Distributed Optimization in Distribution Systems: Use Cases, Limitations, and Research Needs

N. Patari, *Student Member, IEEE,* V. Venkataramanan, *Member, IEEE,* A. K. Srivastava, *Senior Member, IEEE,* D. K. Molzahn, *Senior Member, IEEE,* N. Li, *Member, IEEE,* A. M. Annaswamy, *Fellow, IEEE*

***Abstract*—Electric distribution grid operations typically rely on both centralized optimization and local non-optimal control techniques. As an alternative, distribution system operational practices can consider distributed optimization techniques that leverage communications among various neighboring agents to achieve optimal operation. With the rapidly increasing integration of distributed energy resources (DERs), distributed optimization algorithms are growing in importance due to their potential advantages in scalability, flexibility, privacy, and robustness relative to centralized optimization. Implementation of distributed optimization offers multiple challenges and also opportunities. This paper provides a comprehensive review of the recent advancements in distributed optimization for electric distribution systems and classifications using key attributes. Problem formulations and distributed optimization algorithms are provided for example use cases, including volt/var control, market clearing process, loss minimization, and conservation voltage reduction. Finally, this paper also presents future research needs for the applicability of distributed optimization algorithms in the distribution system.**

***Index Terms*—Optimal Power Flow, Distributed Energy Resources, Active Distribution Systems, Distributed Optimization.**

I. Introduction

ELECTRIC grid operation and control heavily rely on optimal power flow (OPF) techniques to maneuver the system to economic and reliable operating points [1]. Using extensive sensing and communication infrastructures, power grid operators centrally gather all information needed by formulated OPF problems, solve these problems, and send dispatch control to the generators and control devices. The rapidly increasing integration of distributed energy resources (DERs) motivates the application of advanced optimization tools for distribution systems. However, applying OPF solution techniques to distribution systems is challenging for many reasons. Some of these challenges include the huge number of controllable resources, flexible control due to inverter-based resources (IBR), the inapplicability of the DC power flow approximation, limited communication and sensing infrastructures, and privacy concerns with control devices and resource ownership.

N. Patari is with the Washington State University (WSU), Pullman, WA. A. Srivastava was with the WSU and now with the West Virginia University, Morgantown, WV.

V. Venkataramanan and A. M. Annaswamy is with the Massachusetts Institute of Technology, Cambridge, MA, USA.

D. Molzahn is with the Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: molzahn@gatech.edu).

N. Li is with Harvard University, Cambridge, MA, USA.

This work resulted from discussion within IEEE PES WG on “Computational Challenges and Solutions for Implementing Distributed Optimization in the Power System”, Computing and Analytic Methods Subcommittee (CAMS), Analytic Methods for Power Systems (AMPS) Committee. This work is partially supported by the NSF AI Institute for Advances in Optimization (Award 2112533).

Existing centralized schemes for operating and controlling distribution systems span two operational layers: (a) the *physical layer*, where control agents like tap-changing transformers, switched capacitors, reclosers, circuit breakers, etc. are responsible for managing the state of the distribution system, (b) the *cyber layer*, which can be classified into two more sublayers, (i) the *control and management sublayer* where the Advanced Distribution Management System (ADMS) runs an optimization tool and is responsible for making control decisions of power system operations throughout the day and (ii) the *communication sublayer* where commands and measurements are relayed between the physical layer and the control/management sublayer (i.e., the ADMS). Along with these centralized schemes, utilities often use *purely local feedback based* control strategies in power grids. However, these schemes generate non-optimal solutions since they are solely based on local measurements. Stability is also an issue for purely local feedback based schemes as they are unable to consistently regulate the voltage/frequency throughout the system [2]. Complementing this framework, distributed algorithms involving communication and coordination among various agents/control nodes provide the opportunity for optimal control and operation of active distribution grids. In this context, “active” distribution grids refer to the evolving distribution systems that allow prosumers to participate in grid services which may involve power flow from distribution to transmission grid [3]. Distributed algorithms have the following advantages [4]:

- With the inclusion of DERs, the number of physical control devices is rapidly growing. Hence, the ADMS is challenged by the need to communicate with and manage the operation of all deployed control agents that are associated with the new DERs. Additionally, centralized schemes will have mathematical challenges in solving large-scale multi-variate multi-period problems due to complexities associated with inverter-based DERs. Relative to centralized approaches, distributed algorithms have potential scalability advantages for addressing this challenge.
- Distributed algorithms are based on decomposition of the original centralized problem into smaller subproblems having coordination and communication with only limited numbers of neighboring control agents. The decomposed

subproblems enable fast parallel computations [5], [6].

- Distributed algorithms have potential advantages in data privacy given different ownership of DERs at the edge, since the communication sublayer only involves neighboring agents rather than centralized communication with the ADMS.
- Centralized approaches are prone to single-point cyber failures. Distributed algorithms have potential advantages in robustness that can help improve the reliability and resiliency of active distribution systems.
- Since each agent only needs to communicate with its neighbors, distributed algorithms are naturally capable of adapting to changing conditions such as modifications to the network topology and communication infrastructure. Moreover, unlike centralized schemes, no single agent requires full knowledge of all network parameters while computing optimal setpoints.

Relevant works in recent literature on distributed optimization for problems in distributed systems include [4], [6]–[9]. References [6], [7] focus on voltage control in microgrids and distribution systems using decentralized, local, and distributed control schemes. Voltage control in smart grids considering both transmission and distribution grids is reviewed in [8]. References [4], [9] are both broader surveys of distributed approaches.

The major contributions of this work are:

- **Taxonomy:** A comprehensive taxonomy of distributed algorithms in distribution power grids is given, with the classifications based not only on algorithm and application types, but also based on data exchange mechanism, implementation type, communication type, and the underlying type of power system model.
- **Model Relaxations and Approximations:** Relaxations and approximations utilized in distributed optimization formulations are described.
- **Solution and Use Cases:** Two disparate use cases are presented, one of which employs a proximal-dual method while another employs a dual-ascent method that solves the underlying OPF in a distributed manner.
- **Research Needs:** An in-depth discussion on the research needs for implementation of distributed approaches in the real field is provided.

This paper has the following major differences relative to the work in [4], [9]: (i) domain of application, which is focused on the distribution grid, (ii) model formulations, approximations, and relaxations for AC OPF, and (iii) taxonomy of solution algorithms and two representative use cases that demonstrate two common approaches.

The organization of the paper is as follows. Section II formulates the OPF problem in both centralized and distributed settings and discusses various power flow approximations and relaxations used for distribution system analyses. Section III surveys different distributed algorithms applied for optimal control in distribution grids and discusses comparisons among them. Section IV discusses several use cases for distributed optimization algorithms. Section V presents an overview of research needs for field implementations of distributed algorithms in active distribution systems.

II. AC Optimal Power Flow Problems for Active Distribution Systems

Accurately modeling distribution systems requires AC power flow formulations which consist of nonlinear equations involving complex bus voltages, line power flows, and bus power injections. OPF problems which include the nonlinear AC power flow equations and power system operational bounds in their constraints are known as ACOPF problems. While OPF problems are not the only problems in the distribution grid that can be solved with distributed optimization, a large subset of problems does rely on the OPF solution. At a high level, the methods reviewed in this paper are generalizable to these types of problems. However, these problems also have their own unique features in facilitating and/or challenging the distributed algorithm design that are beyond the scope of this review. This section formulates ACOPF problems in both centralized and distributed settings and describes variants of ACOPF problems that have been proposed for distributed applications.

A. Centralized ACOPF Problems

The centralized ACOPF problem optimizes an objective function while satisfying steady-state power flow equations and operational constraints. The power flow equations are typically represented via either the *Bus Injection Model* [4], [10] or the *Branch Flow Model* (also called the *DistFlow Model*) [11]–[13]. In either representation, power flow equations introduce non-convexities in ACOPF problems. The power flow equations for multiphase systems are further complicated by interphase coupling [14], [15]. Along with equalities corresponding to the power flow equations, ACOPF problems include voltage limits (typically $\pm 5\%$ of the nominal voltage [16]), generator bounds, thermal limits on line flows, and constraints associated with legacy devices such as tap-changing transformers and line capacitors. Power injections from DERs are often limited by the apparent power ratings of their interfacing converters. Line flows are limited by the distribution lines' ampacities. Legacy devices are slow-acting in nature since they are geared by mechanical actuators and switches.

