DISTRIBUTED POLICY LEARNING FOR ONLINE SUBMODULAR MAXIMIZATION OVER TIME-VARYING DIRECTED GRAPHS

Anonymous authors
Paper under double-blind review

ABSTRACT

This paper investigates the online submodular maximization problem for multi-agent systems over time-varying directed communication networks. At each round, multiple agents collaboratively maximize a time-varying monotone submodular utility function using only local information, and without prior knowledge of future objectives. This setting has wide applications including multi-robot target tracking, client selection in federated learning, and distributed information fusion in the dynamic environment. Existing approaches, such as OSG and MA-OSMA, typically rely on fixed and connected graphs, which limits their applicability to networks with asymmetric communication and intermittent connectivity arising from packet losses and communication link failures. To address these limitations, we propose a Push-Sum Online Surrogate Mirror Ascent (POSMA) algorithm which utilizes a multilinear extension to transform the original discrete problem into a continuous optimization problem. A push-sum mechanism is further incorporated to compensate for the aggregation bias. Under the B-strongly connected communication, we theoretically confirm that the algorithm achieves an approximation ratio of $\frac{1-e^{-\bar{c}}}{\bar{c}}$ with an average dynamic regret, where $\bar{c}$ bounds the curvature of all objective functions. Moreover, we quantitatively characterize how network connectivity degradation slows the agreement rate and therefore increases online submodular regret by establishing the regret bound $\mathcal{O}\left(\frac{1}{\sqrt{\lambda(1-\lambda)}}\sqrt{\frac{1+C_T}{T}}\right)$, where C_T represents the cumulative variation of the optimal action sets, and $\lambda = \left(1-\frac{1}{N^{NB}}\right)^{1/B}$ characterizes the rate of agreement related to the jointly connected bound B. Finally, we conduct simulations of multi-target tracking to demonstrate the effectiveness of the proposed algorithm and show the benefits compared with the baseline algorithm under fixed and time-varying communication graphs.

1 INTRODUCTION

Submodular maximization in multi-agent systems has attracted growing attention in recent years, as it captures a broad class of optimization problems, including distributed decision-making, resource allocation, and sensor placement, whose utility functions exhibit diminishing returns (Rezazadeh & Kia, 2023; Zhang et al., 2025; Vondrák, 2008). This problem can be applied to a wide range of real-world applications, such as target coverage maximization by mobile sensors or UAVs in multi-agent target tracking (Xu et al., 2023; Zhou et al., 2019; 2022; Corah & Michael, 2021), representative client selection in federated learning (Balakrishnan et al., 2022; Jiménez et al., 2024; Vardhan et al., 2025) and measurement selection in distributed information fusion (Tzoumas et al., 2022; Ye et al., 2021; Rezazadeh & Kia, 2023). For a fixed objective function, classical offline submodular maximization provides approximation guarantees under different constraints. The greedy algorithm attains a tight $(1-\frac{1}{e})$-approximation for maximizing a monotone submodular function under a cardinality constraint (Nemhauser et al., 1978). Under general matroid constraints, however, the greedy algorithm has a $\frac{1}{2}$-approximation guarantee (Fisher et al., 1978), and an improved $(1-\frac{1}{e})$-approximation under matroid constraints was later obtained via continuous greedy optimization of the multilinear extension followed by pipage rounding (Calinescu et al., 2011).

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

In these applications, however, the dynamic and unpredictable nature of the environment causes the utility functions to change over time, requiring agents to continually adapt their policy (Xu et al., 2023; Zhang et al., 2025). Meanwhile, the limited perception capabilities of individual agents restrict the feedback available for subsequent decisions. These considerations motivate the online submodular maximization problem, in which agents interact with the environment over a sequence of rounds without prior knowledge of future objective functions (Streeter & Golovin, 2008; Xu et al., 2023; Zhang et al., 2025). The main challenge of this online setting is to find coordinated actions based on the local information while adapting to dynamic environmental changes. To address this issue, various algorithms have been developed for online submodular maximization (Streeter & Golovin, 2008; Xu et al., 2023; Zhang et al., 2025). Building on the offline approximation guarantees, an online greedy algorithm was proposed for the centralized, single-decision-maker setting (Streeter & Golovin, 2008), and subsequently extended to matroid constraints (Golovin et al., 2014). Online submodular maximization was further advanced through continuous optimization and reductions to offline greedy algorithms (Chen et al., 2018; Harvey et al., 2020; Niazadeh et al., 2023; Wan et al., 2023). These methods primarily address centralized decision-making and do not provide distributed policy updates using local communication (Chen et al., 2018; Harvey et al., 2020; Niazadeh et al., 2023; Wan et al., 2023). Distributed approaches have also been developed for large-scale offline submodular maximization (Mirzasoleiman et al., 2013), decentralized continuous optimization (Mokhtari et al., 2018), and distributed strategy selection (Rezazadeh & Kia, 2023). However, these distributed methods address static objectives and therefore do not establish tracking guarantees for objectives that change over time. To extend such coordination to changing objectives, an online submodular coordination framework with bounded tracking regret was proposed for multi-robot coordination (Xu et al., 2023). Within this online coordination setting, MA-OSMA establishes a dynamic regret guarantee with a near-optimal approximation factor of $\frac{1-e^{-\bar{c}}}{\bar{c}}$ over a network of agents (Zhang et al., 2025), where $\bar{c}$ denotes an upper bound on the curvature of the objective functions.

Nevertheless, existing methods typically rely on fixed communication graphs, while some formulations additionally assume complete graphs or doubly-stochastic communication weights (Mokhtari et al., 2018; Gharesifard & Smith, 2018; Zhang et al., 2025). In practical applications, however, communication disconnections caused by obstacles, asymmetric channels that only allow one-way transmission (Herrera et al., 2025), or link failures (Gerencsér & Hendrickx, 2019) can make the communication graph directed and time-varying. OSG for multi-agent coordination additionally requires agents to obtain the actions of preceding agents before selecting their own actions (Xu et al., 2023). Although time-varying networks have also been considered, the corresponding decentralized online submodular methods still use doubly-stochastic weight matrices and measure adversarial regret against a fixed comparator (Zhu et al., 2021).

Extending these results to time-varying directed communication presents two challenges. **(1) Biased policy aggregation and intermittent information exchange.** MA-OSMA and related methods assume a fixed connected graph with symmetric doubly stochastic weights (Zhang et al., 2025; Mokhtari et al., 2018). The directed edge prevents symmetric weighting, while local out-degree weights alone do not ensure the agreement. Under B-strong connectivity, individual graphs may also be disconnected, requiring information exchange across rounds. These differences prevent direct application of fixed-matrix agreement analyses. **(2) Preserving the approximation ratio and establishing a dynamic regret bound.** Retaining the approximation ratio $\frac{1-e^{-\bar{c}}}{\bar{c}}$ under such communication requires correcting aggregation bias. The dynamic regret analysis must also account for communication errors together with variation in optimal action sets. We therefore study dynamic regret for coordinated local action selection with online submodular objectives over time-varying B-strongly connected digraphs. Our contributions are as follows:

- We propose **Push–Sum Online Surrogate Mirror Ascent (POSMA)**, which integrates the Push–Sum mechanism with weight-adjusted mirror ascent for online submodular maximization over time-varying digraphs. To accommodate these communication graphs, the doubly-stochastic weight matrices used in MA-OSMA (Zhang et al., 2025) and decentralized methods such as Mokhtari et al. (2018) over static communication graphs are replaced with column-stochastic matrices. The resulting communication weights can be constructed using only the out-degree of each agent, without requiring balanced or undirected communication. Under B-strong connectivity, the $\frac{1-e^{-\bar{c}}}{\bar{c}}$ approximation ratio of MA-OSMA (Zhang et al., 2025) is retained despite these relaxed communication requirements.

- We prove that POSMA attains an average dynamic $\frac{1-e^{-\bar{c}}}{\bar{c}}$-regret upper bound $O\big(\frac{1}{\sqrt{\lambda(1-\lambda)}}\sqrt{\frac{1+C_T}{T}}\big)$, where $\lambda = (1-N^{-NB})^{1/B}$ explicitly reveals how B in B-strong connectivity and the number of agents N govern the agreement rate, and C_T measures the cumulative variation of the optimal sequence, while retaining the curvature-dependent $\frac{1-e^{-\bar{c}}}{\bar{c}}$-approximation ratio. We also establish a bound on the expected difference between the aggregated policy estimates of any two agents, showing how the step sizes and the parameter B affect this bound.
- We validate POSMA on multi-target tracking simulations, showing that it consistently outperforms MA-OSMA, MA-OSEA and OSG under directed, time-varying communication graphs while achieving competitive or better performance than the compared baselines on static undirected networks. The comparison evaluates cumulative average utility, target count, and average tracking distance across different target-motion mixtures.

2 Preliminaries and Problem Formulation

2.1 Preliminaries

Notations. $\mathbb{R}$ and $\mathbb{R}_+$ denote the sets of real and nonnegative real numbers, respectively. For a positive integer m, let $[m] = \{1, \dots, m\}$. For a finite set $\mathcal{S}$, $|\mathcal{S}|$ denotes its cardinality and $2^{\mathcal{S}}$ denotes its power set. For a vector $\mathbf{x} = (x_a)_{a\in\mathcal{S}}$ indexed by a finite set $\mathcal{S}$, $[\mathbf{x}]_a$ denotes the entry indexed by $a \in \mathcal{S}$. For $\mathcal{I} \subseteq \mathcal{S}$, $[\mathbf{x}]_{\mathcal{I}}$ denotes the vector formed by the entries indexed by $\mathcal{I}$. The symbol $\langle\cdot,\cdot\rangle$ denotes the inner product, $\|\cdot\|$ denotes a norm, $\|\cdot\|_*$ denotes its dual norm, and $\|\cdot\|_1$ and $\|\cdot\|_2$ denote the ℓ_1- and Euclidean norms, respectively. The symbols ∇h denote the gradient of h. For $v(T) > 0$, $u(T) = O(v(T))$ means that $\frac{|u(T)|}{v(T)}$ is bounded for sufficiently large T, and $u(T) = o(v(T))$ means that $\lim_{T\to\infty} \frac{u(T)}{v(T)} = 0$.

Time-varying B-strongly connected digraphs. Let $\mathcal{G}_t = (\mathcal{N}, \mathcal{E}_t)$ be a time-varying digraph with node set $\mathcal{N} = \{1, \dots, N\}$ and directed edge set $\mathcal{E}_t \subseteq \mathcal{N} \times \mathcal{N}$, where $(j, i) \in \mathcal{E}_t$ denotes an edge from node j to node i. A digraph is strongly connected if a directed path exists from every node to every other node. For a positive integer B, we call the sequence $\{\mathcal{G}_t\}_{t\geq 1}$ B-strongly connected if the union graph $\mathcal{G}[t, t+B-1] = \Big(\mathcal{N}, \bigcup_{\tau=t}^{t+B-1} \mathcal{E}_\tau\Big)$ is strongly connected for every $t \geq 1$ (Lee et al., 2017). Although the graph at a given round may not be strongly connected, B-strong connectivity allows information exchange between any two nodes over successive rounds. Let $\mathbf{A}(t) \in \mathbb{R}^{N\times N}$ be a nonnegative column-stochastic matrix associated with $\mathcal{G}_t$, which satisfying $\sum_{i=1}^{N}[\mathbf{A}(t)]_{ij} = 1$ for each $j \in \mathcal{N}$. Each entry $[\mathbf{A}(t)]_{ij}$ represents the fraction of a message sent from node j to node i, or retained when $i = j$. Following Lee et al. (2017), we use uniform outgoing weights, including a self-weight,

$$[\mathbf{A}(t)]_{ij} = \begin{cases} \frac{1}{d_j(t)}, & i = j \text{ or } (j, i) \in \mathcal{E}_t, \\ 0, & \text{otherwise}, \end{cases}$$

where $d_j(t) = |\{j\} \cup \{i \in \mathcal{N} : (j, i) \in \mathcal{E}_t\}|$ counts node j itself and its outgoing neighbors. Since $1 \leq d_j(t) \leq N$, all diagonal entries are positive and all nonzero entries are at least $\frac{1}{N}$.

Submodularity, monotonicity, and curvature. Consider N agents indexed by $\mathcal{N} = \{1, ..., N\}$. Each agent $i \in \mathcal{N}$ has a nonempty finite action set $\mathcal{V}_i$, such that $\mathcal{V}_i \cap \mathcal{V}_j = \emptyset$ for any $i \neq j$. The ground set of all available actions is $\mathcal{V} = \bigcup_{i\in\mathcal{N}} \mathcal{V}_i$, with $n = |\mathcal{V}|$ and $n_i = |\mathcal{V}_i|$. Let $t \in [T]$ denote the iteration round, where T is a positive integer representing the total number of rounds. For each $t \in [T]$, let $f_t : 2^{\mathcal{V}} \to \mathbb{R}_+$ be a set function. The function f_t is submodular if, for any $\mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{V}$ and $a \in \mathcal{V} \setminus \mathcal{B}$, $f_t(\mathcal{A} \cup \{a\}) - f_t(\mathcal{A}) \geq f_t(\mathcal{B} \cup \{a\}) - f_t(\mathcal{B})$. The function f_t is monotone if $f_t(\mathcal{A}) \leq f_t(\mathcal{B})$ for any $\mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{V}$, and normalized if $f_t(\emptyset) = 0$. The curvature (Zhang et al., 2025; Conforti & Cornuéjols, 1984) of a normalized monotone submodular function is defined as $c_t = 1 - \min_{\substack{\mathcal{A}\subseteq\mathcal{V},\, a\in\mathcal{V}\setminus\mathcal{A} \\ f_t(\{a\})>0}} \frac{f_t(\mathcal{A}\cup\{a\}) - f_t(\mathcal{A})}{f_t(\{a\})}$, with the bound $c_t \in [0, 1]$. Let $\bar{c} \in (0, 1]$ be a uniform upper bound such that $c_t \leq \bar{c}$ for all $t \in [T]$.

2.2 Problem Formulation

In this setting, decisions are made sequentially without prior knowledge of future objective functions. At the round $t \in [T]$, each agent $i \in \mathcal{N}$ maintains a global policy $\mathbf{x}_{t,i} \in [0,1]^n$, with entries indexed by actions in $\mathcal{V}$, and uses the entries corresponding to $\mathcal{V}_i$ to select and execute an action $a_{t,i} \in \mathcal{V}_i$. These executed actions form the joint action set $\mathcal{A}_t = \bigcup_{i\in\mathcal{N}}\{a_{t,i}\} \subseteq \mathcal{V}$, which contains one action from each agent and yields the collective utility $f_t(\mathcal{A}_t)$. After execution, each agent obtains the local marginal gain oracle $\mathcal{O}_t^i(a,\mathcal{A}) = f_t(\mathcal{A}\cup\{a\}) - f_t(\mathcal{A})$, where $a \in \mathcal{V}_i$ and $\mathcal{A} \subseteq \mathcal{V}$, to update its policy for the next round (Xu et al., 2023; Zhang et al., 2025; Rezazadeh & Kia, 2023). These updates aim to coordinate the actions of all agents to maximize the collective utility at each round. The corresponding online submodular optimization problem at the round t is

$$\max_{\mathcal{A}\subseteq\mathcal{V}} f_t(\mathcal{A}), \qquad \text{s.t.} \quad |\mathcal{A}\cap\mathcal{V}_i| \le 1, \quad \forall i \in \mathcal{N}, \tag{1}$$

where $\mathcal{A}$ is a candidate action set, whereas $\mathcal{A}_t$ is the set executed.

Dynamic α-regret (Harvey et al., 2020; Zhang et al., 2025). Note that solving the per-round submodular maximization problem exactly is NP-hard in general (Banihashem et al., 2023), which motivates polynomial-time approximation algorithms. Dynamic α-regret is utilized to evaluate the cumulative utility. Let $\mathcal{A}_t^*$ be the optimal solution of (1) at the round t. For $\alpha \in (0,1]$, define

$$\mathrm{Reg}_\alpha(T) = \alpha\sum_{t=1}^{T} f_t(\mathcal{A}_t^*) - \sum_{t=1}^{T} f_t(\mathcal{A}_t), \tag{2}$$

where α specifies the target fraction of the optimal cumulative utility. The goal is to find approximate solutions with an $o(T)$ expected regret upper bound. Accordingly,

$$\limsup_{T\to\infty} \frac{\mathbb{E}[\mathrm{Reg}_\alpha(T)]}{T} \le 0,$$

where the expectation is over the algorithmic randomness.

2.3 Multilinear Extension, Surrogate Gradient, and Bregman Divergence

This subsection introduces the multilinear extension (Zhang et al., 2025; Krause & Golovin, 2014), the surrogate gradient (Zhang et al., 2025), and the Bregman divergence (Bregman, 1967).

Definition 1 (Multilinear extension). For each $t \in [T]$, the multilinear extension (Zhang et al., 2025) of $f_t : 2^{\mathcal{V}} \to \mathbb{R}_+$ is defined as

$$F_t(\mathbf{x}) = \sum_{\mathcal{A}\subseteq\mathcal{V}} f_t(\mathcal{A}) \prod_{a\in\mathcal{A}} x_a \prod_{a\in\mathcal{V}\setminus\mathcal{A}} (1-x_a) = \mathbb{E}_{\mathcal{R}\sim\mathbf{x}}[f_t(\mathcal{R})], \tag{3}$$

where $\mathbf{x} = (x_a)_{a\in\mathcal{V}} \in [0,1]^n$ and $x_a \in [0,1]$ is the inclusion probability of action a. The notation $\mathcal{R} \sim \mathbf{x}$ means that $\mathcal{R} \subseteq \mathcal{V}$ is sampled by including each $a \in \mathcal{V}$ independently with probability x_a. Here, $\mathcal{A}$ ranges over all subsets of $\mathcal{V}$, whereas $\mathcal{R}$ is a random subset. Based on the extension, the continuous optimization problem of (1) at the round t is

$$\max_{\mathbf{x}\in\mathcal{C}} F_t(\mathbf{x}), \qquad \mathcal{C} = \left\{ \mathbf{x} \in [0,1]^n \;\middle|\; \sum_{a\in\mathcal{V}_i} x_a \le 1, \ \forall i \in \mathcal{N} \right\}. \tag{4}$$

The set $\mathcal{C}$ is a nonempty compact convex feasible set. For every $a \in \mathcal{V}$, the partial derivative (Calinescu et al., 2011) is

$$\frac{\partial F_t(\mathbf{x})}{\partial x_a} = \mathbb{E}_{\mathcal{R}\sim\mathbf{x}}\left[f_t(\mathcal{R}\cup\{a\}) - f_t(\mathcal{R}\setminus\{a\})\right]. \tag{5}$$

Thus, each gradient component is the expected marginal gain of its corresponding action in (1).

