Distributed Coordination of DERs With Storage for Dynamic Economic Dispatch
Cherukuri, Ashish; Cortes, Jorge

Published in:
IEEE Transactions on Automatic Control

DOI:
10.1109/TAC.2017.2731809

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Final author's version (accepted by publisher, after peer review)

Publication date:
2018

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
https://doi.org/10.1109/TAC.2017.2731809

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 05-01-2020
Distributed coordination of DERs with storage for dynamic economic dispatch

Ashish Cherukuri Jorge Cortés

Abstract—This paper considers the dynamic economic dispatch problem for a group of distributed energy resources (DERs) with storage that communicate over a weight-balanced strongly connected digraph. The objective is to collectively meet a certain load profile over a finite time horizon while minimizing the aggregate cost. At each time slot, each DER decides on the amount of generated power, the amount sent to/drawn from the storage unit, and the amount injected into the grid to satisfy the load. Additional constraints include bounds on the amount of power in storage, ramp constraints on the difference in generation across successive time slots, and bounds on the amount of power to be injected and the load for each time slot of the planning horizon. A majority of solution methods to the DED problem are centralized, and the implementation requires a central coordinator that communicates with all generators. MPC methods have also been employed in [11] for the dynamic economic dispatch with storage (DEDS) problem, which adds storage units to the DED problem to lower the total cost and smooth out the generation profile across time. The stochastic version of the DEDS problem adds uncertainty in demand and generation by renewables. Algorithmic solutions for this problem include stochastic MPC [12], dual decomposition [13], and optimal condition decomposition [14] methods. The work [15] proposes an ADMM-based algorithm to solve a variation of the DEDS problem where optimal electrical vehicle charging is the goal. The above-mentioned methods for the DEDS problem are either centralized or need a central coordinator. On the other hand, [16] proposes an ADMM-based distributed algorithm to find the optimizer of a general time-coupled dispatch problem. In comparison, the algorithm proposed here is robust to load variations and intermittent generator commitment.

Statement of contributions: We start with the formulation of the DEDS problem for a group of DERs communicating over a weight-balanced strongly connected digraph. Since the cost functions are convex and all constraints are linear, the problem is convex in its decision variables, which are the power to be injected and the power to be sent to storage by each DER at each time slot. Using exact penalty functions, we reformulate the problem as an equivalent optimization with equality constraints but without inequality ones. The structure of the modified problem guides our design of the provably-correct distributed strategy termed “dynamic average consensus (dac) + Laplacian nonsmooth gradient (L∂) + nonsmooth gradient (∂)” dynamics to solve the DEDS problem starting from any initial condition. The algorithm consists of two interconnected systems, one estimating the mismatch between the injection and the total load at each time slot, and another using this estimate to reduce the mismatch and optimize the total cost of generation while meeting the constraints.

I. INTRODUCTION

The current electricity grid is up for a major transformation to enable the widespread integration of distributed energy resources and flexible loads to improve efficiency and reduce emissions without affecting reliability and performance. This presents the need for novel coordinated control and optimization strategies which, along with suitable architectures, can handle uncertainties and variability, are fault-tolerant and robust, and preserve privacy. With this context in mind, our objective here is to provide a distributed algorithmic solution to the dynamic economic dispatch problem with storage. We see the availability of such strategies as a necessary building block in realizing the vision of the future grid.

Literature review: Static economic dispatch (SED) involves a group of generators collectively meeting a specified load for a single time slot while minimizing the total cost and respecting individual constraints. In recent years, distributed generation has motivated the shift from traditional solutions of the SED problem to decentralized ones, see e.g., [2], [3], [4] and our own work [5], [6]. As argued in [7], [8], the dynamic version of the problem, termed dynamic economic dispatch (DED), results in better grid control as it optimally plans generation across a time horizon, specifically taking into account ramp limits and variability of renewable sources. A majority of solution methods to the DED problem are centralized [7] with recent works employing model predictive control (MPC)-based algorithms [8], [9]. The work [10] proposes a Lagrangian relaxation method to solve the DED problem, but the implementation requires a central coordinator that communicates with all generators. MPC methods have also been employed in [11] for the dynamic economic dispatch with storage (DEDS) problem, which adds storage units to the DED problem to lower the total cost and smooth out the generation profile across time. The stochastic version of the DEDS problem adds uncertainty in demand and generation by renewables. Algorithmic solutions for this problem include stochastic MPC [12], dual decomposition [13], and optimal condition decomposition [14] methods. The work [15] proposes an ADMM-based algorithm to solve a variation of the DEDS problem where optimal electrical vehicle charging is the goal. The above-mentioned methods for the DEDS problem are either centralized or need a central coordinator. On the other hand, [16] proposes an ADMM-based distributed algorithm to find the optimizer of a general time-coupled dispatch problem. In comparison, the algorithm proposed here is robust to load variations and intermittent generator commitment.

Statement of contributions: We start with the formulation of the DEDS problem for a group of DERs communicating over a weight-balanced strongly connected digraph. Since the cost functions are convex and all constraints are linear, the problem is convex in its decision variables, which are the power to be injected and the power to be sent to storage by each DER at each time slot. Using exact penalty functions, we reformulate the problem as an equivalent optimization with equality constraints but without inequality ones. The structure of the modified problem guides our design of the provably-correct distributed strategy termed “dynamic average consensus (dac) + Laplacian nonsmooth gradient (L∂) + nonsmooth gradient (∂)” dynamics to solve the DEDS problem starting from any initial condition. The algorithm consists of two interconnected systems. A first block allows DERs to track, using dac, the mismatch between the current total power injected and the load for each time slot of the planning horizon. A second block has two components, one that minimizes the total cost while keeping the total injection constant (using Laplacian-nonsmooth-gradient dynamics on injection variables and nonsmooth-gradient dynamics on storage variables) and an error-correcting component that uses the mismatch signal estimated by the first block to adjust, exponentially fast, the total injection towards the load at each time slot.