In centralized settings, the ADMS collects measurements from the entire system and solves an ACOPF problem. The centralized ACOPF problem contains both state variables (e.g., voltage phasors, power flows) and control variables (e.g., setpoints for legacy devices and DER outputs). Denote the set of all problem variables as x. The centralized ACOPF problem aims to minimize operational costs (1a) subject to the power flow equations (1b) and operational limits (1c):

$$\min_{x} \quad f(x) \tag{1a}$$

$$\text{subject to} \quad G(x) = 0 \tag{1b}$$

$$H(x) \leq h \tag{1c}$$

B. Distributed ACOPF Problems

Distributed OPF approaches involve two steps:

- *Decomposition*: The original centralized optimization problem is decomposed into several subproblems. Hence,

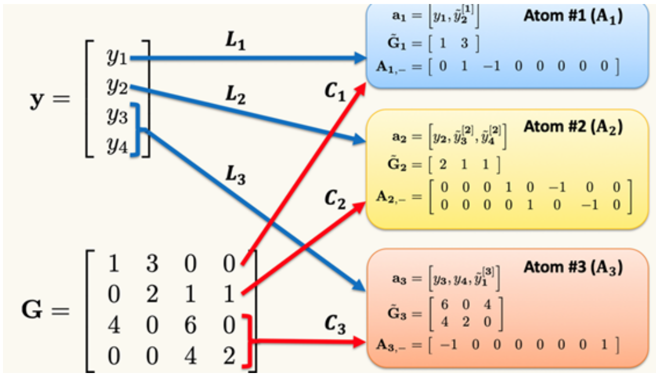


Fig. 1. A representation of the distribution of the objective function G over multiple agents G_j

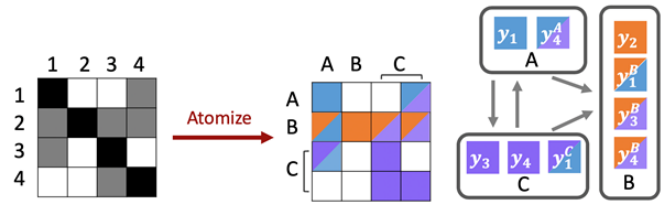


Fig. 2. A representation of the atomization and coupling from the distributed formulation over multiple agents

the centralized problem's objective and constraint functions are decomposed into subproblem-specific objective functions and constraint functions.

- *Coordination*: Each agent solves its subproblem and coordinates with its neighboring agents to share variables of mutual interest. Ultimately, the overall distributed optimization problem is solved when each agent optimizes its own subproblem while reaching consensus regarding values for the shared variables.

To formulate the distributed ACOPF problem, we decompose the centralized ACOPF problem (1) into k subproblems, each of which has an agent that controls the corresponding devices such as inverter-based DERs or legacy components like voltage regulators and shunt capacitors. The set of subproblems is $K = \{1, \ldots, k\}$.

The subproblem $j \in K$ associated with each agent j depends on a subset of the variables x that is denoted as x_j. Each agent performs calculations using a local copy of these variables, which is indicated as $\tilde{x}_j$. The objective, equality constraints, and inequality constraints in the subproblem for agent j are denoted as $f_j(\tilde{x}_j)$, $G_j(\tilde{x}_j)$, and $H_j(\tilde{x}_j)$, respectively. The distributed ACOPF is formulated as:

$$\min \quad \sum_{j \in K} f_j\left(\tilde{x}_j\right) \tag{2a}$$

$$\text{subject to} \quad G_j(\tilde{x}_j) = 0 \qquad j \in K \tag{2b}$$

$$H_j(\tilde{x}_j) \leq h_j \qquad j \in K \tag{2c}$$

$$A\begin{bmatrix}\tilde{x}_1^{\intercal} & \ldots & \tilde{x}_k^{\intercal}\end{bmatrix}^{\intercal} = 0 \tag{2d}$$

where the j^{th} agent solves the corresponding optimization problem defined by objective function (2a), power flow constraints (2b), and operational limits (2c), all of which are functions of agent j's local copy of the variables, $\tilde{x}_j$. This is visually represented in Fig. 1.

The constraint in (2d) represents a *consensus* or *coordination constraint* among neighboring agents. Optimization problems are usually not trivially decomposable, meaning that there are dependencies among different agents, such as a cost function or constraint that depends on variables that are shared with a different agent. With the matrix A constructed appropriately, constraints of the form (2d) address this dependency. Each agent sends information regarding its shared variables to its neighboring agents to reach consensus. The shared variables are often called "coupling" variables. A representation of the distribution and coupling is shown in Fig. 2.

Distributed optimization approaches apply various methods to simplify the coordination constraint (2d) into subproblem-specific formulations such that (2) can be solved in a distributed fashion. Examples of such decomposition-coordination based distributed approaches are the Alternating Direction of Method of Multipliers (ADMM) [17]–[20] and Proximal Atomic Coordination (PAC) [21]. The coupling variable information may include physical states (voltages, branch power flows, etc.) [22], [23], Lagrange multipliers [24], or functions related to reactive power and Lagrange multipliers [25]. Distributed approaches often enforce the coordination constraint (2d) while exploiting network sparsity. Such examples of network-sparsity based approaches are OPTDIST-VC [25] and DIST-OPT [24], [26]–[28].

C. Nonconvexities in ACOPF problems

The set of operating points satisfying the power flow equations (2b) and operational limits (2c) is referred to as the ACOPF problem's "feasible space". However, the power flow equations and other constraints are non-linear, meaning that ACOPF is a non-linear programming (NLP) problem. The major non-linearity lies in the power flow connecting two adjacent buses which is a non-linear function of bus voltages and bus angles [29]. These non-linear power flow equations and operational constraints make the ACOPF feasible space nonconvex [30]–[32]. Moreover, the presence of on-load tap changing transformers and shunt capacitors introduce binary variables in the ACOPF problem, thus adding further non-convexities [15]. Non-convex feasible spaces give rise to the possibility of local solutions rather than a single global optimum [29] as well as the potential for disconnected feasible spaces [30]. Since the ACOPF problem is non-convex, distributed approaches that behave like local solvers may converge to a locally optimal point rather than the global optimum [33], [34]. The ACOPF problem is NP-Hard in general [35] and for tree networks too [36]. Thus, ACOPF problems pose two major challenges:

- Nonconvexity of the ACOPF problem's feasible space resulting in the possibility of finding a local solution instead of a global solution.
- Problem intractability where solution algorithms cannot solve the problem in polynomial time.

D. Approximations and Relaxations in Distributed Algorithms

There are two main approaches for addressing the challenges posed by the non-convexity and NP-hardness of ACOPF problems: (a) Use an off-the-shelf local solver with an initialization that is close to the global optimum. A sufficiently close initialization should enable the algorithm to converge to the global optimum, while a poor initialization may result in failure to converge or convergence to a local optimum. (b) Use convex relaxations or approximations of the power flow equations to convert the centralized non-convex ACOPF problem into a convex programming problem. Once convex, the problem can be solved using any off-the-shelf convex programming solver with polynomial runtime. This can address problems posed by computational intractability.

Convex relaxations enclose the non-convex AC feasible space within a convex space. Once the relaxed ACOPF space is convex, any off-the-shelf convex programming solver can find the globally optimal point. However, it needs to be verified that the solution obtained is feasible, i.e., it must lie within the original ACOPF problem's non-convex feasible space. One advantage of using convex relaxations is that they always provide lower bounds of the original minimization problem of ACOPF. To summarize, convex relaxations applied in distributed algorithms are often one of two types: (a) *Second-Order Cone Programming (SOCP) relaxations* and (b) *Semidefinite Programming (SDP) relaxations*. Many SOCP relaxations replace an equality constraint associated with line losses in the DistFlow equations with a less restrictive inequality constraint. These SOCP relaxations are "exact" (provide the global solution to the original nonconvex ACOPF problem) for radial networks represented via single-phase balanced power flow constraints which also satisfy certain nontrivial technical conditions [37]. SDP relaxations are tighter than certain SOCP relaxations and can have advantages when considering meshed networks and three-phase unbalanced network models. SDP relaxations are typically constructed by reformulating the ACOPF with linear constraints along with a rank constraint on a matrix whose entries represent voltage phasor products. The SDP relaxation is formed by replacing this rank constraint with a weaker positive semidefinite constraint on this matrix.

Unlike convex relaxations, convex approximations do not enclose the non-convex ACOPF space. Instead, they use assumptions regarding the power flow equations to obtain a convex formulation. Convex approximations may greatly reduce the computational effort relative to convex relaxations. In both cases, solution feasibility must be evaluated.

One of the most common approximations is the DC power flow model [38], which assumes (a) lossless lines, (b) voltage magnitudes are close to unity, (c) reactive power is neglected, and (d) angle differences between connected buses are small. Unlike transmission lines, these assumptions are not valid for distribution lines since their resistance-to-inductance (R/X) ratio is high. Accordingly, many distributed algorithms use other linear approximations in distribution system applications, such as the Linearized DistFlow approximation [11].

III. Classification of Distributed Algorithms

This section classifies distributed algorithms used for optimal operation and control of distribution systems into various categories. Fig. 3 presents a taxonomy of distributed algorithms based on their data exchange mechanism, implementation type, power system model, algorithm type, communication paradigm, and application type.

We categorize distributed algorithms based on how data is exchanged with the grid as either (a) *static optimization algorithms* and (b) *dynamic optimization algorithms* (also known as "offline" and "online" algorithms, respectively). In *static optimization algorithms*, control agents communicate with neighboring agents in each optimization iteration and generate control setpoints based on their distributed/atomized optimization problems [10], [17], [17], [19], [21], [39]. Before implementing any actions in the physical system, a solution is obtained through multiple communication rounds among agents with computations performed during each iteration.