Definition 2 (Surrogate function (Zhang et al., 2025; 2026)). For $\mathbf{x} \in \mathcal{C}$ and the curvature bound $\bar{c} \in (0,1]$, the surrogate function $\widehat{F}_t$ is specified, up to an additive constant, by its gradient

$$\nabla\widehat{F}_t(\mathbf{x}) = \int_0^1 e^{\bar{c}(z-1)} \nabla F_t(z\mathbf{x})\,\mathrm{d}z, \tag{6}$$

where $z \in [0,1]$ is an integration variable, and $z\mathbf{x}$ lies on the line segment between $\mathbf{0}$ and $\mathbf{x}$.

Definition 3 (Bregman divergence (Bregman, 1967; Zhang et al., 2025; Jiang & Vandenberghe, 2023)). Let $\mathcal{C} \subseteq \mathbb{R}^n$ be the nonempty compact convex feasible set defined in (4). Let $\psi : \mathcal{C} \to \mathbb{R}$ be a continuously differentiable potential function that is 1-strongly convex with respect to $\|\cdot\|$. For any two vectors $\mathbf{x}, \mathbf{y} \in \mathcal{C}$, the Bregman divergence associated with ψ is

$$\mathcal{D}_\psi(\mathbf{x}, \mathbf{y}) = \psi(\mathbf{x}) - \psi(\mathbf{y}) - \langle \nabla\psi(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle . \tag{7}$$

To let each agent update its global policy only in the entries corresponding to its own action set, we choose a coordinate-separable potential $\psi(\mathbf{x}) = \sum_{a\in\mathcal{V}} g(x_a)$ (Zhang et al., 2025), where x_a is the entry of $\mathbf{x}$ for action a, and $g[0,1] \to \mathbb{R}$ is continuously differentiable. This form separates the divergence across the entries corresponding to each $\mathcal{V}_i$. For each agent $i \in \mathcal{N}$, fix an ordering of $\mathcal{V}_i$. Let $\mathbf{b} = (b_1, \ldots, b_{n_i}) \in [0,1]^{n_i}$ be a candidate for the updated policy entries corresponding to $\mathcal{V}_i$, and $\mathbf{c} = (c_1, \ldots, c_{n_i}) \in [0,1]^{n_i}$ the corresponding reference point after communication. The entries b_k and c_k correspond to the k-th action. The Bregman divergence restricted to the entries corresponding to $\mathcal{V}_i$ is defined as

$$\mathcal{D}_{g,n_i}(\mathbf{b}, \mathbf{c}) = \sum_{k=1}^{n_i} \left[g(b_k) - g(c_k) - g'(c_k)(b_k - c_k) \right], \tag{8}$$

where $g'(c_k) = \left.\frac{\mathrm{d}g(u)}{\mathrm{d}u}\right|_{u=c_k}$ is the first derivative of g at c_k. For $\mathbf{x}, \mathbf{y} \in \mathcal{C}$, $[\mathbf{x}]_{\mathcal{V}_i} = (x_a)_{a\in\mathcal{V}_i} \in [0,1]^{n_i}$ and $[\mathbf{y}]_{\mathcal{V}_i} = (y_a)_{a\in\mathcal{V}_i} \in [0,1]^{n_i}$ denote their entries corresponding to $\mathcal{V}_i$ in the same ordering. Thus, $\mathcal{D}_\psi(\mathbf{x}, \mathbf{y}) = \sum_{i\in\mathcal{N}} \mathcal{D}_{g,n_i}([\mathbf{x}]_{\mathcal{V}_i}, [\mathbf{y}]_{\mathcal{V}_i})$.

3 Push–Sum Online Surrogate Mirror Ascent Algorithm

We propose POSMA for online submodular maximization over time-varying B-strongly connected digraphs (Algorithm 1). Changing paths and intermittent disconnection delay policy exchange, while directed communication can cause unequal influence in aggregation (Nedić & Olshevsky, 2015). Coordination therefore requires exchanges across rounds despite disconnected communication graphs.

Push–Sum supports agreements over time-varying digraphs by exchanging weighted policy estimates and scalar weights, then normalizing the estimates by these weights to correct unequal influence in aggregation (Kempe et al., 2003; Nedić & Olshevsky, 2015). Under B-strong connectivity, repeated exchanges enable coordination even when individual communication graphs are disconnected.

Building on this mechanism, we propose POSMA. At each round, agent i samples an action from its local action set $\mathcal{V}_i$ according to its normalized local policy entries. After execution, it exchanges the weighted policy estimate $\mathbf{q}_{t,i} = w_{t,i}\mathbf{x}_{t,i}$ and weight $w_{t,i}$ with its neighbors, and uses Push–Sum normalization to obtain the aggregated estimate $\mathbf{y}_{t+1,i}$. It then updates the entries corresponding to $\mathcal{V}_i$ using an unbiased surrogate gradient estimate at $\mathbf{x}_{t,i}$, with reference $[\mathbf{y}_{t+1,i}]_{\mathcal{V}_i}$ and Bregman coefficient $\frac{w_{t+1,i}}{\eta_t}$. The remaining entries are copied from $\mathbf{y}_{t+1,i}$, and $\mathbf{q}_{t+1,i} = w_{t+1,i}\mathbf{x}_{t+1,i}$ is formed for the next round.

Remark 1 (Communication and aggregation). MA-OSMA (Zhang et al., 2025) uses $\sum_j W_{ij}\mathbf{x}_{t,j}$ with a fixed symmetric doubly stochastic matrix $\mathbf{W}$. To accommodate time-varying directed communication, POSMA uses column-stochastic $\mathbf{A}(t)$ and normalizes the weighted estimates as $\mathbf{y}_{t+1,i} = \frac{\sum_j [\mathbf{A}(t)]_{ij} w_{t,j}\mathbf{x}_{t,j}}{\sum_j [\mathbf{A}(t)]_{ij} w_{t,j}}$. This corrects aggregation bias (Nedić & Olshevsky, 2015), while B-strong connectivity allows information to propagate across rounds despite intermittent disconnection.

Remark 2 (Updates and approximation ratio). Unlike the coefficient $\frac{1}{\eta_t}$ in MA-OSMA (Zhang et al., 2025), the Bregman term in POSMA uses $\frac{w_{t+1,i}}{\eta_t}$, giving an effective step size $\frac{\eta_t}{w_{t+1,i}}$. This scaling prevents local gradient contributions from being multiplied by $w_{t+1,i}$ in the weighted analysis. Together with the surrogate gradient inequality, it supports the dynamic regret bound with the same approximation ratio $\frac{1-e^{-\bar{c}}}{\bar{c}}$.

Algorithm 1 Push–Sum Online Surrogate Mirror Ascent (POSMA)

Input: Number of iterations T; agents $\mathcal{N}$; local action sets $\{\mathcal{V}_i\}_{i\in\mathcal{N}}$ with $n_i = |\mathcal{V}_i|$ and $\mathcal{V} = \bigcup_{i\in\mathcal{N}} \mathcal{V}_i$; column stochastic matrices $\{\mathbf{A}(t)\}_{t=1}^{T}$ associated with time-varying B-strongly connected digraphs $\mathcal{G}_t = (\mathcal{N}, \mathcal{E}_t)$; curvature parameter $\bar{c} \in (0, 1]$; step sizes $\{\eta_t\}_{t=1}^{T}$; 1-strongly decomposable convex function $\psi(\mathbf{x})$.
Output: Actions $\{a_{t,i}\}$ and strategy vectors $\{\mathbf{x}_{t,i}\}$.
Initialize: For each agent $i \in \mathcal{N}$, set $[\mathbf{x}_{1,i}]_a = \frac{1}{n_i}$ for $a \in \mathcal{V}_i$ and $[\mathbf{x}_{1,i}]_a = 0$ otherwise; $w_{1,i} = 1$ and $\mathbf{q}_{1,i} = \mathbf{y}_{1,i} = \mathbf{x}_{1,i}$.
for $t = 1, \dots, T$ **do**
  for $i \in \mathcal{N}$ **do** ▷ All agents execute in parallel
    Compute $m_{t,i} := \sum_{a\in\mathcal{V}_i} [\mathbf{x}_{t,i}]_a$. ▷ Action selection
    Select an action $a_{t,i} \in \mathcal{V}_i$ with probability $\Pr(a_{t,i} = a) = \frac{[\mathbf{x}_{t,i}]_a}{m_{t,i}}$ and execute.
    Send $\mathbf{q}_{t,i}$ and $w_{t,i}$ to neighbors and receive the messages from neighbors.
    $w_{t+1,i} = \sum_{j=1}^{N} [\mathbf{A}(t)]_{ij} w_{t,j}$ ▷ Push–Sum consensus
    $\mathbf{s}_{t+1,i} = \sum_{j=1}^{N} [\mathbf{A}(t)]_{ij} \mathbf{q}_{t,j}$
    $\mathbf{y}_{t+1,i} = \frac{\mathbf{s}_{t+1,i}}{w_{t+1,i}}$
    Sampling a random number $z_{t,i}$ from $Z_{\bar{c}}$. ▷ Surrogate gradient estimation
    Sampling $\mathcal{R}_{t,i}$ by including each $a \in \mathcal{V}$ independently with probability $z_{t,i}[\mathbf{x}_{t,i}]_a$.
    Compute $[\widetilde{\nabla} F_t(\mathbf{x}_{t,i})]_a = \frac{1-e^{-\bar{c}}}{\bar{c}} [f_t(\mathcal{R}_{t,i} \cup \{a\}) - f_t(\mathcal{R}_{t,i} \setminus \{a\})]$ for any $a \in \mathcal{V}_i$.
    Update $[\mathbf{x}_{t+1,i}]_a = [\mathbf{y}_{t+1,i}]_a$ for all $a \notin \mathcal{V}_i$. ▷ Mirror ascent update
    Update the probabilities of actions in the strategy vector:

$$[\mathbf{x}_{t+1,i}]_{\mathcal{V}_i} = \underset{\substack{\mathbf{b}\in[0,1]^{n_i} \\ \sum_{k=1}^{n_i} b_k \le 1}}{\arg\min} \left(-\left\langle [\widetilde{\nabla} F_t(\mathbf{x}_{t,i})]_{\mathcal{V}_i}, \mathbf{b} \right\rangle + \frac{w_{t+1,i}}{\eta_t} \mathcal{D}_{g,n_i}\left(\mathbf{b}, [\mathbf{y}_{t+1,i}]_{\mathcal{V}_i}\right) \right).$$

    $\mathbf{q}_{t+1,i} = w_{t+1,i} \mathbf{x}_{t+1,i}$ ▷ Weighted strategy update
  end for
end for

4 α-Regret Guarantees

This section establishes the regret and agreement guarantees of POSMA over time-varying B-strongly connected digraphs. We first state the assumptions on the potential function and gradient estimates used in the analysis.

Assumption 1 (Potential function). The potential $\psi : \mathcal{C} \to \mathbb{R}$ is continuously differentiable, 1-strongly convex with respect to $\|\cdot\|$, and separable as $\psi(\mathbf{x}) = \sum_{a\in\mathcal{V}} g(x_a)$.

Assumption 2 (Symmetry and convexity) (Bauschke & Borwein, 2001; Boissonnat et al., 2007). The Bregman divergence satisfies $\mathcal{D}_\psi(\mathbf{x}, \mathbf{y}) = \mathcal{D}_\psi(\mathbf{y}, \mathbf{x})$ for all $\mathbf{x}, \mathbf{y} \in \mathcal{C}$. Moreover, for any $\mathbf{x}, \mathbf{y}_i \in \mathcal{C}$ and weights $\omega_i \ge 0$ with $\sum_{i=1}^{N} \omega_i = 1$, $\mathcal{D}_\psi(\mathbf{x}, \sum_{i=1}^{N} \omega_i \mathbf{y}_i) \le \sum_{i=1}^{N} \omega_i \mathcal{D}_\psi(\mathbf{x}, \mathbf{y}_i)$.

Assumption 3 (Lipschitz continuity). There exists a constant $K > 0$ such that $|\mathcal{D}_\psi(\mathbf{x}, \mathbf{z}) - \mathcal{D}_\psi(\mathbf{y}, \mathbf{z})| \le K\|\mathbf{x} - \mathbf{y}\|$ for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathcal{C}$.

Assumption 4 (Gradient estimates and regularity) (Zhang et al., 2025). For all $t \in [T]$, $i \in \mathcal{N}$, and $\mathbf{x}, \mathbf{y} \in \mathcal{C}$, the gradient estimation satisfy $\mathbb{E}[\widetilde{\nabla} F_t(\mathbf{x}) \mid \mathbf{x}] = \nabla \widehat{F}_t(\mathbf{x})$ and $\mathbb{E}[\|\widetilde{\nabla} F_t(\mathbf{x})\|_*^2] \le G^2$. The second-moment bound also holds for $\widetilde{\nabla} F_t(\mathbf{x}) \odot \mathbf{1}_{\mathcal{V}_i}$. Moreover, $\|\nabla \widehat{F}_t(\mathbf{x}) - \nabla \widehat{F}_t(\mathbf{y})\|_* \le L_g\|\mathbf{x} - \mathbf{y}\|$, $\|(\nabla \widehat{F}_t(\mathbf{x}) - \nabla \widehat{F}_t(\mathbf{y})) \odot \mathbf{1}_{\mathcal{V}_i}\|_* \le L_g\|\mathbf{x} - \mathbf{y}\|$, and $\|\nabla F_t(\mathbf{x})\|_* \le G$, where $G > 0$ and $L_g \ge 0$ are independent of T, and $\|\cdot\|_*$ is the dual norm of $\|\cdot\|$.

Under these assumptions, the following theorem bounds the expected dynamic α-regret in terms of optimal-action variation, communication, and step sizes.

Theorem 1 (Proof in Appendix B). Suppose that Assumptions 1–4 hold. Let each f_t be normalized, monotone, and submodular with curvature $c_t \le \bar{c} \in (0, 1]$, and let $\{\eta_t\}_{t=1}^{T}$ be deterministic, positive,

and nonincreasing. For $\alpha = \frac{1-e^{-\bar{c}}}{\bar{c}}$,

$$\begin{aligned}\mathbb{E}[\mathrm{Reg}_\alpha(T)] \le \frac{NR^2}{\eta_T} + \sum_{t=1}^{T-1} \frac{KNC_2}{\eta_t} |\mathcal{A}^*_{t+1} \Delta \mathcal{A}^*_t| \\ + \frac{2\beta N(L_g D + (1+C_2)G)}{\gamma} \sum_{t=1}^{T} \left(NR\lambda^{t-2} + NG \sum_{s=1}^{t-1} \lambda^{t-s-2} \eta_s \right) \\ + \left[\frac{NG(4L_g D + 4C_2 G)}{2} + \frac{NG^2}{2\gamma} \right] \sum_{t=1}^{T} \eta_t,\end{aligned} \tag{9}$$

where $\mathcal{A}^*_t$ is an optimal action set at the round t, $D = \sup_{\mathbf{x},\mathbf{y}\in\mathcal{C}} \|\mathbf{x}-\mathbf{y}\|$, and $R \ge \max\{\sup_{\mathbf{x}\in\mathcal{C}} \|\mathbf{x}\|, D, \sqrt{KD}\}$.

Remark 3. POSMA extends MA-OSMA from a fixed communication graph with a symmetric doubly stochastic matrix to time-varying directed graphs, while retaining the approximation ratio $\frac{1-e^{-\bar{c}}}{\bar{c}}$ (Zhang et al., 2025). This extension requires accounting for both changing communication paths and unequal influence among agents. Changing communication paths can delay the propagation of policy estimates, so the agreement analysis must account for communication over successive rounds. The latter is addressed through Push–Sum normalization and an effective local step size of $\frac{\eta_t}{w_{t+1,i}}$, implemented by the coefficient $\frac{w_{t+1,i}}{\eta_t}$ in the Bregman term. The regret analysis consequently uses a weighted Bregman potential to account for both communication and local updates.

We next quantify the agreement error by bounding the expected difference between the policy estimates of any two agents.

Theorem 2 (Proof in Appendix C). Under Assumptions 1–4 and the communication model of Section 2.1 with $N \ge 2$, let the step sizes be deterministic and positive. For any $i, j \in \mathcal{N}$ and $2 \le t \le T+1$,

$$\mathbb{E}[\|\mathbf{y}_{t,i} - \mathbf{y}_{t,j}\|] \le \frac{4\beta N}{\gamma} \left(R\lambda^{t-2} + G \sum_{s=1}^{t-2} \lambda^{t-s-2} \eta_s \right), \tag{10}$$

where R is defined in Theorem 1.

Remark 4. The two terms in (10) describe the geometric decay of the initial disagreement and the effect of past local updates, respectively. MA-OSMA characterizes communication through a spectral parameter of its time-invariant matrix (Zhang et al., 2025). Theorem 2 instead uses the matrix-product rate λ and the weight lower bound γ to accommodate time-varying directed graphs. For an infinite run under the same uniform assumptions, a constant step size $\eta_t = \eta > 0$ yields $\limsup_{t\to\infty} \mathbb{E}\|\mathbf{y}_{t,i} - \mathbf{y}_{t,j}\| \le \frac{4\beta NG\eta}{\gamma(1-\lambda)}$.

