exhaustive Search

MSE bounds can be approached, which verifies the validity of Prop.2. Furthermore, Fig.10 also verifies the effectiveness of Alg.4, because it can actually achieve the globally optimal MSE.

VI. Conclusion

This paper conducts a comprehensive research on the joint activation and beamforming design towards power minimization in an IoT sensor network assisted by RIS. Via combining the MM and PDD framework, we have developed an analytically updating and highly parallelized solution. Extensive numerical results verify the effectiveness our solutions and prove the significant benefit from joint sensor activation and RIS deployment.

Appendix

A. Proof of Lemma 1

Proof. Firstly we notice the fact that $\mathsf{Tr}\{\mathbf{X}\mathbf{Y}^{-1}\mathbf{X}^H\}$ is convex in $(\mathbf{X},\mathbf{Y})$ jointly, where $\mathbf{X}\in\mathbb{C}^{M\times N}$ and $\mathbf{Y}\in\mathbb{S}_{++}^{M}$ (i.e. $\mathbf{Y}$ is positive definite). Note that a similar result with $\mathbf{X}$ being specialized to a vector is originally proved in Example 3.4 [49] via epigraph and projection technique. Here we prove a generalized result using a different approach. In fact, according to [49], it suffices to prove that $\mathsf{Tr}\{\mathbf{X}\mathbf{Y}^{-1}\mathbf{X}^H\}$ is convex when restricted to a line segment connecting two arbitrary points $(\mathbf{X}_1,\mathbf{Y}_1)$ and $(\mathbf{X}_2,\mathbf{Y}_2)$, where $\mathbf{X}_1,\ \mathbf{X}_2\in\mathbb{C}^{M\times N}$ and $\mathbf{Y}_1,\ \mathbf{Y}_2\in\mathbb{S}_{++}^{M}$. Towards this end, define $\mathbf{X}(\alpha)\triangleq\alpha\mathbf{X}_1+(1-\alpha)\mathbf{X}_2$ and $\mathbf{Y}(\alpha)\triangleq\alpha\mathbf{Y}_1+(1-\alpha)\mathbf{Y}_2$, where $\alpha\in[0,1]$. Therefore we just need to verify that the function $g(\alpha)\triangleq\mathsf{Tr}\{\mathbf{X}(\alpha)\mathbf{Y}(\alpha)^{-1}\mathbf{X}(\alpha)^H\}$ is convex in α. Since

$$\begin{aligned}\frac{\mathrm{d}g}{\mathrm{d}\alpha}=&\mathsf{Tr}\Big\{\frac{\mathrm{d}\mathbf{X}(\alpha)}{\mathrm{d}\alpha}\mathbf{Y}(\alpha)^{-1}\mathbf{X}^H+\mathbf{X}(\alpha)\mathbf{Y}(\alpha)^{-1}\frac{\mathrm{d}\mathbf{X}(\alpha)}{\mathrm{d}\alpha}^H\\&-\mathbf{X}(\alpha)\mathbf{Y}(\alpha)^{-1}\frac{\mathrm{d}\mathbf{Y}(\alpha)}{\mathrm{d}\alpha}\mathbf{Y}(\alpha)^{-1}\mathbf{X}(\alpha)^H\Big\}.\end{aligned}\tag{61}$$

The second order derivative of $g(\alpha)$ is readily identified as

$$\begin{aligned}\frac{\mathrm{d}^2g}{\mathrm{d}\alpha^2}=&2\mathsf{Tr}\Big\{\Big(\mathbf{X}(\alpha)\mathbf{Y}(\alpha)^{-1}\frac{\mathrm{d}\mathbf{Y}(\alpha)}{\mathrm{d}\alpha}-\frac{\mathrm{d}\mathbf{X}(\alpha)}{\mathrm{d}\alpha}\Big)\mathbf{Y}(\alpha)^{-1}\\&\times\Big(\mathbf{X}(\alpha)\mathbf{Y}(\alpha)^{-1}\frac{\mathrm{d}\mathbf{Y}(\alpha)}{\mathrm{d}\alpha}-\frac{\mathrm{d}\mathbf{X}(\alpha)}{\mathrm{d}\alpha}\Big)^H\Big\}\geq 0.\end{aligned}\tag{62}$$

Therefore, $g(\alpha)$ is convex, which immediately implies the convexity of $\mathsf{Tr}\{\mathbf{X}\mathbf{Y}^{-1}\mathbf{X}^H\}$ in $(\mathbf{X},\mathbf{Y})$.

