Consensus ADMM-Based Distributed Simultaneous Imaging & Communication ⋆

Nishant Mehrotra Ashutosh Sabharwal César A. Uribe

Department of Electrical and Computer Engineering, Rice University, Houston, TX, 77005, USA (e-mail: {nm30, ashu, cauribe}@rice.edu).

Abstract: This paper takes the first steps toward enabling wireless networks to perform both imaging and communication in a distributed manner. We propose *Distributed Simultaneous Imaging and Symbol Detection* (DSISD), a provably convergent distributed simultaneous imaging and communication scheme based on the alternating direction method of multipliers. We show that DSISD achieves similar imaging and communication performance as centralized schemes, with order-wise reduction in computational complexity. We evaluate the performance of DSISD via 2.4 GHz Wi-Fi simulations.

Keywords: ADMM, joint imaging and communication, MIMO, wireless imaging.

1. INTRODUCTION

In recent years, there has been growing interest in performing *imaging* with wireless networks usually tailored for communications (Vakalis et al., 2019; Guan et al., 2021). Wireless networks are ubiquitous and scalable by design, with powerful backend computing capabilities. This makes it possible to perform imaging alongside communication on a *network-wide* scale.

We study the problem of distributed simultaneous imaging and communication. We consider the system configuration shown in Fig. 1, with $N \geq 1$ multiple-input multiple-output (MIMO) base stations receiving transmissions from an uplink user. The base stations aim to: (i) perform uplink user symbol detection (communication) and (ii) estimate the reflective response of scatterers in the environment (imaging). We assume the base stations can cooperate over a static, undirected graph $\mathcal{G}$, which models limited capacity backhaul links between the base stations. Due to the limited capacity of backhaul links, the base stations must perform the two operations (communication and imaging) in a distributed manner with local processing.

We propose *Distributed Simultaneous Imaging and Symbol Detection* (DSISD), a provably convergent distributed algorithm for simultaneous imaging and communication. DSISD is the distributed variant of decode-and-image, a centralized algorithm previously proposed in (Mehrotra and Sabharwal, 2022). We show that DSISD achieves similar imaging and communication performance as decode-and-image, with $O(N)$ reduction in computational complexity. We numerically evaluate the performance of DSISD via 2.4 GHz Wi-Fi simulations.

To the best of our knowledge, we are the first to consider the problem of distributed simultaneous imaging and communication. Prior work has largely focused on communication-only (no imaging) and imaging-only (a-priori known communication data) problems. For instance, distributed algorithms have been proposed for beamforming, symbol detection, and interference alignment (Kumar and Rajawat, 2016; Chen and Tao, 2017; Li et al., 2017), and for radar imaging (Zabolotsky and Mavrychev, 2018; Hu et al., 2021). In the absence of priors on the scatterers in the environment and uplink data symbols, imaging-only and communication-only problems are convex. Nevertheless, we show that simultaneous imaging and communication corresponds to a bi-convex problem. We show that DSISD converges to the stationary points of the bi-convex problem with sublinear rate and guarantees asymptotic consensus across all N base stations.

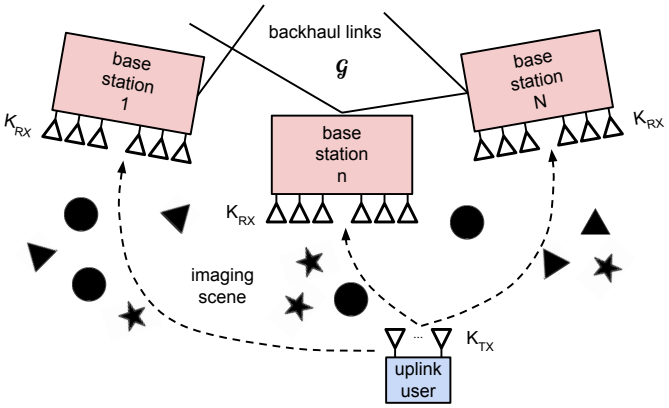


Fig. 1. System configuration with N base stations receiving transmissions from an uplink user. Base stations are connected over static, undirected backhaul network $\mathcal{G}$. Base stations aim to collaboratively image the scene and perform uplink user symbol detection.

This paper is organized as follows. In Section 2, we present the system model and problem formulation for Fig. 1. We present the proposed algorithm, DSISD, and associated theoretical results in Section 4. In Section 5, we show

⋆ This work was partially supported by the NSF Grant CNS-1956297, and the Sustainable Futures Fund #919027.

numerical results and simulations. We conclude the paper in Section 6 with some directions for future work.

2. SYSTEM MODEL

Consider the system shown in Fig. 1 with $N \geq 1$ base stations receiving uplink transmissions. We adopt the system model proposed in (Mehrotra and Sabharwal, 2022), where a similar configuration with a single base station was analyzed. In the sequel, we shall collectively refer to the set of scatterers that reflect uplink transmissions to the base stations as the *imaging scene*. We make the following assumptions about the system operation.

Assumption 1:

- Scatterers in the imaging scene remain static for T symbol durations, i.e., coherence interval.
- The uplink signalling is uni-polarized, with operating wavelength λ.
- All N base stations are equipped with equal number of receive antennas, K_{RX}, and operate synchronously over the same set of time-frequency resources.