Notation: Let \( R \), \( R_{\geq 0} \), \( R_{>0} \), \( Z_{\geq 1} \) denote the real, nonnegative real, positive real, and positive integer numbers, resp. The 2- and \( \infty \)-norm are \( \| \cdot \| \) and \( \| \cdot \|_{\infty} \). Let \( B_\delta(x) \) be the open ball of radius \( \delta > 0 \) centered at \( x \in \mathbb{R}^n \). For
For a symmetric $A \in \mathbb{R}^{n \times n}$, its min and max eigenvalues are $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$. The Kronecker product of $A \in \mathbb{R}^{n \times n}$ and $B \in \mathbb{R}^{p \times q}$ is $A \otimes B \in \mathbb{R}^{np \times nq}$. Let $\Omega_0 = \{0, \ldots, 0\} \in \mathbb{R}^n$, $I_n = (1, \ldots, 1) \in \mathbb{R}^n$, and $I_n \in \mathbb{R}^{n \times n}$ be the identity matrix.

Let $n \in \mathbb{Z}_{\geq 1}$, let $[n] = \{1, \ldots, n\}$. For $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$, $(x; y) \in \mathbb{R}^{n \times m}$ denotes its concatenation. For $x, y \in \mathbb{R}^n$, $x_i$ is the $i$-th component of $x$, and $x \leq y$ denotes $x_i \leq y_i$ for $i \in [n]$. For $h > 0$, $y \in \mathbb{R}^n$ and $k \in [n]$, $y(k) \in \mathbb{R}^n$ contains the $nk-n+1$ to $nk$ components of $y$ (and so $y = (y(1); y(2); \ldots; y(n))$). For $u \in \mathbb{R}$, let $[u]^+ = \max\{0, u\}$. A set-valued map $f : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ associates to each point in $\mathbb{R}^n$ a set in $\mathbb{R}^m$.

II. Preliminaries

This section introduces concepts from graph theory, nonsmooth analysis, differential inclusions, and optimization. A reader already familiar with these concepts can safely skip it.

Graph theory: Following [17], a weighted digraph is a triplet $G = (V, E, A)$, where $V$ is the vertex set, $E \subseteq V \times V$ is the edge set, and $A \in \mathbb{R}^{n \times n}$ is the adjacency matrix with $a_{ij} > 0$ iff $(v_i, v_j) \in E$. A path is an ordered sequence of vertices such that successive vertices form an edge. A digraph is strongly connected if there is a path between any two vertices. For a vertex $v_i$, $N^{\text{out}}(v_i) = \{v_j \in V \mid (v_i, v_j) \in E\}$ is the set of its out-neighbors. The Laplacian matrix is $L = \text{Out} - A$, where $\text{Out} = \text{diagonal matrix defined by } (\text{Out})_{ii} = \sum_{j=1}^{n} a_{ij}$, for all $i \in [n]$. Note that $\text{In} = 0$. If $G$ is strongly connected, then $\lambda_1$ is a simple eigenvalue of $L$. $G$ is weight-balanced if $\lambda_1 > 0$ iff $L + L^T$ is positive semidefinite. If $G$ is weight-balanced and strongly connected, then $\lambda_2(L + L^T)$ is a simple eigenvalue of $L + L^T$ and, for $x \in \mathbb{R}^n$,

$$\lambda_2(L + L^T)\|x - \frac{1}{n}(1^T_n x)1_n\|^2 \leq x^T(L + L^T)x, \quad (1)$$

with $\lambda_2(L + L^T)$ the smallest non-zero eigenvalue of $L + L^T$.

Nonsmooth analysis: Following [18], $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is locally Lipschitz at $x \in \mathbb{R}^n$ if there exist $\epsilon, \lambda \in \mathbb{R}_{>0}$ such that $\|f(y) - f(y')\| \leq \lambda \|y - y'\|$, for all $y, y' \in B(\epsilon)$. A function $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is regular at $x \in \mathbb{R}^n$ if, for all $v \in \mathbb{R}^n$, the right and generalized directional derivatives of $f$ at $x$ along the direction $v$ coincide. A function is regular.

A set-valued map $\mathcal{H} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is upper semicontinuous at $x \in \mathbb{R}^n$ if, for all $\epsilon \in \mathbb{R}_{>0}$, there exists $\delta \in \mathbb{R}_{>0}$ such that $\mathcal{H}(y) \subset \mathcal{H}(x) + B_{\delta}(0)$ for all $y \in B_{\epsilon}(x)$. Also, $\mathcal{H}$ is locally bounded at $x \in \mathbb{R}^n$ if there exist $\epsilon, \delta \in \mathbb{R}_{>0}$ such that $\|x\| \leq \epsilon$ for all $x \in \mathcal{H}(y)$, $y \in B_{\delta}(x)$. Given $f : \mathbb{R}^n \rightarrow \mathbb{R}$ locally Lipschitz, let $\Omega_f$ be the set of points where $f$ is not differentiable. The generalized gradient $\partial f : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ of $f$ is

$$\partial f(x) = \text{co}\{\lim_{t \rightarrow \infty} \nabla f(x_t) \mid x_t \rightarrow x, x_t \in \Omega_f\},$$

where $\text{co}$ is the convex hull and $\mathbb{S} \subset \mathbb{R}^n$ is any set of measure zero.

Differential inclusions: We gather here tools from [18], [6] to analyze the stability properties of differential inclusions, $\dot{x} \in F(x)$, (2)

where $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is a set-valued map. A solution of (2) on $[0,T] \subset \mathbb{R}$ is an absolutely continuous map $x : [0,T] \rightarrow \mathbb{R}^n$ that satisfies (2) for almost all $t \in [0,T]$. If $F$ is locally bounded, upper semicontinuous, taking nonempty, compact, and convex values, then solutions of (2) exist. The set of equilibria of (2) is $\text{Eq}(F) = \{x \in \mathbb{R}^n \mid 0 \in F(x)\}$. For a locally Lipschitz $W : \mathbb{R}^n \rightarrow \mathbb{R}$, the set-valued Lie derivative $\mathcal{L}_W F : W \rightrightarrows \mathbb{R}$ with respect to (2) at $x \in \mathbb{R}^n$ is

$$\mathcal{L}_W F = \{a \in \mathbb{R} \mid \exists v \in F(x) \text{ s.t. } \zeta^T v = a, \forall \zeta \in \partial W(x)\}.$$
Note that $f^*$ is convex as $f$ and $t \mapsto \frac{1}{t}(t^+)$ are convex. Hence, the problem (4) is convex. The following result, see e.g. [20, Proposition 1], identifies conditions under which the solutions of the problems (3) and (4) coincide.