In the following, for ease of notations, we abbreviate the function forms $\widehat{\mathbf{H}}(\mathbf{Z})$, $\mathbf{\Sigma}_{\mathbf{n}}(\mathbf{Z})$ and $\mathbf{Q}(\mathbf{Z})$ to $\widehat{\mathbf{H}}$, $\mathbf{\Sigma}_{\mathbf{n}}$ and $\mathbf{Q}$, respectively. Then we construct an upper-bound of $\mathsf{MSE}(\mathbf{Z})$

as follows

$$\mathsf{MSE}(\mathbf{Z}) = \mathsf{Tr}\big\{\big[\mathbf{I}_K + \widehat{\mathbf{H}}^H \boldsymbol{\Sigma}_{\mathbf{n}}^{-1} \widehat{\mathbf{H}}\big]^{-1}\big\} \tag{63a}$$

$$\overset{(a)}{=} \mathsf{Tr}\big\{\mathbf{I}_K - \widehat{\mathbf{H}}^H \big(\boldsymbol{\Sigma}_{\mathbf{n}} + \widehat{\mathbf{H}}\widehat{\mathbf{H}}^H\big)^{-1} \widehat{\mathbf{H}}\big\} \overset{(b)}{=} -\mathsf{Tr}\big\{\widehat{\mathbf{H}}^H \mathbf{Q}^{-1} \widehat{\mathbf{H}}\big\} + K$$

$$\overset{(c)}{\leq} -\mathsf{Tr}\big\{\widehat{\mathbf{H}}_0^H \mathbf{Q}_0^{-1} \widehat{\mathbf{H}}_0\big\} - 2\mathsf{Re}\big\{\widehat{\mathbf{H}}_0 \mathbf{Q}_0^{-1} \big(\widehat{\mathbf{H}} - \widehat{\mathbf{H}}_0\big)\big\} \tag{63b}$$

$$+ \mathsf{Tr}\big\{\mathbf{Q}_0^{-1} \mathbf{H}_0 \mathbf{H}_0^H \mathbf{Q}_0^{-1} \big(\mathbf{Q} - \mathbf{Q}_0\big)\big\} + K$$

$$\overset{(d)}{=} \mathsf{U}\big(\mathbf{Z}|\mathbf{Z}_0\big) \tag{63c}$$

where (a) is due to matrix-inversion lemma, (b) follows the notation in (12a) and (c) utilizes the conclusion that $-\mathsf{Tr}\{\widehat{\mathbf{H}}^H \mathbf{Q}^{-1} \widehat{\mathbf{H}}\}$ is concave in $(\widehat{\mathbf{H}}, \mathbf{Q})$ and the fact that the first-order Taylor expansion at any point (the point $(\widehat{\mathbf{H}}_0, \mathbf{Q}_0)$ here) of a concave function forms a global upper-bound, and (d) can be obtained after some manipulations.

Next we verify that $\mathsf{U}(\mathbf{Z}|\mathbf{Z}_0)$ satisfies the conditions C1)-C3) in (10). Substituting $\mathbf{Z} = \mathbf{Z}_0$ into (63), we obtain

$$\mathsf{MSE}(\mathbf{Z}_0) = -\mathsf{Tr}\big\{\widehat{\mathbf{H}}_0^H \mathbf{Q}_0^{-1} \widehat{\mathbf{H}}_0\big\} + K = \mathsf{U}(\mathbf{Z}|\mathbf{Z}_0). \tag{64}$$

Therefore, C1) is satisfied. C2) is just reflected in (63b). Note that $\mathsf{MSE}(\mathbf{Z}) = \mathsf{MSE}(\widehat{\mathbf{H}}(\mathbf{Z}), \mathbf{Q}(\mathbf{Z}))$. By chain rules of derivative, we have

$$\partial_{\mathbf{Z}} \mathsf{MSE}(\mathbf{Z}) = \partial_{\widehat{\mathbf{H}}} \mathsf{MSE}\big(\widehat{\mathbf{H}}(\mathbf{Z}), \mathbf{Q}(\mathbf{Z})\big) \partial_{\mathbf{Z}} \widehat{\mathbf{H}}(\mathbf{Z}) \tag{65}$$

$$+ \partial_{\mathbf{Q}} \mathsf{MSE}\big(\widehat{\mathbf{H}}(\mathbf{Z}), \mathbf{Q}(\mathbf{Z})\big) \partial_{\mathbf{Z}} \mathbf{Q}(\mathbf{Z}).$$