In *dynamic optimization algorithms*, each optimization iteration consists of control agents sensing grid variables (e.g., voltages, currents, and power flows), communicating with their neighboring agents, and computing control setpoints based on each agent's local optimization problem. In contrast to static optimization algorithms, these control setpoints are immediately applied to the physical grid as the DER controller references, thus directly affecting the power grid [22], [25]–[27], [42], [44], [45]. The algorithm then operates on the next iteration based on the grid's response to the previous iteration followed by communication and optimization computations.

Distributed optimization can be implemented with shared access to a database (e.g., using cloud computing), hence *federated* [47], or with data access only available locally (e.g., using fog computing), hence *peer-to-peer (P2P)*. *P2P* implementation truly allows distributed optimization, while preserving privacy with no centralized database access [48]. At the same time, *federated* is easier to implement with access to a centralized database and large computing facility, while *P2P* is harder to implement due to higher requirements on communication and computation placed on the distributed computing agents.

Distributed approaches typically use either branch flow based [18], [20], [22], [25], [39], [41], [42], [42], [44], [45] or bus injection based power system models [10], [17], [19], [27], [43]. Since both of these models are non-convex, convex relaxations or approximations (e.g., SDP and SOCP relaxations, the Linearized DistFlow approximation) are usually applied to formulate the problem as a convex optimization problem in order to achieve both computational tractability and convergence guarantees for the distributed algorithms.

Distributed approaches can also be classified into two major types: (a) *Optimization approaches* and (b) *Coordination methods*. Optimization approaches can be classified in turn into two sub-categories: (a) Primal-dual algorithms, and (b) Dual-ascent algorithms. Both of these formulations consist of a dual function formulation with corresponding dual variables associated with constraints. Maximizing the dual function provides a lower bound of the primal problem. Dual ascent

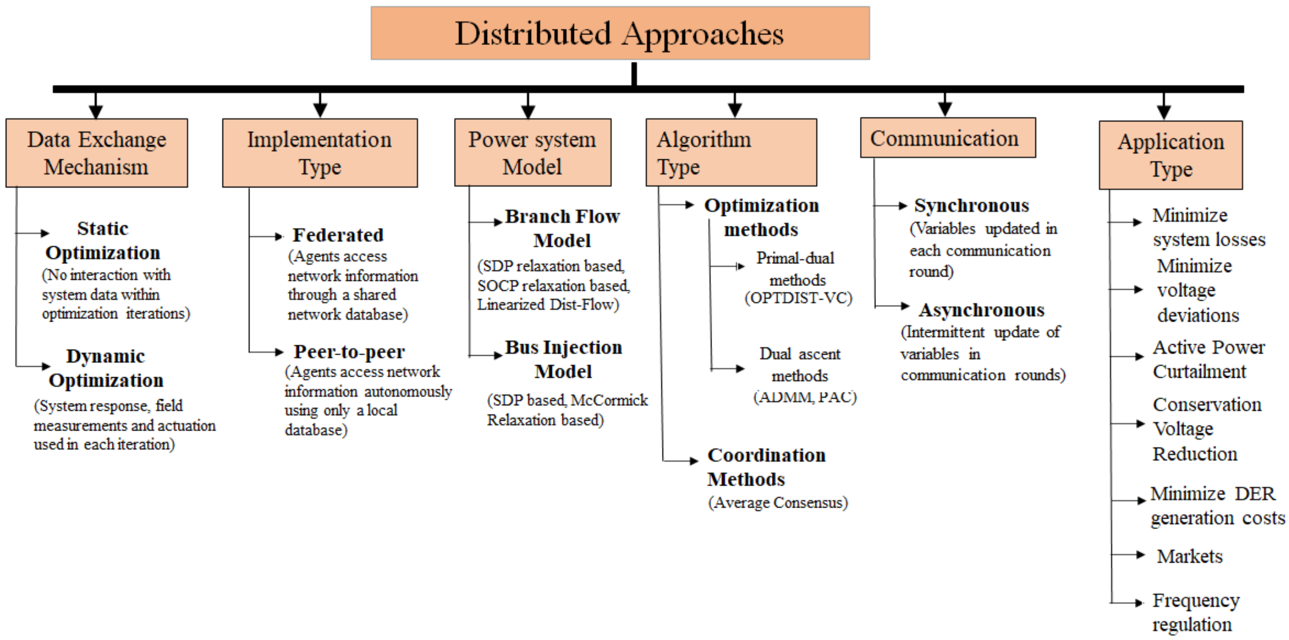


Fig. 3. A Taxonomy of Distributed Approaches.

TABLE I
RELATED WORK ON DISTRIBUTED LAGRANGIAN DUAL BASED OPTIMIZATION ALGORITHMS FOR DISTRIBUTION SYSTEMS

Objective	Reference	Power Flow	Network Model	Approximations /Relaxations	Static /Dynamic	Algorithm	Coupling Variables	Communication
Minimize losses	[10]	BIM	Balanced	SDP rank-1 constraint	Static	Dual-ascent Method	Lagrangian multipliers	Robust under failures
	[27]	BIM	Unbalanced	Losses approximated	Dynamic	Dual decomposition	Lagrangian multipliers	Asynchronous
	[17]	BIM	Unbalanced	SDP rank-1 constraint	Static	ADMM	Bus voltages	Synchronous
	[18]	BFM	Unbalanced	SOCP	Static	ADMM	Branch flows voltages	Synchronous
	[19]	BIM	Unbalanced	SDP rank-1 constraint	Static	ADMM	Voltages	Synchronous
	[39]	BFM	Balanced	Losses ignored	Static	ADMM	Lagrangian Multipliers	Synchronous
	[40]	BFM	Balanced	Losses ignored	Static	ADMM	Lagrangian Mulitpliers	Synchronous
Minimize voltage deviations	[41]	BFM	Unbalanced	SDP rank-1 constraint	Static	ADMM	Bus voltages, Branch power flows	Synchronous
	[22]	BFM	Unbalanced	Losses Ignored	Dynamic	ADMM	Reactive power Voltage	Asynchronous
	[42]	BFM	Unbalanced	Losses ignored	Dynamic	Partial Primal-Dual method	Bus voltages	Asynchronous
Active Power Curtailment	[43]	BIM	Balanced	SDP rank 1 constraint	Static	ADMM	Active power Reactive power	Synchronous
	[44]	BFM	Balanced	Ignore Losses	Dynamic	Dual Ascent Method	Lagrangian multipliers	Robust
CVR	[20]	BFM	Unbalanced	Ignore losses	Static	ADMM	Power flows Bus voltages	Synchronous
Minimize DER generation costs	[26]	BFM	Balanced	Ignore Losses	Dynamic	Primal-Dual Method	Reactive power	Limited communication
	[25]	BFM	Balanced	Ignore losses	Dynamic	Primal-Dual Method	Lagrangian mulitpliers Reactive power	Asynchronous and robust
	[21]	BFM	Balanced	SOCP	Static	Proximal Atomic Coordination	Branch Flows Bus Voltages	Synchronous
	[45]	BFM	Balanced	Ignore losses	Dynamic	Dual-ascent Method	Lagrange multipliers	Asynchronous
	[46]	BFM	Balanced	SOCP	Static	ADMM	Lagrangian Multipliers	Synchronous

algorithms typically solve the dual problem with gradient descent. At each iteration, with the value of the dual variable fixed, the primal problem is completely solved. The resulting primal solution is used to determine the dual variable at the next iteration and so on. The dual ascent method is a precursor to the dual decomposition, method of multipliers, ADMM, and PAC. On the other hand, the primal-dual methods update both the primal and dual variables at each iteration. Both of these approaches have advantages and disadvantages, which are explored more in the use cases. A list of distributed dual methods is presented in Table I. These distributed algorithms are categorized on the basis of the power flow model used, convex relaxations or convex approximations applied, data exchange mechanism, type of coupling or shared variables among the control agents, and the type of communication.

Coordination methods are usually implemented through an average consensus mechanism. In *consensus based methods*, the distributed optimization problem is solved directly in its primal form using communication based coordination. Some examples of consensus based primal methods are the distributed sub-gradient based method [49] and average consensus based methods [50]. Since our focus is more on optimization based approaches, a detailed review of coordination methods is not presented in this paper.

Regarding the communication paradigm, distributed approaches either use *synchronous* or *asynchronous* communication. In *synchronous* communication, control agents share coupling variables during every communication round [17]–[21], [39], [41], [43]. *Asynchronous* communication results due to latency in communication channels, loss of data, and noisy communication channels. Hence, control agents operate on variables shared in previous iterations in case of latency. When updated data is not available, control agents do not have any new inputs for their local optimization algorithms and thereby revert to inputs from previous iterations. Noisy data may result in non-optimal or even infeasible control decisions generated in each optimization iteration. Asynchronous communication may also result in non-convergence of various distributed approaches. Examples of distributed approaches using asynchronous communication are presented in [22], [25]–[27], [42], [44], [45].

Distributed approaches have been considered for many applications related to the optimal operation of distribution grids, including (a) minimizing power losses [10], [17]–[19], [27], [39], (b) minimizing voltage deviations [22], [41], [42], (c) minimizing active power curtailment (APC) [44], [51], (d) performing conservation voltage reduction (CVR) [20], (e) minimizing DER generation costs [21], [25], [26], (f) maximizing social welfare, and (g) regulating the system frequency.

