

# Distributed Generalized Wirtinger Flow for Interferometric Imaging on Networks<sup>★</sup>

Sean M. Farrell Ashok Veeraraghavan  
Ashutosh Sabharwal César A. Uribe

Department of Electrical and Computer Engineering, Rice University,  
Houston, TX 77005, USA

(e-mail: {smf5, vashok, ashu, cauribe}@rice.edu).

**Abstract:** We study the problem of decentralized interferometric imaging over networks, where agents have access to a subset of local radar measurements and can compute pair-wise correlations with their neighbors. We propose a primal-dual distributed algorithm named *Distributed Generalized Wirtinger Flow* (DGWF). We use the theory of low rank matrix recovery to show when the interferometric imaging problem satisfies the Regularity Condition, which implies the Polyak-Łojasiewicz inequality. Moreover, we show that DGWF converges geometrically for smooth functions. Numerical simulations for single-scattering radar interferometric imaging demonstrate that DGWF can achieve the same mean-squared error image reconstruction quality as its centralized counterpart for various network connectivity and size.

**Keywords:** Interferometric imaging, Wirtinger flow, nonconvex optimization.

## 1. INTRODUCTION

The interferometric imaging problem consists of finding an unknown complex signal from cross-correlation measurements. This nonconvex problem arises in many applications such as radar and sonar interferometry, passive electromagnetic imaging, seismic imaging, and radio astronomy (Yonel and Yazici, 2019; Yonel et al., 2020; Vakalis et al., 2019; Garnier and Papanicolaou, 2009). As the scale of these applications increases, solving the interferometric imaging problem requires distributed or decentralized optimization algorithms. In practice, solving the interferometric imaging problem in a distributed manner may be beneficial when 1) sensing devices have limited memory, 2) there are data privacy concerns, 3) communication to a centralized node is unfeasible due to communication constraints. Figure 1 shows a distributed radar interferometric imaging scenario, where a single transmitter illuminates a scene with a static reflectivity function. The edges (dashed lines) represent the communication channels between receivers. The  $i^{th}$  receiver can only access its sampling matrix  $\mathbf{A}_i$  and compute cross-correlation measurements  $\mathbf{d}_{ij}$  with the  $j^{th}$  receiver if a communication channel exists.

Many algorithms used to solve the centralized interferometric imaging problem take inspiration from phase retrieval problems (Candès et al., 2013, 2015; Duchi and Ruan, 2018). A gradient-descent-based iterative low rank matrix recovery approach was proposed in Mason et al. (2015); which *lifts* the solution set to convexify the problem, leading to increased complexity. Recently, Yonel and Yazici (2019) and Yonel et al. (2020) proposed a general-

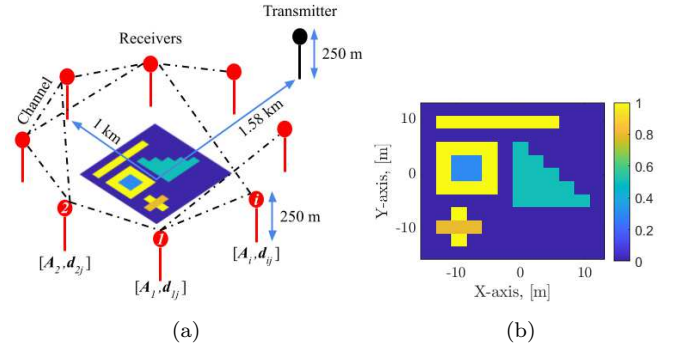


Fig. 1. Illustration of multistatic radar interferometric imaging scene (a). In the distributed network each agent  $i$  can access a subset of the data (i.e.  $[\mathbf{A}_i, \mathbf{d}_{ij}]$ ). Scene reflectivity function used in numerical simulations (b),  $12 \times 12$  image with a pixel spacing of 2.4 m.

ized Wirtinger flow (GWF) algorithm to solve the interferometric problem, which operates in the signal domain resulting in improved computation and memory efficiencies. They show that the GWF algorithm can achieve a linear convergence rate under specific conditions on the measurements (Yonel and Yazici, 2019). Demanet and Jugnon (2017) solve the centralized interferometric imaging problem using a lifting approach with a graph to encode the available interferometric measurements. They establish a relationship between graph connectivity and the robustness of recovery.

Recently, distributed optimization techniques have been proposed to solve the phase retrieval problem. Zhao et al. (2018) introduced the distributed Wirtinger flow algorithm (DWF) to solve the phase retrieval problem using a perturbed proximal primal-dual approach. The authors in Zhao et al. (2018) show that DWF can converge to

<sup>★</sup> This work was partially supported by the NSF Grant CNS-1956297, the ARPA-H Strategic Initiative Seed Fund #916012, and the Sustainable Futures Fund #919027.

an approximate solution at a sublinear rate. Chen et al. (2020) introduced a distributed subgradient method to solve weakly convex and non-smooth problems such as the phase retrieval problem. They report a linear convergence rate, using the gradient norm as a stopping criterion when the function is locally sharp near a minimizer.

In this work, to the best of our knowledge (Demanet and Jugnon, 2017; Yonel and Yazici, 2019; Yonel et al., 2020), we present the first distributed nonconvex primal-dual optimization algorithm to solve the interferometric imaging problem. *The contributions of this paper are:*

- We show that the Polyak-Łojasiewicz (PL) inequality is satisfied for the interferometric imaging problem when its lifted forward model meets the restricted isometry properties of rank-1 positive semi-definite matrices with a sufficiently small restricted isometry constant.
- We show that after proper initialization the Distributed Generalized Wirtinger Flow (DGWF) algorithm converges linearly to a global optimum when the cost function is smooth and PL inequality is satisfied.
- We provide numerical evidence that the DGWF algorithm can achieve comparable image reconstruction quality to centralized methods (Yonel and Yazici, 2019; Yonel et al., 2020) for radar interferometric imaging applications.