At the same time, according to the derivation of $\mathsf{U}(\mathbf{Z}|\mathbf{Z}_0)$ in (63), its derivative in $\mathbf{Z}$ is

$$\partial_{\mathbf{Z}} \mathsf{U}(\mathbf{Z}|\mathbf{Z}_0) = \partial_{\widehat{\mathbf{H}}} \mathsf{MSE}(\widehat{\mathbf{H}}_0, \mathbf{Q}_0) \partial_{\mathbf{Z}} \widehat{\mathbf{H}}(\mathbf{Z}) + \partial_{\mathbf{Q}} \mathsf{MSE}(\widehat{\mathbf{H}}_0, \mathbf{Q}_0) \partial_{\mathbf{Z}} \mathbf{Q}(\mathbf{Z}).$$

Combining the above equation and (65), we obtain

$$\partial_{\mathbf{Z}} \mathsf{MSE}(\mathbf{Z}_0) = \partial_{\mathbf{Z}} \mathsf{U}(\mathbf{Z}_0|\mathbf{Z}_0), \tag{66}$$

i.e. C3) in (10) is also verified. Therefore, the construction of $\mathsf{U}(\mathbf{Z}|\mathbf{Z}_0)$ is indeed consistent with the MM framework. □

B. Proof of Proposition 2

Proof. For an arbitrary RIS phase-shift setting $\boldsymbol{\Phi}$, we use the abbreviation $\widetilde{\mathbf{H}}_\ell$ on behalf of the effective channel $\widetilde{\mathbf{H}}_\ell(\boldsymbol{\Phi}) \triangleq \mathbf{H}_\ell + \mathbf{F}\boldsymbol{\Phi}\mathbf{G}_\ell$ for simplicity. In the following discussion, instead of directly studying MSE, we first dive into investigating the signal-to-noise (SNR) performance, which reads

$$\mathsf{SNR}(\mathcal{B}, \mathbf{w}) = \frac{\mathbb{E}\left\{\left|\left(\mathbf{w}^H \sum_{\ell=1}^L \widetilde{\mathbf{H}}_\ell \mathbf{B}_\ell \mathbf{a}_\ell\right) x\right|^2\right\}}{\mathbb{E}\left\{\left|\mathbf{w}^H \left(\sum_{\ell=1}^L \widetilde{\mathbf{H}}_\ell \mathbf{B}_\ell \mathbf{n}_\ell + \mathbf{n}_0\right)\right|^2\right\}}$$

$$= \frac{\mathbf{w}^H \big[\sum_{i=1}^L \widetilde{\mathbf{H}}_i \mathbf{B}_i \mathbf{a}_i\big] \big[\sum_{j=1}^L \widetilde{\mathbf{H}}_j \mathbf{B}_j \mathbf{a}_j\big]^H \mathbf{w}}{\sigma_0^2 \|\mathbf{w}\|_2^2 + \sum_{\ell=1}^L \mathbf{w}^H \widetilde{\mathbf{H}}_\ell \mathbf{B}_\ell \boldsymbol{\Sigma}_\ell \mathbf{B}_\ell^H \widetilde{\mathbf{H}}_\ell^H \mathbf{w}}, \tag{67}$$

Utilizing the identities $\mathsf{Tr}\{\mathbf{ABCD}^H\} = \mathsf{vec}^H(\mathbf{D})\big[\mathbf{C}^T \otimes \mathbf{A}\big]\mathsf{vec}(\mathbf{B})$ and $\mathsf{Tr}\{\mathbf{AB}\} = \mathsf{Tr}\{\mathbf{BA}\}$ and denoting,

$$\mathbf{X}_{ij} \triangleq \left[\big(\mathbf{a}_i^* \mathbf{a}_j^T\big) \otimes \big(\widetilde{\mathbf{H}}_i^H \mathbf{w}\mathbf{w}^H \widetilde{\mathbf{H}}_j\big)\right],\ i, j \in \mathcal{L}; \tag{68a}$$

$$\mathbf{Y}_\ell \triangleq \left[\boldsymbol{\Sigma}_\ell^* \otimes \big(\widetilde{\mathbf{H}}_\ell^H \mathbf{w}\mathbf{w}^H \widetilde{\mathbf{H}}_\ell\big)\right],\quad \ell \in \mathcal{L},\ c_0 \triangleq \sigma_0^2 \|\mathbf{w}\|_2^2, \tag{68b}$$

$$\mathbf{X} \triangleq [\mathbf{X}_{ij}]_{i,j=1}^L, \mathbf{Y} \triangleq \mathsf{Diag}\big(\{\mathbf{Y}_1, \cdots, \mathbf{Y}_L\}\big),\ i, j \in \mathcal{L}, \tag{68c}$$

we can readily verify that the SNR can be rewritten as

$$\mathsf{SNR} = \frac{\mathbf{b}^H \mathbf{X} \mathbf{b}}{\mathbf{b}^H \mathbf{Y} \mathbf{b} + c_0}. \tag{69}$$