IV. Use cases

This section describes use cases for Lagrangian dual based distributed optimization algorithms in distribution power grids. The different streams of methods used in the literature can be broadly classified per the taxonomy in Fig. I. The vast majority of the papers using distributed optimization techniques utilize two methods – primal-dual or dual-ascent. The two use cases that we will present are based on these two methods.

Section IV-A presents a network-sparsity based primal-dual algorithm for voltage control in active distribution systems called OPTDIST-VC. Section IV-B presents a classical decomposition-coordination based distributed dual algorithm called PAC, which uses a dual-ascent approach. The approaches in Sections IV-A and IV-B have other distinctions as well, which are based on the nature of the distributed optimization. These are enumerated in Table II. OPTDIST-VC uses a linearized power flow model called LinDistFlow, while PAC uses the nonlinear variant, DistFlow. The detailed linearization of these problems are not in the scope of this paper and the reader is referred to [23], [25] for further details.

Apart from this key distinction, various solution techniques can be used to solve the distributed problem. In this specific case, PAC uses a dual-ascent approach, while OPTDIST-VC utilizes a primal-dual solver. The algorithms also differ in their actuations – while PAC actuates once the complete optimization problem is solved, OPTDIST-VC actuates at the end of every timestep (which can be chosen based on the distribution system). While both approaches lead to optimal solutions, there are upsides and downsides to both approaches. While PAC's performance is dependent on the atomization, acceleration constants and other algorithmic parameters, OPTDIST-VC's performance is dependent on the validity of the model's radiality, availability of sufficient number of agents, and choosing an appropriate timestep. These issues are open research problems in the distributed optimization area.

A. Volt-Var Control

We consider a feedback based voltage control strategy where distribution feeder voltages are controlled by varying reactive power injections from DERs. At time t, we denote the vector of controllable reactive power injections as $\mathbf{q(t)} = [q_1(t)\ q_2(t)\ \dots\ q_n(t)]^T$, where n denotes the number of buses in the network. Let v_o be the substation voltage. The distribution feeder voltage vector $\mathbf{v(t)} = [v_1(t)\ v_2(t)\ \dots\ v_n(t)]^T$ can be approximated as [25]:

$$\mathbf{v(q(t))} = \mathbf{X}\,\mathbf{q(t)} + v^{par} = \mathbf{v(t)} \tag{3a}$$

$$v^{par} = \mathbf{X}\,\mathbf{q^u(t)} + \mathbf{R}\,\mathbf{p(t)} + v_o \tag{3b}$$

where $\mathbf{R}$ and $\mathbf{X}$ are the resistance and reactance matrices. v^{par} represents the uncontrollable part of the above equation whereas reactive power injection vector (q^c) represents the controllable part of the equation. Under certain loading conditions, v^{par} remains fixed and q^c can be modified to control feeder bus voltage vector v. Additional details on the parameters are provided in [25]. The vectors of uncontrollable reactive and active power injections are denoted as $\mathbf{q^u(t)}$ and $\mathbf{p(t)}$, respectively. Given $\mathbf{v(t)}$ from (3a), the algorithm seeks optimal controllable reactive power injections $\mathbf{q(t+1)}$ for the next time instant $(t+1)$. The results should satisfy operational constraints on the reactive power injections and voltage magnitudes:

$$\underline{\mathbf{v}}\mathbf{(t)} \leq \mathbf{v(t))} \leq \bar{\mathbf{v}}\mathbf{(t)} \tag{4a}$$

$$\underline{\mathbf{q}}\mathbf{(t)} \leq \mathbf{q(t))} \leq \bar{\mathbf{q}}\mathbf{(t)} \tag{4b}$$

TABLE II
DISTINCTIONS BETWEEN THE USE CASES: OPTDIST-VC AND PAC

Solution approach	OPTDIST-VC	PAC
Formulation	Leverage the underlying unique features of distribution systems to create a naturally distributed problem	Create an “atomized” problem that are solved by several sub-agents
Power flow model	LinDistFlow model, local measurements from the system is used as a surrogate for the power balance constraint	Non-linear DistFlow model
Data exchange mechanism and actuation	Dynamic optimization, optimal reactive power setpoints are directly set as DER power references in the next iteration	Static optimization, actuates once the complete optimization problem is solved
Algorithm used	Primal-dual approach	Dual ascent approach

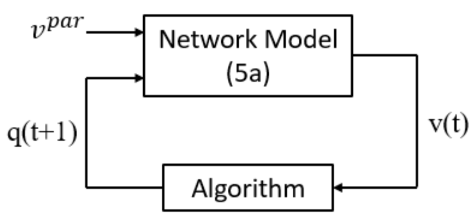
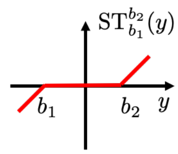


Fig. 4. Feedback structure of the voltage control problem.



$$\mathrm{ST}_{b_1}^{b_2}(y) = \begin{cases} y - b_2 & \text{if } y \geq b_2 \\ y - b_1 & \text{if } y \leq b_1 \\ 0 & \text{otherwise.} \end{cases}$$

Fig. 5. The soft thresholding function.

Fig. 4 shows the feedback structure of voltage control problem.

For a particular time t, the power flow equations (3) and operational constraints (4) can be expressed as (5a) and (5b). The term t is dropped hereafter for notational brevity.

$$G(\mathbf{v}, \mathbf{q}, v^{par}) = 0 \tag{5a}$$

$$H(\mathbf{v}, \mathbf{q}) \leq h \tag{5b}$$

The algorithm drives the network to the optimal point of the following optimization problem under any loading conditions:

$$\min_{q_k} \quad f(\mathbf{q}) \triangleq \sum_{k=1}^{n} f_k(q_k) + \frac{d}{2}\mathbf{q}^T X \mathbf{q} \tag{6a}$$

$$\text{subject to} \quad \text{(5a), (5b)} \tag{6b}$$

The cost function (6a) is the sum of the operating costs f_k and $\frac{1}{2}q^T X q$ which is a network-wide cost, $d \geq 0$ being a weighting parameter. d can be set to zero ($d = 0$) thus ignoring the cost term $\frac{d}{2}q^T X q$. The cost $\frac{1}{2}q^T X q$ approximates the network losses (up to a multiplicative factor and an additive term that does not depend on $\mathbf{q}$), under the assumption that the R/X ratio of the network is constant [28, Lemma 2]. This cost is commonly used in distributed volt-var controller design; see [52], [53].

In order to solve this problem, we next formally introduce the distributed voltage control algorithm known as “Optimal Distributed Feedback Voltage Control” (OPTDIST-VC) [25]. For each bus k, we introduce auxiliary variables, $\hat{q}_k, \xi_k, \bar{\lambda}_k, \underline{\lambda}_k$. At each iteration t, bus k measures the local voltage $v_k(t)$, computes variables $\hat{q}_k(t+1)$, $q_k(t+1)$, $\xi_k(t+1)$, $\bar{\lambda}_k(t+1)$, and $\underline{\lambda}_k(t+1)$, injects the reactive power $q_k(t+1)$, and lastly shares certain variables to its neighboring buses. A detailed description of OPTDIST-VC follows.

OPTDIST-VC:
At time t, each bus k executes the following four steps:
Step 1 (Measuring): Measure the local voltage $v_k(t)$.
Step 2 (Calculating): Calculate $\hat{q}_k(t+1), \xi_k(t+1), \bar{\lambda}_k(t+1), \underline{\lambda}_k(t+1)$ as follows.

$$\begin{aligned} \hat{q}_k(t+1) &= \hat{q}_k(t) - \alpha\Big\{\bar{\lambda}_k(t) - \underline{\lambda}_k(t) + d\hat{q}_k(t) \\ &\quad + \sum_{i\in\mathcal{N}_k}[Y]_{ki}\Big[f_i'(\hat{q}_i(t)) + \mathrm{ST}_{c\underline{q}_i}^{c\bar{q}_i}(\xi_i(t) + c\hat{q}_i(t))\Big]\Big\}, \end{aligned} \tag{7a}$$

$$\xi_k(t+1) = \xi_k(t) + \beta\frac{\mathrm{ST}_{c\underline{q}_k}^{c\bar{q}_k}(\xi_k(t) + c\hat{q}_k(t)) - \xi_k}{c}, \tag{7b}$$

$$\bar{\lambda}_k(t+1) = [\bar{\lambda}_k(t) + \gamma(v_k(t) - \bar{v}_k)]^+, \tag{7c}$$

$$\underline{\lambda}_k(t+1) = [\underline{\lambda}_k(t) + \gamma(\underline{v}_k - v_k(t))]^+, \tag{7d}$$

where $[\cdot]^+$ denotes projection onto the nonnegative orthant and the quantities α, β, γ, and c are positive scalar parameters. For any $b_1 < b_2$, let $\mathrm{ST}_{b_1}^{b_2}(\cdot)$ denote the soft-thresholding function defined as $\mathrm{ST}_{b_1}^{b_2}(y) = \max(\min(y - b_1, 0), y - b_2)$. $\mathcal{N}_k$ is the set of neighbor agents of agent k. (See Fig. 5 for an illustration of this function.)