Proposition 2.3: (Equivalence of (3) and (4)): Assume (3) has nonempty, compact solution set, and satisfies the refined Slater condition. Then, (3) and (4) have the same solutions if $\frac{1}{\epsilon} > \|\lambda\|_{\infty}$, for some Lagrange multiplier $\lambda \in \mathbb{R}_{\geq 0}^m$ of (3).

III. PROBLEM STATEMENT

Consider a network of $n \in \mathbb{Z}_{\geq 1}$ distributed energy resources (DERs) whose communication topology is a strongly connected and weight-balanced digraph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, A)$. For simplicity, assume DERs to be generator units. In our discussion, DERs can also be flexible loads (where the cost function corresponds to the negative of the load utility function). An edge $(i,j)$ represents the capability of unit $j$ to transmit information to unit $i$. Each unit $i$ is equipped with storage capabilities with minimum $C_{i}^{m} \in \mathbb{R}_{\geq 0}$ and maximum $C_{i}^{M} \in \mathbb{R}_{\geq 0}$ capacities. The network collectively aims to meet a power demand profile during a finite-time horizon $\{1, \ldots, h\}$ specified by $L_{t} \in \mathbb{R}_{>0}$, that is, $L_{t+k}$ is the demand at time slot $k \in [h]$. This demand can either be connected to a load requested from an external entity, denoted $L_{w} \geq 0$ for slot $k$, or each DER $i$ might have to satisfy a load at the bus it is connected to, denoted $(L_{b})_{i} \geq 0$ for slot $k$. Thus, for each $k \in [h]$, $L_{t+k} = L_{e} + \sum_{i=1}^{n}(L_{b})_{i}$ is known to an arbitrarily selected unit $r \in [n]$, whereas the demand at bus $i$, $(L_{b})_{i} = (L_{b}^{(1)})_{i}, \ldots, (L_{b}^{(h)})_{i}$, is known to unit $i$. For convenience, $L_{b} = (L_{b}^{(1)}, \ldots, L_{b}^{(h)})$, where $(L_{b})_{i} = (L_{b}^{(1)})_{i}, \ldots, (L_{b}^{(n)})_{i}$ collects the load known to each unit $i \in [h]$. Along with load satisfaction, the group aims to minimize the total cost of generation and to satisfy the constraints for each DER. These elements are explained next.

Each unit $i$ decides at every time slot $k \in [h]$ the amount of power it generates, the portion $j_{i}^{(k)} \in \mathbb{R}$ of it that it injects into the grid to meet the load, and the remaining part $S_{i}^{(k)} \in \mathbb{R}$ that it sends to the storage unit. The power generated by $i$ at $k$ is then $J_{i}^{(k)} + S_{i}^{(k)}$. We denote by $J_{i}^{(k)} = (J_{1}^{(k)}, \ldots, J_{n}^{(k)}) \in \mathbb{R}^{n}$ and $S_{i}^{(k)} = (S_{1}^{(k)}, \ldots, S_{n}^{(k)}) \in \mathbb{R}^{n}$ the collective injected and stored power at time $k$, respectively. The load satisfaction is then expressed as $\mathbf{1}_{n}^T J_{i}^{(k)} = L_{t+k} = L_{e}^{(k)} + \mathbf{1}_{n}^T L_{b}^{(k)}$, for all $k \in [h]$. The cost $f_{i}^{(k)}(J_{i}^{(k)} + S_{i}^{(k)})$ of power generation $J_{i}^{(k)} + S_{i}^{(k)}$ by $i$ at time $k$ is given by $f_{i}^{(k)} : \mathbb{R} \to \mathbb{R}_{\geq 0}$, which we assume convex and continuously differentiable. Given $(J, S)$, the cost incurred by the network at slot $k$ is $f(J + S) = \sum_{i=1}^{n} f_{i}^{(k)}(J_{i}^{(k)} + S_{i}^{(k)})$.

The cumulative cost of generation for the network across the time horizon is $f : \mathbb{R}_{\geq 0}^{nh} \to \mathbb{R}_{\geq 0}$, $f(x) = \sum_{k=1}^{n} f^{(k)}(x^{(k)})$. Given injection $J = (J^{(1)}, \ldots, J^{(h)}) \in \mathbb{R}^{nh}$ and storage $S = (S^{(1)}, \ldots, S^{(h)}) \in \mathbb{R}^{nh}$ values, the total network cost is $f(J + S) = \sum_{k=1}^{h} f^{(k)}(J^{(k)} + S^{(k)})$.

The functions $\{f^{(k)}\}_{k=1}^{h}$ and $f$ are also convex and continuously differentiable. Next, we describe the physical constraints on the DERs. Each unit’s power must belong to the range $[P_{i}^{m}, P_{i}^{M}] \subset \mathbb{R}_{\geq 0}$, representing lower and upper bounds on the power generation at each time slot. Each unit $i$ also respects upper and lower ramp limits: the change in the generation from any time slot $k$ to $k+1$ is upper and lower bounded by $R_{i}^{u}$ and $-R_{i}^{l}$, respectively, with $R_{i}^{u}, R_{i}^{l} \in \mathbb{R}_{\geq 0}$. At each time slot, the power injected into the grid by each unit must be nonnegative, i.e., $J_{i}^{(k)} \geq 0$. Further, the power stored in any storage unit $i$ at any time slot $k \in [h]$ must belong to the range $[C_{i}^{m}, C_{i}^{M}]$. Finally, we assume that at the beginning of the time slot $k = 1$, each storage unit $i$ starts with some stored power $S_{i}^{(0)} \in [C_{i}^{m}, C_{i}^{M}]$. With the above model, the dynamic economic dispatch with storage (DEDS) problem is formally defined by the following convex optimization problem.

minimize $(J, S) \in \mathbb{R}^{nh} \times \mathbb{R}^{nh}$
subject to for $k \in [h]$, $J_{i}^{(k)} = L_{t+k}^{(k)}$, $P_{i}^{m} \leq J_{i}^{(k)} + S_{i}^{(k)} \leq P_{i}^{M}$, $C_{m} \leq S_{i}^{(0)} + \sum_{k=1}^{h} S_{i}^{(k)} \leq C_{M}$, $0_{n} \preceq J_{i}^{(k)}$, $R_{i}^{l} \leq J_{i}^{(k+1)} - J_{i}^{(k)} - S_{i}^{(k)} \leq R_{i}^{u}$. (5f)

We refer to (5b)–(5f) as the load conditions, box constraints, storage limits, injection constraints, and ramp constraints, resp. We denote by $J_{DEDS}$ and $J_{DEDS}^{\ast}$ the feasibility and the solution set of (5), resp., and assume them to be nonempty. Since $J_{DEDS}$ is compact, so is $J_{DEDS}^{\ast}$. Moreover, the refined Slater condition is satisfied as all constraints (5b)–(5f) are affine in the decision variables. Additionally, we assume the DEDS problem satisfies the strong Slater condition with $\rho \in \mathbb{R}_{>0}$ and $(J^{p}, S^{p}) \in \mathbb{R}^{2nh}$. Our aim is to design a distributed algorithm that allows the network to solve (5).