For a constant step size, the regret bound in Theorem 1 separates into terms proportional to $\frac{1}{\eta}$ and $T\eta$, together with a term independent of η. Balancing these contributions gives the following step-size choice.

Corollary 1. Under the conditions of Theorem 1, the positive constant step size minimizing the derived regret upper bound is

$$\eta_t = \eta^* = \sqrt{\frac{NR^2 + KNC_2 C_T}{T\left[\frac{2\beta N^2 G(L_g D + (1+C_2)G)}{\gamma\lambda(1-\lambda)} + 2NG(L_g D + C_2 G) + \frac{NG^2}{2\gamma}\right]}}, \qquad t \in [T], \tag{11}$$

where $C_T := \sum_{t=1}^{T-1} |\mathcal{A}^*_{t+1} \Delta \mathcal{A}^*_t|$ measures the cumulative variation of the optimal action sets. The proof is provided in Appendix D.

Remark 5. When the constants in Theorem 1 are independent of T, Corollary 1 yields an upper bound on the expected average dynamic α-regret of order $O(\sqrt{\frac{1+C_T}{\lambda(1-\lambda)T}})$. This condition permits both the objective functions and the communication graphs to change over time. As the rounds

T progress, the bound converges to zero, implying $\limsup_{T\to\infty} \frac{\mathbb{E}[\mathrm{Reg}_\alpha(T)]}{T} \leq 0$. As in MA-OSMA (Zhang et al., 2025), the step-size choice balances variation of the optimal action sets against accumulated update errors, with the constants for POSMA also accounting for directed communication and Push–Sum weights.

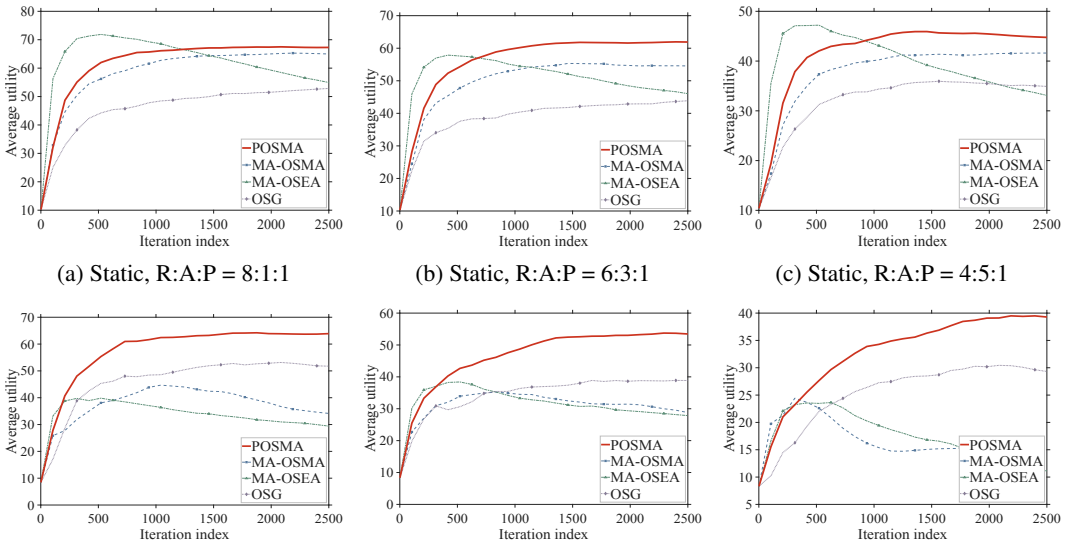
5 Numerical Simulations

We evaluate POSMA through multi-target tracking simulations (Zhang et al., 2025; Corah & Michael, 2021; Xu et al., 2023). The experiments examine the effects of B in B-strong connectivity and the curvature of the objective, compare POSMA with MA-OSMA, MA-OSEA, and OSG, and evaluate the dynamic regret.

Simulation parameters. We simulate 20 agents tracking 30 targets in two dimensions. Agents start uniformly at random within a disk of radius 20 and act every 0.02 seconds for 50 seconds ($T = 2500$ rounds). Each selects from 15 actions combining speeds of $\{5, 10, 15\}$ units/s with four cardinal directions and one diagonal direction. The targets follow Random, Adversarial, and Polyline motion (Zhang et al., 2025), with the ratios of 8:1:1, 6:3:1, and 4:5:1.

For an action $a = (i, \theta, s)$, let $d_{t,j}(a)$ denote the distance between agent i and target j after their movements in round t, with agent i executing a. The resulting tracking utility is $f_t(\mathcal{A}) = \sum_{j=1}^{30} \max_{a\in\mathcal{A}} \frac{1}{d_{t,j}(a)}$, and its cumulative average is $\overline{U}_t = \frac{1}{t}\sum_{\tau=1}^{t} f_\tau(\mathcal{A}_\tau)$. To also assess coverage and tracking distance, we simulate the number of targets within 5 units of at least one agent and the average of the five shortest distances from targets to their nearest agents, both measured after movement in each round. To study objectives with different curvature values, we additionally consider the objective family $f_{t,\bar{c}}(\mathcal{A}) = (1-\bar{c})m_t(\mathcal{A}) + \bar{c}h_t(\mathcal{A})$, where $\bar{c} \in \{0.1, 0.3, 0.5, 0.7, 1.0\}$, $h_t(\mathcal{A}) = \sum_{j=1}^{30} \max_{a\in\mathcal{A}} \frac{1}{d_{t,j}(a)}$ is the coverage objective, and $m_t(\mathcal{A}) = \sum_{a\in\mathcal{A}} h_t(\{a\})$ is a modular function that sums the utility of each action evaluated individually. Thus, $\bar{c} = 0$ gives a modular objective and $\bar{c} = 1$ gives the coverage objective. The parameter $\bar{c}$ controls the curvature of this constructed objective family.

Analysis. We compare POSMA with MA-OSMA, MA-OSEA, and OSG under static undirected and time-varying directed communication in Figure 1.

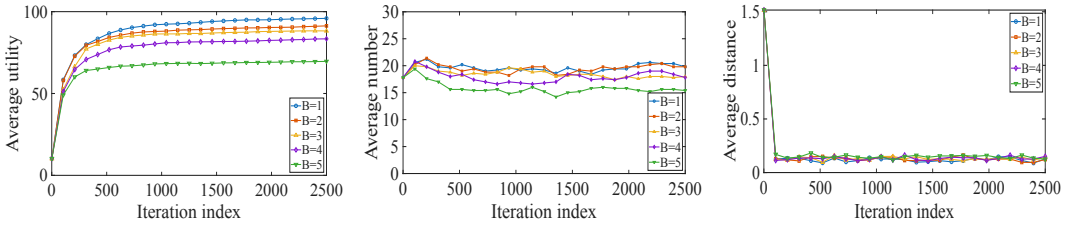


(a) Static, R:A:P = 8:1:1 (b) Static, R:A:P = 6:3:1 (c) Static, R:A:P = 4:5:1

(d) Time-varying, R:A:P = 8:1:1 (e) Time-varying, R:A:P = 6:3:1 (f) Time-varying, R:A:P = 4:5:1

Figure 1: Cumulative average utility of POSMA, MA-OSMA, MA-OSEA, and OSG with $\bar{c} = 1$. Top: a static undirected connected graph. Bottom: time-varying B-strongly connected digraphs with $B = 3$.

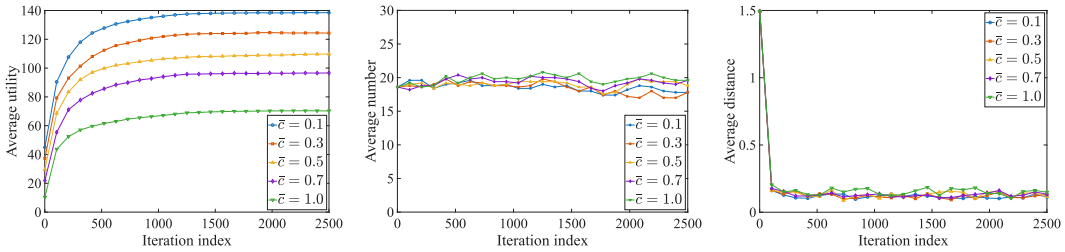
Under static undirected communication, MA-OSEA leads initially but then declines, while POSMA achieves the highest utility in later rounds, followed by MA-OSMA, across all three target ratios. POSMA thus outperforms both methods even in their intended communication setting. Under time-varying directed communication, POSMA also attains the highest final cumulative average utility across all three ratios. Although utility decreases for all four algorithms as the proportion of Adversarial targets increases, POSMA maintains its lead.



(a) Average utility (b) Average number (c) Average distance

Figure 2: Effect of $B \in \{1, 2, 3, 4, 5\}$ in B-strong connectivity with $\bar{c} = 1$ and R:A:P=8:1:1.

We next examine how communication frequency and objective curvature affect POSMA for R:A:P=8:1:1. Figure 2 first compares $B \in \{1, 2, 3, 4, 5\}$ with $\bar{c} = 1$. Larger B means less frequent inter-group communication and slower policy exchange, consistent with the lower final cumulative average utility observed. Nevertheless, the number of covered targets remains within a relatively stable range. The average tracking distance also decreases rapidly and remains low, indicating close tracking of the five nearest targets.



(a) Average utility (b) Average number (c) Average distance

Figure 3: Effect of $\bar{c} \in \{0.1, 0.3, 0.5, 0.7, 1.0\}$ with $B = 3$ and R:A:P=8:1:1.

With $B = 3$, Figure 3 shows higher cumulative average utility for smaller $\bar{c}$, consistent with the larger approximation factor $(1 - e^{-\bar{c}})/\bar{c}$. Further results for the other target proportions, α-regret experiments, and tracking simulations under practically motivated settings are provided in the appendix.

6 Conclusion

We studied online submodular maximization for multi-agent systems over time-varying B-strongly connected digraphs. To address this problem, we proposed POSMA, which combines the multilinear extension, a surrogate gradient, and Push–Sum aggregation using column-stochastic matrices. Under the stated assumptions, POSMA admits an expected average dynamic $\frac{1-e^{-\bar{c}}}{\bar{c}}$-regret upper bound of order $O\big(\frac{1}{\sqrt{\lambda(1-\lambda)}}\sqrt{\frac{1+C_T}{T}}\big)$. To further characterize agreement among agents, we also established a consensus error bound for the local policy estimates. Finally, multi-target tracking simulations complemented the theoretical analysis by demonstrating the effectiveness of POSMA.

AI USE STATEMENT

In this work, we used generative AI tools for developing and debugging simulation code. We have not used generative AI tools for originating the core research ideas and developing the main theoretical derivations. Additionally, we used generative AI tools for polishing the manuscript, improving logical coherence, and suggesting revisions to the presentation of experimental results. We have reviewed all AI-assisted work and revised the outputs as appropriate. We take responsibility for the final content of this work, including all text, claims, and code produced with the assistance of generative AI.

ETHICS STATEMENT

This work studies distributed online optimization through theoretical analysis and numerical simulations. The experiments use simulated agents and targets and do not involve human participants, personal data, or real-world tracking. The assumptions and experimental settings are documented, and the implementation is publicly available to support reproducibility. Any future deployment in real-world tracking applications should consider privacy, safety, and the potential misuse of tracking capabilities.

REPRODUCIBILITY STATEMENT

The algorithm and assumptions are specified in the main text, with detailed proofs provided in the appendix. The simulation code is available at https://anonymous.4open.science/r/Distributed-Policy-Learning-for-Online-Submodular-Maximization-over-Time-Varying-Directed-Graphs-106D/, including experimental parameters, algorithm implementations, and plotting scripts for reproducing the simulation results.

REFERENCES

Ravikumar Balakrishnan, Tian Li, Tianyi Zhou, Nageen Himayat, Virginia Smith, and Jeff Bilmes. Diverse client selection for federated learning via submodular maximization. In *International Conference on Learning Representations*, 2022.

Kiarash Banihashem, Leyla Biabani, Samira Goudarzi, MohammadTaghi Hajiaghayi, Peyman Jabbarzade, and Morteza Monemizadeh. Dynamic non-monotone submodular maximization. In *Advances in Neural Information Processing Systems*, volume 36, pp. 17369–17382, 2023.

Heinz H. Bauschke and Jonathan M. Borwein. Joint and separate convexity of the Bregman distance. In Dan Butnariu, Yair Censor, and Simeon Reich (eds.), *Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications*, volume 8 of *Studies in Computational Mathematics*, pp. 23–36. Elsevier, 2001.

Yatao Bian, Joachim M. Buhmann, and Andreas Krause. Continuous submodular function maximization. *CoRR*, abs/2006.13474, 2020. URL https://arxiv.org/abs/2006.13474.

Jean-Daniel Boissonnat, Frank Nielsen, and Richard Nock. Bregman Voronoi diagrams: Properties, algorithms and applications. Research Report RR-6154, INRIA, 2007. URL https://inria.hal.science/inria-00137865v2.

Lev M Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics*, 7(3):200–217, 1967.

Gruia Calinescu, Chandra Chekuri, Martin Pál, and Jan Vondrák. Maximizing a monotone submodular function subject to a matroid constraint. *SIAM Journal on Computing*, 40(6):1740–1766, 2011.

Gong Chen and Marc Teboulle. Convergence analysis of a proximal-like minimization algorithm using Bregman functions. *SIAM Journal on Optimization*, 3(3):538–543, 1993.

Lin Chen, Hamed Hassani, and Amin Karbasi. Online continuous submodular maximization. In *International Conference on Artificial Intelligence and Statistics*, pp. 1896–1905. PMLR, 2018.

Michele Conforti and Gérard Cornuéjols. Submodular set functions, matroids and the greedy algorithm: Tight worst-case bounds and some generalizations of the Rado–Edmonds theorem. *Discrete Applied Mathematics*, 7(3):251–274, 1984. doi: 10.1016/0166-218X(84)90003-9.

Micah Corah and Nathan Michael. Scalable distributed planning for multi-robot, multi-target tracking. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 437–444. IEEE, 2021.

Marshall L. Fisher, George L. Nemhauser, and Laurence A. Wolsey. An analysis of approximations for maximizing submodular set functions—II. In M. L. Balinski and A. J. Hoffman (eds.), *Polyhedral Combinatorics: Dedicated to the Memory of D. R. Fulkerson*, volume 8 of *Mathematical Programming Study*, pp. 73–87. Springer, 1978. doi: 10.1007/BFb0121195.

Balázs Gerencsér and Julien M Hendrickx. Push-sum with transmission failures. *IEEE Transactions on Automatic Control*, 64(3):1019–1033, 2019.

Bahman Gharesifard and Stephen L Smith. Distributed submodular maximization with limited information. *IEEE Transactions on Control of Network Systems*, 5(4):1635–1645, 2018.

Daniel Golovin, Andreas Krause, and Matthew Streeter. Online submodular maximization under a matroid constraint with application to learning assignments. *arXiv preprint arXiv:1407.1082*, 2014.

Nicholas Harvey, Christopher Liaw, and Tasuku Soma. Improved algorithms for online submodular maximization via first-order regret bounds. In *Advances in Neural Information Processing Systems*, volume 33, pp. 123–133, 2020.

Daniel Perez Herrera, Angelo Rodio, Zheng Chen, and Erik G Larsson. Power allocation for average consensus in wireless networks with link failures. In *European WIRELESS 2025; 30th European Wireless Conference*, pp. 172–177. VDE, 2025.

Ali Jadbabaie, Alexander Rakhlin, Shahin Shahrampour, and Karthik Sridharan. Online optimization: Competing with dynamic comparators. In *Artificial Intelligence and Statistics*, pp. 398–406. PMLR, 2015.

Xin Jiang and Lieven Vandenberghe. Bregman three-operator splitting methods. *Journal of Optimization Theory and Applications*, 196(3):936–972, 2023.

Andrés Catalino Castillo Jiménez, Ege C Kaya, Lintao Ye, and Abolfazl Hashemi. Submodular maximization approaches for equitable client selection in federated learning. *arXiv preprint arXiv:2408.13683*, 2024.

David Kempe, Alin Dobra, and Johannes Gehrke. Gossip-based computation of aggregate information. In *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 482–491. IEEE, 2003.

Andreas Krause and Daniel Golovin. Submodular function maximization. In Lucas Bordeaux, Youssef Hamadi, and Pushmeet Kohli (eds.), *Tractability: Practical Approaches to Hard Problems*, pp. 71–104. Cambridge University Press, 2014.

Soomin Lee, Angelia Nedić, and Maxim Raginsky. Stochastic dual averaging for decentralized online optimization on time-varying communication graphs. *IEEE Transactions on Automatic Control*, 62(12):6407–6414, 2017.

Baharan Mirzasoleiman, Amin Karbasi, Rik Sarkar, and Andreas Krause. Distributed submodular maximization: Identifying representative elements in massive data. In *Advances in Neural Information Processing Systems*, volume 26, 2013.

Aryan Mokhtari, Hamed Hassani, and Amin Karbasi. Decentralized submodular maximization: Bridging discrete and continuous settings. In *International Conference on Machine Learning*, pp. 3616–3625. PMLR, 2018.

Angelia Nedić and Alex Olshevsky. Distributed optimization over time-varying directed graphs. *IEEE Transactions on Automatic Control*, 60(3):601–615, 2015.

George L Nemhauser, Laurence A Wolsey, and Marshall L Fisher. An analysis of approximations for maximizing submodular set functions—I. *Mathematical Programming*, 14(1):265–294, 1978.

Rad Niazadeh, Negin Golrezaei, Joshua Wang, Fransisca Susan, and Ashwinkumar Badanidiyuru. Online learning via offline greedy algorithms: Applications in market design and optimization. *Management Science*, 69(7):3797–3817, 2023.

Navid Rezazadeh and Solmaz S Kia. Distributed strategy selection: A submodular set function maximization approach. *Automatica*, 153:111000, 2023.

Matthew Streeter and Daniel Golovin. An online algorithm for maximizing submodular functions. In *Advances in Neural Information Processing Systems*, volume 21, 2008.

