Near-Optimal Multi-Agent Learning for Safe Coverage Control

Manish Prajapat MANISHP@AI.ETHZ.CH
ETH AI Center
ETH Zurich

Matteo Turchetta MATTEOTU@INF.ETHZ.CH
Department of Computer Science
ETH Zurich

Melanie N. Zeilinger† MZEILINGER@ETHZ.CH
Institute of Dynamic Systems and Control
ETH Zurich

Andreas Krause† KRAUSEA@ETHZ.CH
Department of Computer Science
ETH Zurich

Abstract

In multi-agent coverage control problems, agents navigate their environment to reach locations that maximize the coverage of some density. In practice, the density is rarely known *a priori*, further complicating the original NP-hard problem. Moreover, in many applications, agents cannot visit arbitrary locations due to *a priori* unknown safety constraints. In this paper, we aim to efficiently learn the density to approximately solve the coverage problem while preserving the agents' safety. We first propose a conditionally linear submodular coverage function that facilitates theoretical analysis. Utilizing this structure, we develop MACOPT, a novel algorithm that efficiently trades off the exploration-exploitation dilemma due to partial observability, and show that it achieves sublinear regret. Next, we extend results on single-agent safe exploration to our multi-agent setting and propose SAFEMAC for safe coverage and exploration. We analyze SAFEMAC and give first of its kind results: near optimal coverage in finite time while provably guaranteeing safety. We extensively evaluate our algorithms on synthetic and real problems, including a bio-diversity monitoring task under safety constraints, where SAFEMAC outperforms competing methods.

Keywords: Multi-agent, Submodularity, Coverage control, Safety, Bayesian Optimization

1. Introduction

In multi-agent coverage control (MAC) problems, multiple agents coordinate to maximize coverage over some spatially distributed events. Their applications abound, from collaborative mapping [1], environmental monitoring [2], inspection robotics [3] to sensor networks [4]. In addition, the coverage formulation can address core challenges in cooperative multi-agent RL [5], e.g., *exploration* [6], by providing high-level goals. In these applications, agents often encounter safety constraints that may lead to critical accidents when ignored, e.g., obstacles [7] or extreme weather conditions [8, 9].

. † Authors involved in joint supervision

Deploying coverage control solutions in the real world presents many challenges: (*i*) for a given density of relevant events, this is an *NP hard problem* [10]; (*ii*) such *density* is *rarely known* in practice [2] and must be learned from data, which presents a complex active learning problem as the quantity we measure (the density) differs from the one we want to optimize (its coverage); (*iii*) agents often operate under *safety-critical* conditions, [7–9], that may be *unknown a priori*. This requires cautious exploration of the environment to prevent catastrophic outcomes. While prior work addresses subsets of these challenges (see Section 7), we are not aware of methods that address them jointly.

This work makes the following contributions toward efficiently solving safe coverage control with *a-priori* unknown objectives and constraints. **Firstly**, we model this multi-agent learning task as a *conditionally linear* coverage function. We use the *monotonocity* and the *submodularity* of this function to propose MaCOpt, a new algorithm for the unconstrained setting that enjoys sublinear cumulative regret and efficiently recommends a near-optimal solution. **Secondly**, we extend GoOSE [11], an algorithm for single agent safe exploration, to the multi-agent case. Combining our extension of GoOSE with MaCOpt, we propose SafeMaC, a novel algorithm for safe multi-agent coverage control. We analyze it and show it attains a near-optimal solution in a finite time. **Finally**, we demonstrate our algorithms on a synthetic and two real world applications: safe biodiversity monitoring and obstacle avoidance. We show SafeMaC finds better solutions than algorithms that do not actively explore the feasible region and is more sample efficient than competing near-optimal safe algorithms.

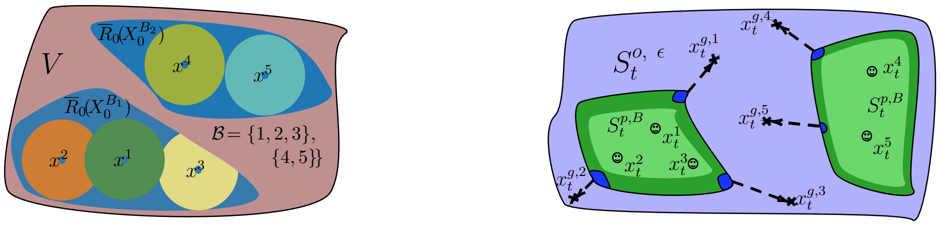
2. Problem Statement

We present the safety-constrained multi-agent coverage control problem that we aim to solve.
Coverage control. Coverage control models situations where we want deploy a swarm of dynamic agents to maximize the coverage of a quantity of interest. Formally, given a finite[1] set of possible locations V, the goal of coverage control is to maximize a function $F : 2^V \to \mathbb{R}$ that assigns to each subset, $X \subseteq V$, the corresponding coverage value. For K agents, the resulting problem is $\arg\max_{X:|X|\leq K} F(X)$.
Sensing region. Depending on the application, we may use different definitions of F. Here, we model cases where agent i at location x^i covers a limited sensing region around it, D^i. While D^i can be any connected subset of V, in practice it is often a ball centered at x^i. Given a function $\rho : V \to \mathbb{R}$ denoting the density of a quantity of interest at each $v \in V$, our coverage objective is

$$F(X; \rho, V) = \sum_{x^i \in X} \sum_{v \in D^{i-}} \rho(v)/N, \tag{1}$$

where $D^{i-} := D^i \setminus D^{1:i-1}$ indicates the elements in V covered by agent i but not by agents $1 : i-1$, and N is the largest number of elements in V covered by a sensing region.
Safety. In many real-world problems, agents cannot go to arbitrary locations due to safety concerns. To model this, we introduce a constraint function $q : V \to \mathbb{R}$ and we consider safe all locations v satisfying $q(v) \geq 0$. Such constraint restricts the space of possible solutions of our problem in two ways. First, it prevents agents from monitoring from unsafe locations. Second, depending on its dynamics, agent i may be unable to safely reach a disconnected safe area

1. Continuous domains can be handled via discretization



(a) Disconnected safe regions

(b) Illustration of multi-agent GoOSE

Figure 1: a) Agents are partitioned into two batches. Agent 1 cover D^1(green), 2 cover D^{2-} (orange) & 3 cover D^{3-}(yellow). b) SafeMaC sets a goal $x_t^{g,i} \forall i$ in the optimistic set. It forms a expander region (dark blue) to safely expand the pessimistic safe set, S_t^p, towards the goal.

starting from x_0^i, see Fig. 1a. We denote with $\bar{R}_{\epsilon_q}(\{x_0^i\})$ the largest safely reachable region starting from x_0^i and with $\mathcal{B}$ a collection of batches of agents such that all agents in the same batch B share the same safely reachable set, $\forall i, j \in B : \bar{R}_{\epsilon_q}(\{x_0^i\}) \cap \bar{R}_{\epsilon_q}(\{x_0^j\}) \neq \emptyset$, see Appendix B for formal definitions. Based on this, we define the safely reachable control problem

$$\sum_{B \in \mathcal{B}} \max_{X^B \in \bar{R}_{\epsilon_q}(X_0^B)} F(X^B; \rho, \bar{R}_{\epsilon_q}(X_0^B)), \tag{2}$$

where, $X_0^B = \{x_0^i\}_{i \in B}$ are the starting locations of all agents in B and $\bar{R}_{\epsilon_q}(X_0^B) = \cup_{i \in B} \bar{R}_{\epsilon_q}(\{x_0^i\})$ indicates the largest safely reachable region from any point x_0^i for all i in B. **Unknown density and constraint**. In practice, the density ρ and the constraint q are often unknown *a priori*. However, the agents can iteratively obtain noisy measurements of their values at target locations. We consider synchronous measurements, i.e., we wait until all agents have collected the desired measurement for the current iteration before moving to the next one. Here, we focus on the high-level problem of choosing informative locations, rather than the design of low-level motion planning. Therefore, our goal is to find an approximate solution to the problem in Eq. (2) preserving safety throughout exploration, i.e., at every location visited by the agents, while taking as few measurements as possible in case the dynamics of the agents are deterministic and known as in [11].

3. Background

Submodularity. Optimizing a function defined over the power set of a finite domain, V, scales combinatorially with the size of V in general. In special cases, we can exploit the structure of the objective to find approximate solutions efficiently. Monotone submodular functions are one example of this.

A set function $F : 2^V \to \mathbb{R}$ is *monotone* if for all $A \subseteq B \subset V$ we have $F(A) \leq F(B)$. It is *submodular* if $\forall v \in V \setminus B$, we have, $F(A \cup \{v\}) - F(A) \geq F(B \cup \{v\}) - F(B)$. In coverage control, this means adding v to A yields a larger or equal relative coverage improvement than adding v to B, if $A \subseteq B$. Crucially, [12] guarantees that the greedy algorithm produces a solution within a factor of $(1 - 1/e)$ of the optimal solution for problems of the type $\arg\max_{X:|X| \leq K} F(X; \rho, V)$, when F is monotone and submodular. In practice, the greedy algorithm often outperforms this worst-case guarantee [13]. The coverage function in Eq. (1) is a conditionally linear, monotone and submodular function (proof in Appendix C), which lets us use the results above to design our algorithm for safe coverage control.

Assumptions. We make regularity assumptions for ρ and q, which let us model them using Gaussian Processes (GP) [14]. For details, see Apx. A.1. In the next sections, we discuss MaCOpt and defer the algorithm, analysis and results of SafeMaC to the Appendix A.

4. MaCOpt: unconstrained multi-agent coverage control

Greedy sensing regions. In sequential optimization, it is crucial to balance exploration and exploitation. GP-UCB [15] is a theoretically sound strategy to strike such a trade-off that works well in practice. Agents evaluate the objective at locations that maximize an upper confidence bound over the objective given by the GP model such that locations with either a high posterior mean (exploitation) or standard deviation (exploration) are visited. We construct a valid confidence upper bound for the coverage $F(X)$ starting from our confidence intervals on ρ, by replacing the true density ρ with its upper bound u_t^ρ in Eq. (1). Next, we apply the greedy algorithm to this upper bound (Line 3 of Algorithm 1) to select K candidate locations for evaluating the density. Unfortunately, this exploration strategy may perform poorly. This is because, to reduce the uncertainty over the coverage F at X, we must learn the density ρ at all locations inside the sensing region, $\cup_{i=1}^K D^i$, rather than simply at X. We say ours is a partial monitoring problem, where the objective F differs from the quantity we measure, i.e., the density ρ. Next, we explain how to choose locations where to observe the density for a given X.
Uncertainty sampling. Given the next locations for the agents, X, we measure the density to learn efficiently about $F(X)$. Intuitively, agent i learns the density where the uncertainty is highest within the area it covers that is not covered by agents $\{1, \dots, i-1\}$, i.e., D_t^{i-} (Line 4).
Stopping criterion. The algorithm should terminate when a near-optimal solution is achieved. Intuitively, this occurs when the uncertainty about the coverage value of the greedy recommendation is low. Formally, we require the sum of the uncertainties over the sampling targets, $w_t = \sum_{i=1}^K u_{t-1}^\rho(x_t^{g,i}) - l_{t-1}^\rho(x_t^{g,i})$, to be below a threshold, ϵ_ρ (Line 2). Importantly, this stopping criterion requires the confidence intervals to shrink only at regions that potentially maximize the coverage.
MaCOpt. Now, we introduce MaCOpt (Algorithm 1). At round t, we select the sensing locations for the agents, X_t, by greedily optimizing the upper confidence bound of the coverage. Then, each agent i collects noisy density measurements at the points of highest uncertainty within D_t^{i-}. Finally, we update our GP over the density and, if the sum of maximum uncertainties within each sensing region is small, we stop the algorithm.

5. Analysis

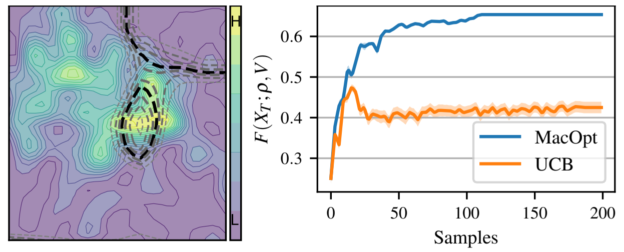
MaCOpt. To measure the progress of MaCOpt, we study its regret, i.e., the difference between its solution and the one we could find if we knew the true density. Since control coverage consists in maximizing a monotone submodular function, we cannot compute the true optimum even for known densities. However, we can efficiently find a solution that is at least $(1 - 1/e)$ within the optimum. Thus, we quantify performance using the following notion of cumulative regret,

$$Reg_{act}(T) = (1 - \frac{1}{e}) \sum_{t=1}^{T} F(X_\star; \rho, V) - \sum_{t=1}^{T} F(X_t; \rho, V), \tag{3}$$

where $F(X_\star; \rho, V)$ is the optimal coverage. We now state one of our main results, which guarantees that the cumulative regret of MaCOpt grows sublinearly in time (proof in Appendix E).

Algorithm 1 MACOPT

1: **Inputs** X_0, ϵ_ρ, V, $GP_\rho, t \leftarrow 1$
2: **while** $w_t > \epsilon_\rho$ **do**
3: $\quad \forall i, x_t^i \leftarrow \arg\max_{x^i} \sum_{v \in D^i \setminus D_t^{1:i-1}} u_{t-1}^\rho(v)$
4: $\quad \forall i, x_t^{g,i} \leftarrow \arg\max_{x \in D_t^{i-}} u_{t-1}^\rho(v) - l_{t-1}^\rho(v)$
5: $\quad w_t \leftarrow \sum_{i=1}^K u_{t-1}^\rho(x_t^{g,i}) - l_{t-1}^\rho(x_t^{g,i})$
6: $\quad \forall i, y_{\rho t}^i = \rho(x_t^{g,i}) + \eta_\rho$, Update GP
7: $\quad t \leftarrow t+1$
8: **Recommend** X_t



(a) Gorilla Nests (b) Coverage MACOPT VS UCB

Figure 2: a) Contours show the Gorilla nests distribution with weather constraints marked by the black dashed line, and its contours with grey dashed line. b) MACOPT in the Gorilla environment.

Theorem 1 *Let $\delta \in (0,1)$ and β_t^ρ as in [16], i.e., ${\beta_t^\rho}^{1/2} = B_\rho + 4\sigma_\rho\sqrt{\gamma_{Kt}^\rho + \ln(1/\delta)}$. With probability at least $1-\delta$, MACOPT's regret defined in Eq. (3) is bounded by $\mathcal{O}(\sqrt{T\beta_T^\rho\gamma_{KT}^\rho})$,*

$$Pr\left\{ Reg_{act}(T) \leq K\sqrt{\frac{8T\beta_T^\rho\gamma_{KT}^\rho}{\log(1+K\sigma_\rho^{-2})}} \right\} \geq 1-\delta. \tag{6}$$

The proof of 1 builds on two key ideas. First, we exploit the conditional linearity of the submodular objective to bound the cumulative regret defined in Eq. (3) with a sum of per agent regrets. Secondly, we bound the per agent regret with the information capacity γ_{KT}^ρ, a quantity that measures the largest reduction in uncertainty about the density that can be obtained from KT noisy evaluations of it. Since γ_{KT}^ρ [17] grows sublinearly with T for commonly used kernels, so does MACOPT's regret in Eq. (6). The immediate corollary of the above theorem, when the MACOPT stopping criteria is reached (Line 2 of Algorithm 1) guarantees a near optimal solution up to ϵ_ρ precision.

Corollary 1 *Let $t_\rho^\star$ be the smallest integer, $\frac{t_\rho^\star}{\beta_{t_\rho^\star}\gamma_{Kt_\rho^\star}} \leq \frac{8K^2}{\log(1+K\sigma^{-2})\epsilon_\rho^2}$, then there exists a $t < t_\rho^\star$ such that w.h.p, MACOPT terminates and achieves, $F(X_t;\rho,V) \geq (1-\frac{1}{e})F(X_\star;\rho,V) - \epsilon_\rho$.*

6. Experiments

We compare MACOPT and SAFEMAC (Appendix A) to existing methods on synthetic and real-world problems. We validate our theoretical claims and observe their superiority. We briefly discuss the Gorilla environment and refer reader to Appendix A, H for the synthetic, the obstacle and the constrained Gorilla experiment setups along with extended empirical analysis.

Gorilla nest environment. We simulate a bio-diversity monitoring task, where we aim to cover areas with high density of gorilla nests with a quadrotor in the Kagwene Gorilla Sanctuary (Fig. 2a). The nest density is obtained by fitting a smooth rate function [18] over Gorilla nest counts [19]. We perform our experiments with $K=3$ agents in a 34×34 grid world. Each agent's disk is defined as the region an agent can reach in $r=5$ steps in the defined grid. We normalize coverage with a maximum value $\sum_{v \in \bar{R}_0(X_0)} \rho(v)/N$.

MACOPT. We compare MACOPT to UCB, a baseline that skips the uncertainty sampling step from Section 4 and obtains measurements at locations as per GREEDY sensing region. Fig. 2b shows comparison in the *gorilla* environment on a day of good weather, i.e., when everywhere is safe. We see that UCB gets stuck in a local optimum as it does not reduce the uncertainty of the density, whereas MaCOpt explores more and achieves a higher coverage value.