Remark 3.1: (Extensions to DEDS formulation): The DEDS formulation can be modified to consider scenarios where only some DERS $\mathcal{V}_{g}$ are equipped with storage and others $\mathcal{V}_{g}$ are not, with $[n] = \mathcal{V}_{gs} \cup \mathcal{V}_{g}$. The formulation can also be extended to consider the cost of storage, inefficiencies, and constraints on (dis)charging of the storage units, as in [11], [13]. These factors either affect the constraint (5d), add additional conditions on the storage variables, or modify the objective function. As long as the resulting cost and constraints are convex in $S$, all these can be treated within (5) without affecting the design methodology. Also, the DEDS formulation does not account for other physical constraints on the power network such as transmission losses and line capacity limits. Our ensuing discussion shows that, even with these omissions, the design of a provably correct distributed algorithm with the communication structure assumed here is challenging.

IV. DISTRIBUTED ALGORITHMIC SOLUTION

We describe here the distributed algorithm that asymptotically solves the DEDS problem. Our design builds on
an equivalent formulation of the optimization using penalty functions (cf. Section IV-A). This reformulation gets rid of the inequality constraints, yielding an optimization whose structure guides our algorithmic design (cf. Section IV-B).

A. Alternative formulation of the DEDS problem: The procedure here follows closely the theory of exact penalty functions outlined in Section II. For an $\epsilon \in \mathbb{R}_{>0}$, consider the modified cost function $f^\epsilon : \mathbb{R}^{nh} \times \mathbb{R}^{nh} \to \mathbb{R}_{\geq 0}$,

$$f^\epsilon(J,S) = f(J + S) + \frac{1}{\epsilon} \left( \sum_{k=1}^{h} 1_n^T \left( [T_1^{(k)}]^+ + [T_2^{(k)}]^+ + [T_3^{(k)}]^+ + [T_4^{(k)}]^+ + [T_5^{(k)}]^+ + \sum_{k=1}^{h-1} 1_n^T \left( [T_6^{(k)}]^+ + [T_7^{(k)}]^+ \right) \right) \right),$$

where

$$T_1^{(k)} = P_m - J^{(k)} - S^{(k)}, \quad T_2^{(k)} = J^{(k)} + S^{(k)} - P^M,$$
$$T_3^{(k)} = C_m - S^{(0)} - \sum_{k=1}^{h} S^{(k)}, \quad T_4^{(k)} = S^{(0)} + \sum_{k=1}^{h} S^{(k)} - C_m, \quad T_5^{(k)} = -J^{(k)},$$
$$T_6^{(k)} = -R - J^{(k+1)} - S^{(k+1)} + J^{(k)} + S^{(k)}, \quad T_7^{(k)} = J^{(k+1)} + S^{(k+1)} - J^{(k)} - S^{(k)} - R^u.$$

This cost contains the penalty terms for all the inequality constraints of the DEDS problem. Note that $f^\epsilon$ is locally Lipschitz, jointly convex in $J$ and $S$, and regular. Thus, the partial generalized gradients $\partial J f^\epsilon$ and $\partial S f^\epsilon$ take nonempty, convex, compact values, and are locally bounded and upper semicontinuous. Consider the modified DEDS problem

$$\min \{ f^\epsilon(J,S) : 1_n^T J^{(k)} = L_k^{(k)} , \forall k \in [h] \}. \quad (7)$$

The difference between the optimizations (5) and (7) is that all inequality constraints of (5) are moved to the objective of (7) in the form of penalty terms. The next result provides a criteria for selecting $\epsilon$ such that the modified DEDS and the DEDS problems have the exact same solutions. This allows us to focus on solving (7). The proof is a direct application of Lemmas 2.2 and 2.3 using the fact that the DEDS problem satisfies the strong Slater condition with $\rho$ and $(J^0, S^0)$.

**Lemma 4.1:** (Equivalence of DEDS and modified DEDS problems): Let $(J^*, S^*) \in F_{\text{DEDS}}$. Then, the optimizers of the problems (5) and (7) are the same for $\epsilon \in \mathbb{R}_{>0}$ satisfying

$$\epsilon < \frac{\rho}{f(J^0 + S^0) - f(J^* + S^*}). \quad (8)$$

From here on, we assume that $\epsilon$ satisfies (8) and so problems (5) and (7) are equivalent. Writing the Lagrangian and the KKT conditions for (7), we obtain the following characterization of the solution set of the DEDS problem

$$F_{\text{DEDS}} = \{ (J,S) \in \mathbb{R}^{2nh} : 1_n^T J^{(k)} = L_k^{(k)} \text{ for all } k \in [h], \quad 0 \in \partial S f^\epsilon(J,S), \text{ and } \exists \nu \in \mathbb{R}^b \text{ such that} \} \nu^{(1)} 1_n, \ldots, \nu^{(h)} 1_n \} = \partial J f^\epsilon(J,S). \quad (9)$$

Recall that $F_{\text{DEDS}}$ is bounded. Next, we stipulate a mild regularity assumption on this set which implies that perturbing it by a small parameter does not result into an unbounded set. This property is of use in our convergence analysis later.

**Assumption 4.2:** (Regularity of $F_{\text{DEDS}}$): For $p \in \mathbb{R}_{\geq 0}$, define the map $p \mapsto F(p) \subset \mathbb{R}^{2nh}$ as

$$F(p) = \{ (J,S) \in \mathbb{R}^{2nh} : 1_n^T J^{(k)} = L_k^{(k)} \leq p \text{ for all } k \in [h], \quad 0 \in \partial S f^\epsilon(J,S) + p B_1(0), \text{ and } \exists \nu \in \mathbb{R}^b \text{ such that} \} \nu^{(1)} 1_n, \ldots, \nu^{(h)} 1_n \in \partial J f^\epsilon(J,S) + p B_1(0) \}. \quad (10)$$

Note that $F(0) = F_{\text{DEDS}}$. Then, there exists a $\bar{p} > 0$ such that $F(p)$ is bounded for all $p \in [0, \bar{p})$.