Vasileios Tzoumas, Ali Jadbabaie, and George J Pappas. Robust and adaptive sequential submodular optimization. *IEEE Transactions on Automatic Control*, 67(1):89–104, 2022.

Harsh Vardhan, Xiaofan Yu, Tajana Rosing, and Arya Mazumdar. Client selection in federated learning with data heterogeneity and network latencies. *arXiv preprint arXiv:2504.01921*, 2025.

Jan Vondrák. Optimal approximation for the submodular welfare problem in the value oracle model. In *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing*, pp. 67–74, 2008.

Zongqi Wan, Jialin Zhang, Wei Chen, Xiaoming Sun, and Zhijie Zhang. Bandit multi-linear DR-submodular maximization and its applications on adversarial submodular bandits. In *International Conference on Machine Learning*, pp. 35491–35524. PMLR, 2023.

Zirui Xu, Hongyu Zhou, and Vasileios Tzoumas. Online submodular coordination with bounded tracking regret: Theory, algorithm, and applications to multi-robot coordination. *IEEE Robotics and Automation Letters*, 8(4):2261–2268, 2023.

Lintao Ye, Nathaniel Woodford, Sandip Roy, and Shreyas Sundaram. On the complexity and approximability of optimal sensor selection and attack for Kalman filtering. *IEEE Transactions on Automatic Control*, 66(5):2146–2161, 2021.

Qixin Zhang, Zongqi Wan, Yu Yang, Li Shen, and Dacheng Tao. Near-optimal online learning for multi-agent submodular coordination: Tight approximation and communication efficiency. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=i8dYPGdB1C.

Qixin Zhang, Yan Sun, Can Jin, Xikun Zhang, Yao Shu, Puning Zhao, Li Shen, and Dacheng Tao. Effective policy learning for multi-agent online coordination beyond submodular objectives. *Advances in Neural Information Processing Systems*, 38:164278–164339, 2026.

Lifeng Zhou, Vasileios Tzoumas, George J Pappas, and Pratap Tokekar. Resilient active target tracking with multiple robots. *IEEE Robotics and Automation Letters*, 4(1):129–136, 2019.

Lifeng Zhou, Vasileios Tzoumas, George J Pappas, and Pratap Tokekar. Distributed attack-robust submodular maximization for multirobot planning. *IEEE Transactions on Robotics*, 38(5):3097–3112, 2022.

Junlong Zhu, Qingtao Wu, Mingchuan Zhang, Ruijuan Zheng, and Keqin Li. Projection-free decentralized online learning for submodular maximization over time-varying networks. *Journal of Machine Learning Research*, 22(51):1–42, 2021.

A USEFUL THEOREMS AND LEMMAS

This section collects the auxiliary results used in the proofs.

Lemma 1 (Lemma 1 of (Zhang et al., 2025); see also (Calinescu et al., 2011; Bian et al., 2020)). Let $f_t : 2^{\mathcal{V}} \to \mathbb{R}_+$ be a normalized, monotone, and submodular function. Its multilinear extension is defined as $F_t(\mathbf{x}) = \sum_{\mathcal{S}\subseteq\mathcal{V}} f_t(\mathcal{S}) \prod_{a\in\mathcal{S}} x_a \prod_{a\in\mathcal{V}\setminus\mathcal{S}} (1-x_a)$. Then the following properties hold.

1. F_t is monotone, i.e.,
$$\frac{\partial F_t(\mathbf{x})}{\partial x_a} \geq 0, \qquad \forall a \in \mathcal{V},\ \mathbf{x} \in [0,1]^{|\mathcal{V}|};$$
2. F_t is concave along any nonnegative direction $\mathbf{d} \in \mathbb{R}_+^{|\mathcal{V}|}$;
3. for every $a \in \mathcal{V}$,
$$\frac{\partial F_t(\mathbf{x})}{\partial x_a} = \mathbb{E}_{\mathcal{R}\sim\mathbf{x}} \left[f_t(\mathcal{R} \cup \{a\}) - f_t(\mathcal{R} \setminus \{a\}) \right];$$
4. for any $\mathbf{x}, \mathbf{y} \in [0,1]^{|\mathcal{V}|}$ satisfying $\mathbf{x} \leq \mathbf{y}$,
$$\nabla F_t(\mathbf{x}) \geq \nabla F_t(\mathbf{y}).$$

Lemma 2 (Lemma 2 of (Zhang et al., 2025); see also (Calinescu et al., 2011; Bian et al., 2020)). Let F_t be the multilinear extension of a monotone submodular function f_t. Then, for any $\mathbf{x}, \mathbf{y} \in [0,1]^{|\mathcal{V}|}$,
$$\langle \mathbf{y} - \mathbf{x}, \nabla F_t(\mathbf{x}) \rangle \geq F_t(\mathbf{x} \vee \mathbf{y}) + F_t(\mathbf{x} \wedge \mathbf{y}) - 2F_t(\mathbf{x}),$$
where $\mathbf{x} \wedge \mathbf{y}$ and $\mathbf{x} \vee \mathbf{y}$ denote the component-wise minimum and maximum, respectively.

Lemma 3 (Lemma 3 of (Zhang et al., 2025)). Let f_t be a normalized monotone submodular function with curvature $c_t \leq \bar{c} \in (0,1]$. For any $\mathbf{x}, \mathbf{y} \in [0,1]^{|\mathcal{V}|}$,
$$F_t(\mathbf{x} \vee \mathbf{y}) \geq (1-\bar{c})(F_t(\mathbf{x}) - F_t(\mathbf{x} \wedge \mathbf{y})) + F_t(\mathbf{y}).$$
Combining this result with Lemma 2 gives
$$\langle \mathbf{y} - \mathbf{x}, \nabla F_t(\mathbf{x}) \rangle \geq F_t(\mathbf{y}) - (1+\bar{c})F_t(\mathbf{x}).$$

In particular, if $\mathbf{x}$ is a stationary point of F_t over $\mathcal{C}$, then
$$F_t(\mathbf{x}) \geq \frac{1}{1+\bar{c}} \max_{\mathbf{y}\in\mathcal{C}} F_t(\mathbf{y}).$$

Lemma 4 (Lemma 4 of (Zhang et al., 2025)). Let f_t be a normalized monotone submodular function with curvature $c_t \leq \bar{c} \in (0,1]$, and let F_t be its multilinear extension. Then, for any $\mathbf{x}, \mathbf{y} \in [0,1]^{|\mathcal{V}|}$,
$$\langle \mathbf{y}, \nabla F_t(\mathbf{x}) \rangle \geq F_t(\mathbf{y}) - \bar{c}F_t(\mathbf{x}).$$

Surrogate gradient inequality (Theorem 2 of Zhang et al., 2025). Let F_t be the multilinear extension of a normalized, monotone submodular function f_t with curvature $c_t \leq \bar{c} \in (0,1]$. Define
$$\nabla \widehat{F}_t(\mathbf{x}) = \int_0^1 e^{\bar{c}(z-1)} \nabla F_t(z\mathbf{x})\, \mathrm{d}z, \qquad \alpha = \frac{1-e^{-\bar{c}}}{\bar{c}}.$$
Then, for any $\mathbf{x}, \mathbf{y} \in [0,1]^{|\mathcal{V}|}$,
$$\left\langle \mathbf{y} - \mathbf{x}, \nabla \widehat{F}_t(\mathbf{x}) \right\rangle \geq \alpha F_t(\mathbf{y}) - F_t(\mathbf{x}).$$

The inequalities in Lemmas 3 and 4 remain valid with the upper bound $\bar{c}$ because F_t is nonnegative and monotone. In particular, Lemma 4 gives
$$\begin{aligned}
\langle \mathbf{y} - \mathbf{x}, \nabla \widehat{F}_t(\mathbf{x}) \rangle &\geq \int_0^1 e^{\bar{c}(z-1)} \left(F_t(\mathbf{y}) - \bar{c}F_t(z\mathbf{x}) - \frac{\mathrm{d}}{\mathrm{d}z} F_t(z\mathbf{x}) \right) \mathrm{d}z \\
&= \alpha F_t(\mathbf{y}) - \left[e^{\bar{c}(z-1)} F_t(z\mathbf{x}) \right]_{z=0}^{z=1} \\
&= \alpha F_t(\mathbf{y}) - F_t(\mathbf{x}),
\end{aligned}$$

where the last equality uses $F_t(\mathbf{0}) = 0$.

Lemma 5 (Lemma 5 of (Zhang et al., 2025); see also (Chen & Teboulle, 1993; Jadbabaie et al., 2015)). Let ψ be a differentiable, 1-strongly convex function with respect to a norm $\|\cdot\|$ on the convex set $\mathcal{C}$, and define the Bregman divergence by

$$\mathcal{D}_\psi(\mathbf{u}, \mathbf{v}) = \psi(\mathbf{u}) - \psi(\mathbf{v}) - \langle \nabla\psi(\mathbf{v}), \mathbf{u} - \mathbf{v}\rangle .$$

Suppose that

$$\mathbf{x}^+ = \arg\min_{\mathbf{u}\in\mathcal{C}} \{\langle \mathbf{b}, \mathbf{u}\rangle + \mathcal{D}_\psi(\mathbf{u}, \mathbf{x})\} .$$

Then, for every $\mathbf{z} \in \mathcal{C}$,

$$\langle \mathbf{x}^+ - \mathbf{z}, \mathbf{b}\rangle \le \mathcal{D}_\psi(\mathbf{z}, \mathbf{x}) - \mathcal{D}_\psi(\mathbf{z}, \mathbf{x}^+) - \mathcal{D}_\psi(\mathbf{x}^+, \mathbf{x}).$$

Equivalently, for the mirror-ascent update

$$\mathbf{x}^+ = \arg\min_{\mathbf{u}\in\mathcal{C}} \left\{ -\langle \mathbf{h}, \mathbf{u}\rangle + \frac{1}{\eta}\mathcal{D}_\psi(\mathbf{u}, \mathbf{x}) \right\},$$

we have

$$\langle \mathbf{z} - \mathbf{x}^+, \mathbf{h}\rangle \le \frac{1}{\eta}\left[\mathcal{D}_\psi(\mathbf{z}, \mathbf{x}) - \mathcal{D}_\psi(\mathbf{z}, \mathbf{x}^+) - \mathcal{D}_\psi(\mathbf{x}^+, \mathbf{x})\right].$$

For the POSMA analysis below, we use Assumptions 1–4, the communication model of Section 2.1 with $N \ge 2$. Assumption 1 concerns the potential function; Assumption 2 gives symmetry and convexity of the Bregman divergence; Assumption 3 gives its Lipschitz continuity in the first argument; and Assumption 4 concerns gradient estimates and regularity. The objective and communication sequences are fixed independently of the algorithmic randomness, and all step sizes are deterministic and positive. Expectations are over the algorithmic randomness.

Throughout, $\bar{c} \in (0, 1]$ is a uniform curvature upper bound, $c_t \le \bar{c}$ for all t, and $\alpha = (1 - e^{-\bar{c}})/\bar{c}$. We distinguish the multilinear-extension gradient ∇F_t, the surrogate gradient $\nabla \widehat{F}_t$, and its stochastic estimator $\widetilde{\nabla} F_t$. In particular, $\widetilde{\nabla} F_t$ estimates $\nabla \widehat{F}_t$, not ∇F_t. In the notation of Assumption 4, for every query point $\mathbf{x} \in \mathcal{C}$,

$$\mathbb{E}\left[\widetilde{\nabla} F_t(\mathbf{x}) \mid \mathbf{x}\right] = \nabla \widehat{F}_t(\mathbf{x}), \qquad \mathbb{E}\left[\|\widetilde{\nabla} F_t(\mathbf{x})\|_*^2\right] \le G^2,$$

and the same second-moment bound holds for the estimator restricted to $\mathcal{V}_i$. We abbreviate

$$\mathbf{g}_{t,i} := \widetilde{\nabla} F_t(\mathbf{x}_{t,i}) \odot \mathbf{1}_{\mathcal{V}_i}.$$

Thus,

$$\mathbb{E}[\mathbf{g}_{t,i} \mid \mathbf{x}_{t,i}] = \nabla \widehat{F}_t(\mathbf{x}_{t,i}) \odot \mathbf{1}_{\mathcal{V}_i}, \qquad \mathbb{E}[\|\mathbf{g}_{t,i}\|_*^2] \le G^2.$$

At each round, the auxiliary gradient samples are freshly drawn according to the sampling rule at $\mathbf{x}_{t,i}$, independently of previously determined quantities conditional on that query point. In particular, $\mathbf{g}_{t,i}$ and $\mathbf{y}_{t+1,i}$ are conditionally independent given $\mathbf{x}_{t,i}$. This sampling property, together with the displayed query-point conditional unbiasedness, justifies the inner-product expectation identity used in Lemma 12. The same G and L_g bound the full and restricted quantities specified in Assumption 4, including $\|\nabla F_t(\mathbf{x})\|_*$. The positive weight lower bound γ is treated as a fixed communication constant independent of T; its inverse is retained in all explicit bounds and absorbed into constants only in the asymptotic notation specified in the proof of Corollary 1.

Lemma 6 (POSMA update). Let

$$\mathcal{C} = \left\{ \mathbf{x} \in [0, 1]^{|\mathcal{V}|} : \sum_{a\in\mathcal{V}_i} x_a \le 1,\ \forall i \in \mathcal{N} \right\}.$$

Under Assumption 1 and the communication model of Section 2.1 with $N \ge 2$, the POSMA update can be written as

$$\mathbf{x}_{t+1,i} = \arg\min_{\mathbf{x}\in\mathcal{C}} \left\{ -\left\langle \frac{\widetilde{\nabla} F_t(\mathbf{x}_{t,i}) \odot \mathbf{1}_{\mathcal{V}_i}}{w_{t+1,i}}, \mathbf{x} \right\rangle + \frac{1}{\eta_t}\mathcal{D}_\psi\left(\mathbf{x}, \mathbf{y}_{t+1,i}\right) \right\}, \tag{12}$$

where $\odot$ denotes coordinate-wise multiplication and $\mathbf{1}_{\mathcal{V}_i}$ is the indicator vector of $\mathcal{V}_i$. The effective local step size is $\eta_t/w_{t+1,i}$. Only the entries corresponding to $\mathcal{V}_i$ are updated by mirror ascent, while the remaining entries are copied from $\mathbf{y}_{t+1,i}$.

Proof. The feasible set is the Cartesian product of the sets associated with the disjoint action sets $\mathcal{V}_k$. By separability of ψ, the objective in (12) separates across these sets. Multiplying the objective by $w_{t+1,i} > 0$ gives the local Bregman coefficient $w_{t+1,i}/\eta_t$ used in POSMA. For $k \neq i$, the linear term is zero, and the corresponding Bregman term is uniquely minimized at $[\mathbf{y}_{t+1,i}]_{\mathcal{V}_k}$. This reference point is feasible by Lemma 10, whose feasibility argument follows directly from the local update and communication rules. Thus (12) is equivalent to the POSMA update. □

Lemma 7 (Update residual). Under Assumptions 1 and 4, and the communication model of Section 2.1 with $N \geq 2$, define the update residual as

$$\mathbf{r}_{t+1,i} = \mathbf{x}_{t+1,i} - \mathbf{y}_{t+1,i}.$$

Then, for every $t \in [T]$ and $i \in \mathcal{N}$,

$$\mathbb{E}\left[\|\mathbf{r}_{t+1,i}\|\right] = \mathbb{E}\left[\|\mathbf{x}_{t+1,i} - \mathbf{y}_{t+1,i}\|\right] \leq \frac{G\eta_t}{w_{t+1,i}}. \tag{13}$$

Consequently, multiplying by $w_{t+1,i}$ and summing,

$$\sum_{i=1}^{N} w_{t+1,i}\, \mathbb{E}\left[\|\mathbf{r}_{t+1,i}\|\right] \leq NG\eta_t, \tag{14}$$

Proof. By Lemma 6, the first-order optimality condition, with the feasible comparison point $\mathbf{y}_{t+1,i} \in \mathcal{C}$ (Lemma 10), gives

$$\frac{w_{t+1,i}}{\eta_t} \left\langle \nabla\psi(\mathbf{x}_{t+1,i}) - \nabla\psi(\mathbf{y}_{t+1,i}), \mathbf{r}_{t+1,i} \right\rangle \leq \left\langle \mathbf{g}_{t,i}, \mathbf{r}_{t+1,i} \right\rangle.$$

By 1-strong convexity and Hölder's inequality,

$$\frac{w_{t+1,i}}{\eta_t} \|\mathbf{r}_{t+1,i}\|^2 \leq \|\mathbf{g}_{t,i}\|_* \|\mathbf{r}_{t+1,i}\|.$$

Dividing by $\|\mathbf{r}_{t+1,i}\|$ when it is nonzero (the zero case is immediate) yields $\|\mathbf{r}_{t+1,i}\| \leq (\eta_t/w_{t+1,i})\|\widetilde{\nabla} F_t(\mathbf{x}_{t,i}) \odot \mathbf{1}_{\mathcal{V}_i}\|_*$. Taking expectations and using $\mathbb{E}[\|\widetilde{\nabla} F_t(\mathbf{x}_{t,i}) \odot \mathbf{1}_{\mathcal{V}_i}\|_*^2] \leq G^2$ (Assumption 4) together with Jensen's inequality proves (13). Multiplying by $w_{t+1,i}$ and summing over i proves (14). □

Lemma 8 (Lemma 13 of (Zhang et al., 2025)). Suppose that f_t is monotone submodular, $\mathbf{x}_{t,i} \in \mathcal{C}$, and $a_{t,i}$ is the action selected by agent i at time t. Conditional on the policy estimates, agents sample independently within their own blocks, with probabilities proportional to $[\mathbf{x}_{t,i}]_a$ for $a \in \mathcal{V}_i$ (an action is sampled uniformly if the block mass is zero). This is the rounding rule used here. Define

$$\mathcal{A}_t = \bigcup_{i \in \mathcal{N}} \{a_{t,i}\}.$$

Then,

$$\mathbb{E}[f_t(\mathcal{A}_t)] \geq \mathbb{E}\left[F_t\left(\sum_{i \in \mathcal{N}} \mathbf{x}_{t,i} \odot \mathbf{1}_{\mathcal{V}_i}\right)\right].$$

Lemma 9 (Corollary 2 of (Nedić & Olshevsky, 2015); see also (Lee et al., 2017)). Under the communication model of Section 2.1 with $N \geq 2$, there is a stochastic vector $\boldsymbol{\phi}(t)$ such that, for every $i, j \in \mathcal{N}$ and $1 \leq s \leq t$,

$$|[\mathbf{A}(t:s)]_{ij} - \phi_i(t)| \leq \beta\lambda^{t-s}, \qquad \mathbf{A}(t:s) = \mathbf{A}(t)\mathbf{A}(t-1)\cdots\mathbf{A}(s),$$

where

$$\beta = 4, \qquad \lambda = (1 - N^{-NB})^{1/B} \in (0,1).$$

With $\mathbf{w}_1 = \mathbf{1}_N$ and $\mathbf{w}_{t+1} = \mathbf{A}(t)\mathbf{w}_t$,

$$\gamma := \inf_{t \geq 1} \min_{i \in \mathcal{N}} w_{t,i} \geq N^{-NB} > 0.$$

Column stochasticity gives $\sum_i w_{t,i} = N$, and hence $0 < \gamma \leq 1$.