Formally, let K_{TX} denote the number of transmit antennas at the uplink user, and let M denote the number of scatterers in the imaging scene. We denote the reflective response of the scatterers by a *scene reflectivity* vector $f \in \mathbb{C}^M$. Furthermore, let X_T denote the $K_{\mathsf{TX}} \times T$ matrix of transmitted uplink symbols, and $N_T^{(n)}$ denote the $K_{\mathsf{RX}} \times T$ matrix of additive noise at the n-th base station, for all $n \in \{1, \cdots, N\}$. Then, within a coherence interval of T symbol durations, the $K_{\mathsf{RX}} \times T$ matrix of receive symbols at the n-th base station is given by

$$Y_T^{(n)} = \underbrace{P_{\mathsf{RX}}^{(n)} \mathsf{diag}\big(f\big)\big(P_{\mathsf{TX}}\big)^\top}_{H_{\mathsf{comm}}^{(n)}} X_T + N_T^{(n)}, \tag{1}$$

where the two matrices P_{TX} and $P_{\mathsf{RX}}^{(n)}$ have sizes $K_{\mathsf{TX}} \times M$ and $K_{\mathsf{RX}} \times M$ respectively The $\big(k, m\big)$-th element of each path delay matrix is a scaled complex exponential that depends on the signalling wavelength λ, and the locations of the k-th antenna and m-th scatterer in the scene,

$$P(k,m) = \frac{\exp(-j\big(\frac{2\pi}{\lambda}\big)\left\|r_k - \tilde{r}_m\right\|_2)}{4\pi\left\|r_k - \tilde{r}_m\right\|_2}, \ \forall P \in \{P_{\mathsf{TX}}, P_{\mathsf{RX}}^{(n)}\},$$

where r_k and $\tilde{r}_m$ denote the position vectors of the k-th antenna and m-th scatterer in the imaging scene.

In the next section, we formulate the distributed simultaneous imaging and communication problem.

3. PROBLEM FORMULATION

Given the system model in (1), the base stations aim to collaboratively perform two functions:

(1) **Imaging:** Estimate the reflectivity vector f, and
(2) **Communication:** Estimate the uplink symbols X_T.

Our goal is to enable both functionalities with only *local* processing of the received symbols $Y_T^{(n)}$ at every base station $n \in \{1, \cdots, N\}$. Specifically, we make the following assumption on the prior knowledge at the base stations.

Assumption 2: The n-th base station has *local* knowledge of received symbols $Y_T^{(n)}$, path delay matrices $\{P_{\mathsf{TX}},\ P_{\mathsf{RX}}^{(n)}\}$ and discrete set $\mathcal{X}_{T-T_1}$ from which uplink data symbols X_{T-T_1} are drawn. Examples of $\mathcal{X}_T$ are $\mathcal{X}_T = \{\pm 1\}^{K_{\mathsf{TX}} \times T}$ and $\mathcal{X}_T = \{(\pm 1 \pm j)/\sqrt{2}\}^{K_{\mathsf{TX}} \times T}$ for binary and quadrature phase-shift keying (BPSK and QPSK).

We assume the first T_1 symbols in $X_T = \big[X_{T_1}, X_{T-T_1}\big]$ to be pilot symbols *known* to all N base stations. The base stations thus aim to collaboratively solve the problem:

$$\min_{X_{T-T_1} \in \mathcal{X}_{T-T_1}, f} \ \tfrac{1}{2} \sum_{n=1}^{N} \left\| Y_T^{(n)} - P_{\mathsf{RX}}^{(n)} \mathsf{diag}\big(f\big)\big(P_{\mathsf{TX}}\big)^\top X_T \right\|_F^2. \tag{P1}$$

Problem (P1) is an *integer constrained bi-convex* problem, since the objective function is non-convex and non-separable in f and X_{T-T_1}, but convex in each variable separately. We simplify the problem by relaxing the integer constraints, i.e., solve the unconstrained problem.

To formulate the distributed version of (P1), we assume the base stations are connected over a backhaul network, modeled as a graph $\mathcal{G}$.

Assumption 3: The graph $\mathcal{G} = \big(\mathcal{V}, \mathcal{E}\big)$ is static and undirected, and every node knows who its neighbors are.

The distributed problem is formulated by defining local variables $f^{(n)}$ and $X_{T-T_1}^{(n)}$, with the constraint that local solutions of nodes connected by an edge in $\mathcal{E}$ are identical,

$$\min_{X_{T-T_1}^{(n)}, f^{(n)}} \ \tfrac{1}{2} \sum_{n=1}^{N} \left\| Y_T^{(n)} - P_{\mathsf{RX}}^{(n)} \mathsf{diag}\big(f^{(n)}\big)\big(P_{\mathsf{TX}}\big)^\top X_T^{(n)} \right\|_F^2$$
$$\text{s.t. } X_{T-T_1}^{(i)} = X_{T-T_1}^{(j)}, \ f^{(i)} = f^{(j)}, \ \forall \big(i,j\big) \in \mathcal{E}. \tag{P2}$$

After solving (P2), nodes perform symbol detection via zero-forcing (Tse and Viswanath, 2005),

$$\hat{X}_{T-T_1}^{(n)} = \Pi_{\mathcal{X}_{T-T_1}}\big(X_{T-T_1}^{(n)}\big), \ \forall n \in \mathcal{V},$$

where $\Pi_{\mathcal{S}}\big(\,\cdot\,\big)$ denotes projection onto set $\mathcal{S}$.

4. MAIN RESULTS

We begin by analyzing first-order optimality conditions in Section 4.1 to characterize the number of uplink pilots T_1 required to solve (P1). We subsequently present the proposed DSISD algorithm in Section 4.2, and derive associated convergence guarantees in Section 4.3.