Firstly, we observe the fact that $\mathbf{X}$ has unit rank, i.e. $\mathbf{X} = \mathbf{d}\mathbf{d}^H$, with $\mathbf{d}$ defined as

$$\mathbf{d} \triangleq \begin{bmatrix} \mathbf{d}_1 \\ \vdots \\ \mathbf{d}_L \end{bmatrix} \triangleq \begin{bmatrix} \mathbf{a}_1^* \otimes \mathbf{H}_1^H \mathbf{w} \\ \vdots \\ \mathbf{a}_L^* \otimes \mathbf{H}_L^H \mathbf{w} \end{bmatrix}. \tag{70}$$

In fact, according to the definition of $\mathbf{X}_{ij}$ in (68a) and utilizing the identity $\big(\mathbf{AB}\big) \otimes \big(\mathbf{CD}\big) = \big(\mathbf{A} \otimes \mathbf{C}\big)\big(\mathbf{B} \otimes \mathbf{D}\big)$, we have

$$\mathbf{X}_{ij} = \big(\mathbf{a}_i^* \mathbf{a}_j^T\big) \otimes \big(\widetilde{\mathbf{H}}_i^H \mathbf{w}\mathbf{w}^H \widetilde{\mathbf{H}}_j\big) = \big(\mathbf{a}_i^* \otimes \widetilde{\mathbf{H}}_i^H \mathbf{w}\big)\big(\mathbf{a}_j^T \otimes \mathbf{w}^H \widetilde{\mathbf{H}}_j\big). \tag{71}$$

Then the j-th block column of $\mathbf{X}$ is given as

$$\mathbf{X}_{:j} = \begin{bmatrix} \mathbf{X}_{1j} \\ \vdots \\ \mathbf{X}_{Lj} \end{bmatrix} = \begin{bmatrix} \big(\mathbf{a}_1^* \otimes \widetilde{\mathbf{H}}_1^H \mathbf{w}\big)\big(\mathbf{a}_j^T \otimes \mathbf{w}^H \widetilde{\mathbf{H}}_j\big) \\ \vdots \\ \big(\mathbf{a}_L^* \otimes \widetilde{\mathbf{H}}_L^H \mathbf{w}\big)\big(\mathbf{a}_j^T \otimes \mathbf{w}^H \widetilde{\mathbf{H}}_j\big) \end{bmatrix} \tag{72}$$

$$= \begin{bmatrix} \big(\mathbf{a}_1^* \otimes \widetilde{\mathbf{H}}_1^H \mathbf{w}\big) \\ \vdots \\ \big(\mathbf{a}_L^* \otimes \widetilde{\mathbf{H}}_L^H \mathbf{w}\big) \end{bmatrix} \big(\mathbf{a}_j^T \otimes \mathbf{w}^H \widetilde{\mathbf{H}}_j\big) = \mathbf{d}\big(\mathbf{a}_j^T \otimes \mathbf{w}^H \widetilde{\mathbf{H}}_j\big),$$

where the last equality utilizes (70). Then the matrix $\mathbf{X}$ can be represented by packing all its column blocks as follows

$$\mathbf{X} = [\mathbf{X}_{:1}, \cdots, \mathbf{X}_{:L}] \tag{73}$$

$$= \Big[\mathbf{d}\big(\mathbf{a}_1^T \otimes \mathbf{w}^H \widetilde{\mathbf{H}}_1\big), \cdots, \mathbf{d}\big(\mathbf{a}_L^T \otimes \mathbf{w}^H \widetilde{\mathbf{H}}_L\big)\Big]$$

$$= \mathbf{d}\Big[\big(\mathbf{a}_1^T \otimes \mathbf{w}^H \widetilde{\mathbf{H}}_1\big), \cdots, \big(\mathbf{a}_L^T \otimes \mathbf{w}^H \widetilde{\mathbf{H}}_L\big)\Big] = \mathbf{d}\mathbf{d}^H.$$