This paper is organized as follows. Section 2 provides the problem formulation. Section 3 introduces the proposed DGWF algorithm. Section 4 shows linear convergence for the DGWF algorithm. Numerical simulations are in Section 5, and Section 6 concludes this paper.

## 2. PROBLEM FORMULATION

Suppose the desired signal is denoted by  $x \in \mathbb{C}^K$ , the interferometric imaging problem can be formulated as:

$$\begin{aligned} &\text{find} && x \\ &\text{subject to} && d_{ij}^s = \langle a_i^s, x \rangle \overline{\langle a_j^s, x \rangle}, \\ &\text{for} && i = 1, 2, \dots, N; j \neq i; s = 1, 2, \dots, S \end{aligned}$$

where  $d_{ij}^s \in \mathbb{C}$  is the  $s^{th}$  cross-correlation sample for the  $i^{th}$  and  $j^{th}$  sensing processes,  $a_i^s \in \mathbb{C}^K$  is the  $s^{th}$  sampling vector for the  $i^{th}$  sensing process,  $\overline{(\cdot)}$  is the complex conjugate, and  $\langle \cdot, \cdot \rangle$  denotes the inner product.

Now, consider a connected network with  $N$  agents defined by the undirected graph  $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$  with  $|\mathcal{V}| = N$  vertices and  $|\mathcal{E}| = E$  edges. We assume the graph  $\mathcal{G}$  does not contain self-loops and let  $L$  denote the graph's Laplacian matrix. Agents try to cooperatively solve the following optimization problem,

$$\min_{x \in \mathbb{C}^K} f(x) := \frac{1}{N} \sum_{i=1}^N f_i(x). \quad (1)$$

Let  $\mathcal{X}^*$  and  $f^*$  denote the optimal set and the minimum function value for optimization Problem (1). In practice the cross-correlation measurements  $d_{ij}^s$  are assumed to be corrupted with additive i.i.d. noise. To recover  $x$ , each agent  $i$  formulates a local least squares error minimization objective function defined as,

$$f_i(x) = \frac{1}{2|\mathcal{N}_i|S_i} \sum_{j \in \mathcal{N}_i} \sum_{s=1}^{S_i} |d_{ij}^s - \langle a_i^s, x \rangle \overline{\langle a_j^s, x \rangle}|^2, \quad (2)$$

which can be thought of as the sensing process associated with the  $i^{th}$  agent of the network (Yonel and Yazici, 2019). Agent  $i$ 's neighbors are defined as  $\mathcal{N}_i \triangleq \{j \in \mathcal{V} | (i, j) \in \mathcal{E}\}$ . The term  $S_i$  denotes the number of cross-correlation samples agent  $i$  measures. *Thus the graph influences the optimization process and the number of cross-correlation measurements that can be computed.*

Next we state two technical assumptions that will help us later prove convergence for the DGWF algorithm.

*Assumption 1.* The undirected graph  $\mathcal{G}$  is connected.

*Assumption 2.* The optimal set  $\mathcal{X}^*$  of (1) is nonempty and  $f^* > -\infty$ .

Exploiting the structure of the graph Laplacian  $L$ , we can thus write the distributed interferometric imaging Problem (1) equivalently as,

$$\begin{aligned} &\min_{\mathbf{x} \in \mathbb{C}^{NK}} \tilde{f}(\mathbf{x}) \\ &\text{s.t. } \mathbf{L}\mathbf{x} = \mathbf{0}_{NK}, \end{aligned} \quad (3)$$

where  $\mathbf{x} = \text{col}(x_1, \dots, x_N)$ ,  $\tilde{f}(\mathbf{x}) = \sum_{i=1}^N f_i(x_i)$ , and  $\mathbf{L} = L \otimes I_K$  where  $L$  is the graph Laplacian and  $\otimes$  is the Kronecker product (Uribe et al., 2021).

## 3. DISTRIBUTED GENERALIZED WIRTINGER FLOW ALGORITHM

Recently, Yi et al. (2021) presented an algorithm for distributed nonconvex optimization that converges linearly if the local and global cost function is smooth and satisfies the PL inequality, respectively. We propose a primal-dual algorithm inspired by Yi et al. (2021) that uses Wirtinger derivatives (Yonel and Yazici, 2019). We call the proposed algorithm the *Distributed Generalized Wirtinger Flow* (DGWF) algorithm.