4.1 Optimality Conditions for Problem (P1)

Let the objective function in (P1) be denoted by

$$\mathcal{L}\big(f, X_{T-T_1}\big) = \frac{1}{2} \sum_{n=1}^{N} \left\| Y_T^{(n)} - P_{\mathsf{RX}}^{(n)} \mathsf{diag}\big(f\big)\big(P_{\mathsf{TX}}\big)^\top X_T \right\|_F^2.$$

The KKT conditions for (P1) correspond to

$$\nabla_f \mathcal{L}\big(f^*, X_{T-T_1}^*\big) = 0, \ \nabla_{X_{T-T_1}} \mathcal{L}\big(f^*, X_{T-T_1}^*\big) = 0,$$

where f^* and $X_{T-T_1}^*$ are minimizers of (P1). Hence, the stationary points of (P1) are given by

$$f^* = \big(\ H_{\mathsf{img}}|_{X_{T-T_1} = X_{T-T_1}^*}\ \big)^\dagger \mathsf{vec}\big(Y_T\big), \tag{2}$$

$$X_{T-T_1}^* = \big(\ H_{\mathsf{comm}}|_{f = f^*}\ \big)^\dagger Y_{T-T_1}, \tag{3}$$

where $(\cdot)^\dagger$ denotes the pseudo-inverse and $\mathsf{vec}(\cdot)$ denotes the vectorization operator. The matrix H_{img} is given by

$$H_{\mathsf{img}} = \big(X_T^\top \otimes I_{NK_{\mathsf{RX}}}\big)\big(P_{\mathsf{TX}} * P_{\mathsf{RX}}\big),$$

Algorithm 1 *Distributed Simultaneous Imaging & Symbol Detection* (DSISD)

Input: P_{TX}, $P_{\mathsf{RX}}^{(n)}$, $Y_T^{(n)}$, $\mathcal{X}_{T-T_1}$, X_{T_1}, K, ρ_X, ρ_f

for node $n = 1$ **to** N in $\mathcal{V}$ {in parallel} **do**
  Initialize $f_0^{(n)} = D_{f,0}^{(n)} = 0, X_{T-T_1,0}^{(n)} = D_{X,0}^{(n)} = 0$
  for iteration $k = 1$ **to** K **do**
    Transmit $f_{k-1}^{(n)}$, $X_{T-T_1,k-1}^{(n)}$ to $\{m : (n,m) \in \mathcal{E}\}$
    Receive $\{f_{k-1}^{(m)},\ X_{T-T_1,k-1}^{(m)},\ \forall m : (n,m) \in \mathcal{E}\}$
    Update reflectivities $f_k^{(n)}$ by solving

$$\min_f \left\{ \begin{array}{l} \frac{\rho_f}{2}\big\| \sum_m L_{\mathcal{G}}(n,m) f_{k-1}^{(m)} + L_{\mathcal{G}}(n,n) f + D_{f,k-1}^{(n)} \big\|_2^2 \\ + \frac{1}{2}\big\| Y_T^{(n)} - P_{\mathsf{RX}}^{(n)} \mathsf{diag}\big(f\big) \big(P_{\mathsf{TX}}\big)^\top X_{T,k-1}^{(n)} \big\|_F^2 \end{array} \right\}$$

    Update uplink symbols $X_{T-T_1,k}^{(n)}$ by solving

$$\min_X \left\{ \begin{array}{l} \frac{\rho_X}{2}\big\| \sum_m L_{\mathcal{G}}(n,m) X_{T-T_1,k-1}^{(m)} + L_{\mathcal{G}}(n,n) X + D_{X,k-1}^{(n)} \big\|_2^2 \\ + \frac{1}{2}\big\| Y_T^{(n)} - P_{\mathsf{RX}}^{(n)} \mathsf{diag}\big(f_k^{(n)}\big) \big(P_{\mathsf{TX}}\big)^\top X \big\|_F^2 \end{array} \right\}$$

    Update dual variables via

$$D_{f,k}^{(n)} = D_{f,k-1}^{(n)} + \sum_m L_{\mathcal{G}}(n,m) f_k^{(m)} + L_{\mathcal{G}}(n,n) f_k^{(n)}$$

$$D_{X,k}^{(n)} = D_{X,k-1}^{(n)} + \sum_m L_{\mathcal{G}}(n,m) X_{T-T_1,k}^{(m)} + L_{\mathcal{G}}(n,n) X_{T-T_1,k}^{(n)}$$

  end for
end for
Output: $\hat{f} = f_K^{(N)}$, $\hat{X}_{T-T_1} = \Pi_{\mathcal{X}_{T-T_1}}\big(X_{T-T_1,K}^{(N)}\big)$

where $\otimes$ and $*$ denote the Kronecker and column-wise Khatri-Rao products, and I_n denotes the identity matrix of size $n \times n$. Matrices P_{RX}, H_{comm} and Y_T are concatenations of local path delay, channel and received symbol matrices across all N base stations.

Note that the values of (2) and (3) are coupled. In the following lemma, we characterize conditions under which stationary points satisfying (2) and (3) are *decoupled*.

Lemma 1. *Let $N_T = 0$. If $X_{T_1}\big(X_{T_1}^*\big)^\dagger = I_{K_{\mathsf{TX}}}$, or equivalently, $T_1 \geq K_{\mathsf{TX}}$ for time-orthogonal pilots, then* (2) *and* (3) *are decoupled.*

Proof. In the absence of noise, f^* from (2) is given by

$$f^* = \big(P_{\mathsf{TX}} * P_{\mathsf{RX}}\big)^\dagger \Big(\big(X_T\big(X_T^*\big)^\dagger\big)^\top \otimes I_{NK_{\mathsf{RX}}}\Big)\big(P_{\mathsf{TX}} * P_{\mathsf{RX}}\big) f.$$

The above expression is decoupled with the recovered uplink symbols X_T^* when $X_T\big(X_T^*\big)^\dagger = I_{K_{\mathsf{TX}}}$. Since data portion of X_T and recovered symbols X_T^* may differ, a sufficient condition is thus $X_{T_1}\big(X_{T_1}^*\big)^\dagger = I_{K_{\mathsf{TX}}}$. For time-orthogonal pilot symbols, this condition is equivalent to

$$\mathsf{rank}\big(X_{T_1}\big) = \min\big\{T_1, K_{\mathsf{TX}}\big\} \geq K_{\mathsf{TX}} \implies T_1 \geq K_{\mathsf{TX}}.$$

□

Following Lemma 1, we assume $T_1 = K_{\mathsf{TX}}$ in all subsequent discussion. We now present the proposed DSISD algorithm for solving (P2).