The equivalent reformulation (7), has a desirable structure: it does not have inequality constraints and the equalities have the special property that their coefficient vector lies in the null space of the Laplacian matrix. In the following section, we see how these facts help in the algorithm design and analysis.

B. The $\text{acc+}(L\partial, \partial)$ coordination algorithm: Here, we present our distributed algorithm and establish its asymptotic convergence to the set of solutions of the DEDS problem starting from any initial condition. Our design combines ideas of Laplacian-gradient dynamics [5] and dynamic average consensus [22]. Consider the set-valued dynamics,

$$\dot{J} \in - (I_b \otimes L) \partial J f^\epsilon(J,S) + \nu_1 z, \quad (10a)$$
$$\dot{S} \in - \partial S f^\epsilon(J,S), \quad (10b)$$
$$\dot{z} = - \alpha z - \beta (I_b \otimes L) z - v + \nu_2 (L_e \otimes e_r + L_b - J), \quad (10c)$$
$$\dot{v} = \alpha \beta (I_b \otimes L) z, \quad (10d)$$

where $\alpha, \beta, \nu_1, \nu_2 \in \mathbb{R}_{>0}$ are design parameters and $e_r \in \mathbb{R}^n$ is the unit vector along the $r$-th coordinate. We refer to (10) as $\text{acc+}(L\partial, \partial)$ dynamics and below we explain its components.

**[Description of $\text{acc+}(L\partial, \partial)$ dynamics]:** The dynamics (10) consists of “dynamic average consensus in $(z, v)$ Laplacian gradient in $J$ + gradient in $S$”, and so we use the terminology $\text{acc+}(L\partial, \partial)$. The $(z,v)$-component corresponds to the dynamic average consensus part. Here, $z^{(k)}$ is the $k$-th coordinate of $z$. The $z$-dynamics have two terms. The first term seeks to minimize $f^\epsilon$ keeping constant the total power generation. The second term gets the feedback of the mismatch between the total generation and the total load from the $z$-dynamics and drives the network towards load satisfaction. Finally, the $S$-component is gradient descent and seeks to minimize $f^\epsilon$ with respect to the variable $S$. From this description, one can see that getting rid of the inequalities of (5) using penalty functions simplifies the design.

For convenience, we denote the right-hand side of (10) by $X_{\text{acc+}(L\partial, \partial)} : \mathbb{R}^{4nh} \to \mathbb{R}^{4nh}$. Note that Eq($X_{\text{acc+}(L\partial, \partial)} = F_{\text{DEDS}}$ and $\partial J f^\epsilon$ and $\partial S f^\epsilon$ are locally bounded, upper semicontinuous and take nonempty convex compact values, the solutions of $X_{\text{acc+}(L\partial, \partial)}$ exist (cf. Section II).

**Remark 4.3:** (Distributed implementation of $\text{acc+}(L\partial, \partial)$ dynamics): For (10), each $i \in [n]$ implements the dynamics of its decision variables, which are $\{J_i^{(k)}, S_i^{(k)}, z_i^{(k)}, v_i^{(k)}\}_{k=1}^\infty$.

That is, for each $k \in [h]$, unit $i$ implements

$$J_i^{(k)} = - \sum_{j \in N^{(i)}} a_{ij} (\zeta_i^{(k)} - \zeta_j^{(k)}) + \nu_1 z_i^{(k)}, \quad (11a)$$
Thus, the entities \( \zeta_i^{(k)} \in \partial g_i f^*(J, S) \) and \( \xi_i^{(k)} \in \partial g_i f^*(J, S) \), for all \( k \in [h] \), only depend on the decision variables of unit \( i \) and so are computable by it. This further implies that (11b) can be implemented by \( i \) using its own state and, to execute (11a), \( i \) needs information from its out-neighbors. Hence, the dynamics can be executed in a distributed manner. For real-time implementation, we discretize (10); selecting a small enough stepsize results in trajectories that follow closely the continuous-time trajectories leading to the optimizers. 

Next, we state our main convergence result. The proof is provided in the Appendix.

**Theorem 4.4:** (Convergence of the \( \text{dac+}(L_0, \vartheta) \) dynamics to the solutions of the DEDS problem): Let \( F_{\text{DEDS}} \) satisfy Assumption 4.2, \( \epsilon \) satisfy (8), and \( \alpha, \beta, \nu_1, \nu_2 > 0 \) satisfy

\[
\frac{\nu_1}{\beta \nu_2 \lambda_2 (L + L^T)} + \frac{\nu_2^2 \lambda_{\text{max}} (L^T L)}{2 \alpha} < \lambda_2 (L + L^T). \tag{12}
\]

Then, any trajectory of (10) starting in \( \mathbb{R}^{nh} \times \mathbb{R}^{nh} \times \mathbb{R}^{nh} \times (H_0)^{h} \) converges to \( F_{\text{DEDS}}^{*} = \{(J, S, z, v) \in F_{\text{DEDS}}^{*} \times \{0\} \times \mathbb{R}^{nh} \mid v = \nu_2 (L_e \otimes e_r + L_b - J)\}. \)

From the first step in the proof of Theorem 4.4, one sees that the mismatch between the network power injection and the load profile converges asymptotically to zero. This guarantees robustness of the algorithm, in the sense that during its execution, the load can vary or agents can join or leave the network (provided that there is always a participating node that knows the external demand), and the dynamics adjusts for these perturbations. We do not expand more on this here due to lack of space, see [6, Section 5.2] for a similar discussion for a different problem setup.

**Remark 4.5:** (General setup for storage: revisited): The \( \text{dac+}(L_0, \vartheta) \) dynamics (10) can be modified to scenarios that include more general descriptions of storage capabilities, as in Remark 3.1. For instance, if only a subset of units have storage capabilities, the only modification is to set the variables \( \{S_{i}^{(k)}\}_{i \in V_S, k \in [h]} \) to zero and execute (10b) only for the variables \( \{k_i^{(k)}\}_{i \in V_S, k \in [h]} \). The resulting strategy converges to the solution set of the corresponding DEDS problem.