Following the algorithm formulation presented in Yi et al. (2021), the augmented Lagrangian function for (3) can be written as,

$$\mathcal{A}(\mathbf{x}, \mathbf{u}) = \tilde{f}(\mathbf{x}) + \frac{\lambda_1}{2} \mathbf{x}^\top \mathbf{L} \mathbf{x} + \lambda_2 \mathbf{u}^\top \mathbf{L}^{1/2} \mathbf{x}, \quad (4)$$

where  $\lambda_1 > 0$  and  $\lambda_2 > 0$  are the regularization parameters,  $\mathbf{u} \in \mathbb{C}^{NK}$  is the dual variable, and the constraint  $\mathbf{L}\mathbf{x} = \mathbf{0}_{NK}$  is replaced with its equivalent  $\mathbf{L}^{1/2} \mathbf{x} = \mathbf{0}_{NK}$  since its nulls are identical (Uribe et al., 2021). Yi et al. (2021) proposed the following first order primal-dual gradient method to solve (4),

$$\mathbf{x}_{t+1} = \mathbf{x}_t - \frac{\eta}{\|\mathbf{x}_0\|} (\lambda_1 \mathbf{L} \mathbf{x}_t + \lambda_2 \mathbf{L}^{1/2} \mathbf{u}_t + \nabla \tilde{f}(\mathbf{x}_t)) \quad (5)$$

$$\mathbf{u}_{t+1} = \mathbf{u}_t + \frac{\eta}{\|\mathbf{x}_0\|} \lambda_2 \mathbf{L}^{1/2} \mathbf{x}_t, \quad \forall \mathbf{x}_t \in \mathbb{C}^{NK}, \quad (6)$$

where  $\eta > 0$  is the step size and  $\|\mathbf{x}_0\|$  is the norm of the initial estimate. If we denote  $\mathbf{v}_t = \mathbf{L}^{1/2} \mathbf{u}_t$ , we can rewrite (5) and (6) as,

$$\mathbf{x}_{t+1} = \mathbf{x}_t - \frac{\eta}{\|\mathbf{x}_0\|} (\lambda_1 \mathbf{L} \mathbf{x}_t + \lambda_2 \mathbf{v}_t + \nabla \tilde{f}(\mathbf{x}_t)) \quad (7)$$

---

**Algorithm 1** Distributed Generalized Wirtinger Flow

---

**Input:** parameters  $\lambda_1 > 0, \lambda_2 > 0$ , and  $\eta > 0$

**Initialize:**  $x_{i,0} \in \mathbb{C}^K$  with (9) and  $v_{i,0} = \mathbf{0}_K, \forall i \in [N]$ .

```

1: for  $t = 0, 1, \dots$  do
2:   for  $i = 1, \dots, N$  in parallel do
3:     Send  $x_{i,t}$  to  $\mathcal{N}_i$  and receive  $x_{j,t}$  from  $j \in \mathcal{N}_i$ ;
     Primal variable update:
4:      $x_{i,t+1} = x_{i,t} - \frac{\eta}{\|x_{i,0}\|} (\lambda_1 \sum_{j \in \mathcal{N}_i} L_{ij} x_{j,t} + \lambda_2 v_{i,t} + \nabla f_i(x_{i,t}))$ ;
     Dual variable update:
5:      $v_{i,t+1} = v_{i,t} + \frac{\eta}{\|x_{i,0}\|} \lambda_2 \sum_{j \in \mathcal{N}_i} L_{ij} x_{j,t}$ ;
6:   end for
7: end for
8: return  $x_t$ 

```

---

$$v_{t+1} = v_t + \frac{\eta}{\|x_0\|} \lambda_2 L x_t, \forall x_t \in \mathbb{C}^{NK}, \sum_{j=1}^N v_{j,0} = \mathbf{0}_K. \quad (8)$$

For initializing the DGWF algorithm we adopt the spectral initialization scheme used in Yonel and Yazici (2019); Yonel et al. (2020). The initial estimate  $x_0 \in \mathbb{C}^K$  is the rank-1, positive semi-definite matrix approximation of the lifted backprojection estimate  $\hat{X} \in \mathbb{C}^{K \times K}$ ,

$$\hat{X} := \frac{1}{N} \sum_{i=1}^N \frac{1}{2|\mathcal{N}_i|S_i} \sum_{j \in \mathcal{N}_i} \sum_{s=1}^{S_i} d_{ij}^s a_i^s (a_j^s)^H + \overline{d_{ij}^s} a_j^s (a_i^s)^H, \quad (9)$$

where  $x_0 = \sqrt{\lambda_0} \mathbf{v}_0$  with  $\lambda_0, \mathbf{v}_0$  being the leading eigenvalue-eigenvector pair of  $\hat{X}$  and  $(\cdot)^H$  denotes the complex conjugate transpose. The lifted formulation of the  $i^{\text{th}}$  and  $j^{\text{th}}$  sensing processes is  $F^s = a_j^s (a_i^s)^H$ . Let  $\mathbf{d} = [d_{ij}^1, d_{ij}^2, \dots, d_{ij}^{S_i}] \in \mathbb{C}^S$  denote the vector of cross-correlated measurements. The lifted forward model  $\mathcal{F} : \mathbb{C}^{K \times K} \rightarrow \mathbb{C}^S$  is defined as,

$$\mathbf{d} = \mathcal{F}(\hat{X}), \quad (10)$$

where  $\mathcal{F}$  is a  $[S \times K^2]$  matrix with  $F^s$  as its rows and  $\hat{X}$  concatenated into a vector. This spectral method initializes the iterates within a bounded set i.e.,  $\mathcal{X} = \{x \mid \|x\|^2 \leq \tau\}$ , where  $\tau$  is a constant dependent on the restricted isometry constant over rank-1 matrices ( $\text{RIC}_{\delta_1}$ ) for the lifted forward model (10) (Yonel and Yazici, 2019, Theorem 4.6). The initial estimate is distributed to all agents, i.e.,  $x_0 = x_{1,0} = x_{2,0} = \dots = x_{i,0}$ .