Next we prove that $\mathbf{d} \in \mathsf{Range}(\mathbf{Y})$. Recall that $\mathbf{d}$ is defined in (70) and $\mathbf{Y} \triangleq \mathsf{Diag}\big(\{\mathbf{Y}_1, \cdots, \mathbf{Y}_L\}\big)$ with $\mathbf{Y}_\ell$ defined in (68b). Therefore it suffices to prove that $\exists \mathbf{z}_\ell$ such that $\mathbf{Y}_\ell \mathbf{z}_\ell = \mathbf{d}_\ell, \forall \ell \in \mathcal{L}$. Assume that $\boldsymbol{\Sigma}_\ell^*$ has the eigenvalue decomposition $\boldsymbol{\Sigma}_\ell^* = \mathbf{V}_\ell \boldsymbol{\Pi}_\ell \mathbf{V}_\ell^H$. Utilizing the identity $\big(\mathbf{AB}\big) \otimes \big(\mathbf{CD}\big) = \big(\mathbf{A} \otimes \mathbf{C}\big)\big(\mathbf{B} \otimes \mathbf{D}\big)$ we have $\mathbf{B}_\ell = \big(\mathbf{V}_\ell \otimes \mathbf{H}_\ell^H \mathbf{w}\big)\boldsymbol{\Pi}_\ell \big(\mathbf{V}_\ell \otimes \mathbf{H}_\ell^H \mathbf{w}\big)^H$. Thus $\mathsf{Range}(\mathbf{Y}_\ell) = \mathsf{span}(\mathbf{V}_\ell \otimes \mathbf{H}_\ell^H \mathbf{w})$. So it is equivalent to show $\exists \mathbf{y}_\ell$ such that $(\mathbf{V}_\ell \otimes \mathbf{H}_\ell^H \mathbf{w})\mathbf{y}_\ell = (\mathbf{V}_\ell \mathbf{y}_\ell) \otimes \big(\mathbf{H}_\ell^H \mathbf{w}\big) = \mathbf{a}_\ell \otimes \mathbf{H}_\ell^H \mathbf{w} = \mathbf{d}_\ell$. It suffices to show that $\exists \mathbf{y}_\ell$ such that $\mathbf{V}_\ell \mathbf{y}_\ell = \mathbf{a}_\ell$, which is obvious since $\mathbf{V}_\ell$ is unitary and $\mathbf{y}_\ell$ can be chosen as $\mathbf{y}_\ell = \mathbf{V}_\ell^H \mathbf{a}_\ell$.

In the sequel, we determine the optimal transmitter $\mathbf{b}^\star$, which maximizes SNR when transmission power is unlimited. Assume that $r_{\mathbf{Y}} \triangleq \mathsf{rank}(\mathbf{Y})$ and $\mathbf{Y}$ has the eigenvalue decomposition $\mathbf{Y} \triangleq \mathbf{U}\boldsymbol{\Xi}\mathbf{U}^H$, with $\boldsymbol{\Xi} \succ 0$ being of size $r_{\mathbf{Y}} \times r_{\mathbf{Y}}$. Since $\mathbf{d} \in \mathsf{Range}(\mathbf{Y})$ as shown above, components of $\mathbf{d}$ living in $\mathsf{Null}(\mathbf{Y})$ will not contribute to the numerator or denominator of the SNR. Therefore without loss of optimality we can assume $\mathbf{d} \in \mathsf{Range}(\mathbf{Y})$, i.e. $\mathbf{b} = \mathbf{U}\boldsymbol{\alpha}$. Since SNR increases when we increase scaling of $\mathbf{b}$, to study the supremum of SNR without power constraint, it is safe to drop the c_0 in the denominator

of SNR. Then the supremum of SNR with unlimited power can be expressed as

$$\sup_{\mathbf{b}} .\mathsf{SNR}(\mathbf{b}|\mathbf{w}) = \sup_{\mathbf{b}} . \frac{\mathbf{b}^H \mathbf{d}\mathbf{d}^H \mathbf{b}}{\mathbf{b}^H \mathbf{Y} \mathbf{b} + c_0} \tag{74}$$
$$\xrightarrow{\{\mathsf{P}_\ell\} \to \infty} \sup_{\mathbf{d}} . \frac{\mathbf{d}^H \mathbf{d}\mathbf{d}^H \mathbf{b}}{\mathbf{b}^H \mathbf{Y} \mathbf{b}} = \max_{\boldsymbol{\alpha}} . \frac{\boldsymbol{\alpha}^H \mathbf{U}^H \mathbf{d}\mathbf{d}^H \mathbf{U} \boldsymbol{\alpha}}{\boldsymbol{\alpha}^H \boldsymbol{\Xi} \boldsymbol{\alpha}}.$$

According to generalized eigenvalue decomposition, we can conclude that the maximum of the above fractional is $\mathbf{d}^H \mathbf{Y}^\dagger \mathbf{d}$ and is achieved when $\boldsymbol{\alpha}^\star = \boldsymbol{\Xi}^{-1} \mathbf{U}^H \mathbf{d}$, i.e. $\mathbf{b}^\star = \mathbf{U} \boldsymbol{\alpha}^\star = \mathbf{Y}^\dagger \mathbf{d}$, where $\mathbf{Y}^\dagger$ is the pseudo-inverse of $\mathbf{Y}$. It can be easily verified that $\mathbf{Y}^\dagger = \mathsf{Diag}\{\mathbf{Y}_1, \cdots, \mathbf{Y}_L\}^\dagger = \mathsf{Diag}\{\mathbf{Y}_1^\dagger, \cdots, \mathbf{Y}_L^\dagger\}$. Thus $\mathbf{d}^H \mathbf{Y}^\dagger \mathbf{d} = \sum_{\ell=1}^L \mathbf{d}_\ell^H \mathbf{Y}_\ell^\dagger \mathbf{d}_\ell^H$.