B Proof of Theorem 1

We first establish feasibility, consensus, and weighted-potential bounds. Throughout this section, the conditions of Theorem 1 hold, and $\mathbf{x}_t^* = \mathbf{1}_{\mathcal{A}_t^*} \in \mathcal{C}$ denotes the indicator vector of a fixed optimal action set at round t. In particular, $F_t(\mathbf{x}_t^*) = f_t(\mathcal{A}_t^*)$. Set $\eta_0 = 0$ and interpret all empty sums as zero.

Lemma 10. Consider the constraint set

$$\mathcal{C} = \left\{ \mathbf{x} \in [0,1]^{|\mathcal{V}|} : \sum_{a \in \mathcal{V}_k} x_a \le 1,\ \forall k \in \mathcal{N} \right\}.$$

Under Assumption 1 and the communication model of Section 2.1 with $N \ge 2$, the policy estimates and aggregated policy estimates generated by POSMA satisfy

$$\mathbf{x}_{t,i} \in \mathcal{C}, \qquad \mathbf{y}_{t+1,i} \in \mathcal{C}, \qquad \forall t \in \{1, \dots, T\},\ i \in \mathcal{N}. \tag{15}$$

Proof. We prove the result by induction on t. The positive lower bound on the Push–Sum weights in Lemma 9 ensures that $w_{t,i} > 0$, so the normalization steps below are well defined.

Base case ($t = 1$). We first verify that $\mathbf{x}_{1,i} \in \mathcal{C}$. By the initialization of POSMA, for every $i \in \mathcal{N}$,

$$[\mathbf{x}_{1,i}]_a = \begin{cases} 1/n_i, & a \in \mathcal{V}_i, \\ 0, & a \notin \mathcal{V}_i, \end{cases} \qquad n_i = |\mathcal{V}_i|.$$

Since the local action sets are pairwise disjoint, for every $k \in \mathcal{N}$,

$$\sum_{a \in \mathcal{V}_k} [\mathbf{x}_{1,i}]_a = \begin{cases} \displaystyle\sum_{a \in \mathcal{V}_i} \frac{1}{n_i} = 1, & k = i, \\ 0, & k \ne i. \end{cases}$$

All coordinates of $\mathbf{x}_{1,i}$ are nonnegative, and every block sum is at most one. Thus, $\mathbf{x}_{1,i} \in \mathcal{C}$ for every $i \in \mathcal{N}$.

Next, we verify that $\mathbf{y}_{2,i} \in \mathcal{C}$. The initial weights and weighted policy estimates satisfy

$$w_{1,j} = 1, \qquad \mathbf{q}_{1,j} = \mathbf{x}_{1,j}, \qquad \forall j \in \mathcal{N}. \tag{16}$$

The first Push–Sum step gives

$$\begin{aligned} \mathbf{s}_{2,i} &= \sum_{j=1}^{N} [\mathbf{A}(1)]_{ij} \mathbf{q}_{1,j}, \\ w_{2,i} &= \sum_{j=1}^{N} [\mathbf{A}(1)]_{ij} w_{1,j}, \\ \mathbf{y}_{2,i} &= \frac{\mathbf{s}_{2,i}}{w_{2,i}}. \end{aligned} \tag{17}$$

Substituting (16) into (17), for any $k \in \mathcal{N}$, we obtain

$$\begin{aligned} \sum_{a \in \mathcal{V}_k} [\mathbf{y}_{2,i}]_a &= \sum_{a \in \mathcal{V}_k} \frac{\sum_{j=1}^{N} [\mathbf{A}(1)]_{ij} [\mathbf{x}_{1,j}]_a}{\sum_{j=1}^{N} [\mathbf{A}(1)]_{ij}} \\ &= \frac{\sum_{j=1}^{N} [\mathbf{A}(1)]_{ij} \left(\sum_{a \in \mathcal{V}_k} [\mathbf{x}_{1,j}]_a \right)}{\sum_{j=1}^{N} [\mathbf{A}(1)]_{ij}}. \end{aligned}$$

Since $\mathbf{x}_{1,j} \in \mathcal{C}$ and $[\mathbf{A}(1)]_{ij} \ge 0$, it follows that

$$\sum_{a \in \mathcal{V}_k} [\mathbf{y}_{2,i}]_a \le \frac{\sum_{j=1}^{N} [\mathbf{A}(1)]_{ij} \cdot 1}{\sum_{j=1}^{N} [\mathbf{A}(1)]_{ij}} = 1. \tag{18}$$

Moreover, every coordinate of $\mathbf{y}_{2,i}$ is nonnegative. Together with (18), this also implies that each coordinate is at most one. Therefore, $\mathbf{y}_{2,i} \in \mathcal{C}$ for every $i \in \mathcal{N}$.

Inductive step. Suppose that, for some $1 \le t < T$,

$$\mathbf{x}_{t,i} \in \mathcal{C}, \qquad \mathbf{y}_{t+1,i} \in \mathcal{C}, \qquad \forall i \in \mathcal{N}.$$

We first prove that $\mathbf{x}_{t+1,i} \in \mathcal{C}$.

For the local coordinates indexed by $\mathcal{V}_i$, POSMA solves the mirror ascent subproblem over

$$\mathcal{C}_i = \left\{ \mathbf{b} \in [0,1]^{n_i} : \sum_{k=1}^{n_i} b_k \le 1 \right\}.$$

Consequently, the local update satisfies

$$[\mathbf{x}_{t+1,i}]_{\mathcal{V}_i} \in \mathcal{C}_i, \qquad \sum_{a \in \mathcal{V}_i} [\mathbf{x}_{t+1,i}]_a \le 1. \tag{19}$$

For any nonlocal block $\mathcal{V}_k$ with $k \neq i$, POSMA copies the corresponding consensus coordinates.

$$[\mathbf{x}_{t+1,i}]_a = [\mathbf{y}_{t+1,i}]_a, \qquad \forall a \in \mathcal{V}_k.$$

By the induction hypothesis, $\mathbf{y}_{t+1,i} \in \mathcal{C}$, and hence

$$\sum_{a \in \mathcal{V}_k} [\mathbf{x}_{t+1,i}]_a = \sum_{a \in \mathcal{V}_k} [\mathbf{y}_{t+1,i}]_a \le 1, \qquad \forall k \neq i. \tag{20}$$

The local and nonlocal coordinates all belong to $[0,1]$. Combining (19) and (20), we conclude that $\mathbf{x}_{t+1,i} \in \mathcal{C}$ for every $i \in \mathcal{N}$.

It remains to prove that $\mathbf{y}_{t+2,i} \in \mathcal{C}$. At the end of round t, the weighted policy estimate is updated as

$$\mathbf{q}_{t+1,j} = w_{t+1,j} \mathbf{x}_{t+1,j}, \qquad \forall j \in \mathcal{N}. \tag{21}$$

The communication step in round $t+1$ then gives

$$\begin{aligned}
\mathbf{s}_{t+2,i} &= \sum_{j=1}^{N} [\mathbf{A}(t+1)]_{ij} \mathbf{q}_{t+1,j}, \\
w_{t+2,i} &= \sum_{j=1}^{N} [\mathbf{A}(t+1)]_{ij} w_{t+1,j}, \\
\mathbf{y}_{t+2,i} &= \frac{\mathbf{s}_{t+2,i}}{w_{t+2,i}}.
\end{aligned}$$

Using (21), for every $k \in \mathcal{N}$, we have

$$\begin{aligned}
\sum_{a \in \mathcal{V}_k} [\mathbf{y}_{t+2,i}]_a &= \frac{\sum_{a \in \mathcal{V}_k} \sum_{j=1}^{N} [\mathbf{A}(t+1)]_{ij} w_{t+1,j} [\mathbf{x}_{t+1,j}]_a}{\sum_{j=1}^{N} [\mathbf{A}(t+1)]_{ij} w_{t+1,j}} \\
&= \frac{\sum_{j=1}^{N} [\mathbf{A}(t+1)]_{ij} w_{t+1,j} \left(\sum_{a \in \mathcal{V}_k} [\mathbf{x}_{t+1,j}]_a \right)}{\sum_{j=1}^{N} [\mathbf{A}(t+1)]_{ij} w_{t+1,j}}.
\end{aligned}$$

We have already established that $\mathbf{x}_{t+1,j} \in \mathcal{C}$. Since $w_{t+1,j} > 0$ and $[\mathbf{A}(t+1)]_{ij} \ge 0$, it follows that

$$\sum_{a \in \mathcal{V}_k} [\mathbf{y}_{t+2,i}]_a \le \frac{\sum_{j=1}^{N} [\mathbf{A}(t+1)]_{ij} w_{t+1,j} \cdot 1}{\sum_{j=1}^{N} [\mathbf{A}(t+1)]_{ij} w_{t+1,j}} = 1. \tag{22}$$

Every coordinate of $\mathbf{y}_{t+2,i}$ is nonnegative, so (22) also ensures that each coordinate is at most one. Thus, $\mathbf{y}_{t+2,i} \in \mathcal{C}$ for every $i \in \mathcal{N}$.

The induction is complete, and (15) holds for all $t \in \{1, \dots, T\}$. The same local feasibility argument at $t = T$ also gives $\mathbf{x}_{T+1,i} \in \mathcal{C}$. □

Lemma 11. Suppose that Assumptions 1 and 4 hold, POSMA follows the communication model of Section 2.1 with $N \geq 2$, and the step sizes are deterministic and positive. Define the average weighted policy estimate by

$$\overline{\mathbf{q}}_t = \frac{1}{N}\sum_{j=1}^{N} \mathbf{q}_{t,j} = \frac{1}{N}\sum_{j=1}^{N} w_{t,j}\mathbf{x}_{t,j}. \tag{23}$$

Let

$$\gamma = \inf_{t\geq 1}\min_{i\in\mathcal{N}} w_{t,i} > 0, \qquad D = \sup_{\mathbf{x},\mathbf{y}\in\mathcal{C}} \|\mathbf{x}-\mathbf{y}\|, \qquad R \geq \max\left\{\sup_{\mathbf{x}\in\mathcal{C}}\|\mathbf{x}\|,\ D,\ \sqrt{KD}\right\}.$$

The positivity of γ follows from Lemma 9. For the uniform outgoing-weight model in Lemma 9, the communication constants can be chosen as

$$\beta = 4, \qquad \lambda = \left(1 - \frac{1}{N^{NB}}\right)^{1/B} \in (0,1).$$

Then, for every $i \in \mathcal{N}$ and $t \in \{1,\dots,T\}$,

$$\mathbb{E}[\|\mathbf{y}_{t+1,i} - \overline{\mathbf{q}}_t\|] \leq \frac{2\beta NR}{\gamma}\lambda^{t-1} + \frac{2\beta NG}{\gamma}\sum_{s=1}^{t-1}\lambda^{t-s-1}\eta_s. \tag{24}$$

The sum is interpreted as zero when $t = 1$.

Proof. Define the local update residual by

$$\mathbf{r}_{t+1,i} = \mathbf{x}_{t+1,i} - \mathbf{y}_{t+1,i}.$$

Thus, $\mathbf{x}_{t+1,i} = \mathbf{y}_{t+1,i} + \mathbf{r}_{t+1,i}$. Multiplying both sides by $w_{t+1,i}$ gives

$$\begin{aligned}\mathbf{q}_{t+1,i} &= w_{t+1,i}\mathbf{x}_{t+1,i} \\ &= w_{t+1,i}\mathbf{y}_{t+1,i} + w_{t+1,i}\mathbf{r}_{t+1,i}.\end{aligned} \tag{25}$$

By the Push–Sum update,

$$w_{t+1,i}\mathbf{y}_{t+1,i} = \mathbf{s}_{t+1,i} = \sum_{j=1}^{N}[\mathbf{A}(t)]_{ij}\mathbf{q}_{t,j}. \tag{26}$$

Substituting (26) into (25), we obtain

$$\mathbf{q}_{t+1,i} = \sum_{j=1}^{N}[\mathbf{A}(t)]_{ij}\mathbf{q}_{t,j} + w_{t+1,i}\mathbf{r}_{t+1,i}. \tag{27}$$

Let $n = |\mathcal{V}|$, and stack the vectors as rows.

$$\mathbf{Q}_t = \begin{bmatrix}\mathbf{q}_{t,1}^\top \\ \vdots \\ \mathbf{q}_{t,N}^\top\end{bmatrix} \in \mathbb{R}^{N\times n}, \qquad \mathbf{R}_{t+1} = \begin{bmatrix}\mathbf{r}_{t+1,1}^\top \\ \vdots \\ \mathbf{r}_{t+1,N}^\top\end{bmatrix} \in \mathbb{R}^{N\times n},$$

$$\mathbf{S}_{t+1} = \begin{bmatrix}\mathbf{s}_{t+1,1}^\top \\ \vdots \\ \mathbf{s}_{t+1,N}^\top\end{bmatrix} \in \mathbb{R}^{N\times n}, \qquad \mathbf{w}_t = (w_{t,1},\dots,w_{t,N})^\top.$$

Equation (27) can then be written as

$$\mathbf{Q}_{t+1} = \mathbf{A}(t)\mathbf{Q}_t + \mathrm{diag}(\mathbf{w}_{t+1})\mathbf{R}_{t+1}. \tag{28}$$

For integers $u \geq v \geq 1$, define

$$\mathbf{A}(u:v) = \mathbf{A}(u)\mathbf{A}(u-1)\cdots\mathbf{A}(v),$$

with the convention $\mathbf{A}(v-1:v)=\mathbf{I}_N$. Recursively expanding (28) from the initialization yields

$$\begin{aligned}\mathbf{Q}_t &= \mathbf{A}(t-1:1)\mathbf{Q}_1 \\ &\quad + \sum_{s=1}^{t-1} \mathbf{A}(t-1:s+1)\,\mathrm{diag}(\mathbf{w}_{s+1})\mathbf{R}_{s+1}.\end{aligned} \tag{29}$$

Since $\mathbf{S}_{t+1}=\mathbf{A}(t)\mathbf{Q}_t$, multiplying (29) by $\mathbf{A}(t)$ gives

$$\begin{aligned}\mathbf{S}_{t+1} &= \mathbf{A}(t:1)\mathbf{Q}_1 \\ &\quad + \sum_{s=1}^{t-1} \mathbf{A}(t:s+1)\,\mathrm{diag}(\mathbf{w}_{s+1})\mathbf{R}_{s+1}.\end{aligned}$$

Taking the ith row and transposing, we obtain

$$\begin{aligned}\mathbf{s}_{t+1,i} &= \sum_{j=1}^{N} [\mathbf{A}(t:1)]_{ij}\mathbf{q}_{1,j} \\ &\quad + \sum_{s=1}^{t-1}\sum_{j=1}^{N} [\mathbf{A}(t:s+1)]_{ij} w_{s+1,j}\mathbf{r}_{s+1,j}.\end{aligned} \tag{30}$$

We next expand the average weighted policy estimate. By (23) and (29),

$$\begin{aligned}\overline{\mathbf{q}}_t^{\top} &= \frac{1}{N}\mathbf{1}_N^{\top}\mathbf{Q}_t \\ &= \frac{1}{N}\mathbf{1}_N^{\top}\mathbf{A}(t-1:1)\mathbf{Q}_1 \\ &\quad + \frac{1}{N}\sum_{s=1}^{t-1}\mathbf{1}_N^{\top}\mathbf{A}(t-1:s+1)\,\mathrm{diag}(\mathbf{w}_{s+1})\mathbf{R}_{s+1}.\end{aligned} \tag{31}$$

The column stochasticity of the communication matrices implies

$$\mathbf{1}_N^{\top}\mathbf{A}(\tau)=\mathbf{1}_N^{\top}, \qquad \mathbf{1}_N^{\top}\mathbf{A}(u:v)=\mathbf{1}_N^{\top}.$$

Consequently, (31) reduces to

$$\overline{\mathbf{q}}_t = \frac{1}{N}\sum_{j=1}^{N}\mathbf{q}_{1,j} + \frac{1}{N}\sum_{s=1}^{t-1}\sum_{j=1}^{N} w_{s+1,j}\mathbf{r}_{s+1,j}. \tag{32}$$

Using $\mathbf{y}_{t+1,i}=\mathbf{s}_{t+1,i}/w_{t+1,i}$, we subtract (32) from the normalized version of (30) to obtain

$$\begin{aligned}\mathbf{y}_{t+1,i}-\overline{\mathbf{q}}_t &= \frac{1}{w_{t+1,i}}\sum_{j=1}^{N}[\mathbf{A}(t:1)]_{ij}\mathbf{q}_{1,j} \\ &\quad + \frac{1}{w_{t+1,i}}\sum_{s=1}^{t-1}\sum_{j=1}^{N}[\mathbf{A}(t:s+1)]_{ij}w_{s+1,j}\mathbf{r}_{s+1,j} \\ &\quad - \frac{1}{N}\sum_{j=1}^{N}\mathbf{q}_{1,j} - \frac{1}{N}\sum_{s=1}^{t-1}\sum_{j=1}^{N} w_{s+1,j}\mathbf{r}_{s+1,j}.\end{aligned}$$