4.2 Distributed Simultaneous Imaging & Symbol Detection

The bi-convexity of the objective function $\mathcal{L}\big(f, X_{T-T_1}\big)$ naturally motivates using an *alternating* procedure to solve (P2). The proposed DSISD algorithm (Algorithm 1) is thus based on consensus ADMM (Boyd et al., 2011).

Let the concatenation of local solutions across all N nodes be denoted by variables $\overline{X}^\top = \big[\big(X_{T-T_1}^{(1)}\big)^\top, \cdots, \big(X_{T-T_1}^{(N)}\big)^\top\big]$ and $\overline{f}^\top = \big[\big(f^{(1)}\big)^\top, \cdots, \big(f^{(N)}\big)^\top\big]$. Moreover, let

$$g\big(\overline{X}, \overline{f}\big) = \tfrac{1}{2} \sum_{n=1}^N \big\| Y_T^{(n)} - P_{\mathsf{RX}}^{(n)} \mathsf{diag}\big(f^{(n)}\big)\big(P_{\mathsf{TX}}\big)^\top X_T^{(n)} \big\|_F^2$$

denote the objective function in (P2). Finally, $L_{\mathcal{G}} \in \mathbb{R}^{N \times N}$ denotes the Laplacian matrix for $\mathcal{G}$.

With above defined notation, (P2) may be recast as

$$\begin{aligned} &\min_{\overline{X}, \overline{f}}\ g\big(\overline{X}, \overline{f}\big) \\ &\text{s.t. } \big(L_{\mathcal{G}} \otimes I_{K_{\mathsf{TX}}}\big)\overline{X} = 0,\ \big(L_{\mathcal{G}} \otimes I_M\big)\overline{f} = 0. \end{aligned} \tag{P3}$$

At every iteration $k \in \{0, 1, \cdots\}$, DSISD performs the following ADMM updates:

$$\overline{f}^{k+1} = \arg\min_{\overline{f}} \mathcal{A}\big(\overline{X}^k, \overline{f}, \overline{D}_X^k, \overline{D}_f^k\big), \tag{A1}$$

$$\overline{X}^{k+1} = \arg\min_{\overline{X}} \mathcal{A}\big(\overline{X}, \overline{f}^{k+1}, \overline{D}_X^k, \overline{D}_f^k\big), \tag{A2}$$

$$\overline{D}_f^{k+1} = \overline{D}_f^k + \big(L_{\mathcal{G}} \otimes I_M\big)\overline{f}^{k+1}. \tag{A3}$$

$$\overline{D}_X^{k+1} = \overline{D}_X^k + \big(L_{\mathcal{G}} \otimes I_{K_{\mathsf{TX}}}\big)\overline{X}^{k+1}, \tag{A4}$$

where $\mathcal{A}\big(\overline{X}, \overline{f}, \overline{D}_X, \overline{D}_f\big)$ denotes the augmented Lagrangian,

$$\begin{aligned} \mathcal{A}\big(\overline{X}, \overline{f}, \overline{D}_X, \overline{D}_f\big) &= g\big(\overline{X}, \overline{f}\big) + \frac{\rho_f}{2}\big\|\big(L_{\mathcal{G}} \otimes I_M\big)\overline{f}\big\|_F^2 \\ &+ \rho_X \big\langle \overline{D}_X, \big(L_{\mathcal{G}} \otimes I_{K_{\mathsf{TX}}}\big)\overline{X} \big\rangle + \rho_f \big\langle \overline{D}_f, \big(L_{\mathcal{G}} \otimes I_M\big)\overline{f} \big\rangle \\ &+ \frac{\rho_X}{2}\big\|\big(L_{\mathcal{G}} \otimes I_{K_{\mathsf{TX}}}\big)\overline{X}\big\|_F^2. \end{aligned}$$

The variables $\overline{D}_X$ and $\overline{D}_f$ denote scaled dual variables, whereas ρ_X and ρ_f denote penalty parameters.

The updates in (A1) to (A4) can be performed locally at every base station since $g\big(\overline{X}, \overline{f}\big)$ and $\mathcal{A}\big(\overline{X}, \overline{f}, \overline{D}_X, \overline{D}_f\big)$ are separable across pairs of local solutions $\big(f^{(n)}, X_T^{(n)}\big)$. Below, we derive performance guarantees for DSISD.

4.3 Performance Guarantees for DSISD

In Theorem 1 and Corollary 1, we characterize convergence guarantees and algorithm complexity for DSISD.

Theorem 1. *Let Assumptions 1, 2, 3 and the conditions in Lemma 1 hold. Then, the updates of DSISD satisfy the following properties:*

Asymptotic consensus: *All N nodes in $\mathcal{G}$ are in consensus asymptotically, i.e.,*

$$\lim_{k \to \infty} \big\|\big(L_{\mathcal{G}} \otimes I_M\big)\overline{f}^{k+1}\big\|_2 = 0,$$

$$\lim_{k \to \infty} \big\|\big(L_{\mathcal{G}} \otimes I_{K_{\mathsf{TX}}}\big)\overline{X}^{k+1}\big\|_F = 0.$$

Convergence to stationary points: *Limit points of iterates $\big(\overline{X}^k, \overline{f}^k, \overline{D}_X^k, \overline{D}_f^k\big)$ converge to a KKT point of* (P3).

Sublinear convergence rate: *Iterates* $(\overline{X}^k, \overline{f}^k, \overline{D}_X^k, \overline{D}_f^k)$ *converge to a KKT point of* (P3) *with rate* $O(1/K)$ *in terms of the optimality gap*

$$Q(\overline{X}^{k+1}, \overline{f}^{k+1}, \overline{D}_X^k, \overline{D}_f^k) = \big\|(L_{\mathcal{G}} \otimes I_{K_{\mathsf{TX}}})\overline{X}^{k+1}\big\|_F^2 + \big\|\nabla_{\overline{f}}\mathcal{A}(\overline{X}^{k+1}, \overline{f}^{k+1}, \overline{D}_X^k, \overline{D}_f^k)\big\|_F^2 + \big\|(L_{\mathcal{G}} \otimes I_M)\overline{f}^{k+1}\big\|_2^2 + \big\|\nabla_{\overline{X}}\mathcal{A}(\overline{X}^{k+1}, \overline{f}^{k+1}, \overline{D}_X^k, \overline{D}_f^k)\big\|_F^2.$$

Proof. Due to space constraints, we only provide a proof sketch with key steps and ideas for the complete proof.