**Remark 4.6:** (Distributed selection of design parameters): The implementation of (10) requires the selection of parameters \( \alpha, \beta, \nu_1, \nu_2, \epsilon \) satisfying (8) and (12). Instead of condition (12), one can check a different, stronger inequality that requires knowledge of the maximum and minimum of various network-wide quantities. In turn, these can be computed in finite time by the units resorting to distributed consensus-based procedures [23] in order to collectively select appropriate values, see e.g., [6, Remark 5.4]. Regarding (8), an upper bound on the denominator of the right-hand side can be computed by aggregating, using consensus, the difference between the max and the min values that each DER’s aggregate cost function takes in its respective feasibility set (neglecting load conditions). The challenge for the units, however, is to estimate \( \rho \). This can be accomplished by considering the optimization “find the largest \( \rho \) for which the DEDS problem satisfies the strong Slater condition” and having the units employ a distributed algorithm to solve it, see e.g., [24].

**V. SIMULATIONS**

We illustrate the application of the \( \text{dac+}(L_0, \vartheta) \) dynamics to solve the DEDS problem for a group of \( n = 10 \) generators with communication defined by a directed ring with bi-directional edges \( \{(1, 5), (2, 6), (3, 7), (4, 8)\} \) (all edge weights are 1). The planning horizon is \( \vartheta = 6 \) and the load profile consists of the external load \( L_e = (1950, 1980, 2700, 2370, 1900, 1850) \) and the load at each generator \( i \) for each slot \( k \) given by \( (L_{b_i}^{(k)}) = 101 \). Thus, for each slot \( k \), \( (L_{b_i}^{(k)}) = \sum_{i=1}^{10} (L_{b_i}^{(k)}) = 550 \) and, so, \( L_i = (2500, 2530, 3250, 2920, 2450, 2400) \). Generators have storage capacities determined by \( C^M = 1001 \) and \( C^S = S^{(0)} = 51 \). The cost function of each unit is quadratic and constant across time. Table I details the cost function coefficients, generation limits, and ramp constraints, which are modified from the data for 39-bus New England system [25]. Figure 1 illustrates the evolution of the total power injected at each time slot and the total cost incurred by the network, respectively. As established in Theorem 4.4 and shown in Figure 2, the total injection asymptotically converges to the load profile \( l \), the total aggregate cost converges to the minimum \( 201092 \) and the converged solution satisfies (5c)-(5f). The number of variables maintained and updated by each generator is linear in the length of the time horizon \( h \), and therefore, at each iteration, the computation time and the communication volume increase linearly with \( h \).

**VI. CONCLUSIONS**

We have studied the DEDS problem for a group of generators with storage capabilities that communicate over a strongly connected, weight-balanced digraph. Using exact
penalty functions, we have provided an alternative problem formulation, upon which we have built to design the distributed dac+(LΩ, 0) dynamics. This dynamics provably converges to the set of solutions of the problem from any initial condition. In future work we plan to extend the scope of our formulation to include power flow equations, constraints on the power lines, various losses, uncertainty of the available data (loads, costs, and generator availability), and develop online and opportunistic state-triggered implementations. We also intend to explore the use of our dynamics as a building block in solving grid control problems across different time scales (e.g., implementations at long time scales on high-inertia generators and at short time scales on low-inertia generators in the face of highly-varying demand) and hierarchical levels (e.g., in multi-layer architectures where aggregators at one layer coordinate their response to a request for power production, and feed their decisions as load requirements to the devices in lower layers).

**REFERENCES**


**APPENDIX**

**Proof of Theorem 4.4:** For convenience, let $\mathcal{M}_g = \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \times (H_0)^b$ and $\mathcal{M}_e = \prod_{k=1}^n H_{L_2(k)}^n \times (H_0)^b \times (H_0)^b$. We divide the proof into three broad steps.

**Step 1: Characterizing the ω-limit set:** We show that the ω-limit set of any trajectory of (10) with initial condition $(J_0, S_0, z_0, v_0) \in \mathcal{M}_e$ belongs to $\mathcal{M}_g$. For this, write (10d) as

$$\dot{v}(k) = \alpha J L_z(k)$$

for all $k \in [b]$. Note that $\int_0^b \dot{v}(k) = \alpha J L_z(k) = 0$ for all $k \in [b]$ because $G$ is weight-balanced. Therefore, the initial condition $v_0 \in (H_0)^b$
implies that \( v(t) \in \mathbb{H}_0^b \) for all \( t \geq 0 \) along any trajectory of (10) starting at \((J_0, S_0, z_0, v_0)\). Now, if \( \zeta \in \partial f^\ast(J, S) \) then, from (10a) and (10c), we get for any \( k \in [b] \)
\[
    j^{(k)} = -L\zeta^{(k)} + \nu_2 z^{(k)}, \\
    \dot{z}^{(k)} = -\alpha z^{(k)} - \beta Lz^{(k)} - v_2(L_e^{(k)}e_s + L_b^{(k)}) - j^{(k)}.
\]
Let \( \xi_k = \mathbf{1}_n^T J^{(k)} - L^{(k)} \). Then, from the above equations we get
\[
    \ddot{\xi}_k = \mathbf{1}_n^T j^{(k)} = \nu_2 \mathbf{1}_n^T z^{(k)}. \\
\]
Further, we have
\[
    \ddot{\xi}_k = \nu_2 \mathbf{1}_n^T z^{(k)} = -\alpha \mathbf{1}_n^T z^{(k)} + \nu_1 v_2 (L_1^T - \mathbf{1}_n^T J^{(k)}),
\]
forming a second-order linear system for \( \xi_k \). The LaSalle Invariance Principle [26] with the function \( \nu_1 v_2 \|\xi\|_2 + \|\xi\|^2 \)
implies that as \( t \to \infty \) we have \( \xi_k(t) \to 0 \) and so \( \xi_k \) is bounded and \( \xi_k \to 0 \) as \( t \to \infty \).