Grouping the initial terms and the residual terms gives

$$\begin{aligned}\mathbf{y}_{t+1,i}-\overline{\mathbf{q}}_t &= \sum_{j=1}^{N}\left(\frac{[\mathbf{A}(t:1)]_{ij}}{w_{t+1,i}} - \frac{1}{N}\right)\mathbf{q}_{1,j} \\ &\quad + \sum_{s=1}^{t-1}\sum_{j=1}^{N} w_{s+1,j}\left(\frac{[\mathbf{A}(t:s+1)]_{ij}}{w_{t+1,i}} - \frac{1}{N}\right)\mathbf{r}_{s+1,j}.\end{aligned} \tag{33}$$

We now bound the coefficients in (33). By Lemma 9, there exists a stochastic vector $\boldsymbol{\phi}(t)$ such that

$$|[\mathbf{A}(t:u)]_{ij} - \phi_i(t)| \le \beta\lambda^{t-u}, \qquad 1 \le u \le t.$$

In particular,

$$\begin{aligned} |[\mathbf{A}(t:s+1)]_{ij} - \phi_i(t)| &\le \beta\lambda^{t-s-1}, \quad 1 \le s \le t-1, \\ |[\mathbf{A}(t:1)]_{ik} - \phi_i(t)| &\le \beta\lambda^{t-1}. \end{aligned} \tag{34}$$

Since $\mathbf{w}_1 = \mathbf{1}_N$, the weight recursion yields

$$\mathbf{w}_{t+1} = \mathbf{A}(t:1)\mathbf{1}_N, \qquad w_{t+1,i} = \sum_{k=1}^{N} [\mathbf{A}(t:1)]_{ik} \ge \gamma. \tag{35}$$

For a residual term indexed by s, we have

$$\left| \frac{[\mathbf{A}(t:s+1)]_{ij}}{w_{t+1,i}} - \frac{1}{N} \right| = \frac{\left| [\mathbf{A}(t:s+1)]_{ij} - \frac{1}{N} w_{t+1,i} \right|}{w_{t+1,i}}.$$

Substituting (35) and adding and subtracting $\phi_i(t)$ in the numerator, we obtain

$$\begin{aligned} &[\mathbf{A}(t:s+1)]_{ij} - \frac{1}{N} w_{t+1,i} \\ &\quad = [\mathbf{A}(t:s+1)]_{ij} - \frac{1}{N} \sum_{k=1}^{N} [\mathbf{A}(t:1)]_{ik} \\ &\quad = [\mathbf{A}(t:s+1)]_{ij} - \phi_i(t) + \frac{1}{N} \sum_{k=1}^{N} \left(\phi_i(t) - [\mathbf{A}(t:1)]_{ik} \right). \end{aligned}$$

The triangle inequality and $w_{t+1,i} \ge \gamma$ therefore imply

$$\begin{aligned} \left| \frac{[\mathbf{A}(t:s+1)]_{ij}}{w_{t+1,i}} - \frac{1}{N} \right| &\le \frac{1}{\gamma} \left| [\mathbf{A}(t:s+1)]_{ij} - \phi_i(t) \right| \\ &\quad + \frac{1}{N\gamma} \sum_{k=1}^{N} \left| \phi_i(t) - [\mathbf{A}(t:1)]_{ik} \right|. \end{aligned}$$

Applying (34) gives

$$\begin{aligned} \left| \frac{[\mathbf{A}(t:s+1)]_{ij}}{w_{t+1,i}} - \frac{1}{N} \right| &\le \frac{\beta}{\gamma} \left(\lambda^{t-s-1} + \lambda^{t-1} \right) \\ &\le \frac{2\beta}{\gamma} \lambda^{t-s-1}. \end{aligned} \tag{36}$$

The second inequality follows from $0 < \lambda < 1$ and $t - 1 \ge t - s - 1$.

The initial coefficient is bounded in the same way:

$$\begin{aligned} \left| \frac{[\mathbf{A}(t:1)]_{ij}}{w_{t+1,i}} - \frac{1}{N} \right| &\le \frac{1}{\gamma} \left| [\mathbf{A}(t:1)]_{ij} - \phi_i(t) \right| \\ &\quad + \frac{1}{N\gamma} \sum_{k=1}^{N} \left| \phi_i(t) - [\mathbf{A}(t:1)]_{ik} \right| \\ &\le \frac{2\beta}{\gamma} \lambda^{t-1}. \end{aligned} \tag{37}$$

Taking norms in (33) and applying the triangle inequality, we obtain

$$\begin{aligned} \|\mathbf{y}_{t+1,i} - \overline{\mathbf{q}}_t\| \le &\sum_{j=1}^{N} \left| \frac{[\mathbf{A}(t:1)]_{ij}}{w_{t+1,i}} - \frac{1}{N} \right| \|\mathbf{q}_{1,j}\| \\ &+ \sum_{s=1}^{t-1} \sum_{j=1}^{N} w_{s+1,j} \left| \frac{[\mathbf{A}(t:s+1)]_{ij}}{w_{t+1,i}} - \frac{1}{N} \right| \|\mathbf{r}_{s+1,j}\|. \end{aligned}$$

Substituting (36) and (37) into this inequality gives

$$\|\mathbf{y}_{t+1,i} - \overline{\mathbf{q}}_t\| \le \frac{2\beta}{\gamma}\lambda^{t-1}\sum_{j=1}^{N}\|\mathbf{q}_{1,j}\| + \frac{2\beta}{\gamma}\sum_{s=1}^{t-1}\lambda^{t-s-1}\sum_{j=1}^{N} w_{s+1,j}\|\mathbf{r}_{s+1,j}\|. \tag{38}$$

For the prescribed communication sequence, the weights are deterministic. Taking expectations over the algorithmic sampling in (38) therefore yields

$$\mathbb{E}[\|\mathbf{y}_{t+1,i} - \overline{\mathbf{q}}_t\|] \le \frac{2\beta}{\gamma}\lambda^{t-1}\sum_{j=1}^{N}\|\mathbf{q}_{1,j}\| + \frac{2\beta}{\gamma}\sum_{s=1}^{t-1}\lambda^{t-s-1}\sum_{j=1}^{N} w_{s+1,j}\mathbb{E}[\|\mathbf{r}_{s+1,j}\|]\,. \tag{39}$$

By Lemma 7, the *weighted* residual satisfies

$$w_{s+1,j}\,\mathbb{E}[\|\mathbf{r}_{s+1,j}\|] = w_{s+1,j}\,\mathbb{E}[\|\mathbf{x}_{s+1,j} - \mathbf{y}_{s+1,j}\|] \le G\eta_s, \tag{40}$$

since $w_{s+1,j}$ cancels against the $1/w_{s+1,j}$ factor in (13). Moreover, column stochasticity preserves the total Push–Sum weight.

$$\begin{aligned}\sum_{j=1}^{N} w_{s+1,j} &= \mathbf{1}_N^\top \mathbf{A}(s)\mathbf{w}_s \\ &= \mathbf{1}_N^\top \mathbf{w}_s \\ &= \mathbf{1}_N^\top \mathbf{w}_1 = N.\end{aligned} \tag{41}$$

Combining (40) and (41), we have

$$\sum_{j=1}^{N} w_{s+1,j}\mathbb{E}[\|\mathbf{r}_{s+1,j}\|] \le NG\eta_s.$$

Consequently, (39) becomes

$$\mathbb{E}[\|\mathbf{y}_{t+1,i} - \overline{\mathbf{q}}_t\|] \le \frac{2\beta}{\gamma}\lambda^{t-1}\sum_{j=1}^{N}\|\mathbf{q}_{1,j}\| + \frac{2\beta NG}{\gamma}\sum_{s=1}^{t-1}\lambda^{t-s-1}\eta_s. \tag{42}$$

Finally, by the initialization, $w_{1,j} = 1$ and $\mathbf{q}_{1,j} = \mathbf{x}_{1,j}$. Since $\mathbf{x}_{1,j} \in \mathcal{C}$ by Lemma 10, the definition of R implies

$$\|\mathbf{q}_{1,j}\| = \|\mathbf{x}_{1,j}\| \le R, \qquad \sum_{j=1}^{N}\|\mathbf{q}_{1,j}\| \le NR. \tag{43}$$

Substituting (43) into (42), we conclude that

$$\mathbb{E}[\|\mathbf{y}_{t+1,i} - \overline{\mathbf{q}}_t\|] \le \frac{2\beta NR}{\gamma}\lambda^{t-1} + \frac{2\beta NG}{\gamma}\sum_{s=1}^{t-1}\lambda^{t-s-1}\eta_s.$$

This proves (24). □

Lemma 12. Suppose that the conditions of Theorem 1 hold. Let $\mathbf{x}_t^* = \mathbf{1}_{\mathcal{A}_t^*}$, and let $\overline{\mathbf{q}}_t$ be defined in Lemma 11. Assumption 4 gives, for every $i \in \mathcal{N}$,

$$\left\|\left(\nabla\widehat{F}_t(\mathbf{u}) - \nabla\widehat{F}_t(\mathbf{v})\right)\odot\mathbf{1}_{\mathcal{V}_i}\right\|_* \le L_g\|\mathbf{u}-\mathbf{v}\|, \qquad \mathbf{u},\mathbf{v}\in\mathcal{C}. \tag{44}$$

Then the following bound holds.

$$\begin{aligned}
&\left(\frac{1-e^{-\bar{c}}}{\bar{c}}\right)\sum_{t=1}^{T}F_t(\mathbf{x}_t^*)-\sum_{t=1}^{T}\mathbb{E}\left[F_t(\overline{\mathbf{q}}_t)\right]\\
&\le\sum_{t=1}^{T}\sum_{i\in\mathcal{N}}\frac{w_{t+1,i}}{\eta_t}\mathbb{E}[\mathcal{D}_\psi(\mathbf{x}_t^*,\mathbf{y}_{t+1,i})-\mathcal{D}_\psi(\mathbf{x}_t^*,\mathbf{x}_{t+1,i})]\\
&\quad+\frac{2\beta N(L_gD+G)}{\gamma}\sum_{t=1}^{T}\left(NR\lambda^{t-2}+NG\sum_{s=1}^{t-1}\lambda^{t-s-2}\eta_s\right)\\
&\quad+\frac{NG(4\gamma L_gD+G)}{2\gamma}\sum_{t=1}^{T}\eta_t.
\end{aligned}\tag{45}$$

The weighted potential term matches the update (12). The bounds below retain the residual contribution of that update.

Proof. By Lemma 10, $\mathbf{x}_{t,i}\in\mathcal{C}$. Moreover, Lemma 11 gives $w_{t,i}>0$ and $\sum_{i=1}^{N}w_{t,i}=N$. Since $\mathcal{C}$ is convex,

$$\overline{\mathbf{q}}_t=\sum_{i=1}^{N}\frac{w_{t,i}}{N}\mathbf{x}_{t,i}\in\mathcal{C},\qquad\|\mathbf{x}_t^*-\overline{\mathbf{q}}_t\|\le D.\tag{46}$$

Applying the surrogate-function inequality (Theorem 2 of (Zhang et al., 2025)) at $\overline{\mathbf{q}}_t$ yields

$$\left(\frac{1-e^{-\bar{c}}}{\bar{c}}\right)F_t(\mathbf{x}_t^*)-F_t(\overline{\mathbf{q}}_t)\le\left\langle\nabla\widehat{F}_t(\overline{\mathbf{q}}_t),\mathbf{x}_t^*-\overline{\mathbf{q}}_t\right\rangle.\tag{47}$$

Split this inner product using $\sum_i\mathbf{1}_{\mathcal{V}_i}=\mathbf{1}_{|\mathcal{V}|}$ and $\mathbf{y}_{t+1,i}$ into the three terms

$$\begin{aligned}
&\left\langle\nabla\widehat{F}_t(\overline{\mathbf{q}}_t),\mathbf{x}_t^*-\overline{\mathbf{q}}_t\right\rangle\\
&=\underbrace{\left\langle\nabla\widehat{F}_t(\overline{\mathbf{q}}_t)-\sum_{i\in\mathcal{N}}\left(\nabla\widehat{F}_t(\mathbf{x}_{t,i})\odot\mathbf{1}_{\mathcal{V}_i}\right),\mathbf{x}_t^*-\overline{\mathbf{q}}_t\right\rangle}_{①}\\
&\quad+\underbrace{\sum_{i\in\mathcal{N}}\left\langle\nabla\widehat{F}_t(\mathbf{x}_{t,i})\odot\mathbf{1}_{\mathcal{V}_i},\mathbf{x}_t^*-\mathbf{y}_{t+1,i}\right\rangle}_{②}\\
&\quad+\underbrace{\sum_{i\in\mathcal{N}}\left\langle\nabla\widehat{F}_t(\mathbf{x}_{t,i})\odot\mathbf{1}_{\mathcal{V}_i},\mathbf{y}_{t+1,i}-\overline{\mathbf{q}}_t\right\rangle}_{③}.
\end{aligned}\tag{48}$$

To bound the first term, set $X_t:=\sum_i\mathbb{E}\|\mathbf{x}_{t,i}-\overline{\mathbf{q}}_t\|$. For $t\ge2$, the mass recursion in Lemma 11 gives

$$\mathbb{E}\|\overline{\mathbf{q}}_t-\overline{\mathbf{q}}_{t-1}\|\le\frac{1}{N}\sum_i w_{t,i}\mathbb{E}\|\mathbf{r}_{t,i}\|\le G\eta_{t-1}.$$

Inserting $\mathbf{y}_{t,i}$ and $\overline{\mathbf{q}}_{t-1}$ and using Lemmas 7 and 11, we obtain

$$X_t\le\frac{2\beta N}{\gamma}\left(NR\lambda^{t-2}+NG\sum_{s=1}^{t-2}\lambda^{t-s-2}\eta_s\right)+NG(1+\gamma^{-1})\eta_{t-1}.$$

Since $2\beta N/\lambda\ge1-\gamma$, the additional summand at $s=t-1$ absorbs $NG(\gamma^{-1}-1)\eta_{t-1}$. Therefore,

$$X_t\le\frac{2\beta N}{\gamma}\left(NR\lambda^{t-2}+NG\sum_{s=1}^{t-1}\lambda^{t-s-2}\eta_s\right)+2NG\eta_{t-1}.\tag{49}$$

Set $\eta_0 = 0$ and interpret empty sums as zero. For $t = 1$, $X_1 \le ND \le NR$, so (49) also holds. Hölder's inequality and (44) give $\mathbb{E}[①] \le L_g D X_t$. For the third term, unbiasedness and the second-moment bound in Assumption 4 imply $\|\nabla \widehat{F}_t(\mathbf{x}_{t,i}) \odot \mathbf{1}_{\mathcal{V}_i}\|_* \le G$. Applying Lemma 11 and using $0 < \lambda < 1$ yields the following two estimates.

$$\mathbb{E}[①] \le \frac{2\beta N L_g D}{\gamma}\left(NR\lambda^{t-2} + NG\sum_{s=1}^{t-1}\lambda^{t-s-2}\eta_s\right) + 2NL_g DG\eta_{t-1}, \tag{50}$$

$$\mathbb{E}[③] \le \frac{2\beta NG}{\gamma}\left(NR\lambda^{t-2} + NG\sum_{s=1}^{t-1}\lambda^{t-s-2}\eta_s\right) \tag{51}$$

Bounding ②. By the conditional unbiasedness in Assumption 4,

$$\mathbb{E}[\mathbf{g}_{t,i} \mid \mathbf{x}_{t,i}] = \nabla \widehat{F}_t(\mathbf{x}_{t,i}) \odot \mathbf{1}_{\mathcal{V}_i}.$$

Since the fresh estimator and $\mathbf{y}_{t+1,i}$ are conditionally independent given $\mathbf{x}_{t,i}$, and $\mathbf{x}_t^*$ is fixed independently of the algorithmic sampling, the tower property gives

$$\begin{aligned}&\mathbb{E}[\langle \mathbf{g}_{t,i}, \mathbf{x}_t^* - \mathbf{y}_{t+1,i}\rangle \mid \mathbf{x}_{t,i}] \\ &= \langle \mathbb{E}[\mathbf{g}_{t,i} \mid \mathbf{x}_{t,i}], \mathbf{x}_t^* - \mathbb{E}[\mathbf{y}_{t+1,i} \mid \mathbf{x}_{t,i}]\rangle \\ &= \mathbb{E}\Big[\Big\langle \nabla \widehat{F}_t(\mathbf{x}_{t,i}) \odot \mathbf{1}_{\mathcal{V}_i}, \mathbf{x}_t^* - \mathbf{y}_{t+1,i}\Big\rangle \mid \mathbf{x}_{t,i}\Big].\end{aligned}$$

Taking expectations and summing over i therefore gives

$$\mathbb{E}[②] = \sum_{i\in\mathcal{N}} \mathbb{E}\Big[\Big\langle \widetilde{\nabla} F_t(\mathbf{x}_{t,i}) \odot \mathbf{1}_{\mathcal{V}_i}, \mathbf{x}_t^* - \mathbf{y}_{t+1,i}\Big\rangle\Big],$$

which we split at $\mathbf{x}_{t+1,i}$ as $\mathbb{E}[②] = \mathbb{E}[④] + \mathbb{E}[⑤]$ with $④ = \sum_i \langle \widetilde{\nabla} F_t(\mathbf{x}_{t,i}) \odot \mathbf{1}_{\mathcal{V}_i}, \mathbf{x}_t^* - \mathbf{x}_{t+1,i}\rangle$ and $⑤ = \sum_i \langle \widetilde{\nabla} F_t(\mathbf{x}_{t,i}) \odot \mathbf{1}_{\mathcal{V}_i}, \mathbf{x}_{t+1,i} - \mathbf{y}_{t+1,i}\rangle$.