Result 1: Asymptotic Consensus

We seek to prove

$$\lim_{k\to\infty} \big\|\overline{D}_f^{k+1} - \overline{D}_f^k\big\|_2 = 0, \ \lim_{k\to\infty} \big\|\overline{D}_X^{k+1} - \overline{D}_X^k\big\|_2 = 0.$$

We show the above result via two intermediate steps.

Step 1: We first derive upper bounds on successive dual update norms, $\big\|\overline{D}_f^{k+1} - \overline{D}_f^k\big\|_2$ and $\big\|\overline{D}_X^{k+1} - \overline{D}_X^k\big\|_2$. Consider the first-order optimality condition for (A2),

$$\nabla_{\overline{X}} g(\overline{X}^{k+1}, \overline{f}^{k+1}) + \rho_X \big(L_{\mathcal{G}}^\top \otimes I_{K_{\mathsf{TX}}}\big)\overline{D}_X^{k+1} = 0.$$

Thus, an upper bound on $\big\|\overline{D}_X^{k+1} - \overline{D}_X^k\big\|_F^2$ is given by

$$\big\|\overline{D}_X^{k+1} - \overline{D}_X^k\big\|_F^2 \leq \frac{1/\rho_X^2}{\lambda_{\min}\big(L_{\mathcal{G}}^\top L_{\mathcal{G}}\big)} \big\|\nabla_{\overline{X}} g(\overline{X}^k, \overline{f}^k) - \nabla_{\overline{X}} g(\overline{X}^{k+1}, \overline{f}^{k+1})\big\|_F^2,$$

where we have used the fact that $\frac{1}{2}\big\|(L_{\mathcal{G}} \otimes I_{K_{\mathsf{TX}}})\overline{X}\big\|_F^2$ is $\lambda_{\min}\big(L_{\mathcal{G}}^\top L_{\mathcal{G}}\big)$-strongly convex.

On appropriate substitutions, the right hand side can be upper bounded in terms of successive *primal* update norms, $\big\|\overline{X}^{k+1} - \overline{X}^k\big\|_F^2$, $\big\|\overline{f}^{k+1} - \overline{f}^k\big\|_F^2$. Thus, we next show $\lim_{k\to\infty} \big\|\overline{f}^{k+1} - \overline{f}^k\big\|_2 = 0$ and $\lim_{k\to\infty} \big\|\overline{X}^{k+1} - \overline{X}^k\big\|_F = 0$.

Step 2: We equivalently show $\sum_{k=1}^{\infty} \big\|\overline{X}^{k+1} - \overline{X}^k\big\|_F^2 < \infty$ and $\sum_{k=1}^{\infty} \big\|\overline{f}^{k+1} - \overline{f}^k\big\|_2^2 < \infty$ via two sub-results.

(i) First, we show that the augmented Lagrangian is upper bounded in terms of the primal update norms as

$$\mathcal{A}(\overline{X}^{k+1}, \overline{f}^{k+1}, \overline{D}_X^k, \overline{D}_f^k) - \mathcal{A}(\overline{X}^k, \overline{f}^k, \overline{D}_X^k, \overline{D}_f^k) \leq -\frac{\lambda_{\min}\big(L_{\mathcal{G}}^\top L_{\mathcal{G}}\big)}{2}\big(\rho_X \big\|\overline{X}^{k+1} - \overline{X}^k\big\|_F^2 + \rho_f \big\|\overline{f}^{k+1} - \overline{f}^k\big\|_2^2\big).$$

We show the above by upper bounding the left hand side and using the first-order optimality conditions corresponding to (A1) and (A2), the convexity of the objective function $g(\overline{X}, \overline{f})$ in $\overline{X}$ (resp. $\overline{f}$) given fixed $\overline{f}$ (resp. $\overline{X}$), as well as the $\lambda_{\min}\big(L_{\mathcal{G}}^\top L_{\mathcal{G}}\big)$ strong convexity of the terms $\frac{1}{2}\big\|(L_{\mathcal{G}} \otimes I_{K_{\mathsf{TX}}})\overline{X}\big\|_F^2$ and $\frac{1}{2}\big\|(L_{\mathcal{G}} \otimes I_{K_{\mathsf{M}}})\overline{f}\big\|_2^2$.

(ii) Next, we show that the augmented Lagrangian is lower bounded at every iteration, i.e.,

$$\sum_{k=1}^{K} \mathcal{A}(\overline{X}^{k+1}, \overline{f}^{k+1}, \overline{D}_X^k, \overline{D}_f^k) > -\infty,$$
$$\sum_{k=1}^{K} \mathcal{A}(\overline{X}^k, \overline{f}^k, \overline{D}_X^k, \overline{D}_f^k) > -\infty.$$

To show the above result, we utilize the dual updates in (A3) and (A4), as well as the equality

$$\langle C, D\rangle = \frac{1}{2}\Big(\|C + D\|_F^2 - \|C\|_F^2 - \|D\|_F^2\Big),$$

for any arbitrary C and D.

On taking the limit $K \to \infty$ in (i) and (ii) above, we obtain $\sum_{k=1}^{\infty} \big\|\overline{X}^{k+1} - \overline{X}^k\big\|_F^2 < \infty$ and $\sum_{k=1}^{\infty} \big\|\overline{f}^{k+1} - \overline{f}^k\big\|_2^2 < \infty$.