Step 3 (Injecting Reactive Power): Set reactive power injection at time $t+1$ as

$$q_k(t+1) = [\hat{q}_k(t+1)]_{\underline{q}_k}^{\bar{q}_k}, \tag{8}$$

where $[\cdot]_{\underline{q}_k}^{\bar{q}_k}$ denotes projection onto the set $[\underline{q}_k, \bar{q}_k]$.

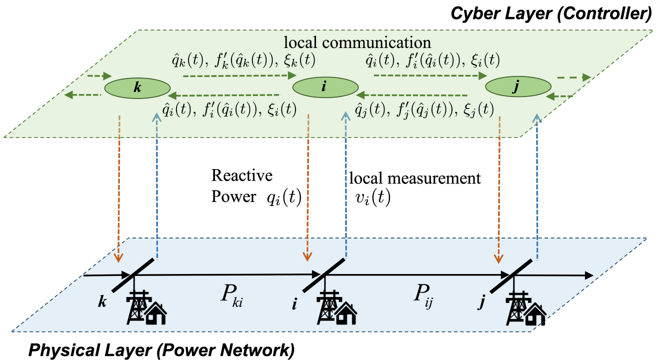


Fig. 6. Information Flow of OPTDIST-VC.

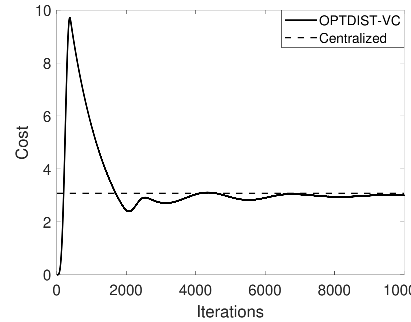


Fig. 7. Convergence charateristics (OPTDIST-VC vs centralized)

Step 4 (Communicating): Send values $f'_k(\hat{q}_k(t+1)) + \mathrm{ST}^{c\bar{q}_k}_{c\underline{q}_k}(\xi_k(t+1) + c\hat{q}_k(t+1))$ to all neighboring buses.

OPTDIST-VC is a primal-dual gradient algorithm [54]–[57] for an augmented Lagrangian [58], in which $\hat{q}_k(t)$ is the primal variable, $\xi_k(t)$, $\bar{\lambda}_k(t)$, and $\underline{\lambda}_k(t)$ are dual variables, and α, β, and γ are step sizes. Optimization problem (6) in OPTDIST-VC resembles (2) where $\tilde{x}_k(t) = [\hat{q}_k(t), v_k(t), \xi_k(t), \bar{\lambda}_k(t), \underline{\lambda}_k(t)]^T$. However, the coordination constraints among neighboring agents is taken care of by utilizing network sparsity. As can be observed in (7a), the term $\sum_{i\in\mathcal{N}_k}[Y]_{ki}\Big[f'_i(\hat{q}_i(t)) + \mathrm{ST}^{c\bar{q}_i}_{c\underline{q}_i}(\xi_i(t) + c\hat{q}_i(t))\Big]$ only consists of calculation of auxiliary variables belonging to set $\mathcal{N}_k$ where $\mathcal{N}_k$ denotes the set of neighbor agents of agent k. Hence, agent k needs to communicate only with neighboring agents to calculate $\hat{q}_k(t+1)$.

Fig. 6 shows the information exchange between different buses and between the cyber layer (controller) and physical layer (network model) under OPTDIST-VC. The only interaction between the cyber layer and the physical layer is through voltage measurement $v_k(t)$ (Step 1) and reactive power injection $q_k(t)$ (Step 3) as shown in Fig. 6. However, all other steps of OPTDIST-VC are performed entirely inside the cyber layer, including calculation of auxilliary variables (Step 2) and communication among neighbor agents (Step 4). We make a few comments regarding OPTDIST-VC:

- $q_k(t)$ and $v_k(t)$ are physical quantities (reactive power injection and voltage) being exchanged between cyber layer and physical layer, while $(\hat{q}_k(t), \xi_k(t), \bar{\lambda}_k(t), \underline{\lambda}_k(t))$ are "digital" variables stored in the controller's memory.
- Variable $\hat{q}_k(t)$ is the desired amount of reactive power to be injected by physical DER controller at bus k. However, $\hat{q}_k(t)$ may violate the reactive power capacity constraint. Therefore, we set $q_k(t)$ to the projection of $\hat{q}_k(t)$ onto the capacity constraint.
- The update equation (7a) for the desired reactive power injection $\hat{q}_k(t)$ drives $\hat{q}_k(t)$ towards the superposition of the gradient of f and certain "correction directions", related to $\bar{\lambda}_k(t) - \underline{\lambda}_k(t)$ and $\xi_k(t)$, which directs $\hat{q}_k(t)$ to satisfy the constraints. Due to the superposition of the two directions, $\hat{q}_k(t)$ will be driven to minimize f and also avoid constraint violations.
- The variables $\xi_k(t)$ and $\{\bar{\lambda}_k(t), \underline{\lambda}_k(t)\}$ are Lagrange multipliers associated with violations of the reactive power limits and voltage limits, respectively.

In OPTDIST-VC, for any $c > 0$, when α, β, and γ are small enough and satisfy mild conditions, $\mathbf{q}(t)$ will converge to the unique optimizer of (6). This is proved in [25]. Fig. 7 shows convergence characteristics of OPTDIST-VC.

Online dynamic optimization techniques such as OPTDIST-VC are essentially feedback-based approaches where the nonlinear power system acts as the plant whereas controllers are designed to operate in a distributed manner, as shown in Fig. 4. The reactive power setpoints calculated following Steps 1-4 as presented are then input back to the plant. Hence, if the algorithm receives voltage measurements at time instant t, it facilitates communication among agents at time instant $t+1$ and generates reactive power setpoints at $t+1$ which are then input to the plant model. In Fig. 6, consider the case for controller located in the cyber layer corresponding to the physical layer at bus i. The controller i receives local voltage measurement $v_i(t)$ which represents Step 1. The reactive power setpoints $q_i(t+1)$ are calculated based on shared variables obtained at time instant t. This represents Step 2 and Step 3. The controller i then shares the local variables $\hat{q}_i(t)$, $f'_i(\hat{q}_i(t))$, and $\xi_i(t)$ with the neighboring controllers at bus k and bus j. Note that since the reactive power setpoints of DERs as calculated in Step 2 are input back to the plant at time instant $(t+1)$, these setpoints should be within the capacity constraints of the DER inverter. Hence, the reactive power setpoints are projected back to the capacity constraint set as in Step 3 and then the resultant reactive power setpoint is input back to the system. This avoids infeasible operating points within optimization operations. Almost all dynamic optimization approaches utilize this projection operator in each optimization operation when power setpoints are input back to the power system model [22], [26], [42], [44].

B. Retail Markets using Distributed Algorithms

The Proximal Atomic Coordination (PAC) Algorithm is a recently developed distributed optimization algorithm [21], [59] with enhanced privacy-preserving capabilities. The algorithm leverages local data and measurements as well as structured communications between immediate neighbours to recover the optimal actuation required to minimize a global objective subject to network constraints. The network is represented as a directed graph $\mathbf{\Gamma}_{\rm D} = \langle \mathcal{B}, \mathcal{T}_{\rm D} \rangle$, where $j \in \mathcal{B}$ represents the nodes and $\mathcal{T}_{\rm D} \subseteq \mathcal{T}$ represents the directed edges.

A general optimization problem (1) subject to equality and inequality constraints can be decomposed (or atomized) into j different optimization problems, where a_j is the local atomic decision vector for node j. This is done as per the decomposition profile discussed in [21], [59], which has local copies of coupling variables between two neighboring nodes j and i, in either x_j and x_i or constraint matrices G_j or H_j. Additional equality constraints, termed "coordination constraints", are introduced to enforce the copies to coincide with the true value of the coupling variables at convergence:

$$Aa = 0, \tag{9}$$

where A represents the adjacency matrix of $\Gamma_{\rm D}$. The atomized problem then takes on the form of (2), and can be solved with the fully distributed PAC algorithm.