By the update (12), Lemma 5 applies with step size $\eta_t / w_{t+1,i}$ and $\mathbf{h} = \widetilde{\nabla} F_t(\mathbf{x}_{t,i}) \odot \mathbf{1}_{\mathcal{V}_i}$, giving

$$\begin{aligned}&\Big\langle \widetilde{\nabla} F_t(\mathbf{x}_{t,i}) \odot \mathbf{1}_{\mathcal{V}_i}, \mathbf{x}_t^* - \mathbf{x}_{t+1,i}\Big\rangle \\ &\le \frac{w_{t+1,i}}{\eta_t}\Big(\mathcal{D}_\psi(\mathbf{x}_t^*, \mathbf{y}_{t+1,i}) - \mathcal{D}_\psi(\mathbf{x}_t^*, \mathbf{x}_{t+1,i}) - \mathcal{D}_\psi(\mathbf{x}_{t+1,i}, \mathbf{y}_{t+1,i})\Big).\end{aligned} \tag{52}$$

Since ψ is 1-strongly convex, $\mathcal{D}_\psi(\mathbf{x}_{t+1,i}, \mathbf{y}_{t+1,i}) \ge \frac{1}{2}\|\mathbf{x}_{t+1,i} - \mathbf{y}_{t+1,i}\|^2$, so

$$\begin{aligned}\mathbb{E}[④] \le &\sum_{i\in\mathcal{N}} \frac{w_{t+1,i}}{\eta_t}\mathbb{E}[\mathcal{D}_\psi(\mathbf{x}_t^*, \mathbf{y}_{t+1,i}) - \mathcal{D}_\psi(\mathbf{x}_t^*, \mathbf{x}_{t+1,i})] \\ &- \sum_{i\in\mathcal{N}} \frac{w_{t+1,i}}{2\eta_t}\mathbb{E}\big[\|\mathbf{x}_{t+1,i} - \mathbf{y}_{t+1,i}\|^2\big].\end{aligned} \tag{53}$$

For ⑤, Hölder's inequality and Young's inequality with parameter $a = \eta_t / w_{t+1,i}$ give

$$\Big\langle \widetilde{\nabla} F_t(\mathbf{x}_{t,i}) \odot \mathbf{1}_{\mathcal{V}_i}, \mathbf{x}_{t+1,i} - \mathbf{y}_{t+1,i}\Big\rangle \le \frac{w_{t+1,i}}{2\eta_t}\|\mathbf{x}_{t+1,i} - \mathbf{y}_{t+1,i}\|^2 + \frac{\eta_t}{2w_{t+1,i}}\|\widetilde{\nabla} F_t(\mathbf{x}_{t,i}) \odot \mathbf{1}_{\mathcal{V}_i}\|_*^2.$$

Using $\mathbb{E}[\|\widetilde{\nabla} F_t(\mathbf{x}_{t,i}) \odot \mathbf{1}_{\mathcal{V}_i}\|_*^2] \le G^2$ and $w_{t+1,i} \ge \gamma$ (Lemma 9),

$$\mathbb{E}[⑤] \le \sum_{i\in\mathcal{N}} \frac{w_{t+1,i}}{2\eta_t}\mathbb{E}\big[\|\mathbf{x}_{t+1,i} - \mathbf{y}_{t+1,i}\|^2\big] + \frac{NG^2\eta_t}{2\gamma}. \tag{54}$$

Adding (53) and (54) cancels the squared-residual terms exactly, so

$$\mathbb{E}[②] \le \sum_{i\in\mathcal{N}} \frac{w_{t+1,i}}{\eta_t}\mathbb{E}[\mathcal{D}_\psi(\mathbf{x}_t^*, \mathbf{y}_{t+1,i}) - \mathcal{D}_\psi(\mathbf{x}_t^*, \mathbf{x}_{t+1,i})] + \frac{NG^2\eta_t}{2\gamma}. \tag{55}$$

The noise contribution is therefore $NG^2\eta_t/(2\gamma)$.

Combining the three terms. Taking expectations in (47) and (48), and applying (50), (51), and (55), we obtain

$$\begin{aligned}
&\left(\frac{1-e^{-\bar{c}}}{\bar{c}}\right) F_t(\mathbf{x}_t^*) - \mathbb{E}[F_t(\overline{\mathbf{q}}_t)] \\
&\quad \le \sum_{i\in\mathcal{N}} \frac{w_{t+1,i}}{\eta_t} \mathbb{E}[\mathcal{D}_\psi(\mathbf{x}_t^*, \mathbf{y}_{t+1,i}) - \mathcal{D}_\psi(\mathbf{x}_t^*, \mathbf{x}_{t+1,i})] \\
&\qquad + \frac{2\beta N(L_g D + G)}{\gamma} \left(NR\lambda^{t-2} + NG \sum_{s=1}^{t-1} \lambda^{t-s-2}\eta_s \right) \\
&\qquad + 2NL_g DG\eta_{t-1} + \frac{NG^2\eta_t}{2\gamma}.
\end{aligned} \tag{56}$$

Using $\eta_0 = 0$ and $\sum_{t=1}^T \eta_{t-1} \le \sum_{t=1}^T \eta_t$,

$$\begin{aligned}
2NL_g DG \sum_{t=1}^T \eta_{t-1} + \frac{NG^2}{2\gamma} \sum_{t=1}^T \eta_t &\le \left(2NL_g DG + \frac{NG^2}{2\gamma}\right) \sum_{t=1}^T \eta_t \\
&= \frac{NG(4\gamma L_g D + G)}{2\gamma} \sum_{t=1}^T \eta_t.
\end{aligned} \tag{57}$$

Summing (56) over $t = 1, \dots, T$ and using (57) proves (45). □

Lemma 13. Under the conditions of Theorem 1,

$$\sum_{t=1}^T \sum_{i=1}^N \frac{w_{t+1,i}}{\eta_t} \mathbb{E}\left[\mathcal{D}_\psi(\mathbf{x}_t^*, \mathbf{y}_{t+1,i}) - \mathcal{D}_\psi(\mathbf{x}_t^*, \mathbf{x}_{t+1,i})\right] \le \frac{NR^2}{\eta_T} + KN \sum_{t=1}^{T-1} \frac{\|\mathbf{x}_{t+1}^* - \mathbf{x}_t^*\|}{\eta_t}. \tag{58}$$

The weighted form on the left-hand side is precisely the potential term in Lemma 12.

Proof. Define the weighted potential

$$\Phi_t(\mathbf{u}) := \sum_{i=1}^N w_{t,i}\, \mathcal{D}_\psi(\mathbf{u}, \mathbf{x}_{t,i}), \qquad \mathbf{u} \in \mathcal{C}. \tag{59}$$

By Assumption 3 and the definition of R, $\mathcal{D}_\psi(\mathbf{u}, \mathbf{v}) = |\mathcal{D}_\psi(\mathbf{u}, \mathbf{v}) - \mathcal{D}_\psi(\mathbf{v}, \mathbf{v})| \le K\|\mathbf{u} - \mathbf{v}\| \le KD \le R^2$ for $\mathbf{u}, \mathbf{v} \in \mathcal{C}$. Since $\mathcal{D}_\psi(\mathbf{u}, \cdot) \le R^2$ on $\mathcal{C}$ and $\sum_i w_{t,i} = N$ (Lemma 9), we have $0 \le \Phi_t(\mathbf{u}) \le NR^2$. By Assumption 3 (Lipschitz continuity of $\mathcal{D}_\psi$ in its first argument with constant K),

$$|\Phi_t(\mathbf{u}) - \Phi_t(\mathbf{v})| \le NK\|\mathbf{u} - \mathbf{v}\|, \qquad \mathbf{u}, \mathbf{v} \in \mathcal{C}. \tag{60}$$

By the Push–Sum update $\mathbf{y}_{t+1,i} = \mathbf{s}_{t+1,i}/w_{t+1,i}$ with $\mathbf{s}_{t+1,i} = \sum_j [\mathbf{A}(t)]_{ij}\mathbf{q}_{t,j} = \sum_j [\mathbf{A}(t)]_{ij} w_{t,j}\mathbf{x}_{t,j}$, the point $\mathbf{y}_{t+1,i}$ is a convex combination (with weights $[\mathbf{A}(t)]_{ij} w_{t,j}/w_{t+1,i}$) of $\{\mathbf{x}_{t,j}\}_j$. Convexity of $\mathcal{D}_\psi(\mathbf{u}, \cdot)$ from Assumption 2 and column stochasticity under the communication model of Section 2.1 then give

$$\begin{aligned}
\sum_{i=1}^N w_{t+1,i}\, \mathcal{D}_\psi(\mathbf{u}, \mathbf{y}_{t+1,i}) &\le \sum_{i=1}^N \sum_{j=1}^N [\mathbf{A}(t)]_{ij} w_{t,j}\, \mathcal{D}_\psi(\mathbf{u}, \mathbf{x}_{t,j}) \\
&= \sum_{j=1}^N w_{t,j}\, \mathcal{D}_\psi(\mathbf{u}, \mathbf{x}_{t,j}) = \Phi_t(\mathbf{u}).
\end{aligned} \tag{61}$$

Hence, for $\mathbf{u} = \mathbf{x}_t^*$,

$$\sum_{i=1}^N w_{t+1,i} \left[\mathcal{D}_\psi(\mathbf{x}_t^*, \mathbf{y}_{t+1,i}) - \mathcal{D}_\psi(\mathbf{x}_t^*, \mathbf{x}_{t+1,i})\right] \le \Phi_t(\mathbf{x}_t^*) - \Phi_{t+1}(\mathbf{x}_t^*). \tag{62}$$

Dividing by η_t, taking expectations, and summing over $t = 1, \dots, T$,

$$\sum_{t=1}^{T}\sum_{i=1}^{N}\frac{w_{t+1,i}}{\eta_t}\mathbb{E}\left[\mathcal{D}_\psi(\mathbf{x}_t^*, \mathbf{y}_{t+1,i}) - \mathcal{D}_\psi(\mathbf{x}_t^*, \mathbf{x}_{t+1,i})\right] \le \sum_{t=1}^{T}\frac{\mathbb{E}[\Phi_t(\mathbf{x}_t^*)] - \mathbb{E}[\Phi_{t+1}(\mathbf{x}_t^*)]}{\eta_t}. \tag{63}$$

Inserting $\Phi_{t+1}(\mathbf{x}_{t+1}^*)$ and telescoping,

$$\begin{aligned}\sum_{t=1}^{T}\frac{\mathbb{E}[\Phi_t(\mathbf{x}_t^*)] - \mathbb{E}[\Phi_{t+1}(\mathbf{x}_t^*)]}{\eta_t} &= \frac{\mathbb{E}[\Phi_1(\mathbf{x}_1^*)]}{\eta_1} - \frac{\mathbb{E}[\Phi_{T+1}(\mathbf{x}_T^*)]}{\eta_T} \\ &\quad + \sum_{t=2}^{T}\left(\frac{1}{\eta_t} - \frac{1}{\eta_{t-1}}\right)\mathbb{E}[\Phi_t(\mathbf{x}_t^*)] \\ &\quad + \sum_{t=2}^{T}\frac{\mathbb{E}[\Phi_t(\mathbf{x}_t^*)] - \mathbb{E}[\Phi_t(\mathbf{x}_{t-1}^*)]}{\eta_{t-1}}.\end{aligned} \tag{64}$$

Because $\{\eta_t\}$ is nonincreasing, the coefficients $1/\eta_t - 1/\eta_{t-1}$ are nonnegative. Thus the middle sum is bounded above by $NR^2(1/\eta_T - 1/\eta_1)$. The last sum is bounded using (60). Discarding only the nonpositive terminal potential, we obtain

$$\begin{aligned}&\sum_{t=1}^{T}\frac{\mathbb{E}[\Phi_t(\mathbf{x}_t^*)] - \mathbb{E}[\Phi_{t+1}(\mathbf{x}_t^*)]}{\eta_t} \\ &\quad\le \frac{NR^2}{\eta_1} + NR^2\left(\frac{1}{\eta_T} - \frac{1}{\eta_1}\right) + NK\sum_{t=2}^{T}\frac{\|\mathbf{x}_t^* - \mathbf{x}_{t-1}^*\|}{\eta_{t-1}} \\ &\quad= \frac{NR^2}{\eta_T} + NK\sum_{t=1}^{T-1}\frac{\|\mathbf{x}_{t+1}^* - \mathbf{x}_t^*\|}{\eta_t},\end{aligned} \tag{65}$$

where the equality simplifies the reciprocal-step terms and reindexes the last sum. Combining (63)–(65) proves (58). □

Lemma 14. Under the conditions of Theorem 1, the cumulative approximation gap for the policy formed from the entries corresponding to the individual action sets satisfies

$$\begin{aligned}&\left(\frac{1 - e^{-\bar{c}}}{\bar{c}}\right)\sum_{t=1}^{T}F_t(\mathbf{x}_t^*) - \sum_{t=1}^{T}\mathbb{E}\left[F_t\left(\sum_{i\in\mathcal{N}}\mathbf{x}_{t,i}\odot\mathbf{1}_{\mathcal{V}_i}\right)\right] \\ &\le \frac{NR^2}{\eta_T} + KN\sum_{t=1}^{T-1}\frac{\|\mathbf{x}_{t+1}^* - \mathbf{x}_t^*\|}{\eta_t} \\ &\quad + \frac{2\beta N(L_g D + (1 + C_2)G)}{\gamma}\sum_{t=1}^{T}\left(NR\lambda^{t-2} + NG\sum_{s=1}^{t-1}\lambda^{t-s-2}\eta_s\right) \\ &\quad + \left[\frac{NG(4L_g D + 4C_2 G)}{2} + \frac{NG^2}{2\gamma}\right]\sum_{t=1}^{T}\eta_t.\end{aligned} \tag{66}$$

The bound follows by combining the weighted telescope with the state-disagreement estimate from Lemma 12.

Proof. By the gradient bound in Assumption 4, the multilinear extension F_t is G-Lipschitz continuous on $\mathcal{C}$. Therefore,

$$\begin{aligned}&\left|F_t\left(\sum_{i\in\mathcal{N}}\mathbf{x}_{t,i}\odot\mathbf{1}_{\mathcal{V}_i}\right) - F_t(\overline{\mathbf{q}}_t)\right| \\ &\quad\le G\left\|\sum_{i\in\mathcal{N}}(\mathbf{x}_{t,i} - \overline{\mathbf{q}}_t)\odot\mathbf{1}_{\mathcal{V}_i}\right\| \\ &\quad\le GC_2\sum_{i\in\mathcal{N}}\|\mathbf{x}_{t,i} - \overline{\mathbf{q}}_t\|.\end{aligned}$$

By (49),

$$\sum_i \mathbb{E}\|\mathbf{x}_{t,i} - \overline{\mathbf{q}}_t\| \le \frac{2\beta N}{\gamma}\left(NR\lambda^{t-2} + NG\sum_{s=1}^{t-1}\lambda^{t-s-2}\eta_s\right) + 2NG\eta_{t-1}.$$

Taking expectations and summing over $t = 1, \dots, T$,

$$\begin{aligned}&\left|\sum_{t=1}^{T}\mathbb{E}\left[F_t\left(\sum_{i\in\mathcal{N}}\mathbf{x}_{t,i}\odot\mathbf{1}_{\mathcal{V}_i}\right) - F_t(\overline{\mathbf{q}}_t)\right]\right| \\ &\le \frac{2\beta NGC_2}{\gamma}\sum_{t=1}^{T}\left(NR\lambda^{t-2} + NG\sum_{s=1}^{t-1}\lambda^{t-s-2}\eta_s\right) + 2NC_2G^2\sum_{t=1}^{T}\eta_t.\end{aligned} \tag{67}$$

Combining the final bound of Lemma 12 with Lemma 13 gives

$$\begin{aligned}&\left(\frac{1-e^{-\bar{c}}}{\bar{c}}\right)\sum_{t=1}^{T}F_t(\mathbf{x}_t^*) - \sum_{t=1}^{T}\mathbb{E}[F_t(\overline{\mathbf{q}}_t)] \\ &\le \frac{NR^2}{\eta_T} + KN\sum_{t=1}^{T-1}\frac{\|\mathbf{x}_{t+1}^* - \mathbf{x}_t^*\|}{\eta_t} \\ &\quad + \frac{2\beta N(L_gD+G)}{\gamma}\sum_{t=1}^{T}\left(NR\lambda^{t-2} + NG\sum_{s=1}^{t-1}\lambda^{t-s-2}\eta_s\right) \\ &\quad + \frac{NG(4\gamma L_gD+G)}{2\gamma}\sum_{t=1}^{T}\eta_t.\end{aligned} \tag{68}$$

Adding (67) to (68) via the triangle inequality through $F_t(\overline{\mathbf{q}}_t)$ merges the two λ-geometric sums into coefficient $2\beta N(L_gD+G)/\gamma + 2\beta NGC_2/\gamma = 2\beta N(L_gD + (1+C_2)G)/\gamma$, and merges the two η_t coefficients into

$$\frac{NG(4\gamma L_gD+G)}{2\gamma} + 2NC_2G^2 = \frac{NG(4L_gD+4C_2G)}{2} + \frac{NG^2}{2\gamma},$$

proving (66). □

By Lemma 8, we have

$$\mathbb{E}\left[f_t\left(\bigcup_{i\in\mathcal{N}}\{a_{t,i}\}\right)\right] \ge \mathbb{E}\left[F_t\left(\sum_{i\in\mathcal{N}}\mathbf{x}_{t,i}\odot\mathbf{1}_{\mathcal{V}_i}\right)\right]. \tag{69}$$

For $\mathbf{x}_t^* = \mathbf{1}_{\mathcal{A}_t^*}$, the definition of C_2 gives, for $t = 1, \dots, T-1$,

$$\begin{aligned}\left\|\mathbf{x}_{t+1}^* - \mathbf{x}_t^*\right\| &\le C_2\left\|\mathbf{1}_{\mathcal{A}_{t+1}^*} - \mathbf{1}_{\mathcal{A}_t^*}\right\|_1 \\ &= C_2\left|\mathcal{A}_{t+1}^* \,\Delta\, \mathcal{A}_t^*\right|.\end{aligned} \tag{70}$$

Combining Lemma 14 with (69) and (70), we can prove Theorem 1, that is,

$$\begin{aligned}&\left(\frac{1-e^{-\bar{c}}}{\bar{c}}\right)\sum_{t=1}^{T}f_t(\mathcal{A}_t^*) - \sum_{t=1}^{T}\mathbb{E}\left[f_t\left(\bigcup_{i\in\mathcal{N}}\{a_{t,i}\}\right)\right] \\ &\le \frac{NR^2}{\eta_T} + \sum_{t=1}^{T-1}\frac{KNC_2}{\eta_t}\left|\mathcal{A}_{t+1}^* \,\Delta\, \mathcal{A}_t^*\right| \\ &\quad + \frac{2\beta N(L_gD+(1+C_2)G)}{\gamma}\sum_{t=1}^{T}\left(NR\lambda^{t-2} + NG\sum_{s=1}^{t-1}\lambda^{t-s-2}\eta_s\right) \\ &\quad + \left[\frac{NG(4L_gD+4C_2G)}{2} + \frac{NG^2}{2\gamma}\right]\sum_{t=1}^{T}\eta_t.\end{aligned} \tag{71}$$

The left-hand side is $\mathbb{E}[\mathrm{Reg}_\alpha(T)]$, which proves Theorem 1. □

C Proof of Theorem 2

Theorem 2. Under Assumptions 1–4 and the communication model of Section 2.1 with $N \geq 2$, with deterministic positive step sizes, for any $i, j \in \mathcal{N}$ and $2 \leq t \leq T+1$, the consensus error satisfies

$$\mathbb{E}[\|\mathbf{y}_{t,i} - \mathbf{y}_{t,j}\|] \leq \frac{4\beta N}{\gamma} \left(R\lambda^{t-2} + G \sum_{s=1}^{t-2} \lambda^{t-s-2} \eta_s \right). \tag{72}$$

This statement follows from the consensus estimate in Lemma 11.