7. Related work

Our work is at the intersection of multiple fields. This section highlights the most relevant connections to them but is not an exhaustive overview (we reference surveys whenever possible).
Bayesian optimization. In BO, an agent evaluates an objective at a sequence of inputs to maximize it [20]. In contrast, in our setting the quantity we measure differs from our objective. Partial monitoring [21] addresses this kind of issues with randomized algorithms rooted in information theory [22, 23]. In coverage control with unknown density, this challenge is often addressed by learning the density uniformly over the domain [24, 25]. In contrast, MaCOpt learns the density only at promising locations.
Coverage control. MaC with known densities is a well-studied NP hard [26] problem. Many algorithms use efficient heuristics to converge quickly to a local optimum. One popular strategy is Lloyd's algorithm [27], which has been studied in different settings, e.g., with known densities [28, 29], *a-priori* unknown densities [25, 30–32], taking into account agent's dynamics and constraints [33], or in case of non-identical robots [34]. These methods apply to continuous state and action space and show convergence to a local optimum but lack optimality guarantees [24, 25, 33] and do not provide sample complexity bounds. Moreover, their extension to non-convex, disconnected domains is not trivial [35].
Submodular optimization. Online submodular maximization aims at optimizing unknown submodular functions from noisy measurements. It has multiple applications, including optimization of numerical solvers [36] and information gathering [37]. Mainly related to ours is the work in [38], which proposes an algorithm for contextual news recommendation for linear user preferences with strong regret guarantees. In contrast to that setting, we consider dynamic agents and have access to partial feedback.
Safety. Depending on the safety formulation and the assumptions, many algorithms have been proposed for safe learning in dynamical systems [39–48]. In the setting related to us, i.e., one with *a-priori* unknown constraints, there exists safe exploration algorithms that leverage regularity and establish safety and optimality guarantee for the BO case [49, 50] and further extended to the MDP case [51, 52]. All these approaches may be sample inefficient as they may explore the constraint in regions not relevant to the objective. GoOSE [11] addresses this problem for both BO and MDP cases. The only work in this context that addresses multi-agent problems is [53]. However, they have a different objective and do not have safety guarantees.

8. Conclusion

We present two novel algorithms for multi-agent coverage control in unconstrained (MaCOpt) and safety critical environments (SafeMaC). We show MaCOpt achieves sublinear cumulative regret, despite the challenge of partial observability. Moreover, we prove SafeMaC achieves near optimal coverage in finite time while navigating safely. We demonstrate the superiority of our algorithms in terms of sample efficiency and coverage in real-world applications such as safe biodiversity monitoring. We dedicated this paper to choosing informative goal locations. In future, we plan to extend this work to plan informative trajectories as well.

References

[1] Miquel Kegeleirs, Giorgio Grisetti, and Mauro Birattari. Swarm slam: Challenges and perspectives. *Frontiers in Robotics and AI*, 8, 2021. ISSN 2296-9144. doi: 10.3389/frobt.2021.618268. URL `https://www.frontiersin.org/article/10.3389/frobt.2021.618268`.

[2] Alan Mainwaring, David Culler, Joseph Polastre, Robert Szewczyk, and John Anderson. Wireless sensor networks for habitat monitoring. In *Proceedings of the 1st ACM International Workshop on Wireless Sensor Networks and Applications*, WSNA '02, page 88–97, New York, NY, USA, 2002. Association for Computing Machinery. ISBN 1581135890. doi: 10.1145/570738.570751. URL `https://doi.org/10.1145/570738.570751`.

[3] Mahmoud Tavakoli, Gonçlo Cabrita, Ricardo Faria, Lino Marques, and Anibal T. de Almeida. Cooperative multi-agent mapping of three-dimensional structures for pipeline inspection applications. *The International Journal of Robotics Research*, 31(12): 1489–1503, 2012. doi: 10.1177/0278364912461536. URL `https://doi.org/10.1177/0278364912461536`.

[4] Bang Wang. *Coverage Control in Sensor Networks*. Springer Publishing Company, Incorporated, 1st edition, 2010. ISBN 1849960585.

[5] Ryan Lowe, Yi Wu, Aviv Tamar, Jean Harb, Pieter Abbeel, and Igor Mordatch. Multi-agent actor-critic for mixed cooperative-competitive environments. *CoRR*, abs/1706.02275, 2017. URL `http://arxiv.org/abs/1706.02275`.

[6] Iou-Jen Liu, Unnat Jain, Raymond A Yeh, and Alexander Schwing. Cooperative exploration for multi-agent deep reinforcement learning. In *International Conference on Machine Learning*, pages 6826–6836. PMLR, 2021.

[7] Pericle Salvini, Diego Paez Granados, and Aude Billard. Safety concerns emerging from robots navigating in crowded pedestrian areas. *International Journal of Social Robotics*, 14, 03 2022. doi: 10.1007/s12369-021-00796-4.

[8] Yuliya Averyanova and E. Znakovskaja. Weather hazards analysis for small uass durability enhancement. In *2021 IEEE 6th International Conference on Actual Problems of Unmanned Aerial Vehicles Development (APUAVD)*, pages 41–44, 2021. doi: 10.1109/APUAVD53804.2021.9615440.

[9] Mozhou Gao, Chris H Hugenholtz, Thomas A Fox, Maja Kucharczyk, Thomas E Barchyn, and Paul R Nesbit. Weather constraints on global drone flyability. *Scientific reports*, 11(1):1–13, 2021.

[10] Andreas Krause and Carlos Guestrin. Submodularity and its applications in optimized information gathering. *ACM Trans. Intell. Syst. Technol.*, 2(4), jul 2011. ISSN 2157-6904. doi: 10.1145/1989734.1989736. URL `https://doi.org/10.1145/1989734.1989736`.

[11] Matteo Turchetta, Felix Berkenkamp, and Andreas Krause. Safe exploration for interactive machine learning, 2019.

[12] George Nemhauser, Laurence Wolsey, and M. Fisher. An analysis of approximations for maximizing submodular set functions. *Mathematical Programming*, 14:265–294, 12 1978. doi: 10.1007/BF01588971.

[13] Jure Leskovec, Andreas Krause, Carlos Guestrin, Christos Faloutsos, Jeanne VanBriesen, and Natalie Glance. Cost-effective outbreak detection in networks. In *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '07, page 420–429, New York, NY, USA, 2007. Association for Computing Machinery. ISBN 9781595936097. doi: 10.1145/1281192.1281239. URL https://doi.org/10.1145/1281192.1281239.

[14] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005. ISBN 026218253X.

[15] Niranjan Srinivas, Andreas Krause, Sham M. Kakade, and Matthias W. Seeger. Information-theoretic regret bounds for gaussian process optimization in the bandit setting. *IEEE Transactions on Information Theory*, 58(5):3250–3265, 2012. doi: 10.1109/TIT.2011.2182033.

[16] Sayak Ray Chowdhury and Aditya Gopalan. On kernelized multi-armed bandits. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 844–853. PMLR, 06–11 Aug 2017. URL https://proceedings.mlr.press/v70/chowdhury17a.html.

[17] Sattar Vakili, Kia Khezeli, and Victor Picheny. On information gain and regret bounds in gaussian process bandits. In Arindam Banerjee and Kenji Fukumizu, editors, *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pages 82–90. PMLR, 13–15 Apr 2021. URL https://proceedings.mlr.press/v130/vakili21a.html.

[18] Mojmír Mutný and Andreas Krause. Sensing cox processes via posterior sampling and positive bases. *CoRR*, abs/2110.11181, 2021. URL https://arxiv.org/abs/2110.11181.

[19] Neba Funwi-gabga and Jorge Mateu. Understanding the nesting spatial behaviour of gorillas in the kagwene sanctuary, cameroon. *Stochastic Environmental Research and Risk Assessment*, 26, 08 2011. doi: 10.1007/s00477-011-0541-1.

[20] Bobak Shahriari, Kevin Swersky, Ziyu Wang, Ryan P. Adams, and Nando de Freitas. Taking the human out of the loop: A review of bayesian optimization. *Proceedings of the IEEE*, 104(1):148–175, 2016. doi: 10.1109/JPROC.2015.2494218.

[21] Tor Lattimore and Csaba Szepesvári. *Partial Monitoring*, page 423–451. Cambridge University Press, 2020. doi: 10.1017/9781108571401.046.

[22] Tor Lattimore and Csaba Szepesvári. An information-theoretic approach to minimax regret in partial monitoring. In *Conference on Learning Theory*, pages 2111–2139. PMLR, 2019.

[23] Johannes Kirschner, Tor Lattimore, and Andreas Krause. Information directed sampling for linear partial monitoring. In *Conference on Learning Theory*, pages 2328–2369. PMLR, 2020.

[24] Lai Wei, Andrew McDonald, and Vaibhav Srivastava. Regret analysis of distributed gaussian process estimation and coverage. *CoRR*, abs/2101.04306, 2021. URL `https://arxiv.org/abs/2101.04306`.

[25] Andrea Carron, Marco Todescato, Ruggero Carli, Luca Schenato, and Gianluigi Pillonetto. Multi-agents adaptive estimation and coverage control using gaussian regression. In *2015 European Control Conference (ECC)*, pages 2490–2495, 2015. doi: 10.1109/ECC.2015.7330912.

[26] Andreas Krause, Ajit Singh, and Carlos Guestrin. Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies. *J. Mach. Learn. Res.*, 9:235–284, jun 2008. ISSN 1532-4435.

[27] S. Lloyd. Least squares quantization in pcm. *IEEE Transactions on Information Theory*, 28(2):129–137, 1982. doi: 10.1109/TIT.1982.1056489.

[28] Jorge Cortes, Sonia Martinez, Timur Karatas, and Francesco Bullo. Coverage control for mobile sensing networks. *IEEE Transactions on robotics and Automation*, 20(2): 243–255, 2004.

[29] Francois Lekien and Naomi Ehrich Leonard. Nonuniform coverage and cartograms. *SIAM Journal on Control and Optimization*, 48(1):351–372, 2009. doi: 10.1137/070681120. URL `https://doi.org/10.1137/070681120`.

[30] Yunfei Xu and Jongeun Choi. Adaptive sampling for learning gaussian processes using mobile sensor networks. *Sensors (Basel, Switzerland)*, 11:3051–66, 12 2011. doi: 10.3390/s110303051.

[31] Wenhao Luo and Katia Sycara. Adaptive sampling and online learning in multi-robot sensor coverage with mixture of gaussian processes. pages 6359–6364, 05 2018. doi: 10.1109/ICRA.2018.8460473.

[32] Alessia Benevento, María Santos, Giuseppe Notarstefano, Kamran Paynabar, Matthieu Bloch, and Magnus Egerstedt. Multi-robot coordination for estimation and coverage of unknown spatial fields. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 7740–7746, 2020. doi: 10.1109/ICRA40945.2020.9197487.

[33] Andrea Carron and Melanie N. Zeilinger. Model predictive coverage control∗∗this work was supported by the swiss national centre of competence in research nccr digital fabrication, and by the eth career seed grant 19-18-2. *IFAC-PapersOnLine*, 53(2):6107–6112, 2020. ISSN 2405-8963. doi: https://doi.org/10.1016/j.ifacol.2020.12.1686. URL `https://www.sciencedirect.com/science/article/pii/S2405896320322898`. 21st IFAC World Congress.

[34] Soobum Kim, María Santos, Luis Guerrero-Bonilla, Anthony Yezzi, and Magnus Egerstedt. Coverage control of mobile robots with different maximum speeds for time-sensitive applications. *IEEE Robotics and Automation Letters*, 7(2):3001–3007, 2022. doi: 10.1109/LRA.2022.3146593.

[35] Francesco Bullo, Ruggero Carli, and Paolo Frasca. Gossip coverage control for robotic networks: Dynamical systems on the space of partitions. *SIAM Journal on Control and Optimization*, 50(1):419–447, 2012.

[36] Matthew Streeter, Daniel Golovin, and Stephen F Smith. Combining multiple heuristics online. In *AAAI*, pages 1197–1203, 2007.

[37] 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.

[38] Yisong Yue and Carlos Guestrin. Linear submodular bandits and their application to diversified retrieval. In J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 24. Curran Associates, Inc., 2011. URL https://proceedings.neurips.cc/paper/2011/file/33ebd5b07dc7e407752fe773eed20635-Paper.pdf.

[39] Lukas Hewing, Kim P Wabersich, Marcel Menner, and Melanie N Zeilinger. Learning-based model predictive control: Toward safe learning in control. *Annual Review of Control, Robotics, and Autonomous Systems*, 3:269–296, 2020.

[40] Matteo Turchetta, Andrey Kolobov, Shital Shah, Andreas Krause, and Alekh Agarwal. Safe reinforcement learning via curriculum induction. *Advances in Neural Information Processing Systems*, 33:12151–12162, 2020.

[41] Felix Berkenkamp, Matteo Turchetta, Angela Schoellig, and Andreas Krause. Safe model-based reinforcement learning with stability guarantees. *Advances in neural information processing systems*, 30, 2017.

[42] Yinlam Chow, Ofir Nachum, Edgar Duenez-Guzman, and Mohammad Ghavamzadeh. A lyapunov-based approach to safe reinforcement learning. *Advances in neural information processing systems*, 31, 2018.

[43] Jaime F Fisac, Anayo K Akametalu, Melanie N Zeilinger, Shahab Kaynama, Jeremy Gillula, and Claire J Tomlin. A general safety framework for learning-based control in uncertain robotic systems. *IEEE Transactions on Automatic Control*, 64(7):2737–2752, 2018.

[44] Joshua Achiam, David Held, Aviv Tamar, and Pieter Abbeel. Constrained policy optimization. In *International conference on machine learning*, pages 22–31. PMLR, 2017.

[45] Jeremy Coulson, John Lygeros, and Florian Dorfler. Distributionally robust chance constrained data-enabled predictive control. *IEEE Transactions on Automatic Control*, 2021.

[46] Lukas Brunke, Melissa Greeff, Adam W Hall, Zhaocong Yuan, Siqi Zhou, Jacopo Panerati, and Angela P Schoellig. Safe learning in robotics: From learning-based control to safe reinforcement learning. *Annual Review of Control, Robotics, and Autonomous Systems*, 5, 2021.

[47] Alex Ray, Joshua Achiam, and Dario Amodei. Benchmarking safe exploration in deep reinforcement learning. *arXiv preprint arXiv:1910.01708*, 7:1, 2019.

[48] Jan Leike, Miljan Martic, Victoria Krakovna, Pedro A Ortega, Tom Everitt, Andrew Lefrancq, Laurent Orseau, and Shane Legg. Ai safety gridworlds. *arXiv preprint arXiv:1711.09883*, 2017.

[49] Yanan Sui, Alkis Gotovos, Joel Burdick, and Andreas Krause. Safe exploration for optimization with gaussian processes. In Francis Bach and David Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 997–1005, Lille, France, 07–09 Jul 2015. PMLR. URL `https://proceedings.mlr.press/v37/sui15.html`.

[50] Yanan Sui, Vincent Zhuang, Joel W. Burdick, and Yisong Yue. Stagewise safe bayesian optimization with gaussian processes, 2018. URL `https://arxiv.org/abs/1806.07555`.

[51] Matteo Turchetta, Felix Berkenkamp, and Andreas Krause. Safe exploration in finite markov decision processes with gaussian processes. *Advances in Neural Information Processing Systems*, 29, 2016.

[52] Akifumi Wachi and Yanan Sui. Safe reinforcement learning in constrained markov decision processes. In *ICML*, pages 9797–9806, 2020. URL `http://proceedings.mlr.press/v119/wachi20a.html`.

[53] Zheqing Zhu, Erdem Bıyık, and Dorsa Sadigh. Multi-agent safe planning with gaussian processes. In *International Conference on Intelligent Robots and Systems (IROS)*, pages 6260–6267, 10 2020. doi: 10.1109/IROS45743.2020.9341169.

[54] Bernhard Schlkopf, Alexander J. Smola, and Francis Bach. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. The MIT Press, 2018. ISBN 0262536579.

[55] Open weather. `https://openweathermap.org/`.

[56] Andreas Krause and Daniel Golovin. Submodular function maximization. *Tractability*, 3:71–104, 2014.

[57] Maximilian Balandat, Brian Karrer, Daniel Jiang, Samuel Daulton, Ben Letham, Andrew G Wilson, and Eytan Bakshy. Botorch: a framework for efficient monte-carlo bayesian optimization. *Advances in neural information processing systems*, 33:21524–21538, 2020.

[58] Jacob Gardner, Geoff Pleiss, Kilian Q Weinberger, David Bindel, and Andrew G Wilson. Gpytorch: Blackbox matrix-matrix gaussian process inference with gpu acceleration. *Advances in neural information processing systems*, 31, 2018.

[59] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

Part I
Appendix

Table of Contents

Appendix A. SafeMaC

A.1 Background

Goal-oriented safe exploration. GoOSE [11] is a single-agent safe exploration algorithm that extends unconstrained methods to safety-critical cases. Concretely, it maintains an under and an over approximation of the feasible set, known as the pessimistic and optimistic safe sets. It preserves safety by restricting the agent to the pessimistic safe set. It efficiently explores the objective by letting the original unconstrained algorithm recommend locations within the optimistic safe set. If such recommendations are provably safe, the agent evaluates the objective there. Otherwise, it evaluates the constraint at a sequence of safe locations to prove that such recommendation is either safe, which allows it to evaluate the objective, or unsafe, which triggers the unconstrained algorithm to provide a new recommendation.