The algorithm is based on a distributed linearized variant of the proximal method of multipliers [60], [61] and is stated below (see [21], [59] for further details):

$$a_j\left[\tau+1\right] = \underset{a_j}{\text{argmin}} \left\{ \begin{array}{c} \mathcal{L}_j\left(a_j, \bar{\mu}_j\left[\tau\right], \bar{\nu}\left[\tau\right]\right) \\ +\frac{1}{2\rho}\left\|a_j - a_j\left[\tau\right]\right\|_2^2 \end{array} \right\}, \tag{10a}$$

$$\mu_j\left[\tau+1\right] = \mu_j\left[\tau\right] + \rho\gamma_j \tilde{G}_j a_j\left[\tau+1\right], \tag{10b}$$

$$\bar{\mu}_j\left[\tau+1\right] = \mu_j\left[\tau+1\right] + \rho\hat{\gamma}_j[\tau+1]\tilde{G}_j a_j\left[\tau+1\right], \tag{10c}$$

$$\text{Communicate } a_j \text{ for all } j \in [K] \text{ with neighbors,} \tag{10d}$$

$$\nu_j\left[\tau+1\right] = \nu_j\left[\tau\right] + \rho\gamma_j\left[B\right]^{O_j} a\left[\tau+1\right], \tag{10e}$$

$$\bar{\nu}_j\left[\tau+1\right] = \nu_j\left[\tau+1\right] + \rho\hat{\gamma}_j[\tau+1]\left[B\right]^{O_j} a\left[\tau+1\right], \tag{10f}$$

$$\text{Communicate } \bar{\nu}_j \text{ for all } j \in [K] \text{ with neighbors,} \tag{10g}$$

where $\rho > 0$ is the common step-size and $\gamma_j, \hat{\gamma}_j[\tau] > 0$ are two over-relaxation terms with $\gamma_j > \hat{\gamma}_j[\tau] > 0$. In (10), μ are the dual variables corresponding to equality constraints and ν are the dual variables corresponding to coordination constraints. As shown in [21], [59], the primal variables a and dual variables μ and ν converge to the optimal solution a^*, μ^*, and ν^*, with rate $o(1/\tau)$, where τ is the number of algorithmic iterations, while maintaining complete privacy of the dual variables, which can be interpreted as shadow prices within a market.

The above distributed optimization problem is relevant in a market with an objective function corresponding to the DER's generation costs. In a competitive economic environment, such as a market involving DERs, the distributed algorithm must also preserve the privacy of each agents' information by limiting the dissemination of any sensitive information and protecting any sensitive data that is shared. We consider any information about an agent's computations that can be used to the competitive advantage of other agents as sensitive. We have limited the objectives of rogue agents to the use of sensitive information to sabotage the global convergence properties of the overall distributed algorithm, in contrast to other adversarial scenarios. Thus, sensitive information includes the cost functions $f_j(x_j)$, operating constraints, and the dual variables μ_j and ν_j. Under the PAC framework, the cost functions, operating constraints, and dual variables μ_j (which correspond to market prices) are not shared between neighbours and are thus kept private. The "protected" dual variable $\hat{\nu}_j$ is communicated between neighbours, but the "true" value ν_j cannot be recovered by a rouge agent due to the use of a time-varying rate $\hat{\gamma}_j[\tau]$, which is unique and private to each atom j. Information regarding the trajectory of ν_j may be used to sabotage the overall convergence of PAC through deliberate manipulations of the coordination constraints.

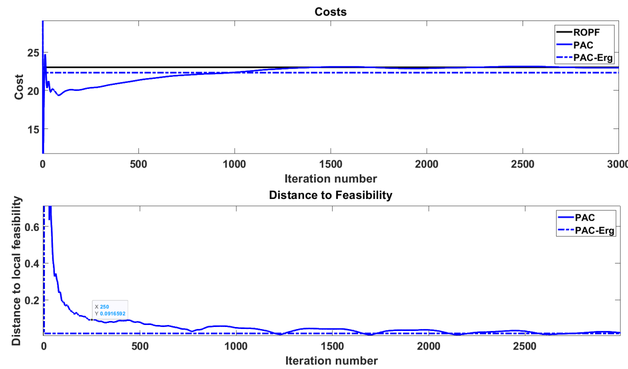


Fig. 8. Convergence characteristics for the PAC algorithm showing (a) costs comparison between centralized OPF and PAC and (b) distance to feasibility.

The main assumptions of this algorithm are that the cost and constraints are convex but both can be nonlinear. The PAC algorithm also assumes a radial, balanced three-phase system, and does not demonstrate the same convergence characteristic for unbalanced meshed networks. However, in contrast to the dynamic optimization presented in Section IV-A, this algorithm does not make any assumptions regarding the sparsity of the power system topology. Hence, the PAC algorithm can be extended to study meshed systems, if the coordination constraints can be appropriately formulated. Also, while performance worsens for unbalanced meshed networks, convergence can still be obtained if an appropriate time interval is chosen.

A short discussion on the simulation results using the PAC algorithm is presented below, and further details are provided in [21], [23], [59]. PAC is used to solve a retail market problem where the objective function maximizes social welfare. The algorithm is initialized using a flat start. To show the performance of the distributed optimization algorithm, two plots are presented in Fig. 8. The first plot shows the cost for the global atomic variable $\mathbf{a}[\tau]$ at every iteration τ. We compare this cost with that of the optimal solution obtained from the central solver, $f^{|OPF|}(\mathrm{x}^*)$. The second plot shows how close each atomic variable $\mathbf{a}_j[\tau]$ is to satisfying the local constraints, $\tilde{G}_j a_j[\tau] = 0$. From Fig. 8, observe that PAC exhibits decaying oscillatory behavior, with a reasonably accurate result achieved at around the 250 iteration mark. An algorithm for maximizing the convergence rate for the PAC algorithm is in [59, Theorem 4.7, 4.8]. The convergence rate depends on multiple factors including system size, convexity of the problem, atomization of the problem and so on. Further research is underway to determine the sensitivity of these parameters on the PAC convergence.

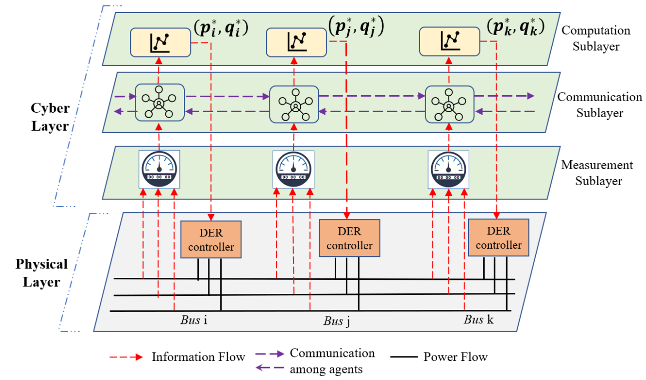


Fig. 9. An online distributed algorithm architecture.

C. Communication Requirements for Distributed Optimization

Distributed optimization techniques depend on the exchange of data between the various agents. An important aspect of distributed optimization is the communication network topology used to connect the power grid components, both agents and computation devices. For centralized optimization, the communication topologies have different Quality of Service (QoS) requirements as compared to distributed optimization algorithms. Typically, distributed optimization schemes need more communication infrastructure due to the higher number of agents [62]. However, QoS requirements can vary, and may even be less stringent than centralized optimization depending on how the distributed optimization is set up [63]. Also, the synchronous/asynchronous nature of the optimization algorithm is another important consideration. For synchronous optimization algorithms, the communication becomes more important, as the iteration cannot converge without data from all distributed nodes. For asynchronous schemes, this constraint can be relaxed as a solution can potentially be found even if data is not available simultaneously.

The choice of communication scheme often involves a trade-off between convergence rates and communication costs [64]. Distributed optimization techniques also involve iterative methods, and hence data needs to be communicated for every iteration and possibly even between iterations (such as in the PAC algorithm in steps (10d) and (10g)). Appropriately balancing this trade-off is essential for effective implementations of distributed optimization algorithms [65]. In the two use cases studied in Section IV-A and IV-B, this difference becomes evident. For a dynamic optimization scheme such as OPTDIST-VC (Section IV-A), the actuation is applied to the power grid control components at every time t. Conversely, for static optimization algorithms such as PAC (Section IV-B), the actuation is only applied after the algorithm converges. The communication scheme and infrastructure need to be designed considering these requirements. However, the communication between the distributed agents is crucial for both PAC and OPTDIST-VC, and further research is needed to study the effect of communication system effects and on system performance.

In addition to the communication scheme, the choice of communication network topology is an important consideration for distributed optimization. Studies such as [66], [67] explore various communication network topologies and their impacts on power system operational decisions, such as reconfiguration. These studies provide simulation-based empirical analyses regarding the effect of latency on power system control algorithms. Similar studies have also been performed for distributed optimization approaches by Guo et al. [68]. These studies explore the tradeoffs between communication latency, computational speed, and convergence time for various distributed optimization algorithms. Berahas et al. [64] propose a new metric to balance communication and computation requirements for distributed optimization algorithms, finding the optimal balance for one algorithm.

In addition to asynchronous methods, software-based methods adapted from distributed and fault-tolerant computing can also be applied to address challenges associated with latency. An example of this type of approach is to use Software Defined Networking to reroute packets in the presence of traffic congestion [69]. Other methods for addressing latency and handling data management requirements for distributed optimization algorithms include adopting lightweight protocols that more quickly transfer information [63] and moving to a cloud-based infrastructure instead of hierarchical/centralized communication topologies [70].