*Remark 3.* The spectral initialization (9) computes a rank-1 approximation to a low rank matrix recovery problem using all the local cross-correlation measurements stored at every agent. The lifted backprojection estimate  $\hat{X}$  is an average over all local lifted backprojection estimates. Thus, one possible way to create a distributed initialization method would be to run a distributed averaging algorithm over the local lifted backprojection estimates (Olshevsky and Tsitsiklis, 2009). Analysis of how distributed initialization schemes effect the DGWF algorithm is left for future work.

The DGWF algorithm using spectral initialization (9) and updates (7), (8) is shown in pseudo-code as Algorithm 1.

The gradient  $\nabla f_i(x_{i,t})$  is explicitly found by computing the Wirtinger derivative using the local information agent  $i$  can access (Yonel and Yazici, 2019). For the cross-

correlation between agent  $i$  and agent  $j$  with  $(i, j) \in \mathcal{E}$  the gradient is

$$\begin{aligned} \nabla f_i(x_{i,t}) &= \nabla \left[ \frac{1}{2|\mathcal{N}_i|S_i} \sum_{j \in \mathcal{N}_i} \sum_{s=1}^{S_i} |d_{ij}^s - \langle a_i^s, x_{i,t} \rangle \overline{\langle a_j^s, x_{i,t} \rangle}|^2 \right] \\ &= \frac{1}{2|\mathcal{N}_i|S_i} \sum_{j \in \mathcal{N}_i} \sum_{s=1}^{S_i} e_{ij}^s (a_j^s (a_i^s)^H x_{i,t}) + e_{ij}^s (a_i^s (a_j^s)^H x_{i,t}) \end{aligned} \quad (11)$$

where  $e_{ij}^s = ((a_i^s)^H x_{i,t} (x_{i,t})^H a_j^s - d_{ij}^s)$ .

In the next section, we show that when the cost function is smooth and satisfies the PL inequality, the DGWF algorithm converges linearly to a minimum.

#### 4. CONVERGENCE ANALYSIS

We follow Yonel and Yazici (2019) geometric analysis of the interferometric imaging problem to show when the lifted forward model's restricted isometry properties over rank-1 positive semi-definite matrices satisfy the Regularity Condition (RC). Then we extend our analysis to show that the RC implies the PL inequality. Finally, convergence of the DGWF algorithm is established.

The RC bounds the local smoothness and curvature of a function, meaning that the function's gradients are *well behaved*.

*Definition 4.* (Regularity Condition): A function  $f$  satisfies the  $\text{RC}(\alpha, \beta)$  with  $\alpha, \beta > 0$  if,

$$\langle \nabla f(z), z - x^* \rangle \geq \frac{1}{\alpha} \|\nabla f(z)\|^2 + \frac{1}{\beta} \|z - x^*\|^2, \quad (12)$$

for all  $z \in \mathbb{R}^{2K}$  and  $x^*$  being a minimizer of  $f$ .

Next, the PL inequality is defined. The PL inequality does not require convexity but implies that every stationary point is a global minimizer.

*Definition 5.* (Polyak-Lojasiewicz Inequality): The function  $f$  satisfies the PL inequality if, for some  $\mu > 0$ ,

$$\frac{1}{2} \|\nabla f(z)\|^2 \geq \mu(f(z) - f^*), \quad \forall z \in \mathbb{R}^{2K}, \quad (13)$$

where  $f : \mathbb{R}^{2K} \rightarrow \mathbb{R}$  is continuously differentiable and  $f^* = \min_{z \in \mathbb{R}^{2K}} f(z)$ . When this condition holds the function is  $\mu$ -PL.

We begin by establishing the relationship between the lifted forward model and the RC using the restricted isometry property (RIP). The RIP arose from the compressed sensing field as a way to measure how close a matrix is to an orthonormal system given a certain degree of sparsity (Candes and Tao, 2005).

*Definition 6.* (Restricted isometry property): Let  $\mathcal{A} : \mathbb{C}^{M \times K} \rightarrow \mathbb{C}^S$  denote a linear operator. Without loss of generality assume  $M \leq K$ . For every  $1 \leq r \leq M$ , the  $r$ -restricted isometry constant (RIC) is defined as the smallest  $\delta_r < 1$  such that

$$(1 - \delta_r) \|X\|_F^2 \leq \|\mathcal{A}(X)\|^2 \leq (1 + \delta_r) \|X\|_F^2 \quad (14)$$

holds for all matrices  $X$  of rank at most  $r$ , where  $\|X\|_F = \sqrt{\text{Tr}(X^H X)}$  denotes the Forbenius norm.

From (Yonel and Yazici, 2019, Theorem 4.6), when the lifted forward model  $\mathcal{F}$  satisfies the RIP condition over

rank-1 positive semi-definite matrices with a  $\text{RIC}_{\delta_1} < 0.214$  then Definition (4) surely holds with any  $\alpha, \beta$  satisfying,

$$\frac{1}{\alpha \|x^*\|^2} + \frac{c^2(\delta_1) \|x^*\|^2}{\beta} \leq h(\delta_1) := (1 - \delta_2)(1 - \epsilon)(2 - \epsilon), \quad (15)$$

where  $\epsilon^2 = (2 + \delta_1)(1 - \sqrt{1 - \frac{\delta_1}{1 + \delta_1}}) + \frac{\delta_1^2}{8}$ ,  $\delta_2 = \frac{\sqrt{2(2 + \epsilon)\delta_1}}{\sqrt{(1 - \epsilon)(2 - \epsilon)}}$ , and  $c(\delta_1) = (2 + \epsilon)(1 + \epsilon)(1 + \delta_1)$ .

#### 4.1 RC & PL Inequality

In this subsection we restate (Yazdani and Hale, 2021, Lemma 1) in Lemma 7 which shows that the RC implies the PL inequality.