Assumptions. To guarantee safety, GoOSE makes two main assumptions. First, it assumes there is an initial set of safe locations, X_0, from where the agent can start exploring. Second, it assumes the constraint is sufficiently well-behaved, so that we can use data to infer the safety of unvisited locations. Formally, it assumes the domain V is endowed with a positive definite kernel $k^q(\cdot,\cdot)$, and that the constraint's norm in the associated *Reproducing Kernel Hilbert Space* [54] is bounded, $\|q\|_{k^q} \leq B_q$. This lets us use Gaussian Processes (GPs) [14] to construct high-probability confidence intervals for q. We specify the GP prior over q through a mean function, which we assume to be zero everywhere w.l.o.g., $\mu(v) = 0, \forall v \in V$, and a kernel function, k, that captures the covariance between different locations. If we have access to T measurements, at $V_T = \{v_t\}_{t=1}^T$ perturbed by i.i.d. Gaussian noise, $y_T = \{q(v_t) + \eta\}_{t=1}^T$ with $\eta \sim \mathcal{N}(0, \sigma^2)$, we can compute the posterior mean and covariance over the constraint at unseen locations v, v' as $\mu_T(v) = k_T^\top(v)(K_T + \sigma^2 I)^{-1} y_T$ and $k_t(v, v') = k(v, v') - k_T^\top(v)(K_T + \sigma^2 I)^{-1} k_T(v')$, where $k_T(v) = (k(v_1, v), ..., k(v_T, v))$, K_T is a the positive definite kernel matrix $[k(v, v')]_{v,v' \in V_T}$ and $I \in \mathbb{R}^{T \times T}$ denotes the identity matrix. In this work, we make the same assumptions about the safe seed and the regularity of q and ρ.

Approximations of the feasible set. Based on the GP posterior above, GoOSE builds monotonic confidence intervals for the constraint at each iteration t as $l_t^q(v) \coloneqq \max\{l_{t-1}^q(x), \mu_{t-1}^q(v) - \beta_t^q \sigma_{t-1}^q(v)\}$ and $u_t^q(v) \coloneqq \min\{u_{t-1}^q(x), \mu_{t-1}^q(v) + \beta_t^q \sigma_{t-1}^q(v)\}$, which contain the true constraint function for every $v \in V$ and $t \geq 1$, with high probability if β_t^q is selected as in [16] or Section 5. GoOSE uses these confidence intervals within a set $S \subseteq V$ together with the Lipschitz continuity of q to define operators that determine which locations are safe in a worst and best case scenarios,

$$p_t(S) = \{v \in V, | \exists z \in S : l^q(z) - L_q d(v, z) \geq 0\}, \quad (4)$$

$$o_t^{\epsilon_q}(S) = \{v \in V, | \exists z \in S : u^q(z) - \epsilon_q - L_q d(v, z) \geq 0\}. \quad (5)$$

Notice the pessimistic operator relies on the lower bound, l^q, while the optimistic one on the upper bound, u^q. Moreover, the optimistic one uses a margin ϵ_q to exclude "barely" safe locations as the agent might get stuck learning about them. Finally, to disregard locations the agent could not safely reach or from where it could not safely return, GoOSE introduces the $R^{\text{ergodic}}(\cdot,\cdot)$ operator. $R^{\text{ergodic}}(p_t(S), S)$ indicates locations in S or locations in $p_t(S)$ reachable from S and from where the agent can return to S along a path contained in $p_t(S)$. Combining $p_t(S)$ and $R^{\text{ergodic}}(\cdot,\cdot)$, GoOSE defines the pessimistic and ergodic operator $\tilde{P}_t(\cdot)$, which it uses to update the pessimistic safe set. Similarly, it defines $\tilde{O}_t(\cdot)$ using $o_t^{\epsilon_q}(\cdot)$ to compute the optimistic safe set.

A.2 SafeMaC: safety-constrained multi-agent coverage control

Intuition. We adopt a perspective similar to GoOSE as we separate the exploration of the safe set from the maximization of the coverage. Given an over and under approximation of the safe set (whose computation is discussed later), we want to explore optimistically optimal goals for each agent, similar to MaCOpt. To this end, we find the maximizers of the density upper bound in the optimistic safe set with the Greedy algorithm. Then, we define sampling goals to learn the coverage at those locations.

Phases of SafeMaC. Coverage values depend both on the density and the feasible region (Eq. (2)). Thus, there are two sensible sampling goals given a disk assignment: i) *optimistic coverage*: if we are uncertain about the density within the disks, we target locations with the highest density uncertainty (Line 6 of Algorithm 2); ii) *optimistic exploration*: if we know the density within the disk but there are locations under it that we cannot classify as either safe (in S^p) or unsafe (in $V \setminus S^{o,\epsilon_q}$), we target those with the highest constraint uncertainty among them (Line 8). If all the goal locations are safe with high probability, which can only happen during *optimistic coverage*, we safely evaluate the density there (Line 19). Otherwise, we explore the constraint with a goal directed strategy that aims at classifying them as either safe or unsafe similar to GoOSE (Line 9-12). In case this changes the topological connection of the optimistic feasible set, we recompute the disks as this may change Greedy's output (Line 15-17). We repeat this loop until we know the feasibility of all the points under the disks recommended by Greedy and their density uncertainty is low (Line 4). Next, we explain how the multiple agents coordinate their individual safe regions to evaluate a goal (MaCOpt *in batches*), how the agents progress toward their goals (*safe expansion*) and finally we describe SafeMaC *convergence*.

MaCOpt in batches. In the multi-agent setting of GoOSE (see Fig. 1b), each agent i maintains $S_t^{p,i}$ a pessimistic (or $S_t^{o,\epsilon_q,i}$ an optimistic) belief of the safe locations, obtained by iteratively applying $\tilde{P}_t(\cdot)$ the pessimistic (or $\tilde{O}_t(\cdot)$ the optimistic) ergodic operators (see Section 3) to the previous pessimistic belief $S_{t-1}^{p,i}$ (Line 11 of Algorithm 2). Since the agents cannot navigate to an arbitrary location in the constrained case, SafeMaC computes coverage maximizers on a restricted region, obtained by ignoring the known unsafe locations. To denote such a restricted region, we define a union set $S_t^{u,i} \coloneqq S_t^{o,\epsilon_q,i} \cup S_t^{p,i}$, which is the largest set known to be optimistically or pessimistically safe up to time t. Moreover, if the agents are topologically disconnected, they cannot travel from one safe region to another and the best strategy for any batch of agents is to maximize coverage locally. For this, we form a collection of batches $\mathcal{B}_t$, such that any batch $B \in \mathcal{B}_t$ contains agents that lie in topologically connected regions determined by the union set (Line 13-14). SafeMaC computes a Greedy solution for each $B \in \mathcal{B}_t$ in their corresponding $S_t^{u,B} \coloneqq \cup_{i\in B} S_t^{u,i}$. This is the largest set where the agents can find an optimistically safe path to travel. Analogous to $\mathcal{B}_t$, we define $\mathcal{B}_t^p$ as collection of batches where any $B \in \mathcal{B}_t^p$ contains agents which are topologically connected in pessimistic set and $S_t^{p,B} \coloneqq \cup_{i\in B} S_t^{p,i}$.

Safe expansion. Safe expansion is the sub-routine inspired by GoOSE for goal-oriented exploration of the safe set that we use to learn about the feasibility of sampling targets.

It uses a heuristic h to assign priority scores p to points that are optimistically but not pessimistically safe. Those determine locations whose feasibility is relevant to learn that of the sampling targets (Line 2 of Algorithm 4). A simple and effective choice for the heuristic is the inverse of the distance to the targets. Then, it identifies safe locations where the constraint is not yet known ϵ_q-accurately (Line 3). Among them, it determines the α-immediate expanders, i.e., those that could potentially add locations with priority α to the pessimistic set, $G_t^{\epsilon_q}(\alpha) = \{v \in W_t^{\epsilon_q} | \exists z \in A_t(\alpha) : u_t^q - L_q d(v, x) \geq 0\}$. In Line 4, it selects the non-empty α-expander set with the highest priority. In Line 6 - 7, the agent evaluates the constraint at the location with the highest uncertainty in this set (see [11] for details).
SafeMaC convergence. The *optimistic coverage* phase switches to *optimistic exploration* phase, when density uncertainty is under the disks is low ($w_t \leq \epsilon_\rho$). In the exploration, either the topological connection of the optimistic feasible set changes or will classify the uncertain region as pessimistically safe. In the former case, SafeMaC will recompute a new coverage location and switch to the coverage phase. Alternatively, if the uncertain region is pessimistically safe, SafeMaC is said to be converged since the density uncertainty in the exploration phase is already low. The phases show an interesting dynamics; SafeMaC continuously iterates between the *optimistic exploration* and the *optimistic coverage* phase until we know about the feasibility of the disk and their uncertainty is low. In the worst case, SafeMaC might explore the entire environment. In this case the sample complexity will be similar to a two-stage algorithm, where we explore the whole domain and then optimize coverage in the resulting known environment.However, in practice, SafeMaC is much better than this worst case.

A.3 Analysis

SafeMaC. This section presents our main result for safety-constrained multi-agent coverage control. In particular, Theorem 2 (proof in Appendix F) guarantees that SafeMaC safely achieves near-optimal safe coverage in finite time.

Theorem 2 *Let $\delta \in (0,1)$ and β_t^ρ as in [16], i.e., $\beta_t^{\rho 1/2} = B_\rho + 4\sigma_\rho\sqrt{\gamma_{Kt}^\rho + 1 + \ln(1/\delta)}$ and $t_\rho^\star$ be the smallest integer such that $\frac{t_\rho^\star}{\beta_{t_\rho^\star}\gamma_{Kt_\rho^\star}} \geq \frac{8K^2}{\log(1+K\sigma^{-2})\epsilon_\rho^2}$. Let β_t^q and $t_q^\star$ be defined analogously. Then, there exists $t < t_q^\star + t_\rho^\star$, such that with probability at least $1-\delta$*

$$\sum_{B \in \mathcal{B}_t} F(X_t^B; \rho, \bar{R}_0(X_0^B)) \geq (1 - \frac{1}{e}) \sum_{B \in \mathcal{B}} F(X_\star^B; \rho, \bar{R}_{\epsilon_q}(X_0^B)) - \epsilon_\rho. \tag{7}$$

The theoretical analysis has two components: (*i*) we show SafeMaC's coverage is near-optimal at convergence (Lem. 12), and (*ii*) we prove it converges in finite time. Since SafeMaC learns the constraint *and* the density, we must bound the sample complexity for both to prove (*ii*). For the constraint, we extend the results for single-agent GoOSE to our multi-agent setting (Appendix G).

For the density, we use results from Theorem 1 to show that, within a coverage phase, the cumulative regret is sublinear. Next, we use additivity of the information gain (Lem. 16) between any pair of coverage phases to bound the sample complexity of density for the subsequent coverage phases. Combining these results, we obtain Theorem 2.
Intermediate recommendation. Theorem 2 guarantees SafeMaC converges to a safe and near-optimal solution. Can it also make sensible recommendations before the stopping

Algorithm 2 SafeMac

Inputs X_0, L_q, ϵ_ρ, V, GP_ρ, GP_q
$\forall i,\ S_0^{p,i} \leftarrow X_0,\ S_0^{o,\epsilon_q,i} \leftarrow V,\ t \leftarrow 0$
$X_1, w_1 \leftarrow$ Greedy$(u_0^\rho, l_0^\rho, [K], V)$
while $\forall i, (S_t^{o,\epsilon_q,i} \setminus S_t^{p,i}) \cap D_t^i \neq \emptyset$ or $w_t > \epsilon_\rho$ **do**
  if $w_t > \epsilon_\rho$ **then**
    $\forall i, x_t^{g,i} \leftarrow \arg\max_{x \in D_t^{i-}} u_{t-1}^\rho(v) - l_{t-1}^\rho(v)$
  else
    $\forall i, x_t^{g,i} \leftarrow \arg\max_{x \in (S_{t-1}^{o,\epsilon_q,i} \setminus S_{t-1}^{p,i}) \cap D_t^i} u_{t-1}^q(v) - l_{t-1}^q(v)$
  if *for any* $i \in [K], x_t^{g,i} \notin S_t^{p,i}$ **then**
    SE$(S_{t-1}^{o,\epsilon_q,i}, S_{t-1}^{p,i}, x_t^{g,i})$
    $S_t^{p,i} \leftarrow \tilde{P}_t(S_{t-1}^{p,i}), S_t^{o,\epsilon_q,i} \leftarrow \tilde{O}_t^{\epsilon_q}(S_{t-1}^{p,i})$
    $t \leftarrow t+1$
  $\forall i,\ \mathcal{B}_t'(i) = \{j \in [K] | S_t^{u,i} \cap S_t^{u,j} \neq \emptyset\}$
  $\mathcal{B}_t = \bigcup_{i \in [K]} \mathcal{B}_t'(i)$
  if *for any* $B \in \mathcal{B}_t, S_t^{u,B} \neq S_{t-1}^{u,B}$ **then**
    $X_t, w_t \leftarrow$ Greedy$(u_{t-1}^\rho, u_{t-1}^\rho, B, S_t^{u,B})$
    $\forall i, x_t^{g,i} \leftarrow \arg\max_{x \in D_t^{i-}} u_{t-1}^\rho(v) - l_{t-1}^\rho(v)$
  if $\forall i, x_t^{g,i} \in S_t^{p,i}$ and $w_t > \epsilon_\rho$ **then**
    $\forall i, y_{\rho_t}^i = \rho(x_t^{g,i}) + \eta_\rho$, Update GP
    update GP i.e, compute u_t^ρ, l_t^ρ
    $t \leftarrow t+1$
    $X_t, w_t \leftarrow$ Greedy$(u_{t-1}^\rho, u_{t-1}^\rho, B, S_{t-1}^{u,B})$

Recommend X_t

Algorithm 3 Greedy UCB (Greedy)

Inputs $u_{t-1}^\rho, l_{t-1}^\rho, B, S_t^u$
for $i = 1, 2, ..., |B|$ **do**
  $x_t^i \leftarrow \arg\max_{x^i} \sum_{v \in D^i \setminus D_t^{1:i-1} \cap S_t^u} u_{t-1}^\rho(v)$
  $x_t^{g,i} \leftarrow \arg\max_{v \in D^i \setminus D_t^{1:i-1} \cap S_t^u} u_{t-1}^\rho(v) - l_{t-1}^\rho(v)$
$w_t \leftarrow \sum_{i=1}^{|B|} u_{t-1}^\rho(x_t^{g,i}) - l_{t-1}^\rho(x_t^{g,i})$
Return X_t^B, w_t

Algorithm 4 Safe Expansion (SE)

Inputs $S_t^{o,\epsilon_q}, S_t^p, x_t^g$
$A_t(p) \leftarrow \{v \in S_t^{o,\epsilon_q} \setminus p_t(S_t^p) | h(v) = p\}$
$W_t^{\epsilon_q} \leftarrow \{v \in S_t^p | u_t^q(v) - l_t^q(v) > \epsilon_q\}$
$\alpha^\star \leftarrow \max_\alpha\ s.t.\ |G_t^{\epsilon_q}(\alpha)| > 0$
if Optimization problem feasible **then**
  $v_t \leftarrow \arg\max_{v \in G_t^{\epsilon_q}(\alpha^\star)} u_t^q(v) - l_t^q(v)$
  Update GP with $y_t = q(v_t) + \eta_q$

criteria are met? Ideally, such recommendations should (i) be safely reachable and (ii) ensure a minimum coverage. To satisfy (i), they should be in the pessimistic safe set, S_t^p. To satisfy (ii), their coverage should be computed according to $F(\cdot; l_{t-1}^{\rho}, S_t^p)$, i.e., assuming a worst-case density, l_{t-1}^{ρ}, and a worst-case feasible set, S_t^p. If the greedy recommendation X_t is in S_t^p, we can recommend it at intermediate steps. However, this is not always the case and we need an alternative. To this end, we compute $X_t^{l,B}$, i.e., the greedy solution w.r.t. the worst-case objective, $F(\cdot; l_{t-1}^{\rho}, S_t^{p,B})\, \forall B \in \mathcal{B}_t^p$. At any time T, SafeMaC recommends the best of either strategy up to time T according to the worst-case objective.

In Appendix F.1, we show that such recommendation is also near optimal at convergence.

A.4 SafeMaC experiments

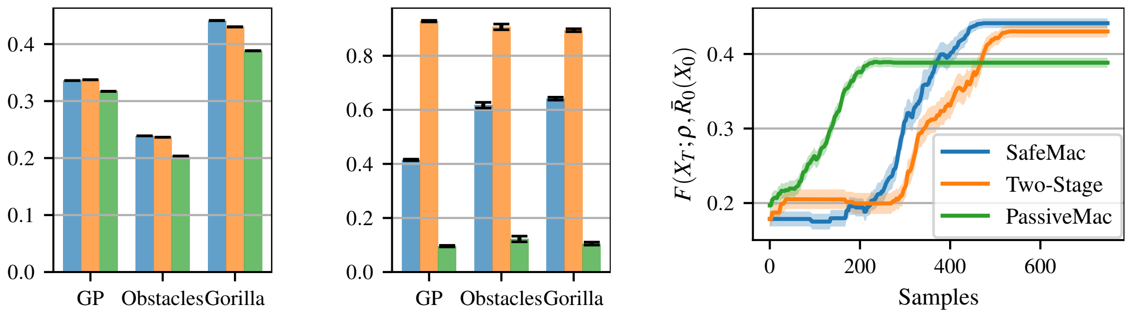
Environments. Below, we present the 3 environments we consider.

i) In *synthetic data*, both the density ρ and the constrain q are sampled from a GP with zero mean and Matérn Kernel with $\nu = 2.5$, scale $\sigma_k = 1$, and lengthscale $l = 2$. The observations are perturbed by i.i.d noise $\mathcal{N}(0, 10^{-3})$.

ii) In *obstacles*, we sample maps with several block-shaped obstacles (Fig. 5a) and we aim to maximize coverage while avoiding dangerous collisions. At v, each agent senses the distance to the nearest obstacle $d_m(v)$, which could be given by sensors such 1D-Lidars. We use $q'(v) = 1/(1 + \exp(-1.5 d_m(v)))$, to map the distance between $[0, 3]$ and saturate the constraint value for large distances, and we set $q(v) = q'(v) - 0.5$ to avoid collisions. The density is sampled from the same GP as the synthetic case.

iii) In *gorilla nest*, the Kagwene Gorilla Sanctuary (Fig. 2a) has regions affected by adverse weather (e.g. rain and storms) which are unsafe for the drone due to higher chances of crashes and should be avoided. This forms a constrained case of gorilla nest environment explained in Section 6. As a proxy for bad weather, we use the cloud coverage data over the KGS from OpenWeather [55].