By the above assumption, the eigenvalue decomposition of $\mathbf{Y}_\ell$ can be determined as

$$\mathbf{Y}_\ell = \mathbf{Z}_\ell \Big(\|\mathbf{H}_\ell^H \mathbf{w}\|_2^2 \boldsymbol{\Pi}_\ell \Big) \mathbf{Z}_\ell^H, \tag{75}$$

where $\mathbf{Z}_\ell \triangleq \mathbf{V}_\ell \otimes \frac{\mathbf{H}_\ell^H \mathbf{w}}{\|\mathbf{H}_\ell^H \mathbf{w}\|_2}$ is a slim matrix having orthonormal columns. Therefore the eigenvalue decomposition of $\mathbf{Y}_\ell^\dagger$ can be readily obtained as

$$\mathbf{Y}_\ell^\dagger = \mathbf{Z}_\ell \Big(\|\mathbf{H}_\ell^H \mathbf{w}\|_2^{-2} \boldsymbol{\Pi}_\ell^{-1} \Big) \mathbf{Z}_\ell^H. \tag{76}$$

Recall that $\mathbf{d}_\ell = \mathbf{a}_\ell \otimes (\mathbf{H}_\ell^H \mathbf{w})$. Thus we can obtain

$$\mathbf{Z}_\ell^H \mathbf{d}_\ell = \Big[\mathbf{V}_\ell \otimes \frac{\mathbf{H}_\ell^H \mathbf{w}}{\|\mathbf{H}_\ell^H \mathbf{w}\|_2} \Big]^H \Big[\mathbf{a}_\ell^* \otimes (\mathbf{H}_\ell^H \mathbf{w}) \Big] \tag{77}$$
$$= \Big(\mathbf{V}_\ell^H \mathbf{a}_\ell^* \Big) \otimes \Big(\frac{\mathbf{w}^H \mathbf{H}_\ell}{\|\mathbf{H}_\ell^H \mathbf{w}\|_2} \mathbf{H}_\ell^H \mathbf{w} \Big) = \|\mathbf{H}_\ell^H \mathbf{w}\|_2 \mathbf{V}_\ell^H \mathbf{a}_\ell^*.$$

Combining the above results in (76) and (77) we obtain

$$\mathbf{d}_\ell^H \mathbf{Y}_\ell^\dagger \mathbf{d}_\ell = \mathbf{a}_\ell^T \mathbf{V}_\ell \boldsymbol{\Pi}_\ell^{-1} \mathbf{V}_\ell^H \mathbf{a}_\ell^* = \mathbf{a}_\ell^T \boldsymbol{\Sigma}_\ell^{-*} \mathbf{a}_\ell^*. \tag{78}$$

Summing up $\mathbf{d}_\ell^H \mathbf{Y}_\ell \mathbf{d}_\ell$ for all ℓ, we can obtain the maximal SNR. Then utilizing the well known result that $\mathsf{MSE}^\star = \big(1 + \mathsf{SNR}^\star\big)^{-1}$, we can obtain $\mathsf{MSE}_{\inf}$ in (58). □

REFERENCES

[1] I. F. Akyildiz, A. Kak, and S. Nie, "6g and beyond: The future of wireless communications systems," *IEEE Access*, vol. 8, pp. 133 995–134 030, 2020.

[2] *Introduction to Bluetooth Low Energy*, [Online], Available: https://cdn-learn . adafruit . com / downloads / pdf / introduction - to - bluetooth - low - energy.pdf, 2019.

[3] *Zigbee Specification*, [Online], Available: http://www.zigbee.org/wp-content/uploads/2014/11/docs-05-3474-20-0csg-zigbeespecification.pdf, 2012.

[4] M. B. Yassein, W. Mardini, and A. Khalil, "Smart homes automation using z-wave protocol," in *IEEE Int. Conf. Eng. MIS(ICEMIS)*, 2016.

[5] T. Adame, A. Bel, B. Bellalta, J. Barcelo, and M. Oliver, "IEEE802.11AH: The WiFi approach for M2M communications," *IEEE Wireless Commun.*, vol. 21, no. 6, pp. 144–152, Dec. 2014.