*Lemma 7.* (Yazdani and Hale (2021), Lemma 1). Let  $f$  have a Lipschitz continuous gradient with Lipschitz constant  $L_f$  and the set  $\mathcal{X}^* = \{x^* \in \mathbb{R}^{2K} | \nabla f(x^*) = 0\}$  is nonempty and finite. If  $f$  is  $\text{RC}(\alpha, \beta)$ , then  $f$  is  $1/(\beta^2 L_f)$ -PL.

#### 4.2 DGWF Algorithm Convergence Analysis

Problem (2) is not complex differentiable due to the mapping from  $\mathbb{C}^K$  to  $\mathbb{R}$  (Kreutz-Delgado, 2009). In the DGWF algorithm Wirtinger derivatives are used because they provide an elegant way to compute partial derivatives that are differentiable in the complex domain. However, for the convergence analysis it is more convenient to work in the real-valued domain. Thus we equivalently reformulate (2) to the real-valued domain where it is differentiable (Kreutz-Delgado, 2009). Without loss of generality, we use (2) for the  $i^{\text{th}}$  and  $j^{\text{th}}$  sensing processes and define  $\tilde{x} \triangleq [\text{Re}(x); \text{Im}(x)] \in \mathbb{R}^{2K}$ ,  $d_R^s = \text{Re}(d_{ij}^s) \in \mathbb{R}$ ,  $d_I^s = \text{Im}(d_{ij}^s) \in \mathbb{R}$ , and

$$\tilde{A}_R^s \triangleq \begin{bmatrix} \tilde{A}_1^s & \tilde{A}_2^s \\ -\tilde{A}_2^s & \tilde{A}_1^s \end{bmatrix} \in \mathbb{R}^{2K \times 2K}, \tilde{A}_I^s \triangleq \begin{bmatrix} \tilde{A}_2^s & -\tilde{A}_1^s \\ \tilde{A}_1^s & \tilde{A}_2^s \end{bmatrix} \in \mathbb{R}^{2K \times 2K},$$

with  $\tilde{A}_1^s = \text{Re}(a_i^s)^\top \text{Re}(a_j^s) + \text{Im}(a_i^s)^\top \text{Im}(a_j^s)$ , and  $\tilde{A}_2^s = -\text{Re}(a_i^s)^\top \text{Im}(a_j^s) + \text{Im}(a_i^s)^\top \text{Re}(a_j^s)$ , for  $s = 1, 2, \dots, S$ . We can equivalently rewrite (2) in terms of  $\tilde{x}$  as

$$f(\tilde{x}) \triangleq \frac{1}{2S} \sum_{s=1}^S ((d_R^s - \tilde{x}^\top \tilde{A}_R^s \tilde{x})^2 + (d_I^s - \tilde{x}^\top \tilde{A}_I^s \tilde{x})^2). \quad (16)$$

*Lemma 8.* Function  $f(\tilde{x})$  has Lipschitz gradient on the set  $\tilde{\mathcal{X}} \triangleq \{\tilde{x} | \|\tilde{x}\|^2 \leq \tau\}$  with Lipschitz constant given by

$$L_f \triangleq \frac{1}{S} \sum_{s=1}^S \left( d_R^s \sigma_{\max}(\tilde{A}_R^s) + d_I^s \sigma_{\max}(\tilde{A}_I^s) + 3\tau \left( \sigma_{\max}^2(\tilde{A}_R^s) + \sigma_{\max}^2(\tilde{A}_I^s) \right) \right). \quad (17)$$

**Proof.** Using the definition of the Lipschitz continuity, for  $\forall u, v \in \tilde{\mathcal{X}}$  we have

$$\begin{aligned} & \|\nabla f(u) - \nabla f(v)\| \\ & \leq \frac{1}{S} \left\| \sum_{s=1}^S \left( (d_R^s - u^\top \tilde{A}_R^s u) \tilde{A}_R^s u - (d_R^s - v^\top \tilde{A}_R^s v) \tilde{A}_R^s v \right. \right. \end{aligned}$$

$$\begin{aligned} & \left. + (d_I^s - u^\top \tilde{A}_I^s u) \tilde{A}_I^s u - (d_I^s - v^\top \tilde{A}_I^s v) \tilde{A}_I^s v \right) \right\| \\ & \leq \frac{1}{S} \sum_{s=1}^S \left( d_R^s \sigma_{\max}(\tilde{A}_R^s) + \|u\|^2 \sigma_{\max}^2(\tilde{A}_R^s) \right) \|u - v\| \\ & \quad + \frac{1}{S} \sum_{s=1}^S \left\| u^\top \tilde{A}_R^s u \tilde{A}_R^s v - u^\top \tilde{A}_R^s v \tilde{A}_R^s v \right\| \\ & \quad + \frac{1}{S} \sum_{s=1}^S \left\| u^\top \tilde{A}_I^s u \tilde{A}_I^s v - u^\top \tilde{A}_I^s v \tilde{A}_I^s v \right\| \\ & \quad + \frac{1}{S} \sum_{s=1}^S \left( d_I^s \sigma_{\max}(\tilde{A}_I^s) + \|u\|^2 \sigma_{\max}^2(\tilde{A}_I^s) \right) \|u - v\| \\ & \quad + \frac{1}{S} \sum_{s=1}^S \left\| u^\top \tilde{A}_I^s u \tilde{A}_I^s v - u^\top \tilde{A}_I^s v \tilde{A}_I^s v \right\| \\ & \quad + \frac{1}{S} \sum_{s=1}^S \left\| u^\top \tilde{A}_I^s v \tilde{A}_I^s v - v^\top \tilde{A}_I^s v \tilde{A}_I^s v \right\| \\ & \leq \frac{1}{S} \sum_{s=1}^S \left( d_R^s \sigma_{\max}(\tilde{A}_R^s) + d_I^s \sigma_{\max}(\tilde{A}_I^s) \right. \\ & \quad \left. + 3\tau \left( \sigma_{\max}^2(\tilde{A}_R^s) + \sigma_{\max}^2(\tilde{A}_I^s) \right) \right) \|u - v\|. \end{aligned}$$