SafeMaC. We compare SafeMaC with two baselines, i) a two-stage algorithm [52], that first fully explores the feasible region, and then uses MaCOpt to maximize the coverage ii) PassiveMaC, a baseline inspired by [49] that runs MaCOpt in the pessimistic set and passively measures the constraint in the process. Figs. 3a and 3b show the coverage at convergence and the number of samples to converge for SafeMaC and the two baselines across all the environments. PassiveMaC converges quickly but gets stuck in a local optimum as it does not actively explore the constraint. SafeMaC and Two-Stage converge to much higher coverage values. However, SafeMaC is more sample efficient thanks to its goal-oriented exploration. The results are averaged over 50 instances produced using different seeds and samples for every environment. Fig. 3c shows the coverage value of the intermediate safe recommendations (Section 5) in the *gorilla* environment as a function of the number of samples. It confirms the previous results: SafeMaC finds solutions comparable to Two-Stage more efficiently and PassiveMaC gets stuck in a local optimum.



(a) Coverage at convergence (b) Total number of samples (c) Coverage on Gorilla nests (Rainy day)

Figure 3: Comparison of SafeMac with PassiveMac and Two-Stage in all environments at convergence (a) and (b) and during optimization for the gorilla environment in (c).

Appendix B. Definitions

B.1 Notations

<u>Problem Formulation</u>

F	$\triangleq$	Submodular function, $F : 2^V \to \mathbb{R}$
V	$\triangleq$	Domain
v	$\triangleq$	An element in the domain V
$F(X; \rho, V)$	$\triangleq$	Coverage objective defined in Eq. (1)
i	$\triangleq$	Agent index
ρ	$\triangleq$	Density function, $\rho : V \to \mathbb{R}$
q	$\triangleq$	Constraint function, $q : V \to \mathbb{R}$
D^i	$\triangleq$	Sensing region around agent i
$D^{1:i}$	$\triangleq$	$\cup_{j=1}^{i} D^j$, union of sensing regions of agents $1 : i$
D^{i-}	$\triangleq$	$D^i \setminus D^{1:i-1}$, region occupied by agent i, but not by $1 : i-1$ agents
$\tilde{D}^i$	$\triangleq$	Sensing region occupied by greedy optimal location of agent i
$\tilde{D}^{i-}$	$\triangleq$	$\tilde{D}^i \setminus D^{1:i-1}$
N	$\triangleq$	Largest number of elements in D^i for any $x^i \in V$
K	$\triangleq$	Total number of agents

<u>Batch Operation</u>

B	$\triangleq$	A batch of agents, $\{1, 2 \ldots \lvert B \rvert\}$
$\mathcal{B}'_t(i)$	$\triangleq$	$\{j \in [K] \lvert S_t^{u,i} \cap S_t^{u,j} \neq \emptyset\}$, agents connected in union set with agent i
$\mathcal{B}_t$	$\triangleq$	$\bigcup_{i \in [K]} \mathcal{B}'_t(i)$. Collection of batches sharing the union set.
$\mathcal{B}$	$\triangleq$	Collection of batches sharing the largest reachable set $(\bar{R}_{\epsilon_q}(X_0^B))$
$\mathcal{B}_t^p$	$\triangleq$	Collection of batches sharing the pessimistic set

X Notations		
x_t^i	$\triangleq$	Planned location of agent i at time t
$x_t^{g,i}$	$\triangleq$	Goal of agent i at time t, defined by Line 6 and Line 8 in Algorithm 2
$\tilde{x}^i$	$\triangleq$	Greedy optimal location of agent i, Eq. (19)
X_t	$\triangleq$	$\cup_{i\in[K]}\{x_t^i\}$, A set of agents at time t
X_t^B	$\triangleq$	$\cup_{i\in B}\{x_t^i\}$, A set of agents in batch B at time t
$X_\star^B$	$\triangleq$	Optimal location of agents in batch B
$X_\star$	$\triangleq$	$\cup_{B\in\mathcal{B}}X_\star^B$
$X^{1:i}$	$\triangleq$	A set of agents 1 to i
$x_{1:T}^{g,1:K}$	$\triangleq$	A set of $1:K$ agents' goal locations up to time T
		Density (ρ) and Constraint (q) GP
l_t^q	$\triangleq$	Lower confidence bound of the constraint at time t
u_t^q	$\triangleq$	Upper confidence bound of the constraint at time t
β_t^q	$\triangleq$	Scaling, defined as per [16]
L_q	$\triangleq$	Lipschitz constant
ϵ_q	$\triangleq$	Statistical confidence up to which constraint function q is learnt
$d(v,z)$	$\triangleq$	Distance metric
σ_q	$\triangleq$	Standard deviation of constraint observations noise
σ_t^q	$\triangleq$	Posterior standard deviation of constraint GP
B_q	$\triangleq$	Norm bound of the constraint function, $\|q\|_{k^q} \le B_q$
η_q	$\triangleq$	Noise in constraint observations
l_t^ρ	$\triangleq$	Lower confidence bound of the density at time t
u_t^ρ	$\triangleq$	Upper confidence bound of the density at time t
β_t^ρ	$\triangleq$	Scaling, defined as per [16]
w_t	$\triangleq$	$\sum_{i=1}^K u_{t-1}^\rho(x_t^{g,i}) - l_{t-1}^\rho(x_t^{g,i})$, sum of highest uncertainty below disks
ϵ_ρ	$\triangleq$	Accuracy threshold for learning the density, $w \le \epsilon_\rho$
σ_t^ρ	$\triangleq$	Posterior standard deviation of denisty GP
σ_ρ	$\triangleq$	Standard deviation of density observations noise
B_ρ	$\triangleq$	Norm bound of the density function, $\|\rho\|_{k^\rho} \le B_\rho$
δ	$\triangleq$	$\in (0,1)$ for high probability argument
$H(y_A)$	$\triangleq$	Shannon entropy
$I(y_A;\rho)$	$\triangleq$	$H(y_A) - H(y_A\|\rho)$, Information gain
γ	$\triangleq$	Information capacity
γ_{KT}^ρ	$\triangleq$	$\sup_{A\subset V} I(Y_A;\rho)$, A is set of KT obs. $\gamma_{KT}^\rho \coloneqq \gamma_{KT_\rho}, \rho$ is clear in T.
γ_{KT}^q	$\triangleq$	$\sup_{A\subset V} I(Y_A;q)$, A is set of KT obs. $\gamma_{KT}^q \coloneqq \gamma_{KT_q}, q$ is clear in T.
Tr	$\triangleq$	Trace of a Matrix

K^ρ	$\triangleq$	Posterior kernel matrix with density observations
$\lambda_{i,t}$	$\triangleq$	Eigenvalue of the kernel matrix
η_ρ	$\triangleq$	Noise in the density observations

Time

t	$\triangleq$	Any round of the algorithm
T	$\triangleq$	Time at which the algorithm gets terminated
$t^\star_q$	$\triangleq$	Maximum number of constraint observations
$t^\star_\rho$	$\triangleq$	Maximum number of density observations
$t^{\star 1}_\rho$	$\triangleq$	Maximum number of density observations for the first coverage phase
$\delta t^{\star n}_\rho$	$\triangleq$	Maximum number of density obs. from $(n-1)^{th}$ to n^{th} coverage phase
δt^n_ρ	$\triangleq$	Number of density obs. from $(n-1)^{th}$ to n^{th} coverage phase
t^n_ρ	$\triangleq$	Number of density obs. till n^{th} coverage phase

GoOSE and Safe Expansion

$p_t(S)$	$\triangleq$	pessimistic operator $\{v \in V, \lvert \exists z \in S : l^q_t(z) - L_q d(v,z) \geq 0\}$
$o^{\epsilon_q}_t(S)$	$\triangleq$	optimistic operator $\{v \in V, \lvert \exists z \in S : u^q_t(z) - \epsilon_q - L_q d(v,z) \geq 0\}$
$\tilde{P}_t(\cdot)$	$\triangleq$	Pessimistic expansion operator
$\tilde{O}_t(\cdot)$	$\triangleq$	Optimistic expansion operator
$\bar{R}_{\epsilon_q}(\{x^i_0\})$	$\triangleq$	Maximum safely reachable set up to ϵ_q, Eq. (12)
$\bar{R}_{\epsilon_q}(X^B_0)$	$\triangleq$	$\cup_{i\in B}\bar{R}_{\epsilon_q}(\{x^i_0\})$
$S^{p,i}_t$	$\triangleq$	Pessimistic set of agent i, $\tilde{P}_t(S^{p,i}_{t-1})$
$S^{p,B}_t$	$\triangleq$	$\cup_{i\in B} S^{p,i}_t$
S^p_t	$\triangleq$	Pessimistic set of all K agents
$S^{o,\epsilon_q,i}_t$	$\triangleq$	Optimistic set of agent i, $\tilde{O}^{\epsilon_q}_t(S^{p,i}_{t-1})$
$S^{o,\epsilon_q,B}_t$	$\triangleq$	$\cup_{i\in B} S^{o,\epsilon_q,i}_t$
S^{o,ϵ_q}_t	$\triangleq$	Optimistic set of all K agents
$S^{u,i}_t$	$\triangleq$	Union set, $S^{o,\epsilon_q,i}_t \cup S^{p,i}_t$
$S^{u,B}_t$	$\triangleq$	$\cup_{i\in B} S^{u,i}_t$
S^u_t	$\triangleq$	Union set of all K agents
$R^{\text{safe}}_{\epsilon_q}(S)$	$\triangleq$	True safety constraint operator, Eq. (8)
$R^{\text{reach}}_n(S)$	$\triangleq$	n step reachability in the graph, Eq. (9)
$\tilde{R}^{\text{reach}}(S)$	$\triangleq$	$\lim_{n\to\infty} R^{\text{reach}}_n(S)$
$R^{\text{n}}_{\epsilon_q}(S)$	$\triangleq$	n step safely reachable set in the graph, Eq. (12)
$\bar{R}_{\epsilon_q}(S)$	$\triangleq$	$\lim_{n\to\infty} R^{\text{n}}_{\epsilon_q}(S)$
$W^{\epsilon_q}_t$	$\triangleq$	Set of locations whose safety is not ϵ_q-accurate, Algorithm 4
$G^{\epsilon_q}_t(\alpha)$	$\triangleq$	A set of potential immediate expanders, Algorithm 4
p	$\triangleq$	Priority, Algorithm 4
$h(v)$	$\triangleq$	Heuristic function, Algorithm 4
$A_t(\alpha)$	$\triangleq$	Subset of locations with equal priority, Algorithm 4

Regret

$F(X)$	$\triangleq$	$F(X;\rho,V)$, short notation when ρ and V are obvious
$\Delta(x^i \vert X^{1:i-1};\rho,V)$	$\triangleq$	Marginal coverage gain by agent i, Eq. (17)
$\Delta(x^i \vert X^{1:i-1})$	$\triangleq$	$\Delta(x^i \vert X^{1:i-1};\rho,V)$, short notation when ρ and V are obvious
$Reg_{act}(T)$	$\triangleq$	Actual regret in unconstrained case, Eq. (3)
OPT_l^i	$\triangleq$	Per agent cumulative optimal, Eq. (21)
$Reg^i(T)$	$\triangleq$	Per agent regret, Eq. (22)
OPT	$\triangleq$	$\sum_{t=1}^{T} F(X_\star)$
r_t^{act}	$\triangleq$	Simple actual regret, constrained case, Eq. (29)
r_t^{O}	$\triangleq$	Simple actual regret in union set, constrained case, Eq. (29)
r_t	$\triangleq$	Simple per agent regret, constrained case, Eq. (29)
$Reg_{act}^{\mathrm{O}}(T)$	$\triangleq$	Cumulative actual regret, Eq. (30)
$Reg_l^{\mathrm{O}}(T)$	$\triangleq$	Sum of cumulative per agent regret, Eq. (30)

B.2 GoOSE operators

We denote with $\mathcal{G}=(V,\mathcal{E})$ the undirected graph describing the dependency among locations, V indicates the vertices of the graph, i.e., the state space of the problem and $\mathcal{E}\subseteq V\times V$ denotes the edges. In our setting, there are K identical agents having the same transition dynamics. Each agent can have a separate $\tilde{R}_{\epsilon_q}(\{x_0^i\})$.

The baseline as per true safety constraint operator:

$$R_{\epsilon_q}^{\text{safe}}(S) = S \cup \{v \in V\backslash S, |\exists z \in S : q(z) - \epsilon_q - L_q d(v,z) \geq 0\} \tag{8}$$

Now, we define reachability operator as all the locations that can be reached starting from set S.

$$\begin{aligned} R^{\text{reach}}(S) &= S \cup \{v \in V\backslash S, |\exists z \in S : (z,v) \in \mathcal{E}\}, \\ R_n^{\text{reach}}(S) &= R_n^{\text{reach}}(R_{n-1}^{\text{reach}}(S)) \text{ with } R_1^{\text{reach}}(S) = R^{\text{reach}}(S) && (9) \\ \tilde{R}^{\text{reach}}(S) &= \lim_{n\to\infty} R_n^{\text{reach}}(S), && (10) \end{aligned}$$

For defining $\bar{R}_{\epsilon_q}(S)$,

$$\begin{aligned} R_{\epsilon_q}(S) &= R_{\epsilon_q}^{\text{safe}}(S) \cap \tilde{R}^{\text{reach}}(S) \\ R_{\epsilon_q}^{\text{n}}(S) &= R_{\epsilon_q}(R_{\epsilon_q}^{\text{n-1}}(S)) \ \text{ with } R_{\epsilon_q}^{1}(S) = R_{\epsilon_q}(S) && (11) \\ \bar{R}_{\epsilon_q}(S) &= \lim_{n\to\infty} R_{\epsilon_q}^{\text{n}}(S) && (12) \end{aligned}$$

Optimistic and pessimistic constrain satisfaction operators:

$$\begin{aligned} o_t^{\epsilon_q}(S) &= \{v \in V, |\exists z \in S : u_t^q(z) - \epsilon_q - L_q d(v,z) \geq 0\} \\ p_t^{\epsilon_q}(S) &= \{v \in V, |\exists z \in S : l_t^q(z) - \epsilon_q - L_q d(v,z) \geq 0\} \end{aligned}$$

In this section, for simplicity, we have considered an undirected graph. This results in the same reachability and returnability operators since the edges are bidirectional. The extension to the directed graph is easy by using the reachability, the returnability and the ergodic operator. (Appendix A of Turchetta et al. [11] does it for the directed graph, so we did not repeat it here)

The optimistic and pessimistic expansion operators are given by,

$$
\begin{aligned}
O_t^{\epsilon_q}(S) &= o_t^{\epsilon_q}(S) \cap \tilde{R}^{\text{reach}}(S) \\
O_t^{\epsilon_q,n}(S) &= O_t^{\epsilon_q}(O_t^{\epsilon_q,n-1}(S)) \text{ with } O_t^{\epsilon_q,1}(S) = O_t^{\epsilon_q}(S) \\
\tilde{O}_t^{\epsilon_q}(S) &= \lim_{n\to\infty} O_t^{\epsilon_q,n}(S)
\end{aligned}
$$

Pessimistic expansion operator

$$
\begin{aligned}
P_t^{\epsilon_q}(S) &= p_t^{\epsilon_q}(S) \cap \tilde{R}^{\text{reach}}(S) \\
P_t^{\epsilon_q,n}(S) &= P_t^{\epsilon_q}(P_t^{\epsilon_q,n-1}(S)) \text{ with } P_t^{\epsilon_q,1}(S) = P_t^{\epsilon_q}(S) \\
\tilde{P}_t^{\epsilon_q}(S) &= \lim_{n\to\infty} P_t^{\epsilon_q,n}(S)
\end{aligned}
$$

This gives the optimistically and pessimistically, safe and reachable set respectively as:

$$
\begin{aligned}
S_t^{o,\epsilon_q} &= \tilde{O}_t^{\epsilon_q}(S_{t-1}^p) \\
S_t^p &= \tilde{P}_t^0(S_{t-1}^p)
\end{aligned}
$$

Now in our setting with K agents, we denote with $S_t^{o,\epsilon_q,i}$ and $S_t^{p,i}$, the optimistic and the pessimistic set respectively of agent i. The union set for any agent i is defined as,

$$
S_t^{u,i} := S_t^{o,\epsilon_q,i} \cup S_t^{p,i} \tag{13}
$$

B.3 Batching operation

For a set of agents, we partition them in batches, such that each batch B contains the agents that share at least a node in the union set. The total collection of batches, $\mathcal{B}$, is defined as,

$$
\mathcal{B}_t = \bigcup\nolimits_{i\in[K]} \mathcal{B}_t'(i) \quad where \quad \mathcal{B}_t'(i) = \{j \in [K] \,|\, S_t^{u,i} \cap S_t^{u,j} \neq \emptyset\} \tag{14}
$$

Analogous to $\mathcal{B}_t$, we define $\mathcal{B}_t^p$ (or $\mathcal{B}$) as collection of batches where any $B \in \mathcal{B}_t^p$ (or $\mathcal{B}$) contains agents which are topologically connected in the pessimistic (or maximum safely reachable) set. Precisely,

$$
\mathcal{B}_t^p = \bigcup\nolimits_{i\in[K]} \mathcal{B}_t'(i) \quad where \quad \mathcal{B}_t'(i) = \{j \in [K] \,|\, S_t^{p,i} \cap S_t^{p,j} \neq \emptyset\} \tag{15}
$$

$$
\mathcal{B} = \bigcup\nolimits_{i\in[K]} \mathcal{B}'(i) \quad where \quad \mathcal{B}'(i) = \{j \in [K] \,|\, \bar{R}_{\epsilon_q}(\{x_0^i\}) \cap \bar{R}_{\epsilon_q}(\{x_0^j\}) \neq \emptyset\} \tag{16}
$$

The resulting batch collection are mutually exclusive that is $\forall\ B_1, B_2 \in \mathcal{B}_t, B_1 \neq B_2, B_1 \cap B_2 = \emptyset$ and also, $\sum_{B\in\mathcal{B}_t} |B| = K$.