[6] L. Vangelista, A. Zanella, and M. Zorzi, "Long-range IoT technologies: The dawn of LoRa," *IEEE Trans. Wireless Commun.*, pp. 51–58, 2015, Cham, Switzerland: Springer.

[7] *SigFox—the global communications service provider for the Internet of Things (IoT)*, [Online], Available: http://www.sigfox.com/.

[8] J. Xiao, S. Cui, Z. Luo, and A. J. Goldsmith, "Linear coherent decentralized estimation," *IEEE Trans. Signal Process.*, vol. 56, no. 2, pp. 757–770, 2008.

[9] A. S. Behbahani, A. M. Eltawil, and H. Jafarkhani, "Decentralized estimation under correlated noise," *IEEE Trans. Signal Process.*, vol. 62, no. 21, pp. 5603–5614, 2014.

[10] F. Jiang, J. Chen, and A. L. Swindlehurst, "Optimal power allocation for parameter tracking in a distributed amplify-and-forward sensor network," *IEEE Trans. Signal Process.*, vol. 62, no. 9, pp. 2200–2211, 2014.

[11] J. Akhtar and K. Rajawat, "Distributed sequential estimation in wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 1, pp. 86–100, 2018.

[12] Y. Liu and J. Li, "Linear precoding to optimize throughput, power consumption and energy efficiency in MIMO wireless sensor networks," *IEEE Trans. Commun.*, vol. 66, no. 5, pp. 2122–2136, 2018.

[13] Y. Liu, J. Li, and H. Wang, "Robust linear beamforming in wireless sensor network," *IEEE Trans. Commun.*, vol. 67, no. 6, pp. 4450–4463, 2019.

[14] S. Zhang, S. Liu, V. Sharma, and P. K. Varshney, "Optimal sensor collaboration for parameter tracking using energy harvesting sensors," *IEEE Trans. Signal Process.*, vol. 66, no. 12, pp. 3339–3353, 2018.

[15] V. V. Mai, W. Shin, and K. Ishibashi, "Wireless power transfer for distributed estimation in sensor networks," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 3, pp. 549–562, 2017.

[16] L. Zhu, J. Zhang, Z. Xiao, X. Cao, D. O. Wu, and X.-G. Xia, "Millimeter-wave NOMA with user grouping, power allocation and hybrid beamforming," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5065–5079, 2019.

[17] I. Dey, D. Ciuonzo, and P. K. Varshney, "Wideband collaboratively spectrum sensing using massive MIMO decision fusion," *IEEE Trans. Wireless Commun.*, vol. 19, no. 8, pp. 5246–5260, Aug. 2020.

[18] K. P. Lee, A. Kumar, S. Srivastava, A. K. Jagannatham, and L. Hanzo, "Bayesian learning-based linear decentralized sparse parameter estimation in MIMO wireless sensor networks relying on imperfect csi," *IEEE Trans. Commun.*, 2021, Early Access.

[19] X. Wang, G. Li, and P. K. Varshney, "Detection of sparse stochastic signals with quantized measurements in sensor networks," *IEEE Trans. Signal Process.*, vol. 67, no. 8, pp. 2210–2220, Apr. 2019.

[20] X. Cheng, P. Khanduri, B. Chen, and P. K. Varshney, "Joint collaboration and compression design for distributed sequential estimation in a wireless wensor network," *IEEE Trans. Signal Process.*, vol. 69, pp. 5448–5462, Sep. 2021.

[21] X. Zhai, X. Chen, and Y. Cai, "Power minimization for massive MIMO over-the-air computation with two-timescale hybrid beamforming," *IEEE Wireless Commun. Lett.*, vol. 10, no. 4, pp. 873–877, Apr. 2021.

[22] X. Cao, G. Zhu, J. Xu, and K. Huang, "Optimized power control for over-the-air computation in fading channels," *IEEE Trans. Wireless Commun.*, vol. 19, no. 11, pp. 7498–7513, Nov. 2021.

[23] Z. Wang, Y. Shi, J. Xu, and K. Huang, "Wireless-powered over-the-air computation in intelligent reflecting surface-aided IoT networks," *IEEE Internet of Things J.*, vol. 8, no. 3, pp. 1585–1598, Feb. 2021.

[24] M. Shirvanimoghaddam, M. Dohler, and S. J. Johnson, "Massive nonorthogonal multiple access for cellular IoT: Potentials and limitations," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 55–61, Sep. 2017.

[25] M. R. Palattella, N. Accettura, X. Vilajosana, L. A. Grieco, G. Boggia, and M. Dohler, "Standardized protocol stack for the Internet of (important) Things," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 1389–1406, Sep. 2013, 3rd Quart.