**Step 2: Applying the refined LaSalle Invariance Principle:** Consider the change of coordinates \( D : \mathbb{R}^{4n_h} \to \mathbb{R}^{4n_h} \),
\[
    (J, S, \omega_1, \omega_2) = D(J, S, z, v) = (J, S, z, v + \alpha z - \nu_2 (L_e \otimes e_r + L_b - J)).
\]
In these coordinates, the set-valued map (10) takes the form
\[
    X_{\text{d},(\omega,\beta)}(J, S, \omega_1, \omega_2) := \{(-I_h \otimes L)\zeta_1 + \nu_1 \omega_1, -\zeta_2, -\beta (I_h \otimes L)\omega_1 - \omega_2, \nu_1 \nu_2 \omega_1 - \alpha \omega_2 - \nu_2 (I_h \otimes L) \zeta_1 \} \in \mathbb{R}^{4n_h} | \zeta_1 \in \partial_f^\ast(J, S), \zeta_2 \in \partial S^\ast(J, S) \}.
\]
This transformation helps in identifying the LaSalle-type function for the dynamics. We now focus on proving that, in the new coordinates, the trajectories of (10) converge to
\[
    \mathcal{F}_{\text{aug}} = D(\mathcal{F}_{\text{aug}}) = D(\mathcal{F}_{\text{DES}}) \times \{0\} \times \{0\}.
\]
Note that \( D(\mathcal{M}_o) = \mathcal{M}_o \) and so, from the property of the \( \omega \)-limit set of trajectories above, we get that \( t \mapsto (J(t), S(t), \omega_1(t), \omega_2(t)) \) starting in \( D(\mathcal{M}_b) \) belongs to \( \mathcal{M}_o \).

Next, we show the hypotheses of Proposition 2.1 are satisfied, where \( \mathcal{M}_o \) plays the role of \( S \subseteq \mathbb{R}^{4n_h} \) and \( V : \mathbb{R}^{4n_h} \to \mathbb{R}_{\geq 0} \),
\[
    V(J, S, \omega_1, \omega_2) = f^\ast(J, S) + \frac{1}{2} (\nu_1 v_2 \|\omega_1\|^2 + \|\omega_2\|^2).
\]
plays the role of \( W \), resp. Let \((J, S, \omega_1, \omega_2) \in \mathcal{M}_o\) then any element of \( \mathcal{L}_{X_{\text{d},(\omega,\beta)}} V(J, S, \omega_1, \omega_2) \) can be written as
\[
    -\zeta_1^\ast (I_h \otimes L) \zeta_1 + \nu_1 \zeta_1^\ast \omega_1 - \|\zeta_2\|^2 - \beta \nu_1 \nu_2 \omega_1^T (I_h \otimes L) \omega_1 - \alpha \|\omega_2\|^2 - \nu_2 \omega_2^T (I_h \otimes L) \zeta_1,
\]
where \( \zeta_1 \in \partial_f^\ast(J, S) \) and \( \zeta_2 \in \partial S^\ast(J, S) \). Since the digraph \( \mathcal{G} \) is strongly connected and weight-balanced, we use (11) and \( I_n^T \omega_1 = 0 \) to bound the above expression as
\[
    -\frac{1}{2} \beta \nu_1 \nu_2 \omega_2^T (I_h \otimes L^T) \omega_1 - \|\zeta_2\|^2 - \frac{1}{2} \beta \nu_1 \nu_2 \omega_2^T (I_h \otimes L^T) \omega_1 - \alpha \|\omega_2\|^2 - \nu_2 \omega_2^T (I_h \otimes L) \zeta_1 = \gamma^T M \gamma - \|\zeta_2\|^2,
\]
where \( \eta = (\eta^{(1)}; \ldots; \eta^{(b)}) \) with \( \eta^{(k)} = \zeta^{(k)} - \frac{1}{n} (I_n^T \zeta^{(k)}) 1_n \), the vector \( \gamma = (\eta^{(1)}; \omega_1; \omega_2) \), and the matrix
\[
    M = \begin{bmatrix} -\frac{1}{2} \beta \nu_2 (I_h \otimes L^T) \nu_1 B^T & C \\
    B^T & C \end{bmatrix},
\]
with
\[
    B^T = \begin{bmatrix} -\frac{1}{2} \nu_1 \nu_2 (I_h \otimes L^T) \nu_1 B^T & C \\
    0 & -\alpha I_n \end{bmatrix},
\]
Resorting to the Schur complement [19], \( M \in \mathbb{R}^{3n_h \times 3n_h} \) is neg. definite if \( -\frac{1}{2} \beta \nu_1 (I_h \otimes L^T) I_n - B^T C^{-1} B \), that equals
\[
    -\frac{1}{2} \beta \nu_1 (I_h \otimes L^T) I_n + \frac{\nu_1}{2} \beta \nu_2 \nu_1 (I_h \otimes L^T) I_n + \frac{\nu_1}{2} \nu_2 (I_h \otimes L)^T (I_h \otimes L),
\]
is negative definite, which follows from (12). Hence, for any \((J, S, \omega_1, \omega_2) \in \mathcal{M}_o\), we have
\[
    \max_{V(J, S, \omega_1, \omega_2)} V(J, S, \omega_1, \omega_2) \leq 0
\]
and also
\[
    0 \in \mathcal{L}_{X_{\text{d},(\omega,\beta)}} V(J, S, \omega_1, \omega_2) \text{ if } \eta = \zeta_2 = \omega_1 = \omega_2 = 0,
\]
which means \( \zeta^{(k)} \in \text{span}\{1_n\} \) for each \( k \in [b] \). Consequently,
using (9), we deduce that \((J, S)\) is a solution of (7) and so, \((J, S, \omega_1, \omega_2) \in \mathcal{T}_{\text{aug}}\). Since, \(\mathcal{T}_{\text{aug}}\) belongs to a level set of \(\mathcal{V}\), we conclude that Proposition 2.1(i) holds. Further, using [6, Lemma A.1] one can show that Proposition 2.1(ii) holds too (we omit the details due to space constraints).