For any batch B we can define their combined union set, pessimistic set and the maximum safely reachable set as ,

$$
S_t^{u,B} := \cup_{i\in B} S_t^{u,i}, \quad S_t^{p,B} := \cup_{i\in B} S_t^{p,i}, \quad \bar{R}_{\epsilon_q}(X_0^B) = \cup_{i\in B} \bar{R}_{\epsilon_q}(\{x_0^i\}).
$$

Appendix C. Disk Coverage as a submodular function

Set functions Function $F : 2^V \rightarrow \mathbb{R}$ that assign each subset $A \subseteq V$ a value $F(A)$.

Discrete Derivative For a set function $F : 2^V \rightarrow R$, $A \subseteq V$, and $e \in V$, let $\Delta_F(e|A) := F(A \cup \{e\}) - F(A)$ is discrete derivative of F at A with respect to e.

Submodular functions A function F(.) is a submodular if, $\forall A \subseteq B \subseteq V$ and $\forall e \in V \backslash B$

$$F(A \cup \{e\}) - F(A) \geq F(B \cup \{e\}) - F(B),$$
$$\Delta_F(e|A) \geq \Delta_F(e|B).$$

For the disk coverage function $F(A)$, defined in Eq. (1),

$$F(X; \rho, V) = \sum_{x^i \in X} \sum_{v \in D^{i-}} \rho(v)/N,$$

We can write marignal gain as,

$$\begin{aligned}
F(A \cup \{e\}) - F(A) &= \sum_{x^i \in A \cup \{e\}} \sum_{v \in D^{i-}} \rho(v)/N - \sum_{x^i \in A} \sum_{v \in D^{i-}} \rho(v)/N \\
&= \sum_{x^i \in A} \sum_{v \in D^{i-}} \rho(v)/N + \sum_{x^i \in \{e\}} \sum_{v \in D^i \backslash D^{1:|A|}} \rho(v)/N - \sum_{x^i \in A} \sum_{v \in D^{i-}} \rho(v)/N \\
&= \sum_{x^i \in \{e\}} \sum_{v \in D^i \backslash D^{1:|A|}} \rho(v)/N \\
&\geq \sum_{x^i \in \{e\}} \sum_{v \in D^i \backslash D^{1:|B|}} \rho(v)/N \quad \textit{(Since, } A \subseteq B, |D^i \setminus D^{1:|A|}| \geq |D^i \setminus D^{1:|B|}| \\
&= \sum_{x^i \in B \cup \{e\}} \sum_{v \in D^{i-}} \rho(v)/N - \sum_{x^i \in B} \sum_{v \in D^{i-}} \rho(v)/N \\
&= F(B \cup \{e\}) - F(B)
\end{aligned}$$
$$\implies F(A \cup \{e\}) - F(A) \geq F(B \cup \{e\}) - F(B)$$

This shows that the coverage function defined in Eq. (1) is a Submodular function.

Monotonicity is directly implied by the definition of $F(A)$, as an additive function of ρ. Since, $\rho(v) \geq 0, \forall v \in V \implies F(A) \leq F(B)$, if $A \subseteq B$.

Appendix D. Agent wise regret bound

In this section, we upper bound the actual ("greedy") regret with the per agent regret in the unconstrained and the constrained case. The proof methodology to bound with per agent regret is motivated from [38]. We first define marginal gain and agent-wise regret. Then we give a proposition for the submodularity rate equation, which will be central to our lemmas. Finally, we bound the actual regret with the sum of per agent regret for unconstrained and then constrained case in

Marginal coverage gain:

$$\begin{aligned}\Delta(x_t^i|X_t^{1:i-1};\rho,V) &= F(X_t^{1:i-1}\cup\{x_t^i\};\rho,V) - F(X_t^{1:i-1};\rho,V)\\ &= \sum_{x_t^i\in X_t^{1:i}}\sum_{v\in D_t^{i-}}\rho(v)/N - \sum_{x_t^i\in X_t^{1:i-1}}\sum_{v\in D_t^{i-}}\rho(v)/N\\ &= \sum_{v\in D_t^{i-}}\rho(v)/N \end{aligned} \tag{17}$$

Using, $X^{1:0}=\{\emptyset\}, F(X^{1:0})=0$, it follows that,

$$\sum_{i=1}^{K}\Delta(x_t^i|X_t^{1:i-1};\rho,V) = F(X_t^{1:K};\rho,V) \tag{18}$$

Tilde Notations:

$$\tilde{x}_t^i = \arg\max_{x_t^i}\Delta(x_t^i|X_t^{1:i-1};\rho,V) \tag{19}$$

Proposition 1 (Eq. (3-7), [56], Submodular rate equation) *For a monotone Submodular function F the following holds,*

$$\max_{x^i} F(X^{1:i-1}\cup\{x^i\}) - F(X^{1:i-1}) \geq \frac{F(X_\star)-F(X^{1:i-1})}{K}, \tag{20}$$

where $X^{1:i}$ is the set of i agents being picked greedily and K is the number of agents in $X_\star$.

Proof Let $X_\star=\{x_\star^1,\ldots,x_\star^K\}$

$$\begin{aligned}F(X_\star) &\leq F(X_\star\cup X^{1:i-1}) && \text{(With monotonicity of } F\text{)}\\ &= F(X^{1:i-1}) + \sum_{j=1}^{K}\Delta(x_\star^j|X^{1:i-1}\cup\{x_\star^1,\ldots,x_\star^{j-1}\}) && \text{(Telescopic sum)}\\ &\leq F(X^{1:i-1}) + \sum_{x\in X_\star}\Delta(x|X^{1:i-1}) && \text{(Follows by Submodularity of } F\text{)}\\ &\leq F(X^{1:i-1}) + \sum_{x\in X_\star}(F(X^{1:i})-F(X^{1:i-1}))\\ & && \text{(since, } x^i \text{ is added greedily to maximize } \Delta(x|X^{1:i-1})\text{)}\\ &\leq F(X^{1:i-1}) + K(F(X^{1:i})-F(X^{1:i-1})) && (K \text{ agents in } X_\star)\end{aligned}$$

$$\implies \frac{F(X_\star) - F(X^{1:i-1})}{K} \leq F(X^{1:i}) - F(X^{1:i-1})$$

The proposition follows directly since x^i is added greedily to $X^{1:i-1}$. ■

D.1 Unconstrained case

Note that for unconstrained case domain V and utility ρ is obvious, so for convenience we use short hand notation, i.e, $F(\cdot;\rho,V) = F(\cdot)$ and $\Delta(\cdot;\rho,V) = \Delta(\cdot)$.
Locally optimal gain. Let us define OPT_l^i as the local optimal coverage gained by agent i, given all the locations of agents $1 : i-1$, formally given by,

$$OPT_l^i = \sum_{t=1}^{T} \Big(\max_{x_t^i} F(X_t^{1:i-1} \cup \{x_t^i\}) - F(X_t^{1:i-1}) \Big) = \sum_{t=1}^{T} \Delta(\tilde{x}_t^i | X_t^{1:i-1}) \tag{21}$$

We denote with OPT, the optimal coverage, precisely $OPT = \sum_{t=1}^{T} F(X_\star)$.
Per agent regret. Let us define local regret, as the difference in coverage gain in picking state $\tilde{x}_t^i$ vs the picked location x_t^i (this disparity is due to not knowing the actual density)

$$Reg^i(T) = \sum_{t=1}^{T} \Delta(\tilde{x}_t^i | X_t^{1:i-1}) - \sum_{t=1}^{T} \Delta(x_t^i | X_t^{1:i-1}) = OPT_l^i - \sum_{t=1}^{T} \Delta(x_t^i | X_t^{1:i-1}) \tag{22}$$

Actual regret. The actual regret is given by,

$$Reg_{act}(T) = \big(1 - \frac{1}{e}\big) \sum_{t=1}^{T} F(X_\star) - \sum_{t=1}^{T} F(X_t) = \big(1 - \frac{1}{e}\big) OPT - \sum_{t=1}^{T} F(X_t) \tag{23}$$

To prove. In this section we aim to show that actual regret bounded by sum of per agent regret, precisely,

$$Reg_{act}(T) \leq \sum_{i=1}^{K} Reg^i(T)$$

$$\sum_{i=1}^{K} Reg^i(T) \geq \big(1 - \frac{1}{e}\big) OPT - \sum_{t=1}^{T} F(X_t^{1:K}) \qquad \text{(Using defi. of } Reg_{act}(T) \text{ from Eq. (23))}$$

Lemma 2 *For all K agents' local per agent regret $Reg^i(T)$, we have,*

$$\sum_{t=1}^{T} \Delta(x_t^i | X_t^{1:i-1}) \geq \frac{1}{K} \Big(OPT - \sum_{t=1}^{T} F(X_t^{1:i-1}) \Big) - Reg^i(T) \tag{24}$$

Proof

$$\Delta(\tilde{x}_t^i | X_t^{1:i-1}) = \max_{x_t^i} F(X_t^{1:i-1} \cup \{x^i\}) - F(X_t^{1:i-1}) \qquad \text{(Using definition)}$$

$$\geq \frac{F(X_\star) - F(X_t^{1:i-1})}{K} \qquad \text{(Using Eq. (20) from Lem. 1)}$$

$$OPT_l^i \geq \frac{1}{K}\Big(\sum_{t=1}^{T} F(X_\star) - \sum_{t=1}^{T} F(X_t^{1:i-1})\Big) \qquad \text{(Sum over time)}$$

$$= \frac{1}{K}\Big(OPT - \sum_{t=1}^{T} F(X_t^{1:i-1})\Big) \qquad \text{(Using definition of } OPT\text{)}$$

$$\sum_{t=1}^{T} \Delta(x_t^i | X_t^{1:i-1}) \geq \frac{1}{K}\Big(OPT - \sum_{t=1}^{T} F(X_t^{1:i-1})\Big) - Reg^i(T)$$

$$\text{(Using def. of } Reg^i(T) \text{ Eq. (22))}$$

■

Lemma 3 *For any time t, X_t being the recommended location by* MACOPT*, we have*

$$\sum_{t=1}^{T} F(X_t^{1:K}) \geq \big(1 - \frac{1}{e}\big)OPT - \sum_{i=1}^{K} Reg^i(T) \tag{25}$$

And using definition of $Reg_{act}(T)$ from Eq. (23), this further implies that,

$$Reg_{act}(T) \leq \sum_{i=1}^{K} Reg^i(T) \tag{26}$$

Proof The proof is similar to the Lemma 2 from [38]. We begin to prove by induction,

$$OPT - \sum_{t=1}^{T} F(X_t^{1:i}) \leq \Big(1 - \frac{1}{K}\Big)^i OPT + \sum_{m=1}^{i} Reg_l^m(T) \tag{27}$$

Our main goal, i.e, Eq. (25) can be proved by substituting $i = K$ and using the inequality $(1 - 1/K)^K < 1/e$ in Eq. (27).

For i = 0, corresponds to no agent case. So it's trivial.

Let's consider gap to optimal value, when i elements are already selected,

$$\delta^i = OPT - \sum_{t=1}^{T} F(X_t^{1:i}) \qquad \text{(LHS of Eq. (27))}$$

$$= OPT - \sum_{t=1}^{T} \sum_{m=1}^{i} \Delta(x_t^m | X_t^{1:m-1}) \qquad \text{(Sum marginal gain; Using Eq. (18))}$$

$$\delta^{i-1} = OPT - \sum_{t=1}^{T} \sum_{m=1}^{i-1} \Delta(x_t^m | X_t^{1:m-1})$$

$$\implies \delta^i = \delta^{i-1} - \sum_{t=1}^{T} \Delta(x_t^i | X_t^{1:i-1}) \qquad \text{(Subtract } \delta^{i-1} \text{ from } \delta^i\text{)}$$

$$\implies \sum_{t=1}^{T} \Delta(x_t^i | X_t^{1:i-1}) = \delta^{i-1} - \delta^i \tag{28}$$

This says that the gap to optimal reduces by $\sum_{t=1}^{T} \Delta(x_t^i | X_t^{1:i-1})$ after adding element $x_t^i \ \forall\ t$.

$$\begin{aligned}
\sum_{t=1}^{T} \Delta(x_t^i | X_t^{1:i-1}) &\geq \frac{1}{K}(\delta^{i-1}) - Reg^i(T) && \text{(From Eq. (24) and } \delta^i \text{ definition)} \\
\implies \delta^{i-1} - \delta^i &\geq \frac{1}{K}(\delta^{i-1}) - Reg^i(T) && \text{(From Eq. (28))} \\
\implies \delta^i &\leq \Big(1 - \frac{1}{K}\Big)\delta^{i-1} + Reg^i(T) \\
&\leq \Big(1 - \frac{1}{K}\Big)^2 \delta^{i-2} + \sum_{m=1}^{2} Reg^i(T) \\
& && \text{(Subs } \delta^{i-1}\text{, Doing the telescopic bound)} \\
&\vdots \\
&\leq \Big(1 - \frac{1}{K}\Big)^i \delta^0 + \sum_{m=1}^{i} Reg^i(T) \\
&= \Big(1 - \frac{1}{K}\Big)^i OPT + \sum_{m=1}^{i} Reg^i(T) \\
OPT - \sum_{t=1}^{T} F(X_t^{1:i}) &\leq \Big(1 - \frac{1}{K}\Big)^i OPT + \sum_{m=1}^{i} Reg_l^m(T) && \text{(Using } \delta^i \text{ definition)}
\end{aligned}$$

Hence proved. ■

D.2 Constrained case

Simple regret. We define for a particular t, simple regret r_t^{act} and per agent local regret r_t respectively as:

$$\begin{aligned}
r_t^{act} &= (1 - \frac{1}{e}) \sum_{B \in \mathcal{B}} F(X_\star^B; \rho, \bar{R}_{\epsilon_q}(X_0^B)) - \sum_{B \in \mathcal{B}_t} \sum_{i \in B} \Delta(x_t^i | X_t^{1:i-1}; \rho, S_t^{u,B}), \\
r_t^{\mathrm{O}} &= (1 - \frac{1}{e}) \sum_{B \in \mathcal{B}_t} F(X_\star^B; \rho, S_t^{u,B}) - \sum_{B \in \mathcal{B}_t} \sum_{i \in B} \Delta(x_t^i | X_t^{1:i-1}; \rho, S_t^{u,B}) \\
r_t &= \sum_{B \in \mathcal{B}_t} \sum_{i \in B} \Delta(\tilde{x}^i | X_t^{1:i-1}; \rho, S_t^{u,B}) - \Delta(x_t^i | X_t^{1:i-1}; \rho, S_t^{u,B})
\end{aligned} \tag{29}$$

Cumulative regret. The actual cumulative regret $Reg^{\mathrm{O}}_{act}(T)$ and the per agent cumulative regret $Reg^{\mathrm{O}}_{l}(T)$ are respectively given by,

$$Reg^{\mathrm{O}}_{act}(T) = \sum_{t=1}^{T} r_t^{act} \text{ and } Reg^{\mathrm{O}}_{l}(T) = \sum_{t=1}^{T} r_t \tag{30}$$

On bounding per batch regret.

Optimal coverage in a batch B

$$\begin{aligned}
OPT_t &= F(X^B_\star; \rho, S^{u,B}_t) \\
OPT^i_t &= \max_{x^i} F(X^{1:i-1}_t \cup \{x^i\}; \rho, S^{u,B}_t) - F(X^{1:i-1}_t; \rho, S^{u,B}_t) \\
&= \max_{x^i} \Delta(x^i | X^{1:i-1}_t; \rho, S^{u,B}_t) = \Delta(\tilde{x}^i | X^{1:i-1}_t; \rho, S^{u,B}_t) \\
r^i_B(t) &= \Delta(\tilde{x}^i | X^{1:i-1}_t; \rho, S^{u,B}_t) - \Delta(x^i_t | X^{1:i-1}_t; \rho, S^{u,B}_t) \qquad (31)
\end{aligned}$$

To prove:

$$F(X^B_t; \rho, S^{u,B}_t) \geq \big(1 - \frac{1}{e}\big) OPT_t - \sum_{i \in B} r^i_B(t) \tag{32}$$

Proposition 4 *Let K_B be the number of agents in batch B and for all such agents per agent regret is $r^i_B(t)$. Then the following holds,*

$$\Delta(x^i_t | X^{1:i-1}_t; \rho, S^{u,B}_t) \geq \frac{1}{K_B}\Big(OPT_t - F(X^{1:i-1}_t; \rho, S^{u,B}_t)\Big) - r^i_B(t) \tag{33}$$

Proof

$$\begin{aligned}
\Delta(\tilde{x}^i_t | X^{1:i-1}_t; \rho, S^{u,B}_t) &= \max_{x^i_t} F(X^{1:i-1}_t \cup \{x^i\}; \rho, S^{u,B}_t) - F(X^{1:i-1}_t; \rho, S^{u,B}_t) \\
&\hspace{20em} \text{(Using definition)} \\
&\geq \frac{F(X_\star; \rho, S^{u,B}_t) - F(X^{1:i-1}_t; \rho, S^{u,B}_t)}{K_B} \\
&\hspace{20em} \text{(Using Eq. (20) from Lem. 1)} \\
OPT^i_t &\geq \frac{1}{K_B}\Big(OPT_t - F(X^{1:i-1}_t; \rho, S^{u,B}_t)\Big) \\
&\hspace{20em} \text{(Using definition of } OPT_t \text{ and } OPT^i_t) \\
\Delta(x^i_t | X^{1:i-1}_t; \rho, S^{u,B}_t) &\geq \frac{1}{K_B}\Big(OPT_t - F(X^{1:i-1}_t; \rho, S^{u,B}_t)\Big) - r^i_B(t) \\
&\hspace{20em} \text{(Using def. of } r^i_B(t) \text{ Eq. (31))}
\end{aligned}$$

■

Lemma 5 *For any time t, X^B_t being the recommended location by* SafeMac *in the union set $S^{u,B}_t$, we have*

$$F(X^B_t; \rho, S^{u,B}_t) \geq \big(1 - \frac{1}{e}\big) OPT_t - \sum_{i \in B} r^i_B(t), \tag{34}$$

Proof The proof is similar to the Lemma 2 from [38]. We begin to prove by induction,

$$OPT_t - F(X_t^{1:i}; \rho, S_t^{u,B}) \leq \Big(1 - \frac{1}{K_B}\Big)^i OPT_t + \sum_{m=1}^{i} r_B^i(t) \tag{35}$$

For i = 0, corresponds to no agent case. So it's trivial.