*Proposition 9.* Let the lifted forward model (10) satisfy the RIP condition over rank-1 positive semi-definite matrices with a  $\text{RIC}_{\delta_1} \leq 0.214$  (Yonel and Yazici, 2019, Theorem 4.6) and Assumptions 1-2 hold. Moreover, let  $\{\tilde{x}_k\}$  be the sequence generated by Algorithm 1 with  $\lambda_1 > 0$ ,  $\lambda_2 > 0$ ,  $\eta > 0$  applied to (16). Then, there exists a  $c_1 > 0$ , and  $\kappa \in (0, 1)$  such that,

$$\|\tilde{x}_t - \tilde{x}^*\|^2 + N(f(\tilde{x}_t) - f^*) \leq (1 - \kappa)^t c_1, \quad \forall t \in \mathbb{N}_0. \quad (18)$$

where  $\tilde{x}_t = \frac{1}{N} (\mathbf{1}_N^\top \otimes \mathbf{I}_K) \tilde{\mathbf{x}}_t$  and  $\tilde{\mathbf{x}}_t = \mathbf{1}_N \otimes \tilde{x}_t$ .

**Proof.** It follows from Lemma 8 that (16) is smooth with Lipschitz constant (17). Moreover, if the lifted forward model (10) satisfy the RIP condition over rank-1 positive semi-definite matrices with a  $\text{RIC}_{\delta_1} \leq 0.214$  (Yonel and Yazici, 2019, Theorem 4.6) it holds that (16) is RC, and thus the PL inequality holds from Lemma 7. Finally, the desired result follows from (Yi et al., 2021, Theorem 2).

*Remark 10.* Due to space limitations the constants  $\lambda_1$ ,  $\lambda_2$ ,  $\eta$ ,  $c_1$  and  $\kappa$  are not explicitly stated. However they can be directly derived from (Yi et al., 2021, Theorem 1).

## 5. NUMERICAL SIMULATIONS

We consider the system model in Fig. 1, where  $N$  independent receivers encircle a region of interest to be imaged with  $p_i^{rx} \in \mathbb{R}^3$  being the spatial position of the  $i^{\text{th}}$  receiver. A single transmitter of opportunity at a known spatial position  $p^{tx} \in \mathbb{R}^3$  illuminates the imaging scene. The imaging scene domain  $D$  can be discretized into  $K$  voxels at positions  $x \in \mathbb{R}^{K \times 3}$ . The reflectivity function  $\rho(x) \in \mathbb{C}^K$  for the imaging scene is assumed to be nonfluctuating. Assuming the Born approximation (single bounce) and free-space propagation the frequency domain Green's function solution to the Helmholtz equation for receiver  $i$  can be written as (Yonel and Yazici, 2019),

$$h_i(\omega_s) := J(\omega_s) \int_D e^{-i\omega_s \frac{\phi_i(x)}{c_0}} \alpha_i(x, p^{tx}) \rho(x) dx, \quad (19)$$

where  $\omega_s$  is the  $s^{th}$  discretized temporal frequency of the transmitted signal,  $J(\omega_s)$  is the transmitted signal power,  $c_0$  is the wave velocity,  $\alpha_i(x, p^{tx})$  models the propagation attenuation and hardware gains,  $\rho(x)$  is the scene reflectivity at position  $x \in D$ . The  $\phi_i(x) = |x - p_i^{rx}| + |x - p^{tx}|$  is the bi-static delay term due to propagation. The radar interferometric imaging model for the  $s^{th}$  temporal frequency sample of the  $i^{th}$  and  $j^{th}$  receivers in the multistatic radar setup can be formulated as,

$$d_{ij}(\omega_s) = \sum_{k=1}^K \alpha_i(x_k, p^{tx}) e^{-i\omega_s(|x_k - p_i^{rx}| + |x_k - p^{tx}|)/c_0} \rho_k \times J(\omega_s) \overline{J(\omega_s)} \sum_{k'=1}^K \overline{\alpha_j(x_{k'}, p^{tx})} e^{i\omega_s(|x_{k'} - p_j^{rx}| + |x_{k'} - p^{tx}|)/c_0} \overline{\rho_{k'}}. \quad (20)$$

From (20) the linear sampling vectors  $a_i^s$  and  $a_j^s$  can be formulated as,

$$a_i^s = \left[ J(\omega_s) \alpha_i(x_k, p^{tx}) e^{-i\omega_s(|x_k - p_i^{rx}| + |x_k - p^{tx}|)/c_0} \right]_{k=1}^K \quad (21)$$

$$a_j^s = \left[ J(\omega_s) \alpha_j(x_k, p^{tx}) e^{-i\omega_s(|x_k - p_j^{rx}| + |x_k - p^{tx}|)/c_0} \right]_{k=1}^K. \quad (22)$$