V. Research Needs and Path Forward

As illustrated in Fig. 9, a dynamic optimization based distributed algorithm operates on two layers: (a) the *physical layer* where the power network and power-electronic DER controllers interface with the physical grid and (b) the *cyber layer*, where cyber information is used to compute optimal setpoints that are fed back to the physical grid. In contrast to centralized optimization techniques which require a centralized control layer (the ADMS as discussed in the introduction), the communication and computing abilities of the distributed agents are merged into the cyber layer. The cyber layer has three sublayers of operation: (i) the *measurement sublayer*, which consists of sensors which measure power flows, power injections, and bus voltages from the grid, (ii) the *communication sublayer*, where control agents communicate problem variables (Lagrange multipliers or power system states) with neighboring agents, and (iii) the *computation sublayer*, where optimal active and reactive power setpoints (p_i^*, q_i^*) are computed using both the shared and local variables. These setpoints are input to fast-acting DER controllers which change system states in the physical grid.

Although distributed algorithms have certain advantages with respect to both centralized and purely local strategies, distributed algorithms may encounter errors in all three sublayers, potentially leading to non-optimal or even infeasible solutions. Possible source of these errors include:

- Noisy or incomplete data from sensors in the measurement layer.
- Failures transmitting data among neighboring agents in the communication layer.

- Modeling errors or inaccurate choices of algorithm parameters in the computation layer.
- Cyberattacks from malicious agents that compromise one or more of the measurement, communication, and computation layers.

Future research needs include both the theoretical foundations and practical considerations of implementing distributed algorithms:

- Practical implementations of distributed algorithms need to be robust to noise as well as communication and computation failures in order to ensure reliable operation.
- Distributed algorithms must be fast enough to cope with rapid changes in power grid conditions. Many existing algorithms can require thousands of iterations to converge to acceptable accuracy, suggesting that further improvements in convergence rates are needed.
- Distributed algorithms must be robust to failures and errors in the measurement layer, the communication layer, and the computation layer.
- Communication requirements should be simple and limited enough to be implemented via existing communication channels, such as power line communication [26].
- Methods for appropriately selecting algorithm parameters require more thorough study.
- The scalability of distributed algorithms needs to be demonstrated using increasingly large test cases with many DERs under diverse operational conditions and realistic communication infrastructures.
- Operation resulting from distributed algorithms should avoid implementing excessive switching and control actions. This is particularly important for dynamic distributed optimization algorithms.
- Theory regarding convergence guarantees withing reasonable timeframes is needed to provide mathematical rigor.
- The computational requirements for supporting federated or P2P optimization while meeting privacy, communication, and hardware controller requirements need further investigation.
- Modeling, analysis, and mitigation techniques for cyber attacks are needed to ensure acceptable operation of power systems managed using distributed algorithms.

VI. Conclusions

Distributed control algorithms provide multiple complementary advantages relative to traditional centralized and local control approaches in terms of computation, communication, privacy, flexibility, and scalability with increasing DERs at the edge. However, distributed control approaches often require several iterations and communication rounds to reach convergence, which can make them unsuitable for practical implementations in a federated or P2P manner. Existing work also lack a thorough analysis of parametric sensitivity towards algorithm performance and communication requirements for practical implementation. This paper presents a review of distributed algorithms found in the literature, a new taxonomy using key attributes, and a comparison of some use cases. Finally, future research needs for practical implementation of such distributed algorithms are also discussed.

References

[1] M. B. Cain, R. P. O'Neill, and A. Castillo, "History of optimal power flow and formulations," *Federal Energy Regulatory Commission (OPF Paper 1)*, pp. 1–36, 2012.

[2] S. Bolognani, R. Carli, G. Cavraro, and S. Zampieri, "On the need for communication for voltage regulation of power distribution grids," *IEEE Transactions on Control of Network Systems*, vol. 6, no. 3, pp. 1111–1123, 2019.

[3] S. Repo, S. Lu, T. Pöhö, D. Della Giustina, G. Ravera, J. M. Selga, and F. A.-C. Figuerola, "Active distribution network concept for distributed management of low voltage network," in *IEEE PES ISGT Europe 2013*.

[4] D. K. Molzahn, F. Dörfler, H. Sandberg, S. H. Low, S. Chakrabarti, R. Baldick, and J. Lavaei, "A survey of distributed optimization and control algorithms for electric power systems," *IEEE Transactions on Smart Grid*, vol. 8, no. 6, pp. 2941–2962, 2017.

[5] Y. Wang, P. Yemula, and A. Bose, "Decentralized communication and control systems for power system operation," *IEEE Transactions on Smart Grid*, vol. 6, no. 2, pp. 885–893, 2014.

[6] Y. Wang, S. Wang, and L. Wu, "Distributed optimization approaches for emerging power systems operation: A review," *Electric Power Systems Research*, vol. 144, pp. 127–135, 2017.

[7] K. E. Antoniadou-Plytaria, I. N. Kouveliotis-Lysikatos, P. S. Georgilakis, and N. D. Hatziargyriou, "Distributed and decentralized voltage control of smart distribution networks: Models, methods, and future research," *IEEE Transactions on Smart Grid*, vol. 8, no. 6, pp. 2999–3008, 2017.

[8] H. Sun, Q. Guo, J. Qi, V. Ajjarapu, R. Bravo, J. Chow, Z. Li, R. Moghe, E. Nasr-Azadani, U. Tamrakar, G. N. Taranto, R. Tonkoski, G. Valverde, Q. Wu, and G. Yang, "Review of challenges and research opportunities for voltage control in smart grids," *IEEE Transactions on Power Systems*, vol. 34, no. 4, pp. 2790–2801, 2019.

[9] T. Yang, X. Yi, J. Wu, Y. Yuan, D. Wu, Z. Meng, Y. Hong, H. Wang, Z. Lin, and K. H. Johansson, "A survey of distributed optimization," *Annual Reviews in Control*, vol. 47, pp. 278–305, 2019.

[10] B. Zhang, A. Y. Lam, A. D. Domínguez-García, and D. Tse, "An optimal and distributed method for voltage regulation in power distribution systems," *IEEE Transactions on Power Systems*, vol. 30, no. 4, pp. 1714–1726, 2014.

[11] M. Baran and F. F. Wu, "Optimal sizing of capacitors placed on a radial distribution system," *IEEE Transactions on Power Delivery*, vol. 4, no. 1, pp. 735–743, 1989.

[12] M. Farivar and S. H. Low, "Branch flow model: Relaxations and convexification–Part I," *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 2554–2564, 2013.

[13] L. Gan, N. Li, U. Topcu, and S. Low, "Branch flow model for radial networks: Convex relaxation," in *51st IEEE Conference on Decision and Control (CDC)*, IEEE, Dec. 2012.

[14] L. Gan and S. H. Low, "Convex relaxations and linear approximation for optimal power flow in multiphase radial networks," in *18th Power Systems Computation Conference (PSCC)*, pp. 1–9, 2014.

[15] R. R. Jha, A. Dubey, C.-C. Liu, and K. P. Schneider, "Bi-level volt-var optimization to coordinate smart inverters with voltage control devices," *IEEE Transactions on Power Systems*, vol. 34, pp. 1801–1813, May 2019.

[16] "Electric power systems and equipment voltage ratings (60 hertz)," *ANSI Standard Publication no. ANSI C84.1-1995*.

[17] B. A. Robbins, H. Zhu, and A. D. Domínguez-García, "Optimal tap setting of voltage regulation transformers in unbalanced distribution systems," *IEEE Transactions on Power Systems*, vol. 31, no. 1, pp. 256–267, 2015.

[18] W. Zheng, W. Wu, B. Zhang, H. Sun, and Y. Liu, "A fully distributed reactive power optimization and control method for active distribution networks," *IEEE Transactions on Smart Grid*, vol. 7, no. 2, pp. 1021–1033, 2016.

[19] E. Dall'Anese, H. Zhu, and G. B. Giannakis, "Distributed optimal power flow for smart microgrids," *IEEE Transactions on Smart Grid*, vol. 4, no. 3, pp. 1464–1475, 2013.

[20] Q. Zhang, K. Dehghanpour, and Z. Wang, "Distributed CVR in unbalanced distribution systems with PV penetration," *IEEE Transactions on Smart Grid*, vol. 10, no. 5, pp. 5308–5319, 2019.

[21] J. Romvary, *A proximal atomic coordination algorithm for distributed optimization in distribution grids*. PhD thesis, Massachusetts Institute of Technology, 2018.

[22] H. J. Liu, W. Shi, and H. Zhu, "Distributed voltage control in distribution networks: Online and robust implementations," *IEEE Transactions on Smart Grid*, vol. 9, no. 6, pp. 6106–6117, 2017.

[23] R. Haider, S. Baros, Y. Wasa, J. Romvary, K. Uchida, and A. M. Annaswamy, "Toward a retail market for distribution grids," *IEEE Transactions on Smart Grid*, vol. 11, pp. 4891–4905, Nov. 2020.

[24] S. Magnússon, G. Qu, and N. Li, "Distributed optimal voltage control with asynchronous and delayed communication," *IEEE Transactions on Smart Grid*, vol. 11, pp. 3469–3482, July 2020.

[25] G. Qu and N. Li, "Optimal distributed feedback voltage control under limited reactive power," *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 315–331, 2019.

[26] S. Magnússon, G. Qu, C. Fischione, and N. Li, "Voltage control using limited communication," *IEEE Transactions on Control of Network Systems*, vol. 6, no. 3, pp. 993–1003, 2019.