Let's consider gap to optimal value, when i elements are already selected,

$$\begin{aligned}
\delta^i &= OPT_t - F(X_t^{1:i}; \rho, S_t^{u,B}) && \text{(LHS of Eq. (35))}\\
&= OPT_t - \sum_{m=1}^{i} \Delta(x_t^m | X_t^{1:m-1}; \rho, S_t^{u,B}) && \text{(sum of marginal gain)}\\
\delta^{i-1} &= OPT_t - \sum_{m=1}^{i-1} \Delta(x_t^m | X_t^{1:m-1}; \rho, S_t^{u,B})\\
\implies \delta^i &= \delta^{i-1} - \Delta(x_t^i | X_t^{1:i-1}; \rho, S_t^{u,B}) && \text{(Subtract } \delta^{i-1} \text{ from } \delta^i)\\
\implies \quad & \Delta(x_t^i | X_t^{1:i-1}; \rho, S_t^{u,B}) = \delta^{i-1} - \delta^i && (36)
\end{aligned}$$

This says that the gap to optimal reduces by $\Delta(x_t^i | X_t^{1:i-1}; \rho, S_t^{u,B})$ after adding element x_t^i.

$$\begin{aligned}
\Delta(x_t^i | X_t^{1:i-1}; \rho, S_t^{u,B}) &\geq \frac{1}{K_B}(\delta^{i-1}) - r_B^i(t) && \text{(From Eq. (33) and } \delta^i \text{ definition)}\\
\implies \delta^{i-1} - \delta^i &\geq \frac{1}{K_B}(\delta^{i-1}) - r_B^i(t) && \text{(From Eq. (28))}\\
\implies \delta^i &\leq \Big(1 - \frac{1}{K_B}\Big)\delta^{i-1} + r_B^i(t)\\
&\leq \Big(1 - \frac{1}{K_B}\Big)^2 \delta^{i-2} + \sum_{m=1}^{2} r_B^i(t)\\
& && \text{(Subs } \delta^{i-1}\text{, Doing the telescopic bound)}\\
&\vdots\\
&\leq \Big(1 - \frac{1}{K_B}\Big)^i \delta^{0} + \sum_{m=1}^{i} r_B^i(t)\\
&= \Big(1 - \frac{1}{K_B}\Big)^i OPT_t + \sum_{m=1}^{i} r_B^i(t)\\
OPT_t - F(X_t^{1:i}; \rho, S_t^{u,B}) &\leq \Big(1 - \frac{1}{K_B}\Big)^i OPT_t + \sum_{m=1}^{i} r_B^i(t) && \text{(Using } \delta^i \text{ definition)}
\end{aligned}$$

Our main goal, i.e, Eq. (34) can be proved by substituting $i = K$ and using the inequality $(1 - 1/K)^K < 1/e$ in Eq. (35). Hence proved. ■

On combining all the batches.

Lemma 6 *For any time t, X_t being the location recommended by* SafeMaC*, we have*

$$r_t^{act} \leq r_t^{\mathrm{O}} \leq r_t \tag{37}$$

This further implies that,

$$Reg_{act}^{\mathrm{O}}(T) \leq Reg_l^{\mathrm{O}}(T) \tag{38}$$

Proof For a batch B of agents, using Eq. (35) from Lem. 5 and substituting $r_B^i(t)$ from Eq. (31) we know that,

$$\begin{aligned}(1 - \frac{1}{e})F(X_\star^B; \rho, S_t^{u,B}) - \sum_{i \in B} \Delta(x_t^i | X_t^{1:i-1}; \rho, S_t^{u,B}) \\ \leq \sum_{i \in B} \Delta(\tilde{x}^i | X_t^{1:i-1}; \rho, S_t^{u,B}) - \Delta(x_t^i | X_t^{1:i-1}; \rho, S_t^{u,B})\end{aligned}$$

By summing over all the $B \in \mathcal{B}_t$, we get

$$\begin{aligned}r_t^{\mathrm{O}} = (1 - \frac{1}{e}) \sum_{B \in \mathcal{B}_t} F(X_\star^B; \rho, S_t^{u,B}) - \sum_{B \in \mathcal{B}_t} \sum_{i \in B} \Delta(x_t^i | X_t^{1:i-1}; \rho, S_t^{u,B}) \\ \leq \sum_{B \in \mathcal{B}_t} \sum_{i \in B} \Delta(\tilde{x}^i | X_t^{1:i-1}; \rho, S_t^{u,B}) - \Delta(x_t^i | X_t^{1:i-1}; \rho, S_t^{u,B})\end{aligned} \tag{39}$$

Note that in Eq. (29), both the $X_\star^B$ represents optimal agent's location in their respective coverage set i.e, $\bar{R}_{\epsilon_q}(x_0^i)$ and $S_t^{u,B}$, hence both the $X_\star^B$ are different. Since, $\bigcup_{i \in B} \bar{R}_{\epsilon_q}(\{x_0^i\}) \subseteq S_t^{o,\epsilon_q,B} \subseteq S_t^{u,B} \implies \sum_{B \in \mathcal{B}} F(X_\star^B; \rho, \bar{R}_{\epsilon_q}(X_0^B)) \leq \sum_{B \in \mathcal{B}_t} F(X_\star^B; \rho, S_t^{u,B})$,

Moreover on using Eq. (29), Eq. (39) and we can conclude,

$$r_t^{act} \leq r_t^{\mathrm{O}} \leq r_t.$$

This further implies Eq. (38) using definition in Eq. (30). Hence Proved ■

Appendix E. Proof. for Theorem 1 (MaCOpt)

Theorem 1 *Let $\delta \in (0,1)$ and β_t^ρ as in [16], i.e., ${\beta_t^\rho}^{1/2} = B_\rho + 4\sigma_\rho\sqrt{\gamma_{Kt}^\rho + \ln(1/\delta)}$. With probability at least $1-\delta$, MaCOpt's regret defined in Eq. (3) is bounded by $\mathcal{O}(\sqrt{T\beta_T^\rho\gamma_{KT}^\rho})$,*

$$Pr\left\{ Reg_{act}(T) \leq K\sqrt{\frac{8T\beta_T^\rho\gamma_{KT}^\rho}{\log(1+K\sigma_\rho^{-2})}} \right\} \geq 1-\delta. \tag{6}$$

Proof The proof for Theorem 1 goes in the following steps:

1. We first exploit the conditional linearity of the submodular objective to bound the cumulative regret defined in Eq. (3) with a sum of per agent regrets ($\sum_{i=1}^K Reg^i(T)$). Precisely, we show $Reg_{act}(T) \leq \sum_{i=1}^K Reg^i(T)$ in Lem. 3.
2. We next bound the per agent regret with the information capacity γ_{KT}^ρ, a quantity that measures the largest reduction in uncertainty about the density that can be obtained from KT noisy evaluations of it.
    - For this, We quantify the information MaCOpt acquires through the noisy density observations in Lem. 7, through *the information gain* $I(y_A;\rho) = H(y_A) - H(y_A|\rho)$, where H denotes the Shannon entropy and A is the set of locations evaluated by MaCOpt.
    - Next we bound the per agent regret $Reg^i(T)$ with the information gain Lem. 8-9 which is in turn bounded by the information capacity.

Finally, Theorem 1 is a direct consequence of Lem. 3 and Lem. 9. ■

In the end of the section, we proof Corollary 1 which guarantees near optimal result in finite time.

Proposition 7 *The information gain for the points observed by* MaCOpt *can be expressed as:*

$$I(Y_{x_{1:T}^{g,1:K}};\rho) = \frac{1}{2}\sum_{t=1}^{T}\log(\det(I+\sigma_\rho^{-2}K_{x_t^{g,1:K}}^\rho)) = \frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{K}\log(1+\sigma_\rho^{-2}\lambda_{i,t}),$$

where $x_{1:T}^{g,1:K}$ is the set of goal locations set by MaCOpt *for all $1:K$ agents up to time T. $K_{x_t^{g,1:K}}^\rho$ is the positive definite kernel matrix formed by the observed locations and $\lambda_{i,t}$ represents eigenvalue of the matrix.*

Proof We can precisely quantify this notion through *the information gain*

$$I(Y_{x_{1:T}^{g,1:K}};\rho) = H(Y_{x_{1:T}^{g,1:K}}) - H(Y_{x_{1:T}^{g,1:K}}|\rho) \tag{40}$$

where H denotes the Shannon entropy. It can be defined as,

$$H(Y_{x_{1:T}^{g,1:K}}) = H(Y_T^{1:K}|Y_{x_{1:T-1}^{g,1:K}}) + H(Y_{x_{1:T-1}^{g,1:K}}) \qquad \text{(Defined } Y_T^{1:K} := \{y_T^1, y_T^2, ..., y_T^K\}\text{)}$$

$$= \frac{1}{2}\log(\det(2\pi e(\sigma^2 I + K^{\rho}_{x_T^{g,1:K}}))) + H(Y^{1:K}_{T-1}|Y_{x^{g,1:K}_{1:T-2}}) + ... \tag{41}$$

$$= \frac{1}{2}K\log(2\pi e\sigma^2) + \frac{1}{2}\log(\det(I + \sigma_\rho^{-2}K^{\rho}_{x_T^{g,1:K}})) + H(Y^{1:K}_{T-1}|Y_{x^{g,1:K}_{1:T-2}}) + ... \tag{42}$$

$$= \frac{1}{2}\sum_{t=1}^{T} K\log(2\pi e\sigma^2) + \frac{1}{2}\sum_{t=1}^{T}\log(\det(I + \sigma_\rho^{-2}K^{\rho}_{x_t^{g,1:K}})) \tag{43}$$

For Eq. (41), we used that, $Y_T^{1:K} \sim \mathcal{N}(\mu^{\rho}_{T-1}(x_T^{g,1:K}), \sigma^2 I + K^{\rho}_{x_T^{g,1:K}})$ is jointly a multivariate Gaussian. Eq. (42) follows by simplifying det, precisely, $\frac{1}{2}\log(\det(2\pi e(\sigma^2 I + K^{\rho}_{x_T^{g,1:K}}))) = \frac{1}{2}\log((2\pi e\sigma^2)^K \det(I + \sigma_\rho^{-2}K^{\rho}_{x_T^{g,1:K}}))$ and finally Eq. (43) by recursively repeating above 2 steps till $t = 1$. $H(Y_{x^{g,1:K}_{1:T}}|\rho) = \frac{1}{2}\sum_{t=1}^{T} K\log(2\pi e\sigma^2)$ is the entropy because of the noise. On substituting this, with Eq. (43) in Eq. (40) we obtain,

$$\begin{aligned}
I(Y_{x^{g,1:K}_{1:T}};\rho) &= \frac{1}{2}\sum_{t=1}^{T}\log(\det(I + \sigma_\rho^{-2}K^{\rho}_{x_t^{g,1:K}})) \\
&= \frac{1}{2}\sum_{t=1}^{T}\log(\prod_{i=1}^{K}(1 + \sigma_\rho^{-2}\lambda_{i,t})) && \text{(Using Eq. 45)} \\
&= \frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{K}\log(1 + \sigma_\rho^{-2}\lambda_{i,t}) && (44)
\end{aligned}$$

Hence Proved. ■

Log mat inequality:

$$\begin{aligned}
\log(\det(I + \sigma_\rho^{-2}K^{\rho})) &= \log(\det(RR^\top + \sigma_\rho^{-2}R\Lambda R^\top)) && (K^\rho = R\Lambda R^\top, RR^\top = I) \\
&= \log(\det(R(I + \sigma_\rho^{-2}\Lambda)R^\top)) \\
&= \log(\det(RR^\top)) + \log(\det(I + \sigma_\rho^{-2}\Lambda)) && (\text{k is dimension of } K^\rho) \\
&= \log(\prod_{i=1}^{k}(1 + \sigma_\rho^{-2}\lambda_i)) && (45)
\end{aligned}$$

Lemma 8 *Till any time T, if $|\rho(v) - \mu^{\rho}_{t-1}(v)| \leq \beta_t^{1/2}\sigma^{\rho}_{t-1}(v)$ for all $v \in V$, then the agent wise cumulative regret $Reg^i(T)$, is bounded by $\sum_{t=1}^{T} 2\sqrt{\beta_t^{\rho}}\max_{v\in D_t^{i-}}\sigma^{\rho}_{t-1}(v)$ for agent i.*

Proof For notation convenience: $D_t^{i-} := D_t^i \backslash D_t^{1:i-1}$ and $\tilde{D}_t^{i-} := \tilde{D}_t^i \backslash D_t^{1:i-1}$

In MACOPT x_t^i is defined such that,

$$x_t^i = \arg\max_{v} \sum_{v\in D_t^{i-}} \mu^{\rho}_{t-1}(v) + \sqrt{\beta_t^{\rho}}\sigma^{\rho}_{t-1}(v) \tag{46}$$

Due to our picking strategy,

$$\sum_{v\in\tilde{D}_t^{i-}} \rho(v) \le \sum_{v\in\tilde{D}_t^{i-}} \big(\mu_{t-1}^{\rho}(v) + \sqrt{\beta_t^{\rho}}\sigma_{t-1}^{\rho}(v)\big) \le \sum_{v\in D_t^{i-}} \big(\mu_{t-1}^{\rho}(v) + \sqrt{\beta_t^{\rho}}\sigma_{t-1}^{\rho}(v)\big) \tag{47}$$

This first inequality follows due to upper bound and the second one follows based on how x_t^i is picked (Eq. (46)).

$$\begin{aligned}
Reg^i(T) &= \sum_{t=1}^{T} \Delta(\tilde{x}_t^i|X_t^{1:i-1}) - \sum_{t=1}^{T} \Delta(x_t^i|X_t^{1:i-1}) && \text{(with definition Eq. (22))}\\
&= \sum_{t=1}^{T} \Big(\sum_{v\in\tilde{D}_t^{i-}} \rho(v) - \sum_{v\in D_t^{i-}} \rho(v)\big)\Big)/N && \text{(Using defi. } \Delta(.|X_t^{1:i-1}) \text{ Eq. (17))}\\
&\le \sum_{t=1}^{T} \Big(\sum_{v\in D_t^{i-}} \mu_{t-1}^{\rho}(v) + \sqrt{\beta_t^{\rho}}\sigma_{t-1}^{\rho}(v) - \sum_{v\in D_t^{i-}} \rho(v)\Big)/N && \text{(From Eq. (47))}\\
&\le \sum_{t=1}^{T} \Big(\sum_{v\in D_t^{i-}} \mu_{t-1}^{\rho}(v) + \sqrt{\beta_t^{\rho}}\sigma_{t-1}^{\rho}(v) - \sum_{v\in D_t^{i-}} \mu_{t-1}^{\rho}(v) - \sqrt{\beta_t^{\rho}}\sigma_{t-1}^{\rho}(v)\Big)/N \\
& && \text{(Since, } \rho(v) \ge \mu_{t-1}^{\rho}(v) - \sqrt{\beta_t^{\rho}}\sigma_{t-1}^{\rho}(v)\ \forall\ v)\\
&= \sum_{t=1}^{T} 2\sqrt{\beta_t^{\rho}} \sum_{v\in D_t^{i-}} \sigma_{t-1}^{\rho}(v)/N \le \sum_{t=1}^{T} 2\sqrt{\beta_t^{\rho}} \max_{v\in D_t^{i-}} \sigma_{t-1}^{\rho}(v) && (48)
\end{aligned}$$

The last inequality follows since $\sum_{v\in D_t^{i-}} \sigma_{t-1}^{\rho}(v) \le N \max_{v\in D_t^{i-}} \sigma_{t-1}^{\rho}(v)$ and $|D_t^{i-}| \le N$. ■

Lemma 9 *Let $\delta \in (0,1)$ and let $\beta_t^{\rho\,1/2} = B_\rho + 4\sigma_\rho\sqrt{\gamma_{Kt}^{\rho} + 1 + \ln(1/\delta)}$. Then for K agents, $\forall T \ge 1$ the following holds with probability $1-\delta$,*

$$(\sum_{i=1}^{K} Reg^i(T))^2 \le \frac{8TK^2\beta_T^{\rho} I(Y_{x_{1:T}^{g,1:K}}; \rho)}{\log(1+K\sigma_\rho^{-2})} \le \frac{8TK^2\beta_T^{\rho}\gamma_{KT}}{\log(1+K\sigma_\rho^{-2})}$$