Combing (20) - (22) the radar interferometric imaging model can be vectorized as,

$$d_{ij}(\omega_s) = \langle a_i^s, \rho \rangle \langle a_j^s, \rho \rangle, \quad s = 1, \dots, S; i = 1, \dots, N, j \neq i \quad (23)$$

The simulated true scene reflectivity function and multistatic radar setup is shown in Fig. 1. Unless otherwise stated the following assumptions and parameters are assumed: Born approximation,  $S = 64$  frequency samples,  $N = 35$  receivers,  $t_{max} = 4000$  for GWF, DGWF max iterations, stepsize  $\eta_t = \min(1 - \exp(-t/\tau_0), 0.01)$  where  $\tau_0 = 3300$ , center frequency 12 GHz, bandwidth 60 MHz, image is  $12 \times 12$  ( $K = 144$ ) with spacing  $\Delta = 2.4$  m since Fourier resolution limit  $\Delta_{res} = 2.5$  m, graph  $\mathcal{G}$  is small-world with a 0.1 probability of connection,  $\lambda_1$  and  $\lambda_2 = 1$ , signal-to-noise ratio is 50 dB, and transmitter and receiver gains are 100 dB.

The mean square error (MSE):  $MSE \triangleq \frac{1}{K} \sum_{k=1}^K (\rho_{i,k} - \rho_k^*)^2$  is used to quantitatively determine the accuracy of the reconstructed image  $\rho_i$  to the true image  $\rho^*$ . For the proposed DGWF algorithm the consensus error  $\triangleq \sum_{i=1}^N \|\rho_i - \frac{1}{N} \sum_{i=1}^N \rho_i\|^2$ . For each simulation the DGWF algorithm is compared to the centralized GWF algorithm (Yonel and Yazici, 2019; Yonel et al., 2020) using the same cross-correlation measurements as defined by the graph.

### 5.1 Effect of Graph Connectivity

We examine the effects graph connectivity has on the DGWF algorithm's reconstruction performance. Since the graph specifies how agents communicate and form cross-correlation measurements, we expect the speed of convergence and image reconstruction quality to increase as

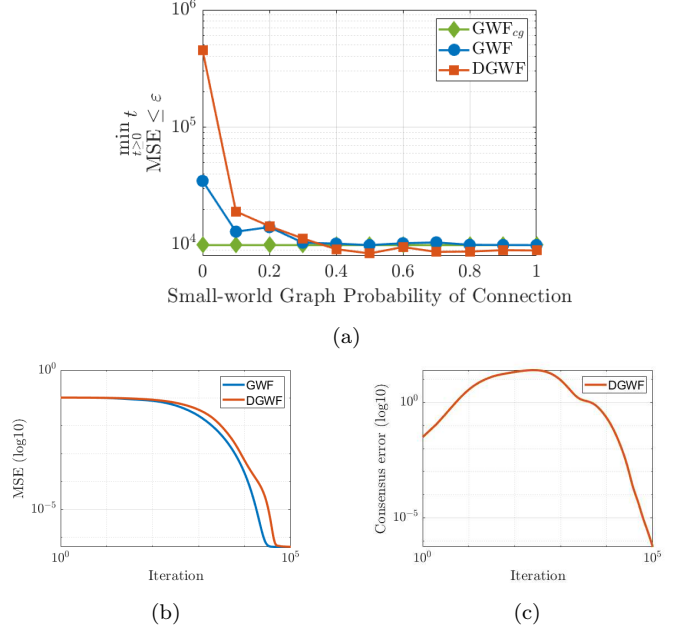


Fig. 2. Comparison between the number of iterations  $t$  to reach an MSE  $\varepsilon = 10^{-5}$  and the connectivity probability of a small-world graph with 35 agents (a). GWF<sub>cg</sub> is GWF using a complete graph. MSE (b) and consensus error (c) per iteration.

graph connectivity increases. To test the relationship between graph connectivity and reconstruction performance, we vary the probability of connection for small-world graphs (Watts and Strogatz, 1998) consisting of 35 agents and record the number of iterations it takes DGWF and GWF to each achieve an MSE of  $10^{-5}$ . For an additional benchmark, we compare DGWF and GWF against GWF<sub>cg</sub>, which is the GWF algorithm using a complete graph. Figure 2a shows that decreased graph connectivity results in longer convergence times for GWF and DGWF. However, after the small-world network's connectivity parameter is greater than 0.4 there is a negligible change in the number of iterations required to reach the desired reconstruction accuracy.

Additionally, we run the DGWF and GWF algorithms for  $10^5$  iterations. Figure 2b shows that DGWF takes longer to converge compared to GWF because DGWF needs to distribute local information between neighboring agents. The DGWF algorithm reaches a competitive MSE relative to the GWF algorithm without centralized data processing. Figure 2c shows that the proposed DGWF algorithm reaches consensus.