Result 2: Convergence to Stationary Points

The KKT conditions corresponding to (P3) are

$$\nabla_{\overline{X}} g(\overline{X}^*, \overline{f}^*) + \rho_X \big(L_{\mathcal{G}}^\top \otimes I_{K_{\mathsf{TX}}}\big)\overline{D}_X^* = 0,$$
$$\nabla_{\overline{f}} g(\overline{X}^*, \overline{f}^*) + \rho_X \big(L_{\mathcal{G}}^\top \otimes I_M\big)\overline{D}_f^* = 0,$$
$$\big(L_{\mathcal{G}} \otimes I_{K_{\mathsf{TX}}}\big)\overline{X}^* = 0, \ \big(L_{\mathcal{G}} \otimes I_M\big)\overline{f}^* = 0.$$

We aim to show that the limit points corresponding to Algorithm 1 satisfy the above KKT conditions. To that end, observe that in the limit $k \to \infty$, the update steps (A1) and (A2) satisfy

$$\nabla_{\overline{X}}\mathcal{A} = 0 \implies \nabla_{\overline{X}} g(\overline{X}^*, \overline{f}^*) + \rho_X \big(L_{\mathcal{G}}^\top \otimes I_{K_{\mathsf{TX}}}\big)\overline{D}_X^* = 0,$$
$$\nabla_{\overline{f}}\mathcal{A} = 0 \implies \nabla_{\overline{f}} g(\overline{X}^*, \overline{f}^*) + \rho_f \big(L_{\mathcal{G}}^\top \otimes I_M\big)\overline{D}_f^* = 0.$$

In addition, since $\lim_{k\to\infty} \big\|(L_{\mathcal{G}} \otimes I_M)\overline{f}^{k+1}\big\|_2 = 0$ and $\lim_{k\to\infty} \big\|(L_{\mathcal{G}} \otimes I_{K_{\mathsf{TX}}})\overline{X}^{k+1}\big\|_F = 0$ (c.f., Result 1), convergence to stationary points follows.

Result 3: Sublinear Convergence Rate

To derive convergence rates, we bound the optimality gap defined in the theorem statement. Per Result 1, for large enough $\xi > 0$, the optimality gap is upper bounded as

$$Q(\overline{X}^{k+1}, \overline{f}^{k+1}, \overline{D}_X^k, \overline{D}_f^k) \leq \xi\big\|\overline{X}^{k+1} - \overline{X}^k\big\|_F^2 + \xi\big\|\overline{f}^{k+1} - \overline{f}^k\big\|_F^2.$$

Moreover, for some large enough $\nu > 0$, Result 1 implies

$$\mathcal{A}(\overline{X}^{k+1}, \overline{f}^{k+1}, \overline{D}_X^k, \overline{D}_f^k) - \mathcal{A}(\overline{X}^k, \overline{f}^k, \overline{D}_X^k, \overline{D}_f^k) \leq -\nu\big\|\overline{X}^{k+1} - \overline{X}^k\big\|_F^2 - \nu\big\|\overline{f}^{k+1} - \overline{f}^k\big\|_F^2.$$

On averaging over indices $k = 1, \cdots, K$, we obtain

$$\frac{1}{K}\sum_{k=1}^{K} Q(\overline{X}^{k+1}, \overline{f}^{k+1}, \overline{D}_X^k, \overline{D}_f^k) \leq \frac{\xi}{K\nu}\big(\mathcal{A}(\overline{X}^1, \overline{f}^1, \overline{D}_X^1, \overline{D}_f^1) - \mathcal{A}(\overline{X}^{K+1}, \overline{f}^{K+1}, \overline{D}_X^1, \overline{D}_f^1)\big).$$

In other words, the convergence in terms of the optimality gap is sublinear with rate $O(1/K)$. □

Remark 1. *To the best of our knowledge, we are not aware of any existing convergence analysis for ADMM that is directly applicable to our problem, which is distributed, integer constrained, bi-convex, and non-separable. On the one hand, in (Xu and Yin, 2013), the authors consider the* centralized *bi-convex and non-separable function class and show sublinear convergence assuming the objective function satisfies the Kurdyka–Łojasiewicz (KL) inequality. On the other hand, in (Hong et al., 2016), the authors consider the distributed non-convex but* separable *function class and show convergence to stationary points. Neither*

analysis applies to our problem. However, in the absence of integer constraints, our problem is closely related to matrix factorization (MF). Hence, we have adapted prior convergence analysis for MF (Hajinezhad et al., 2016; Hajinezhad and Shi, 2018; Hong et al., 2017) to our problem. Since our problem is not identical to MF, our results have certain minor differences. Specifically, convergence to stationary points for MF only holds under certain regimes on the ADMM penalty parameters (e.g., $\rho > 1$). No such conditions are required in our results.

Remark 2. *We remark that the applicability of ADMM to bi-convex problems is well-known from (Boyd et al., 2011). Since strong duality does not hold in bi-convex problems, we can only demonstrate approximate convergence to KKT points. To do so, we have used the optimality gap function from (Hajinezhad et al., 2016; Hajinezhad and Shi, 2018; Hong et al., 2017) since it quantifies both first-order optimality conditions and the consensus error.*

We now characterize the algorithm complexity for DSISD.

Corollary 1. *Let the same assumptions as in Theorem 1 hold, and let $\epsilon > 0$ be a desired accuracy. Then, the algorithm complexity for DSISD is $O\big(2|\mathcal{E}|\big(M+K_{\mathsf{TX}}(T-T_1)+\big(M+K_{\mathsf{RX}}T\big)\big(MNK_{\mathsf{RX}}T\big)\big)/\epsilon\big)$.*

Proof. The communication complexity is given by

$$O\big(2|\mathcal{E}|\big(M+K_{\mathsf{TX}}(T-T_1)\big)/\epsilon\big),$$

which corresponds to the total number of real entries transferred across the network $\mathcal{G}$ over $K = O\big(1/\epsilon\big)$ iterations as per the convergence rate in Theorem 1.