Proof By sum over all the K agents from Lem. 8, we get

$$\sum_{i=1}^{K} Reg^i(T) \le \sum_{i=1}^{K}\sum_{t=1}^{T} 2\sqrt{\beta_t^{\rho}} \max_{v\in D_t^{i-}} \sigma_{t-1}^{\rho}(v) \tag{49}$$

Let's consider,

$$\begin{aligned}
w_t &:= 2\sqrt{\beta_t^{\rho}} \sum_{i=1}^{K} \max_{v\in D_t^{i-}} \sigma_{t-1}^{\rho}(v) && \text{(part of Eq. (49))}\\
w_t^2 &= 4\beta_t^{\rho}\Big(\sum_{i=1}^{K} \max_{v\in D_t^{i-}} \sigma_{t-1}^{\rho}(v)\Big)^2 && \text{(Square operation)}
\end{aligned}$$

$$\leq 4\beta_t^\rho K \sum_{i=1}^{K} \left(\sigma_{t-1}^\rho(x_t^{g,i})\right)^2 \qquad \text{(Cauchy-Schwarz inequality, } x_t^{g,i} = \arg\max_{v\in D_t^{i-}} \sigma_{\rho t-1}^2(v))$$

$$= 4\beta_t^\rho K \sum_{i=1}^{K} \lambda_{i,t} \qquad \left(\textstyle\sum_{i=1}^{K}(\sigma_{t-1}^\rho(x_t^{g,i}))^2 = Tr(K^\rho) = \sum_{i=1}^{K}\lambda_{i,t}\right)$$

$$= 4\beta_t^\rho K \sum_{i=1}^{K} \sigma_\rho^2 \sigma_\rho^{-2}\lambda_{i,t} \leq 4\beta_t^\rho K \sum_{i=1}^{K} \sigma_\rho^2 C_1 \log(1+\sigma_\rho^{-2}\lambda_{i,t})$$

(Since, $s \leq C_1 \log(1+s)$ for $s \in [0, K\sigma_\rho^{-2}]$, where $C_1 = K\sigma_\rho^{-2}/\log(1+K\sigma_\rho^{-2}) \geq 1$)

(Here, $s = \sigma_\rho^{-2}\lambda_{i,t} \leq \sigma_\rho^{-2}\lambda_{\max} \leq \sigma_\rho^{-2}\sum_i \lambda_{i,t} = \sigma_\rho^{-2} Tr(K^\rho) \leq \sigma_\rho^{-2}K$, $(wlog\ k(v,v) \leq 1)$)

$$\leq \frac{8K^2\beta_t^\rho}{\log(1+K\sigma_\rho^{-2})} \sum_{i=1}^{K} \frac{1}{2}\log(1+\sigma_\rho^{-2}\lambda_{i,t}) \tag{50}$$

From Eq. (49) and r_t definition,

$$\Big(\sum_{i=1}^{K} Reg^i(T)\Big)^2 \leq \Big(\sum_{t=1}^{T} w_t\Big)^2 \leq T\sum_{t=1}^{T} w_t^2 \qquad \text{(Using Cauchy-Schwarz inequality)}$$

$$\leq T\sum_{t=1}^{T} \frac{8K^2\beta_t^\rho}{\log(1+K\sigma_\rho^{-2})} \sum_{i=1}^{K} \frac{1}{2}\log(1+\sigma_\rho^{-2}\lambda_{i,t}) \qquad \text{(Using Eq. (50))}$$

$$= \frac{8TK^2\beta_T^\rho}{\log(1+K\sigma_\rho^{-2})} I(Y_{x_{1:T}^{g,1:K}};\rho)$$

(Since β_t^ρ is non-decreasing, using Eq. (44))

$$\leq \frac{8TK^2\beta_T^\rho\gamma_{KT}}{\log(1+K\sigma_\rho^{-2})} \qquad (\gamma_{KT} = \sup_{x_{1:T}^{g,1:K}\subset V} I(Y_{x_{1:T}^{g,1:K}};\rho))$$

$$\implies \sum_{i=1}^{K} Reg^i(T) \leq \sum_{t=1}^{T} w_t \leq \Big(T\sum_{t=1}^{T} w_t^2\Big)^{1/2} \leq K\sqrt{\frac{8TK^2\beta_T^\rho\gamma_{KT}}{\log(1+K\sigma_\rho^{-2})}} \tag{51}$$

Hence Proved. ■

Theorem 1 follows from Lem. 8, Lem. 9 and Eq. (26),

$$Reg_{act}(T) \leq \sum_{i=1}^{K} Reg^i(T) \leq K\sqrt{\frac{8T\beta_T^\rho\gamma_{KT}}{\log(1+K\sigma_\rho^{-2})}}$$

Proof for the corollary 1:

Corollary 1 *Let $t_\rho^\star$ be the smallest integer, $\frac{t_\rho^\star}{\beta_{t_\rho^\star}\gamma_{Kt_\rho^\star}} \leq \frac{8K^2}{\log(1+K\sigma^{-2})\epsilon_\rho^2}$, then there exists a $t < t_\rho^\star$ such that w.h.p, MaCOpt terminates and achieves, $F(X_t;\rho,V) \geq (1-\frac{1}{e})F(X_\star;\rho,V) - \epsilon_\rho$.*

Proof The proof for the corollary goes in the following 2 steps. First, we show that once $w_t \leq \epsilon_\rho$ implies $F(X_t;\rho,V) \geq (1-\frac{1}{e})F(X_\star;\rho,V) - \epsilon_\rho$. Secondly, in Lem. 10 we show MaCOpt achieves $w_t \leq \epsilon_\rho$, at $t < t_\rho^\star$ where $t_\rho^\star$ be the smallest integer satisfying $\frac{t_\rho^\star}{\beta_{t_\rho^\star}\gamma_{Kt_\rho^\star}} \leq \frac{8K^2}{\log(1+K\sigma^{-2})\epsilon_\rho^2}$.

Similar to steps in Lem. 8 for a fix t, (Eq. (48)), we get

$$\Delta(\tilde{x}^i|X_t^{1:i-1}) - \Delta(x_t^i|X_t^{1:i-1}) \leq 2\sqrt{\beta_t} \max_{v \in D_t^{i-}} \sigma_{t-1}^{\rho}(v)$$

From Eq. (37) (for constrained case) one can show for unconstrained case,

$$\begin{aligned}
(1-\frac{1}{e})F(X_\star;\rho,V) - \sum_i^K \Delta(x_t^i|X_t^{1:i-1}) &\leq \sum_i^K \Delta(\tilde{x}^i|X_t^{1:i-1}) - \Delta(x_t^i|X_t^{1:i-1}) \\
&\leq \sum_i^K 2\sqrt{\beta_t} \max_{v \in D_t^{i-}} \sigma_{\rho_{t-1}}(v) \leq \epsilon_\rho \\
\implies F(X_t;\rho,V) &\geq (1-\frac{1}{e})F(X_\star;\rho,V) - \epsilon_\rho
\end{aligned}$$

■

Lemma 10 *Let $\delta \in (0,1)$ and β_t^ρ as in [16], i.e., $\beta_t^{\rho 1/2} = B_\rho + 4\sigma_\rho\sqrt{\gamma_{Kt}^\rho + 1 + \ln(1/\delta)}$ and $t_\rho^\star$ is the smallest integer such that $\frac{t_\rho^\star}{\beta_{t_\rho^\star}\gamma_{Kt_\rho^\star}} \geq \frac{8K^2}{\log(1+K\sigma^{-2})\epsilon_\rho^2}$, then with probability $1-\delta$ that there exists $t_\rho < t_\rho^\star$ such that $w_{t_\rho+1} \leq \epsilon_\rho$, where $w_t = \sum_{B \in \mathcal{B}_t}\sum_{i \in B} 2\sqrt{\beta_t^\rho} \max_{v \in D_t^{i-}} \sigma_{t-1}^\rho(v) \leq \epsilon_\rho$.*

Proof Since,

$$\begin{aligned}
\frac{t_\rho^\star}{\beta_{t_\rho^\star}\gamma_{Kt_\rho^\star}} &\geq \frac{8K^2}{\log(1+K\sigma^{-2})\epsilon_\rho^2} \\
\implies K\sqrt{\frac{8\beta_{t_\rho^\star}\gamma_{Kt_\rho^\star}}{t_\rho^\star\log(1+K\sigma^{-2})}} &\leq \epsilon_\rho && \text{(Rearranging terms)} \\
\frac{\sum_{t=1}^{t_\rho^\star} w_t}{t_\rho^\star} \leq K\sqrt{\frac{8\beta_{t_\rho^\star}\gamma_{Kt_\rho^\star}}{t_\rho^\star\log(1+K\sigma^{-2})}} &\leq \epsilon_\rho && \text{(From Eq. (51) in Lem. 9)} \\
\implies \min_{t \in [1,t_\rho^\star]} w_t &\leq \epsilon_\rho && \left(\frac{t_\rho^\star \min_{t\in[1,t_\rho^\star]} w_t}{t_\rho^\star} \leq \frac{\sum_{t=1}^{t_\rho^\star} w_t}{t_\rho^\star}\right)
\end{aligned}$$

Hence there exists $t_\rho < t_\rho^\star$, such that $w_{t_\rho+1} \leq \epsilon_\rho$.

■

Appendix F. Proof. for Theorem 2 (SafeMaC)

Theorem 2 *Let $\delta \in (0,1)$ and β_t^ρ as in [16], i.e., $\beta_t^{\rho 1/2} = B_\rho + 4\sigma_\rho\sqrt{\gamma_{Kt}^\rho + 1 + \ln(1/\delta)}$*

and $t_\rho^\star$ be the smallest integer such that $\frac{t_\rho^\star}{\beta_{t_\rho^\star}\gamma_{Kt_\rho^\star}} \geq \frac{8K^2}{\log(1+K\sigma^{-2})\epsilon_\rho^2}$. Let β_t^q and $t_q^\star$ be defined analogously. Then, there exists $t < t_q^\star + t_\rho^\star$, such that with probability at least $1-\delta$

$$\sum_{B\in\mathcal{B}_t} F(X_t^B;\rho,\bar{R}_0(X_0^B)) \geq (1-\frac{1}{e})\sum_{B\in\mathcal{B}} F(X_\star^B;\rho,\bar{R}_{\epsilon_q}(X_0^B)) - \epsilon_\rho. \tag{7}$$

Proof The proof for Theorem 2 goes in the following two steps:

1. SafeMaC's coverage is near-optimal at the convergence
   - We first bound the actual regret with the sum of per agent regret in Lem. 6. Precisely, we show the following (Eq. (38)),

   $$Reg_{act}^{\mathrm{O}}(T_\rho) \leq Reg_l^{\mathrm{O}}(T_\rho)$$

   - Next, we establish in Lem. 11 that the $Reg_l^{\mathrm{O}}(T_\rho)$ grows sublinear with the density measurements.
   - Next, we show that if $w_t < \epsilon_\rho$, the coverage is near optimal (Lem. 12). The condition $w_t < \epsilon_\rho$ will eventually happen since $Reg_l^{\mathrm{O}}(T_\rho)$ is sublinear and hence over time will shrink to zero.
   - Finally using Lem. 17, the near optimality in the pessimistic set can be established at convergence when the 2^{nd} termination condition is satisfied, precisely $\{S_t^{o,\epsilon_q,i}\backslash S_t^{p,i}) \cap D_t^i, \forall i \in [K]\} = \emptyset$

2. SafeMaC converges in a finite time $t < t_q^\star + t_\rho^\star$, where $t_\rho^\star$ be the smallest integer such that $\frac{t_\rho^\star}{\beta_{t_\rho^\star}\gamma_{Kt_\rho^\star}} \geq \frac{8K^2}{\log(1+K\sigma^{-2})\epsilon_\rho^2}$ and $t_q^\star$ be the smallest integer such that $\frac{t_q^\star}{\beta_{t_q^\star}\gamma_{Kt_q^\star}} \geq \frac{C_1|\bar{R}_0(X_0)|}{\epsilon_q^2}$, with $C_1 = 8/\log(1+\sigma_q^{-2})$.
   - Since SafeMaC runs by iterating between the coverage and the exploration phase, we decouple it and analyze both the phases separately. Starting with the *coverage phase*, In Lem. 13, we establish a bound on density samples required to terminate the first coverage phase
   - Next, in the Lem. 14, we show that cumulative regret grows sublinear with the density measurements for any coverage phase and utilizes this to bound the density samples between two consecutive coverage phases in Lem. 15
   - Utilizing the above two statements, we present the sample complexity bound to terminate the n^{th} coverage phase till convergence, using that the information gain is additive for consecutive coverage phases in Lem. 16
   - For the *exploration phase*, the worst case time complexity bound is given by the multi-agent version of the GoOSE in Lem. 22 when the agents safely explore the complete domain. The resulting worst case time bound for SafeMaC is sum of the time bound of the *coverage* and the *exploration* phase.

So, near optimality at convergence in Theorem 2 is a direct consequence of Lem. 12 and Lem. 17 and the finite time argument of Theorem 2 is a direct consequence of Lem. 16 and Lem. 22. ■

Lemma 11 *Let $\delta \in (0,1)$ and β_t^ρ as in [16], i.e., $\beta_t^{\rho 1/2} = B_\rho + 4\sigma\sqrt{\gamma_t^\rho + 1 + \ln(1/\delta)}$. With probability at least $1-\delta$,* SafeMaC*'s sum of per agent regret $Reg_l^{\mathrm{O}}(T_\rho)$ is bounded by $\mathcal{O}(\sqrt{T_\rho \beta_T^\rho \gamma_{KT}^\rho})$. Precisely,*

$$Reg_l^{\mathrm{O}}(T_\rho) \leq K\sqrt{\frac{8T_\rho \beta_t^\rho \gamma_{KT}^\rho}{\log(1+K\sigma^{-2})}}$$

where T_ρ is density samples per agent and $Reg_l^{\mathrm{O}}(T_\rho) = \sum_{t=1}^{T_\rho} r_t$ where $r_t = \sum_{B\in\mathcal{B}_t}\sum_{i\in B}\Delta(\tilde{x}^i|X_t^{1:i-1};\rho,S_t^{u,B}) - \Delta(x_t^i|X_t^{1:i-1};\rho,S_t^{u,B})$

Proof Given.

$$\begin{aligned} Reg_l^{\mathrm{O}}(T_\rho) &= \sum_{t=1}^{T_\rho} r_t \\ &= \sum_{t=1}^{T_\rho}\sum_{B\in\mathcal{B}_t}\sum_{i\in B}\Delta(\tilde{x}^i|X_t^{1:i-1};\rho,S_t^{u,B}) - \Delta(x_t^i|X_t^{1:i-1};\rho,S_t^{u,B}) \end{aligned}$$

WLOG, every batch B, is indexed by iterator $i = 1 \; to \; |B|$ sequentially.
Let $\tilde{x}^i = \arg\max \Delta(x_t^i|X_t^{1:i-1};\rho,S_t^{u,B})$ and $\tilde{D}_t^i$ is a disk around $\tilde{x}^i$. For notation convenience: $D_t^{i-} := D_t^i \backslash D_t^{1:i-1} \cap S_t^{u,B}$ and $\tilde{D}_t^{i-} := \tilde{D}_t^i \backslash D_t^{1:i-1} \cap S_t^{u,B}$

SafeMaC picks the agent at x_t^i greedily in the set B. Following the steps in Lem. 12 we can bound simple agent-wise local regret as r_t or simply from Eq. (56) by summing over all the $B \in \mathcal{B}_t$, we get,

$$\begin{aligned} r_t &= \sum_{B\in\mathcal{B}_t}\sum_{i\in B}\Delta(\tilde{x}_t^i|X_t^{1:i-1};\rho,S_t^{u,B}) - \Delta(x_t^i|X_t^{1:i-1};\rho,S_t^{u,B}) \\ &\leq \sum_{B\in\mathcal{B}_t}\sum_{i\in B} 2\sqrt{\beta_t^\rho}\max_{v\in D_t^{i-}}\sigma_{t-1}^\rho(v) = w_t \quad \text{(From Eq. (56))} \end{aligned}$$

On bounding simple regret.

$$\begin{aligned} r_t \leq w_t &= \sum_{B\in\mathcal{B}_t}\sum_{i\in B} 2\sqrt{\beta_t^\rho}\max_{v\in D_t^{i-}}\sigma_{t-1}^\rho(v) = 2\sqrt{\beta_t^\rho}\sum_{i=1}^{K}\sigma_{t-1}^\rho(x_t^{g,i}) \qquad (x_t^{g,i} = \arg\max_{v\in D_t^{i-}}\sigma_{t-1}^\rho(v)) \\ w_t^2 &\leq 4\beta_t^\rho K\sum_{i=1}^{K}(\sigma_{t-1}^\rho(x_t^{g,i}))^2 \qquad \text{(Using Cauchy–Schwarz inequality)} \\ &= 4\beta_t^\rho K\sum_{i=1}^{K}\lambda_{i,t} \qquad (\textstyle\sum_{i=1}^{K}(\sigma_{t-1}^\rho(x_t^{g,i}))^2 = Tr(K^\rho) = \sum_{i=1}^{K}\lambda_{i,t}) \end{aligned}$$

$$= 4\beta_t^\rho K \sum_{i=1}^{K} \sigma^2 \sigma^{-2} \lambda_{i,t}$$

$$\leq 4\beta_t^\rho K \sum_{i=1}^{K} \sigma^2 C_2 \log(1 + \sigma^{-2}\lambda_{i,t})$$

(Since, $s \leq C_2 \log(1+s)$ for $s \in [0, K\sigma^{-2}]$, where $C_2 = K\sigma^{-2}/\log(1+K\sigma^{-2}) \geq 1$)