[27] S. Bolognani, R. Carli, G. Cavraro, and S. Zampieri, "Distributed reactive power feedback control for voltage regulation and loss minimization," *IEEE Transactions on Automatic Control*, vol. 60, no. 4, pp. 966–981, 2014.

[28] S. Bolognani and S. Zampieri, "A distributed control strategy for reactive power compensation in smart microgrids," *IEEE Transactions on Automatic Control*, vol. 58, no. 11, pp. 2818–2833, 2013.

[29] W. A. Bukhsh, A. Grothey, K. I. McKinnon, and P. A. Trodden, "Local solutions of the optimal power flow problem," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4780–4788, 2013.

[30] D. K. Molzahn, "Computing the feasible spaces of optimal power flow problems," *IEEE Transactions on Power Systems*, vol. 32, no. 6, pp. 4752–4763, 2017.

[31] B. C. Lesieutre and I. A. Hiskens, "Convexity of the set of feasible injections and revenue adequacy in FTR markets," *IEEE Transactions on Power Systems*, vol. 20, no. 4, pp. 1790–1798, 2005.

[32] I. A. Hiskens and R. J. Davy, "Exploring the power flow solution space boundary," *IEEE Transactions on Power Systems*, vol. 16, no. 3, pp. 389–395, 2001.

[33] A. J. Conejo, E. Castillo, R. Minguez, and R. Garcia-Bertrand, *Decomposition techniques in mathematical programming: Engineering and science applications*. Springer Science & Business Media, 2006.

[34] S. Boyd, N. Parikh, E. Chu, B. Paleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, pp. 1–122, January 2011.

[35] D. Bienstock and A. Verma, "Strong NP-hardness of AC power flows feasibility," *Operations Research Letters*, vol. 47, no. 6, pp. 494–501, 2019.

[36] K. Lehmann, A. Grastien, and P. Van Hentenryck, "AC-feasibility on tree networks is NP-hard," *IEEE Transactions on Power Systems*, vol. 31, no. 1, pp. 798–801, 2016.

[37] S. H. Low, "Convex relaxation of optimal power flow–Part II: Exactness," *IEEE Transactions on Control of Network Systems*, vol. 1, pp. 177–189, June 2014.

[38] B. Stott, J. Jardim, and O. Alsaç, "DC Power Flow Revisited," *IEEE Transactions on Power Systems*, vol. 24, pp. 1290–1300, August 2009.

[39] P. Šulc, S. Backhaus, and M. Chertkov, "Optimal distributed control of reactive power via the alternating direction method of multipliers," *IEEE Transactions on Energy Conversion*, vol. 29, no. 4, pp. 968–977, 2014.

[40] S. S. Guggilam, E. Dall'Anese, Y. C. Chen, S. V. Dhople, and G. B. Giannakis, "Scalable optimization methods for distribution networks with high PV integration," *IEEE Transactions on Smart Grid*, vol. 7, no. 4, pp. 2061–2070, 2016.

[41] B. A. Robbins and A. D. Domínguez-García, "Optimal reactive power dispatch for voltage regulation in unbalanced distribution systems," *IEEE Transactions on Power Systems*, vol. 31, no. 4, pp. 2903–2913, 2016.

[42] H. J. Liu, W. Shi, and H. Zhu, "Hybrid voltage control in distribution networks under limited communication rates," *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 2416–2427, 2018.

[43] E. Dall'Anese, S. V. Dhople, B. B. Johnson, and G. B. Giannakis, "Decentralized optimal dispatch of photovoltaic inverters in residential distribution systems," *IEEE Transactions on Energy Conversion*, vol. 29, no. 4, pp. 957–967, 2014.

[44] J. Li, Z. Xu, J. Zhao, and C. Zhang, "Distributed online voltage control in active distribution networks considering PV curtailment," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 10, pp. 5519–5530, 2019.

[45] L. Ortmann, A. Prostejovsky, K. Heussen, and S. Bolognani, "Fully distributed peer-to-peer optimal voltage control with minimal model requirements," *Electric Power Systems Research*, vol. 189, p. 106717, 2020.

[46] Q. Peng and S. H. Low, "Distributed optimal power flow algorithm for radial networks, i: Balanced single phase case," *IEEE Transactions on Smart Grid*, vol. 9, no. 1, pp. 111–121, 2016.

[47] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 10, no. 2, pp. 1–19, 2019.

[48] M. Miranbeigi, P. Kandula, K. Kandasamy, and D. Divan, "Collaborative Volt-VAR control using grid-connected PV inverters," in *IEEE 10th International Symposium on Power Electronics for Distributed Generation Systems (PEDG)*, pp. 252–257, 2019.

[49] A. Maknouninejad and Z. Qu, "Realizing unified microgrid voltage profile and loss minimization: A cooperative distributed optimization and control approach," *IEEE Transactions on Smart Grid*, vol. 5, no. 4, pp. 1621–1630, 2014.

[50] M. Zeraati, M. E. H. Golshan, and J. M. Guerrero, "A consensus-based cooperative control of PEV battery and PV active power curtailment for voltage regulation in distribution networks," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 670–680, 2017.

[51] E. Dall'Anese, S. V. Dhople, and G. B. Giannakis, "Optimal dispatch of photovoltaic inverters in residential distribution systems," *IEEE Transactions on Sustainable Energy*, vol. 5, no. 2, pp. 487–497, 2014.

[52] G. Cavraro, S. Bolognani, R. Carli, and S. Zampieri, "The value of communication in the voltage regulation problem," in *IEEE 55th Conference on Decision and Control (CDC)*, pp. 5781–5786, 2016.

[53] N. Li, G. Qu, and M. Dahleh, "Real-time decentralized voltage control in distribution networks," in *52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 582–588, IEEE, 2014.

[54] A. Nedić and A. Ozdaglar, "Subgradient methods for saddle-point problems," *Journal of Optimization Theory and Applications*, vol. 142, no. 1, pp. 205–228, 2009.

[55] A. Cherukuri, E. Mallada, and J. Cortés, "Asymptotic convergence of constrained primal–dual dynamics," *Systems & Control Letters*, vol. 87, pp. 10–15, 2016.

[56] D. Feijer and F. Paganini, "Stability of primal–dual gradient dynamics and applications to network optimization," *Automatica*, vol. 46, no. 12, pp. 1974–1981, 2010.

[57] G. Qu and N. Li, "On the exponential stability of primal-dual gradient dynamics," *IEEE Control Systems Letters*, vol. 3, no. 1, pp. 43–48, 2018.

[58] D. P. Bertsekas, *Constrained optimization and Lagrange multiplier methods*. Academic Press, 2014.

[59] J. J. Romvary, G. Ferro, R. Haider, and A. M. Annaswamy, "A proximal atomic coordination algorithm for distributed optimization," *IEEE Transactions on Automatic Control*, 2022.

[60] G. Chen and M. Teboulle, *A proximal-based decomposition method for convex minimization problems*. Springer-Verlag, 1994.

[61] N. Parikh and S. Boyd, "Proximal algorithms," *Foundations and Trends in Optimization*, vol. 1, no. 3, pp. 127–239, 2014.

[62] F. Karniavoura and K. Magoutis, "Decision-making approaches for performance QoS in distributed storage systems: A survey," *IEEE Transactions on Parallel and Distributed Systems*, vol. 30, no. 8, pp. 1906–1919, 2019.

[63] H. Gjermundrod, D. E. Bakken, C. H. Hauser, and A. Bose, "GridStat: A flexible QoS-managed data dissemination framework for the power grid," *IEEE Transactions on Power Delivery*, vol. 24, no. 1, pp. 136–143, 2008.

[64] A. S. Berahas, R. Bollapragada, N. S. Keskar, and E. Wei, "Balancing communication and computation in distributed optimization," *IEEE Transactions on Automatic Control*, vol. 64, no. 8, pp. 3141–3155, 2018.

[65] Y. Chow, W. Shi, T. Wu, and W. Yin, "Expander graph and communication-efficient decentralized optimization," in *50th Asilomar Conference on Signals, Systems and Computers*, 2016.

[66] V. Venkataramanan, Y. Zhou, and A. Srivastava, "Analyzing impact of communication network topologies on reconfiguration of networked microgrids," in *North American Power Symposium (NAPS)*, 2016.

[67] Z. Zhang and M. Chow, "Convergence analysis of the incremental cost consensus algorithm under different communication network topologies in a smart grid," *IEEE Transactions on Power Systems*, vol. 27, no. 4, pp. 1761–1768, 2012.

[68] J. Guo, G. Hug, and O. K. Tonguz, "On the role of communications plane in distributed optimization of power systems," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 7, pp. 2903–2913, 2017.

[69] F. Bannour, S. Souihi, and A. Mellouk, "Distributed SDN control: Survey, taxonomy, and challenges," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 1, pp. 333–354, 2017.

[70] M. D. Dikaiakos, D. Katsaros, P. Mehra, G. Pallis, and A. Vakali, "Cloud computing: Distributed internet computing for IT and scientific research," *IEEE Internet Computing*, vol. 13, no. 5, pp. 10–13, 2009.