Proof. By the triangle inequality, we have

$$\begin{aligned} \|\mathbf{y}_{t,i} - \mathbf{y}_{t,j}\| &= \left\| \left(\mathbf{y}_{t,i} - \overline{\mathbf{q}}_{t-1}\right) - \left(\mathbf{y}_{t,j} - \overline{\mathbf{q}}_{t-1}\right) \right\| \\ &\leq \left\| \mathbf{y}_{t,i} - \overline{\mathbf{q}}_{t-1} \right\| + \left\| \mathbf{y}_{t,j} - \overline{\mathbf{q}}_{t-1} \right\|. \end{aligned} \tag{73}$$

Taking expectations in (73) and applying Lemma 11 with index $t-1$ gives

$$\begin{aligned} \mathbb{E}\left[\|\mathbf{y}_{t,i} - \mathbf{y}_{t,j}\|\right] &\leq \frac{2\beta N}{\gamma} \left(R\lambda^{t-2} + G \sum_{s=1}^{t-2} \lambda^{t-s-2} \eta_s \right) \\ &\quad + \frac{2\beta N}{\gamma} \left(R\lambda^{t-2} + G \sum_{s=1}^{t-2} \lambda^{t-s-2} \eta_s \right) \\ &= \frac{4\beta N}{\gamma} \left(R\lambda^{t-2} + G \sum_{s=1}^{t-2} \lambda^{t-s-2} \eta_s \right). \end{aligned}$$

□

D Proof of Corollary 1

According to Theorem 1, let the step size be constant, that is, $\eta_t = \eta$ for all $t \in [T]$. We first bound each term in the regret bound (71).

For the first term,

$$\mathrm{Term}_1 = \frac{NR^2}{\eta}.$$

Let

$$C_T = \sum_{t=1}^{T-1} \left| \mathcal{A}_{t+1}^* \,\Delta\, \mathcal{A}_t^* \right|$$

denote the deviation of the maximizer sequence. Then, the second term satisfies

$$\mathrm{Term}_2 = \sum_{t=1}^{T-1} \frac{KNC_2}{\eta} \left| \mathcal{A}_{t+1}^* \,\Delta\, \mathcal{A}_t^* \right| = \frac{KNC_2 C_T}{\eta}.$$

For the third term (the λ-geometric term),

$$\mathrm{Term}_3 = \frac{2\beta N(L_g D + (1+C_2)G)}{\gamma} \sum_{t=1}^{T} \left(NR\lambda^{t-2} + NG \sum_{s=1}^{t-1} \lambda^{t-s-2} \eta \right).$$

Using $\sum_{t=1}^{T} \lambda^{t-2} \leq 1/(\lambda(1-\lambda))$ and $\sum_{t=1}^{T} \sum_{s=1}^{t-1} \lambda^{t-s-2} \leq T/(\lambda(1-\lambda))$,

$$\mathrm{Term}_3 \leq \frac{2\beta N^2 R(L_g D + (1+C_2)G)}{\gamma\lambda(1-\lambda)} + \frac{2\beta N^2 G(L_g D + (1+C_2)G)}{\gamma\lambda(1-\lambda)} T\eta.$$

For the fourth term, Lemma 14 gives

$$\mathrm{Term}_4 = \left[\frac{NG(4L_g D + 4C_2 G)}{2} + \frac{NG^2}{2\gamma} \right] T\eta.$$

Collecting the coefficient of $1/\eta$, denoted by M_1,

$$M_1 = NR^2 + KNC_2C_T.$$

Collecting all terms proportional to $T\eta$, denoted by M_2,

$$M_2 = \frac{2\beta N^2 G(L_g D + (1+C_2)G)}{\gamma\lambda(1-\lambda)} + \frac{NG(4L_g D + 4C_2 G)}{2} + \frac{NG^2}{2\gamma}.$$

Denote the remaining constant (horizon-independent) term by

$$M_0 = \frac{2\beta N^2 R(L_g D + (1+C_2)G)}{\gamma\lambda(1-\lambda)}.$$

The expected dynamic α-regret therefore satisfies

$$\mathbb{E}[\mathrm{Reg}_\alpha(T)] \le \frac{M_1}{\eta} + M_2 T\eta + M_0. \tag{74}$$

For $M_1, M_2 > 0$, minimizing this upper bound over positive constant step sizes gives

$$\eta^* = \sqrt{\frac{M_1}{M_2 T}} = \sqrt{\frac{NR^2 + KNC_2C_T}{\left(\frac{2\beta N^2 G(L_g D + (1+C_2)G)}{\gamma\lambda(1-\lambda)} + \frac{NG(4L_g D + 4C_2 G)}{2} + \frac{NG^2}{2\gamma}\right) T}}. \tag{75}$$

If C_T is not known in advance, a known deterministic upper bound $V_T \ge C_T$ can replace it in M_1 and the step-size choice. The resulting bound then depends on $1 + V_T$.

Substituting η^* into (74), we obtain

$$\mathbb{E}[\mathrm{Reg}_\alpha(T)] \le 2\sqrt{M_1 M_2 T} + M_0.$$

Dividing both sides by T, the average regret satisfies

$$\frac{\mathbb{E}[\mathrm{Reg}_\alpha(T)]}{T} \le 2\sqrt{\frac{M_1 M_2}{T}} + \frac{M_0}{T}.$$

For the asymptotic simplification below, fix $N, R, K, C_2, L_g, D, G, \beta$ and the positive communication constant $\gamma \in (0, 1]$, independently of T. The hidden constants in $O(\cdot)$ are allowed to depend on γ^{-1} and these fixed parameters, but not on T or C_T. We leave the factor $1/[\lambda(1-\lambda)]$ explicit, where $0 < \lambda < 1$. Hence

$$M_1 = O(1 + C_T), \qquad M_2 = O\left(\frac{1}{\lambda(1-\lambda)}\right), \qquad M_0 = O\left(\frac{1}{\lambda(1-\lambda)}\right).$$

Define $U_T := 2\sqrt{M_1 M_2 T} + M_0 \ge 0$. Consequently,

$$\frac{\max\{\mathbb{E}[\mathrm{Reg}_\alpha(T)], 0\}}{T} \le \frac{U_T}{T} = O\left(\sqrt{\frac{1 + C_T}{\lambda(1-\lambda)T}} + \frac{1}{\lambda(1-\lambda)T}\right). \tag{76}$$

The factors $1/\gamma$ in M_0 and M_2 have not been removed from the explicit bound $U_T = 2\sqrt{M_1 M_2 T} + M_0$; they are absorbed into the hidden constants in (76) because γ is fixed. The finite-horizon contribution M_0/T remains explicit. This simplified notation is not a bound uniform in γ as $\gamma \downarrow 0$.

For the uniform outgoing-weight model specified in Lemma 9, with $N \ge 2$,

$$\beta = 4, \qquad \lambda = \left(1 - \frac{1}{N^{NB}}\right)^{1/B} \in (0, 1).$$

Consequently, for fixed communication and problem parameters, if $C_T = o(T)$, then

$$U_T = 2\sqrt{M_1 M_2 T} + M_0 = o(T), \qquad \limsup_{T\to\infty} \frac{\mathbb{E}[\mathrm{Reg}_\alpha(T)]}{T} \le 0.$$

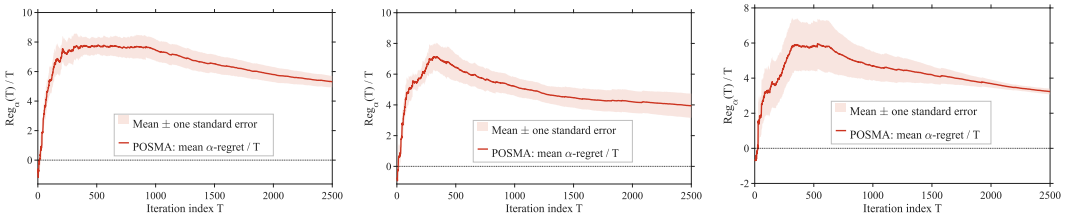
Thus, for fixed problem and communication parameters, the cumulative approximation regret admits a sublinear upper bound whenever $C_T = o(T)$.

E ADDITIONAL SIMULATION RESULTS

This appendix reports additional experiments for the normalized cumulative dynamic α-regret, the effects of the B-strong-connectivity parameter, and the curvature upper bound. Results for R:A:P=8:1:1 are shown together with the corresponding R:A:P=6:3:1 and 4:5:1 settings. We finally consider a scenario that is more representative of practical tracking applications.

E.1 NORMALIZED CUMULATIVE DYNAMIC α-REGRET

Figure 4 reports the normalized cumulative dynamic α-regret, $\mathbb{E}[\mathrm{Reg}_\alpha(T)]/T$, for the three target-mixture settings. It decreases with the horizon and approaches a small value, consistent with the theoretical condition $C_T = o(T)$ for vanishing average regret.

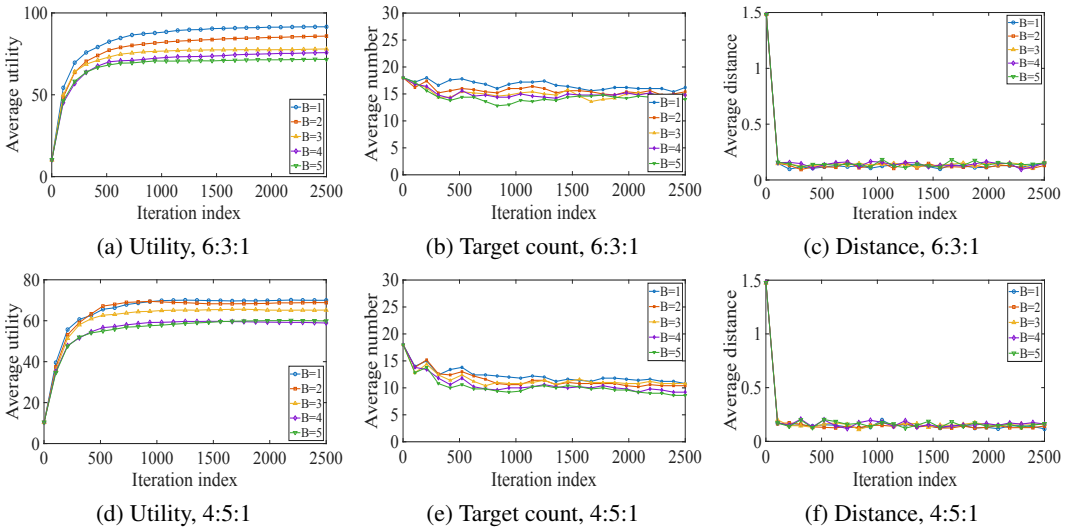


(a) R:A:P = 8:1:1 (b) R:A:P = 6:3:1 (c) R:A:P = 4:5:1

Figure 4: Normalized cumulative dynamic α-regret for the three target-mixture settings.

E.2 EFFECT OF B-STRONG CONNECTIVITY

Figure 5 compares $B \in \{1, 2, 3, 4, 5\}$ with $\bar{c} = 1$. For both target ratios, the cumulative average utility increases and gradually levels off. The $B = 1$ case attains the highest final utility, whereas larger values of B generally reduce performance because information is exchanged less frequently and policy information propagates more slowly across the network. The R:A:P=4:5:1 setting is more difficult than 6:3:1: its larger adversarial-target proportion leads to lower utility and fewer covered targets. After an initial decrease, the target count fluctuates around a stable range, while the average tracking distance remains low.



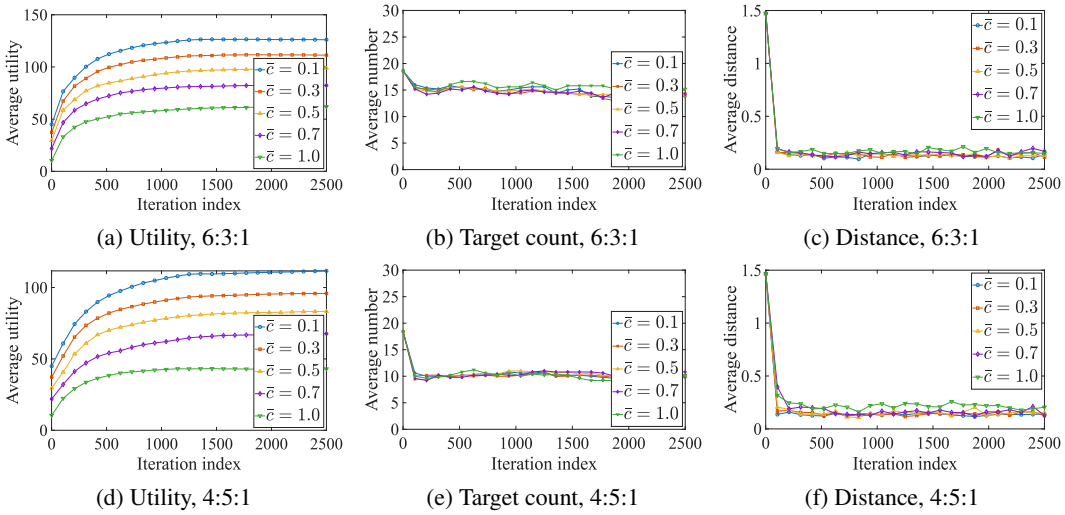
(a) Utility, 6:3:1 (b) Target count, 6:3:1 (c) Distance, 6:3:1

(d) Utility, 4:5:1 (e) Target count, 4:5:1 (f) Distance, 4:5:1

Figure 5: Effect of $B \in \{1, 2, 3, 4, 5\}$ in B-strong connectivity with $\bar{c} = 1$ for R:A:P=6:3:1 and 4:5:1.

E.3 Effect of the curvature upper bound

With $B = 3$ fixed, Figure 6 compares $\bar{c} \in \{0.1, 0.3, 0.5, 0.7, 1.0\}$. For both target ratios, smaller $\bar{c}$ gives higher final cumulative average utility; this is consistent with the larger theoretical approximation factor $(1 - e^{-\bar{c}})/\bar{c}$. The utility curves nevertheless evaluate different simulated objectives and should not be interpreted as a direct measurement of that factor. The target-count curves remain close across curvature values, settling near 15 targets for R:A:P=6:3:1 and near 10 for 4:5:1. Tracking distances decrease rapidly and remain low.



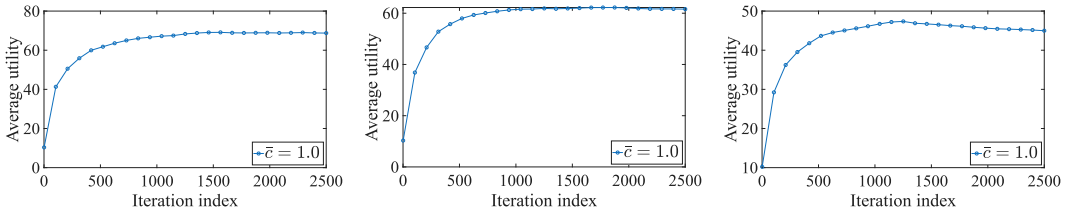
(a) Utility, 6:3:1 (b) Target count, 6:3:1 (c) Distance, 6:3:1

(d) Utility, 4:5:1 (e) Target count, 4:5:1 (f) Distance, 4:5:1

Figure 6: Effect of $\bar{c} \in \{0.1, 0.3, 0.5, 0.7, 1.0\}$ with $B = 3$ for R:A:P=6:3:1 and 4:5:1.

E.4 Tracking in a practically motivated scenario

We evaluate POSMA in a practically motivated scenario with 20 agents and 30 targets initially placed within a 2000-m-radius disk. We run 2500 rounds of 2 s each, with agent speeds in $\{5, 10, 15\}$ m/s. Adversarial targets have trigger and sensing radii of 200 and 800 m, respectively, and escape at 15 m/s for 10 s.



(a) R:A:P = 8:1:1 (b) R:A:P = 6:3:1 (c) R:A:P = 4:5:1

Figure 7: Cumulative average utility with $B = 3$ and $\bar{c} = 1$ over time-varying digraphs.