**Step 3: Showing boundedness of trajectories:** To apply Proposition 2.1, it remains to show that the trajectories starting from \(D(\mathfrak{R}_0)\) are bounded. We reason by contradiction. Assume there exists \(t \rightarrow (J(t), S(t), \omega(t), \omega_2(t))\), with \((J(0), S(0), \omega(0), \omega_2(0)) \in D(\mathfrak{R}_0)\), of \(X_{\lambda_{\omega_1}(\lambda, \theta, \eta)}\) such that \(\|(J(t), S(t), \omega(t), \omega_2(t))\| \to \infty\). Since \(V\) is radially unbounded, this implies \(V(J(t), S(t), \omega(t), \omega_2(t)) \to \infty\). Also, as established above, we know \(L_n^{-1}(J^{(k)}(t) \to L_n^{-1}(t)\) and \(L_n^{-1}(\omega_k^{(k)}(t) \to 0\) for each \(k \in [b]\). Thus, there exist times \(t_m^{(k)}\) with \(t_m \to \infty\) such that for all \(m \in \mathbb{Z}_{\geq 1},\)

\[
\left| L_n^{-1}(\omega_k^{(k)}(t_m)) \right| < 1/m \text{ for all } k \in [b], \quad \text{(A.15)}
\]

\[
\max L_{X_{\lambda_{\omega_1}(\lambda, \theta, \eta)}}(J(t_m), S(t_m), \omega(t_m), \omega_2(t_m)) > 0.
\]

The second inequality existence of the \((\xi_1, m, \xi_2, m) \in (\partial J)^{\circ}(J(t_m), S(t_m)), \partial S^{\circ}(J(t_m), S(t_m)))\), such that

\[
-\xi_1^{-1}(J(t_m), S(t_m), \omega(t_m), \omega_2(t_m)) - \sum_{k=1}^{b} \left| L_n^{-1}(\omega_k^{(k)}(t_m)) \right| < 1/m \text{ for all } m \in \mathbb{Z}_{\geq 1},
\]

for all \(m \in \mathbb{Z}_{\geq 1}\), where we have used (A.14) to write an element of \(L_{X_{\lambda_{\omega_1}(\lambda, \theta, \eta)}}(J(t_m), S(t_m), \omega(t_m), \omega_2(t_m))\). Letting \(\eta_m = \xi_1^{-1}(J(t_m), S(t_m), \omega(t_m), \omega_2(t_m)) + \frac{1}{m} \left| L_n^{-1}(\omega_k^{(k)}(t_m)) \right|\), using (1), and using the relation \(\|\omega_k^{(k)}(t_m)\| \to \infty\), the above inequality can be rewritten as

\[
\gamma_m^T M \gamma_m + \frac{1}{m} \sum_{k=1}^{b} \left| L_n^{-1}(\omega_k^{(k)}(t_m)) \right| > 0, \quad \text{(A.16)}
\]

with \(\gamma_m = (\eta_m; \omega_1(t_m); \omega_2(t_m))\). Using (A.15) on (A.16),

\[
\gamma_m^T M \gamma_m - \sum_{k=1}^{b} L_n^{-1}(\omega_k^{(k)}(t_m)) > 0, \quad \text{(A.17)}
\]

for all \(m \in \mathbb{Z}_{\geq 1}\). Next, we consider two cases, depending on whether the sequence \((J(t_m), S(t_m))\) is (a) bounded or (b) unbounded. In case (a), \((J(t_m), S(t_m))\) is bounded, and hence we have \(M \gamma_m \leq \lambda_{\text{max}}(M)\). This further implies that \(\eta_m \to \infty\) for all \(m \in \mathbb{Z}_{\geq 1}\). Hence, if \(A.19\) holds for all \(m \in \mathbb{Z}_{\geq 1}\), then so is

\[
\sum_{k=1}^{b} U_{k,m} + Z_1 + Z_2 > 0.
\]

Next we show that for each \(k \in K_u\) there exists \(m_k \in \mathbb{Z}_{\geq 1}\) such that \(U_{k,m} + Z_1 + Z_2 m > 0\) for all \(m \geq m_k\). This will lead to the desired contradiction. Assume without loss of generality that \(1_m^{-1}(\omega_k^{(k)}(t) \to \infty\) (reasoning for the case when the sequence approaches negative infinity follows analogously). Then, for \(\lambda_{\text{max}}(M)\),

\[
\lambda_{\text{max}}(M) \|\eta_k\|^2 - \|\xi_2, m\|^2 + \frac{1}{m} \left| L_n^{-1}(\omega_k^{(k)}(t_m)) \right| > 0,
\]

for all \(m \in \mathbb{Z}_{\geq 1}\), we require \(\xi_1^{-1}(J(t_m), S(t_m), \omega(t_m), \omega_2(t_m)) \to \infty\) for all \(i \in [n]\). Indeed, otherwise, recalling that \(\eta_k = \xi_1^{-1}(J(t_m), S(t_m), \omega(t_m), \omega_2(t_m))\), it can be shown that there exist an \(m_k\) such that

\[
\lambda_{\text{max}}(M) \|\eta_k\|^2 < \frac{1}{m} \left| L_n^{-1}(\omega_k^{(k)}(t_m)) \right| + \frac{Z_1 + Z_2}{m} \quad \text{for all } m \geq m_k.
\]

Note that from Lemma A.1 we have \(\|\xi_1^{-1}(J(t_m), S(t_m), \omega(t_m), \omega_2(t_m))\| \leq \frac{1}{m} \text{ for all } i \in [n].\)

With these facts in place, we write

\[
U_{k,m} + Z_1 + Z_2 \geq - \sum_{i=1}^{n} \left( \left| L_n^{-1}(\omega_k^{(k)}(t_m)) \right| + \frac{1}{m} \right) \sum_{i=1}^{n} \left( \omega_i^{(k)}(t_m) \right) + \frac{Z_1 + Z_2}{m}
\]

and deduce that there exists an \(m_k \in \mathbb{Z}_{\geq 1}\) such that the right-hand side of the above expression is negative for all \(m \geq m_k\), which is what we wanted to show.

Finally, consider the case when \(\lambda_{\text{max}}(M)\) is bounded. For (A.19) to be true for all \(m\), we need \(\|\eta_k\| \to 0\) and \(\|\xi_2, m\| \to 0\) as \(m \to \infty\). This further implies that \(\eta_m \to 0\) and, from Assumption 4.2, this is only possible if \(\{J(t_m), S(t_m))\}^{\infty}_{m=1}\) is bounded, a contradiction.

The next result aids the above outlined proof. Due to lack of space, we refer the reader to [27, Lemma 4.3] for its proof.

**Lemma A.1:** (Bound on the difference between \(\partial J^{\circ} f\) and \(\partial f^{\circ}\)). For \((J, S) \in \mathbb{R}^{2n}\), any two elements \(\xi_1, \xi_2 \in \partial f^{\circ}(J, S)\) and \(\xi_2 \in \partial f^{\circ}(J, S)\) satisfy \(\|\xi_1 - \xi_2\| \leq (b + 4)/\epsilon\).