[26] K. S. Adu-Manu, N. Adam, C. Tapparello, H. Ayatollahi, and W. Heinzelman, "Energy-harvesting wireless sensor networks (EH-WSNs): A review," *ACM Trans. Sensor Networks*, vol. 14, no. 2, pp. 1–50, 2018.

[27] "Connectivity preserving localized coverage algorithm for area monitoring using wireless sensor networks," *Computer Communications*, vol. 34, no. 12, pp. 1484–1496, 2011.

[28] "Active node determination for correlated data gathering in wireless sensor networks," *Computer Networks*, vol. 57, no. 5, pp. 1124–1138, 2013.

[29] *CC2652P SimpleLink Multiprotocol 2.4 GHz wireless MCU with integrated power amplifier, Texas Instrum.* [Online], Available: https://www.ti.com/document-viewer/CC2652P/datasheet, 2020.

[30] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," *IEEE Commun. Mag.*, vol. 58, no. 1, pp. 106–112, 2020.

[31] C. Pan, H. Ren, K. Wang, J. F. Kolb, M. Elkashlan, M. Chen, M. Di Renzo, Y. Hao, J. Wang, A. L. Swindlehurst, X. You, and L. Hanzo, "Reconfigurable intelligent surfaces for 6g systems: Principles, applications, and research directions," *IEEE Commun. Mag.*, vol. 59, no. 6, pp. 14–20, 2021.

[32] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4157–4170, 2019.

[33] Q. Wu and R. Zhang, "Beamforming optimization for wireless network aided by intelligent reflecting surface with discrete phase shifts," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1838–1851, 2020.

[34] Y. Liu, J. Zhao, M. Li, and Q. Wu, "Intelligent reflecting surface aided miso uplink communication network: Feasibility and power minimization for perfect and imperfect csi," *IEEE Trans. Commun.*, vol. 69, no. 3, pp. 1975–1989, Mar. 2021.

[35] X. Peng, Y. Shi, J. Zhang, and K. B. Letaief, "Layered group sparse beamforming for cache-enabled green wireless networks," *IEEE Trans. Commun.*, vol. 65, no. 12, pp. 5589–5603, 2017.

[36] Y. Shi, J. Cheng, J. Zhang, B. Bai, W. Chen, H. Zhou, and K. B. Letaief, "Smoothed l_p-minimization for green cloud-ran with user admission control," *IEEE J. Selected Area Commun.*, vol. 34, no. 4, pp. 1022–1035, Apr. 2016.

[37] M. Tao, E. Chen, H. Zhou, and W. Yu, "Content-centric sparse multicast beamforming for cache-enabled cloud ran," *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 6118–6131, Sep. 2016.

[38] D. Ba, B. Babadi, P. L. Purdon, and E. N. Brown, "Convergence and stability of iteratively re-weighted least squares algorithms," *IEEE Trans. Signal Process.*, vol. 62, no. 1, pp. 183–195, Feb. 2014.

[39] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Güntürk, "Iteratively reweighted least squares minimization for sparse recovery," *Commun. Pure Appl. Math.*, vol. 63, no. 1, pp. 1–38, Feb. 2010.

[40] Y. Sun, P. Babu, and D. P. Palomar, "Majorization minimization algorithms in signal processing, communications, and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 794–816, Feb. 2017.

[41] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM J. Optim.*, vol. 23, no. 2, pp. 1126–1153, 2013.

[42] M. Hong and *et. al.*, "A unified algorithmic framework for block-structured optimization involving big data: With applications in machine learning and signal processing," *IEEE Signal Process. Mag.*, vol. 33, no. 1, pp. 57–77, 2016.

[43] Q. Shi and M. Hong, "Penalty dual decomposition method for non-smooth nonconvex optimization—part i: Algorithms and convergence analysis," *IEEE Trans. Signal Process.*, vol. 68, pp. 4108–4122, Jun. 2020.

[44] M. Grant and S. Boyd, *CVX: Matlab software for disciplined convex programming, version 2.1*, http://cvxr.com/cvx, Mar. 2014.

[45] J. Chen, Y.-C. Liang, H. V. Cheng, and W. Yu, *Channel estimation for reconfigurable intelligent surface aided multi-user MIMO systems*, 2019. arXiv: 1912.03619.

[46] Z. Wang, L. Liu, and S. Cui, "Channel estimation for intelligent reflecting surface assisted multiuser communications: Framework, algorithms, and analysis," *IEEE Trans. Wireless Commun.*, Early Access.

[47] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice-Hall, 1993.

[48] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization, Analysis, Algorithms, and Engineering Applications*. Society for Industrial and Applied Mathematics (SIAM), 2001.

[49] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York: Cambridge University Press, 2004.