The computational complexity corresponds to the total cost of least-squares updates in every iteration and equals

$$O\big(N\big(M\big(K_{\mathsf{RX}}T\big)^2+M^2K_{\mathsf{RX}}T+K_{\mathsf{RX}}K_{\mathsf{TX}}^2+K_{\mathsf{TX}}K_{\mathsf{RX}}^2\big)/\epsilon\big).$$

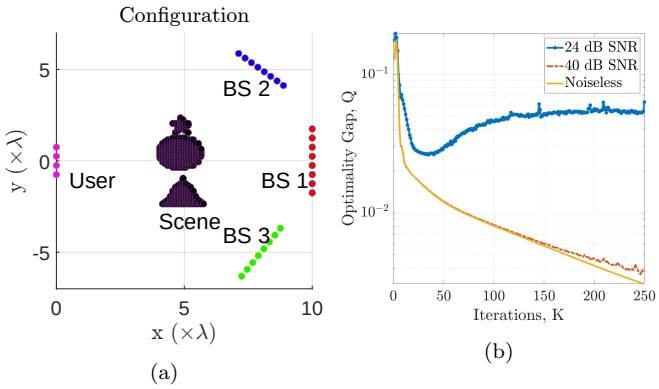
Assuming the number of scatterers in the scene largely dominates the number of transmitting or receiving antennas, i.e., $M \gg K_{\mathsf{TX}}$, K_{RX}, the algorithm complexity is given by the statement in the corollary. □

For comparison, consider decode-and-image. The communication complexity corresponds to every node transferring its $K_{\mathsf{RX}} \times T$ matrix of received symbols $Y_T^{(n)}$ to the central server, and thus equals $O(NK_{\mathsf{RX}}T)$. The computational complexity corresponds to least-squares updates over $N\times$ larger matrices, and equals $O(K_{\mathsf{RX}}NTM(M + K_{\mathsf{RX}}NT))$. Hence, DSISD has $O(N)$ smaller computational complexity compared to decode-and-image.

In the next section, we numerically evaluate the performance of DSISD and compare it with decode-and-image.

5. NUMERICAL EVALUATION

We simulate the 2D planar configuration shown in Fig. 2(a) with $N = 3$ base stations receiving data transmissions from an uplink user in the 2.4 GHz Wi-Fi band ($\lambda = 0.125$ m). We assume uncoded BPSK uplink transmissions, i.e., $\mathcal{X}_T = \{\pm 1\}^{K_{TX}\times T}$, with coherence interval $T = 100$ symbols. The uplink user is equipped with a uniform linear array (ULA) with $K_{\mathsf{TX}} = 4$ antennas. The base stations are each equipped with a ULA with $K_{\mathsf{RX}} = 8$ antennas, and are oriented at angles $0°$, $-30°$ and $45°$ with respect to the uplink user's array.



(a) (b)

Fig. 2. (a) Simulated configuration. (b) DSISD converges sublinearly in terms of optimality gap Q for high receive signal-to-noise ratio (SNR) values.

For simplicity, we consider a star topology $\mathcal{G}$, where the base stations are all connected to a fusion node (not illustrated in Fig. 2(a)). The fusion node only processes the consensus information, i.e., local solutions $f_k^{(n)}$ and $X_{T-T_1,k}^{(n)}$, exchanged by base stations, and does not process the received symbols $Y_T^{(n)}$ directly. The updates at the fusion node correspond to

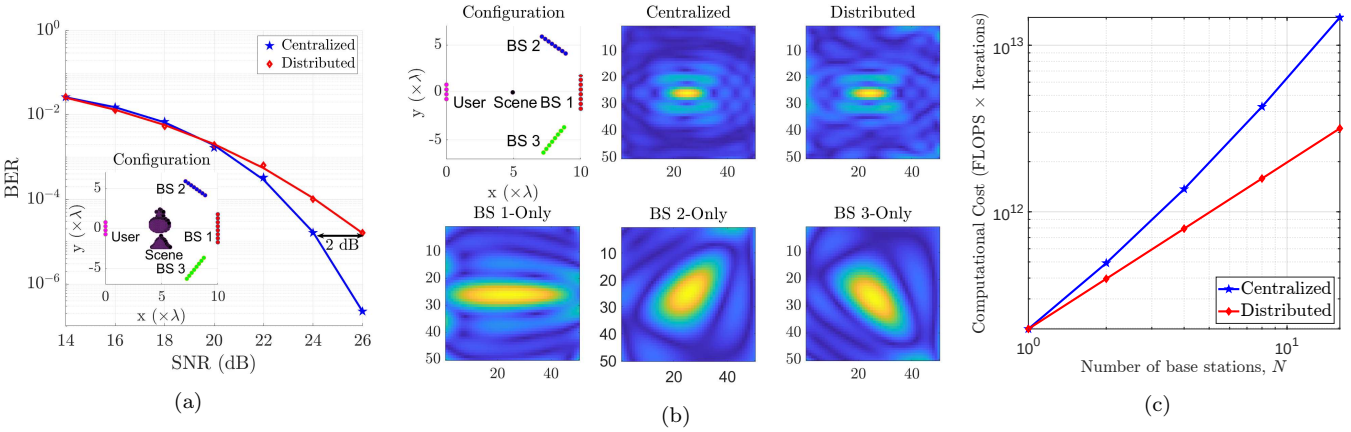
$$\tilde{f}_k = \frac{-\sum_m L_{\mathcal{G}}(N+1,m)f_{k-1}^{(m)}+D_{f,k-1}^{(N+1)}}{L_{\mathcal{G}}(N+1,N+1)},$$

$$\tilde{X}_{T-T_1,k} = \Pi_{\mathcal{X}_{T-T_1}}\left(\frac{-\sum_m L_{\mathcal{G}}(N+1,m)X_{T-T_1,k-1}^{(m)}+D_{X,k-1}^{(N+1)}}{L_{\mathcal{G}}(N+1,N+1)}\right),$$

where the index $(N + 1)$ denotes the fusion node. All remaining steps are identical to Algorithm 1, with $\overline{f}$ and $\overline{X}$ corresponding to the concatenation of local solutions over all $(N + 1)$ nodes in the network.