### 5.2 Effect of Number of Receivers

Next, we analyze the effect the number of receivers has on the convergence behavior for the DGWF and GWF algorithms. We vary the number of agents between 5 and 40 and record the MSE for the DGWF and GWF algorithms after 4000 iterations. Since the number of receivers directly influences the number of cross-correlation measurements each algorithm can compute, we expect that as the number of receivers decreases, the MSE will increase. In Fig. 3a the MSE curves for both the DGWF and GWF algorithms show that decreasing the number of receivers increases

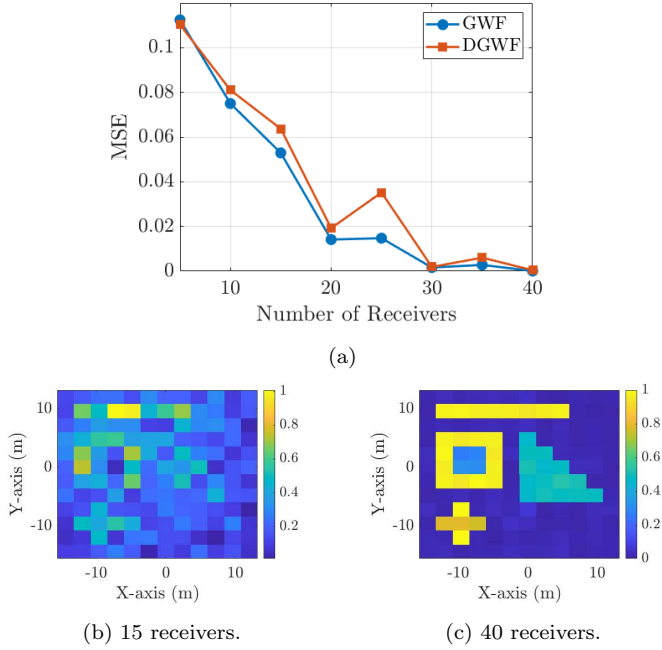


Fig. 3. Comparison of reconstruction MSE vs. the number of receivers (a). DGWF reconstructions after 4000 iterations for 15 (b) and 40 (c) receivers.

the reconstruction MSE. Furthermore, Fig. 3a suggests that after approximately 30 receivers, the addition of more receivers does not significantly decrease the reconstruction error. Example image reconstructions for the DGWF algorithm for 15 and 40 receivers are presented in Fig. 3b-3c. It can be observed that when the number of receivers is below 30, the DGWF algorithm reconstructs noticeably degraded images. However, when the number of receivers exceeds 30, the DGWF algorithm achieves significantly accurate image reconstructions.

## 6. CONCLUSION

We studied the distributed interferometric imaging problem over networks. We combine GWF theory (Yonel and Yazici, 2019) and distributed optimization methods to create the first distributed interferometric imaging algorithm named the Distributed Generalized Wirtinger Flow (DGWF) algorithm. The DGWF algorithm converges linearly to a global optimum with proper initialization and when the cost function is smooth and satisfies the PL inequality. We demonstrate the capabilities of DGWF through numerical radar interferometric imaging simulations. Future work will study convergence analysis for accelerated methods, time-varying directed graphs, and distributed initialization methods.

## REFERENCES

Candes, E. and Tao, T. (2005). Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12), 4203–4215.

Candés, E.J., Li, X., and Soltanolkotabi, M. (2015). Phase retrieval via wirtinger flow: Theory and algorithms. *IEEE Transactions on Information Theory*, 61, 1985–2007.

Candés, E.J., Strohmer, T., and Voroninski, V. (2013). Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming. *Communications on Pure and Applied Mathematics*, 66, 1241–1274.

Chen, S., García, A., and Shahrampour, S. (2020). Distributed projected subgradient method for weakly convex optimization. *ArXiv*, abs/2004.13233.

Demagnet, L. and Jugnon, V. (2017). Convex recovery from interferometric measurements. *IEEE Transactions on Computational Imaging*, 3, 282–295.

Duchi, J.C. and Ruan, F. (2018). Solving (most) of a set of quadratic equalities: composite optimization for robust phase retrieval. *Information and Inference: A Journal of the IMA*, 8(3), 471–529.

Garnier, J. and Papanicolaou, G. (2009). Passive sensor imaging using cross correlations of noisy signals in a scattering medium. *SIAM Journal on Imaging Sciences*, 2, 396–437.

Kreutz-Delgado, K. (2009). The complex gradient operator and the cr-calculus.

Mason, E., Son, I.Y., and Yazici, B. (2015). Passive synthetic aperture radar imaging using low-rank matrix recovery methods. *IEEE Journal of Selected Topics in Signal Processing*, 9, 1570–1582.

Olshevsky, A. and Tsitsiklis, J.N. (2009). Convergence speed in distributed consensus and averaging. *SIAM Journal on Control and Optimization*, 48(1), 33–55.

Uribe, C.A., Lee, S., Gasnikov, A., and Nedić, A. (2021). A dual approach for optimal algorithms in distributed optimization over networks. *Optimization Methods and Software*, 36(1), 171–210.

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

Watts, D.J. and Strogatz, S.H. (1998). Collective dynamics of ‘small-world’ networks. *Nature*, 393, 440–442.

Yazdani, K. and Hale, M. (2021). Asynchronous parallel nonconvex optimization under the polyak łojasiewicz condition. *IEEE Control Systems Letters*, 6, 524–529.

Yi, X., Zhang, S., Yang, T., Chai, T., and Johansson, K.H. (2021). Linear convergence of first- and zeroth-order primal-dual algorithms for distributed nonconvex optimization. *IEEE Transactions on Automatic Control*, 1–1.

Yonel, B., Son, I.Y., and Yazici, B. (2020). Exact multi-static interferometric imaging via generalized wirtinger flow. *IEEE Transactions on Computational Imaging*, 6, 711–726.

Yonel, B. and Yazici, B. (2019). A generalization of wirtinger flow for exact interferometric inversion. *SIAM J. Imaging Sciences*, 12, 2119–2164.

Zhao, Z., Lu, S., Hong, M., and Palomar, D.P. (2018). Distributed optimization for generalized phase retrieval over networks. *2018 52nd Asilomar Conference on Signals, Systems, and Computers*, 48–52.