First, we numerically evaluate the validity of our convergence analysis from Theorem 1. Fig. 2(b) shows the optimality gap for DSISD at various receive signal-to-noise ratio (SNR) values. We observe that the sublinear $O(1/K)$ convergence predicted by Theorem 1 holds in the high SNR regime. In future work, we shall refine our convergence analysis to incorporate the effect of SNR.

In Figs. 3(a), (b) and (c), we show that DSISD achieves similar communication and imaging performance as decode-and-image, with $O(N)$ smaller computational complexity. In Fig. 3(a), we show that DSISD achieves communication bit error rate (BER) within 2 dB of decode-and-image. In Fig. 3(b), we show that the imaging point spread functions (PSFs) for DSISD are similar to those for decode-and-image, with 3 dB main lobe widths of λ and 0.4λ in range (x) and cross-range (y). Moreover, compared to local imaging performed at every base station, DSISD achieves resolution gains equivalent to jointly processing measurements across all $N = 3$ base stations. Finally, Fig. 3(c) shows that the computational complexity (in total number of floating operations) for DSISD is $O(N)$ smaller compared to that of decode-and-image.



(a) (b) (c)

Fig. 3. (a) DSISD achieves communication bit error rate (BER) within 2 dB of centralized performance across a wide range of receive signal-to-noise ratio (SNR) values. (b) DSISD has similar imaging point spread functions (PSFs) as decode-and-image, with 3 dB main lobe widths of λ and 0.4λ in range (x) and cross-range (y). Moreover, compared to local imaging performed at every base station, DSISD achieves resolution gains equivalent to jointly processing measurements across all $N = 3$ base stations. (c) DSISD has $O(N)$ smaller computational complexity (in total number of floating operations) compared to decode-and-image. All plots correspond to $K = 30$ iterations.

6. CONCLUDING REMARKS

We proposed DSISD, a provably convergent distributed algorithm based on consensus ADMM for simultaneous imaging and communication. We showed that DSISD achieves similar imaging and communication performance as centralized schemes with an order-wise reduction in computational complexity. We shall extend our convergence analysis to include the effects of integer constraints and SNR in future work. Moreover, we shall explore accelerated variants of DSISD with faster convergence rates.

REFERENCES

Boyd, S., Parikh, N., and Chu, E. (2011). *Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers*. Now Publishers Inc.

Chen, E. and Tao, M. (2017). ADMM-Based Fast Algorithm for Multi-Group Multicast Beamforming in Large-Scale Wireless Systems. *IEEE Transactions on Communications*, 65(6), 2685–2698.

Guan, J., Paidimarri, A., Valdes-Garcia, A., and Sadhu, B. (2021). 3-D Imaging Using Millimeter-Wave 5G Signal Reflections. *IEEE Transactions on Microwave Theory and Techniques*, 69(6), 2936–2948.

Hajinezhad, D., Chang, T.H., Wang, X., Shi, Q., and Hong, M. (2016). Nonnegative Matrix Factorization Using ADMM: Algorithm and Convergence Analysis. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 4742–4746.

Hajinezhad, D. and Shi, Q. (2018). Alternating Direction Method of Multipliers for a Class of Nonconvex Bilinear Optimization: Convergence Analysis and Applications. *Journal of Global Optimization*, 70(1), 261–288.

Hong, M., Hajinezhad, D., and Zhao, M.M. (2017). Prox-PDA: The Proximal Primal-Dual Algorithm for Fast Distributed Nonconvex Optimization and Learning Over Networks. In D. Precup and Y.W. Teh (eds.), *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, 1529–1538. PMLR.

Hong, M., Luo, Z.Q., and Razaviyayn, M. (2016). Convergence Analysis of Alternating Direction Method of Multipliers for a Family of Nonconvex Problems. *SIAM Journal on Optimization*, 26(1), 337–364.

Hu, R., Mysore Rama Rao, B.S., Murtada, A., Alaee-Kerahroodi, M., and Ottersten, B. (2021). Widely-distributed Radar Imaging Based on Consensus ADMM. In *2021 IEEE Radar Conference (RadarConf21)*, 1–6.

Kumar, S. and Rajawat, K. (2016). Distributed interference alignment for MIMO cellular network via consensus ADMM. In *2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 460–464.

Li, K., Sharan, R.R., Chen, Y., Goldstein, T., Cavallaro, J.R., and Studer, C. (2017). Decentralized Baseband Processing for Massive MU-MIMO Systems. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 7(4), 491–507.

Mehrotra, N. and Sabharwal, A. (2022). On the Degrees of Freedom Region for Simultaneous Imaging & Uplink Communication. *IEEE Journal on Selected Areas in Communications*, 1–1. doi:10.1109/JSAC.2022.3155518.

Tse, D. and Viswanath, P. (2005). *Fundamentals of Wireless Communication*. Cambridge University Press.

Vakalis, S., Gong, L., and Nanzer, J.A. (2019). Imaging With WiFi. *IEEE Access*, 7, 28616–28624.

Xu, Y. and Yin, W. (2013). A Block Coordinate Descent Method for Regularized Multiconvex Optimization with Applications to Nonnegative Tensor Factorization and Completion. *SIAM Journal on Imaging Sciences*, 6(3), 1758–1789. doi:10.1137/120887795.

Zabolotsky, A.A. and Mavrychev, E.A. (2018). Distributed Detection and Imaging in MIMO Radar Network Based on Averaging Consensus. In *2018 IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, 607–611.