(Here, $s = \sigma^{-2}\lambda_{i,t} \leq \sigma^{-2}\lambda_{\max} \leq \sigma^{-2}\sum_i \lambda_{i,t} = \sigma^{-2}Tr(K^\rho) \leq \sigma^{-2}K,\ (wlog\ k(v,v) \leq 1)$)

$$\leq \frac{8K^2\beta_t^\rho}{\log(1+K\sigma^{-2})} \sum_{i=1}^{K} \frac{1}{2} \log(1+\sigma^{-2}\lambda_{i,t}) \tag{52}$$

On bounding cumulative regret with mutual information.

$$\Big(\sum_{t=1}^{T_\rho} w_t\Big)^2 \leq T_\rho \sum_{t=1}^{T_\rho} w_t^2 \quad \text{(Using Cauchy–Schwarz inequality)}$$

$$\leq T_\rho \sum_{t=1}^{T_\rho} \frac{8K^2\beta_t^\rho}{\log(1+K\sigma^{-2})} \sum_{i=1}^{K} \frac{1}{2}\log(1+\sigma^{-2}\lambda_{i,t}) \quad \text{(Using Eq. (52))}$$

$$= \frac{8T_\rho K^2 \beta_T^\rho}{\log(1+K\sigma^{-2})} \sum_{t=1}^{T_\rho}\sum_{i=1}^{K} \frac{1}{2}\log(1+\sigma^{-2}\lambda_{i,t})$$

(Since β_t^ρ is non-decreasing & $\beta_T^\rho := \beta_{T_\rho}^\rho$)

$$= \frac{8T_\rho K^2 \beta_T^\rho I(Y_{x^{g,1:K}_{1:T_\rho}};\rho)}{\log(1+K\sigma^{-2})} \quad \text{(Using Eq. (44))}$$

$$\leq \frac{8T_\rho K^2 \beta_T^\rho \gamma_{KT}^\rho}{\log(1+K\sigma^{-2})} \quad (\gamma_{KT}^\rho = \sup_{X^m_{1:T_\rho} \subset V} I(Y_{X^m_{1:T_\rho}};\rho))$$

$$\implies \sum_{t=1}^{T_\rho} w_t \leq \sqrt{\frac{8T_\rho K^2 \beta_T^\rho \gamma_{KT}^\rho}{\log(1+K\sigma^{-2})}} \tag{53}$$

$$\implies Reg_l^{\mathrm{O}}(T_\rho) \leq \sqrt{\frac{8T_\rho K^2 \beta_T^\rho \gamma_{KT}^\rho}{\log(1+K\sigma^{-2})}} \quad \text{(Since } Reg_l^{\mathrm{O}}(T_\rho) = \textstyle\sum_{t=1}^{T_\rho} r_t \leq \sum_{t=1}^{T_\rho} w_t)$$

■

This lemma nicely connects the near optimal coverage in the reachable set i.e, $(1 - \frac{1}{e})\sum_{B\in\mathcal{B}} F(X_\star^B; \rho, \bar{R}_{\epsilon_q}(X_0^B))$, with the coverage in a possibly disjoint optimistic sets. (Note that the only requirement is that the optimistic set needs to always superset $\bar{R}_{\epsilon_q}(X_0)$.

The agents observe the location only if all the agents can reach the max uncertain point under their disk i.e, $2\sqrt{\beta_t^\rho}\max_{v\in D_t^{i-}} \sigma_{t-1}^\rho(v)$. (Accordingly, information gain is defined, and T_ρ above is a counter when all the agents obtain density measurements).

Lemma 12 (SafeMaC Near-Optimality) *For any $t \geq 1$, if $w_t \leq \epsilon_\rho$ at* SafeMaC*'s recommendation X_t then with high probability,*

$$\sum_{B\in\mathcal{B}_t} F(X_t^B; \rho, S_t^{u,B}) \geq (1 - \frac{1}{e}) \sum_{B\in\mathcal{B}} F(X_\star^B; \rho, \bar{R}_{\epsilon_q}(X_0^B)) - \epsilon_\rho,$$

where $w_t = \sum_{B \in \mathcal{B}_t} \sum_{i \in B} 2\sqrt{\beta_t^\rho} \max_{v \in D_t^{i-}} \sigma_{t-1}^\rho(v)$.

Proof Given. SafeMaC recommends a location for the agent $i \in B$ greedily in the $S_t^{u,B}$ set as per,

$$x_t^i = \arg\max_v \sum_{v \in D_t^{i-}} \mu_{t-1}^\rho(v) + \sqrt{\beta_t^\rho} \sigma_{t-1}^\rho(v) \tag{54}$$

Let $\tilde{x}_t^i = \arg\max \Delta(x_t^i | X_t^{1:i-1}; \rho, S_t^{u,B})$ and $\tilde{D}_t^{i-} := \tilde{D}_t^i \backslash D_t^{1:i-1} \cap S_t^{u,B}$, where $\tilde{D}_t^i$ is a disk around $\tilde{x}^i$. Based on this picking strategy,

$$\begin{aligned}
\sum_{v \in \tilde{D}_t^{i-}} \rho(v) &\leq \sum_{v \in \tilde{D}_t^{i-}} \left(\mu_{t-1}^\rho(v) + \sqrt{\beta_t^\rho} \sigma_{t-1}^\rho(v)\right) && \text{(Follows due to upper confidence bound)} \\
&\leq \sum_{v \in D_t^{i-}} \left(\mu_{t-1}^\rho(v) + \sqrt{\beta_t^\rho} \sigma_{t-1}^\rho(v)\right) && \text{(Since, Eq. (54), } x_t^i \text{ is greedily picked)} \\
\sum_{v \in \tilde{D}_t^{i-}} \rho(v) &\leq \sum_{v \in D_t^{i-}} \left(\mu_{t-1}^\rho(v) + \sqrt{\beta_t^\rho} \sigma_{t-1}^\rho(v)\right) && (55)
\end{aligned}$$

On bounding simple regret. With definition $r_t = \sum_{B \in \mathcal{B}_t} \sum_{i \in B} \Delta(\tilde{x}_t^i | X_t^{1:i-1}; \rho, S_t^{u,B}) - \Delta(x_t^i | X_t^{1:i-1}; \rho, S_t^{u,B})$.
Consider,

$$\begin{aligned}
\Delta(\tilde{x}_t^i | X_t^{1:i-1}; \rho, S_t^{u,B}) &- \Delta(x_t^i | X_t^{1:i-1}; \rho, S_t^{u,B}) \\
&= \Big(\sum_{v \in \tilde{D}_t^{i-}} \rho(v) - \sum_{v \in D_t^{i-}} \rho(v) \Big) / N && \text{(Note } D_t^{i-} \text{ and } \tilde{D}_t^{i-}) \\
&\leq \Big(\sum_{v \in D_t^{i-}} \Big(\mu_{t-1}^\rho(v) + \sqrt{\beta_t^\rho} \sigma_{t-1}^\rho(v)\Big) - \sum_{v \in D_t^{i-}} \rho(v) \Big) / N \\
& && \text{(Using Eq. (55))} \\
&\leq \Big(\sum_{v \in D_t^{i-}} \big(\mu_{t-1}^\rho(v) + \sqrt{\beta_t^\rho} \sigma_{t-1}^\rho(v)\big) - \big(\mu_{t-1}^\rho(v) - \sqrt{\beta_t^\rho} \sigma_{t-1}^\rho(v)\big) \Big) / N \\
& && \text{(Since, } \rho(v) \geq \mu_{t-1}^\rho(v) - \sqrt{\beta_t^\rho} \sigma_{t-1}^\rho(v) \ \forall\ v) \\
&= 2\sqrt{\beta_t^\rho} \sum_{v \in D_t^{i-}} \sigma_{t-1}^\rho(v) / N \\
&\leq 2\sqrt{\beta_t^\rho} \max_{v \in D_t^{i-}} \sigma_{t-1}^\rho(v) && (56)
\end{aligned}$$

The last inequality follows since $\sum_{v \in D_t^{i-}} \sigma_{t-1}^\rho(v) \leq N \max_{v \in D_t^{i-}} \sigma_{t-1}^\rho(v)$ and $|D_t^{i-}| \leq N$.Now,

$$r_t = \sum_{B \in \mathcal{B}_t} \sum_{i \in B} \Delta(\tilde{x}_t^i | X_t^{1:i-1}; \rho, S_t^{u,B}) - \Delta(x_t^i | X_t^{1:i-1}; \rho, S_t^{u,B})$$

$$\leq \sum_{B\in\mathcal{B}_t}\sum_{i\in B} 2\sqrt{\beta_t^\rho} \max_{v\in D_t^{i-}} \sigma_{t-1}^\rho(v) \qquad \text{(from Eq. (56))}$$
$$= w_t \leq \epsilon_\rho$$

From Eq. (37), $(1-\frac{1}{e})\sum_{B\in\mathcal{B}} F(X_\star^B;\rho,\bar{R}_{\epsilon_q}(X_0^B)) - \sum_{B\in\mathcal{B}_t} F(X_t^B;\rho,S_t^{u,B}) = r_t^{act} \leq r_t$

$$\implies \sum_{B\in\mathcal{B}_t} F(X_t^B;\rho,S_t^{u,B}) \geq (1-\frac{1}{e})\sum_{B\in\mathcal{B}} F(X_\star^B;\rho,\bar{R}_{\epsilon_q}(X_0^B)) - \epsilon_\rho$$

■

Proposition 13 *Let $\delta \in (0,1)$ and β_t^ρ as in [16], i.e., ${\beta_t^\rho}^{1/2} = B_\rho + 4\sigma_\rho\sqrt{\gamma_{Kt}^\rho + 1 + \ln(1/\delta)}$ and $t_\rho^{\star 1}$ is the smallest integer such that $\frac{t_\rho^{\star 1}}{\beta_{t_\rho^{\star 1}}^\rho I(Y_{x^{g,1:K}_{1:t_\rho^{\star 1}}};\rho)} \geq \frac{8K^2}{\log(1+K\sigma^{-2})\epsilon_\rho^2}$, then with probability $1-\delta$ that there exists $t_\rho^1 < t_\rho^{\star 1}$ such that $w_{t_\rho^1+1} \leq \epsilon_\rho$, where $w_t = \sum_{B\in\mathcal{B}_t}\sum_{i\in B} 2\sqrt{\beta_t^\rho}\max_{v\in D_t^{i-}} \sigma_{t-1}^\rho(v) \leq \epsilon_\rho$.*

Proof Since,

$$\frac{t_\rho^{\star 1}}{\beta_{t_\rho^{\star 1}}^\rho I(Y_{X^m_{1:t_\rho^{\star 1}}};\rho)} \geq \frac{8K^2}{\log(1+K\sigma^{-2})\epsilon_\rho^2} \tag{57}$$

$$\implies K\sqrt{\frac{8\beta_{t_\rho^{\star 1}}^\rho I(Y_{X^m_{1:t_\rho^{\star 1}}};\rho)}{t_\rho^{\star 1}\log(1+K\sigma^{-2})}} \leq \epsilon_\rho \qquad \text{(Rearranging terms)}$$

$$\frac{\sum_{t=1}^{t_\rho^{\star 1}} w_t}{t_\rho^{\star 1}} \leq K\sqrt{\frac{8\beta_{t_\rho^{\star 1}}^\rho I(Y_{X^m_{1:t_\rho^{\star 1}}};\rho)}{t_\rho^{\star 1}\log(1+K\sigma^{-2})}} \leq \epsilon_\rho \qquad \text{(From Eq. (53) in Lem. 11)}$$

$$\implies \min_{t\in[1,t_\rho^{\star 1}]} w_t \leq \epsilon_\rho \qquad \left(\frac{t_\rho^{\star 1}\min_{t\in[1,t_\rho^{\star 1}]} w_t}{t_\rho^{\star 1}} \leq \frac{\sum_{t=1}^{t_\rho^{\star 1}} w_t}{t_\rho^{\star 1}}\right)$$

Hence there exists $t_\rho^1 < t_\rho^{\star 1}$, such that $w_{t_\rho^1+1} \leq \epsilon_\rho$. ■

For notation convenience we denote with $Reg_l^{\mathrm{O}}(\delta t_\rho^{\star n}) := Reg_l^{\mathrm{O}}(t_\rho^{n-1}+\delta t_\rho^{\star n}) - Reg_l^{\mathrm{O}}(t_\rho^{n-1}) = \sum_{t=t_\rho^{n-1}+1}^{t_\rho^{n-1}+\delta t_\rho^{\star n}} r_t$ and $I(Y_{\delta t_\rho^{\star n}};\rho) = I(Y_{x^{g,1:K}_{t_\rho^{n-1}+1:t_\rho^{n-1}+\delta t_\rho^{\star n}}};\rho)$.

Lemma 14 *Let the coverage phase be terminated for the $(n-1)^{th}$ time at t_ρ^{n-1}, and $\delta t_\rho^{\star n}$ be the maximum number of density measurements required to terminate coverage phase for the*

n^{th} *time. Let* $\delta \in (0,1)$ *and* β_t^ρ *as in [16], i.e.,* $\beta_t^{\rho 1/2} = B_\rho + 4\sigma_\rho\sqrt{\gamma_{Kt}^\rho + 1 + \ln(1/\delta)}$*, then with probability at least* $1-\delta$ *the following inequality holds,*

$$Reg_l^{\mathrm{O}}(\delta t_\rho^{\star n}) \leq \Big(\delta t_\rho^{\star n}\sum_{t=t_\rho^{n-1}+1}^{t_\rho^{n-1}+\delta t_\rho^{\star n}} w_t^2\Big)^{1/2} \leq \sqrt{\frac{8\delta t_\rho^{\star n}K^2\beta^\rho_{t_\rho^{n-1}+\delta t_\rho^{\star n}} I(Y_{\delta t_\rho^{\star n}};\rho)}{\log(1+K\sigma^{-2})}}$$

Proof With definitions,

$$Reg_l^{\mathrm{O}}(\delta t_\rho^{\star n}) = \sum_{t=t_\rho^{n-1}+1}^{t_\rho^{n-1}+\delta t_\rho^{\star n}} r_t \leq \sum_{t=t_\rho^{n-1}+1}^{t_\rho^{n-1}+\delta t_\rho^{\star n}} w_t$$

$$\implies \big(Reg_l^{\mathrm{O}}(\delta t_\rho^{\star n})\big)^2 \leq \sum_{t=t_\rho^{n-1}+1}^{t_\rho^{n-1}+\delta t_\rho^{\star n}} w_t \leq \delta t_\rho^{\star n}\sum_{t=t_\rho^{n-1}+1}^{t_\rho^{n-1}+\delta t_\rho^{\star n}} w_t^2$$

(Using, Cauchy-Schwarz inequality)

Now, the RHS of the inequality can be simplified as,

$$\delta t_\rho^{\star n}\sum_{t=t_\rho^{n-1}+1}^{t_\rho^{n-1}+\delta t_\rho^{\star n}} w_t^2 \leq \delta t_\rho^{\star n}\sum_{t=t_\rho^{n-1}+1}^{t_\rho^{n-1}+\delta t_\rho^{\star n}} \frac{8K^2\beta_t^\rho}{\log(1+K\sigma^{-2})}\sum_{i=1}^{K}\frac{1}{2}\log(1+\sigma^{-2}\lambda_{i,t}) \quad \text{(using Eq. (52))}$$

$$\leq \frac{8\delta t_\rho^{\star n}K^2\beta^\rho_{t_\rho^{n-1}+\delta t_\rho^{\star n}}}{\log(1+K\sigma^{-2})}\sum_{t=t_\rho^{n-1}+1}^{t_\rho^{n-1}+\delta t_\rho^{\star n}}\sum_{i=1}^{K}\frac{1}{2}\log(1+\sigma^{-2}\lambda_{i,t})$$

(since, β_t^ρ is non-decreasing and using definition of mutual information we get,)

$$\implies Reg_l^{\mathrm{O}}(\delta t_\rho^{\star n}) \leq \sum_{t=t_\rho^{n-1}+1}^{t_\rho^{n-1}+\delta t_\rho^{\star n}} w_t \leq \Big(\delta t_\rho^{\star n}\sum_{t=t_\rho^{n-1}+1}^{t_\rho^{n-1}+\delta t_\rho^{\star n}} w_t^2\Big)^{1/2} \leq \sqrt{\frac{8\delta t_\rho^{\star n}K^2\beta^\rho_{t_\rho^{n-1}+\delta t_\rho^{\star n}} I(Y_{\delta t_\rho^{\star n}};\rho)}{\log(1+K\sigma^{-2})}} \quad (58)$$

■

Lemma 15 *Let* $\delta \in (0,1)$ *and* β_t^ρ *as in [16], i.e.,* $\beta_t^{\rho 1/2} = B_\rho + 4\sigma_\rho\sqrt{\gamma_{Kt}^\rho + 1 + \ln(1/\delta)}$ *and* $\delta t_\rho^{\star n}$ *is the smallest integer after* t_ρ^{n-1} *such that* $\frac{\delta t_\rho^{\star n}}{\beta^\rho_{t_\rho^{n-1}+\delta t_\rho^{\star n}} I(Y_{\delta t_\rho^{\star n}};\rho)} \geq \frac{8K^2}{\log(1+K\sigma^{-2})\epsilon_\rho^2}$*, then we know with probability* $1-\delta$ *that there exists* $\delta t_\rho^n < \delta t_\rho^{\star n}$ *such that* $w_{t_\rho^{n-1}+\delta t_\rho^n+1} \leq \epsilon_\rho$*, where* $w_t = \sum_{B\in\mathcal{B}_t}\sum_{i\in B} 2\sqrt{\beta_t^\rho}\max_{v\in D_t^{i-}}\sigma_{t-1}^\rho(v) \leq \epsilon_\rho$.

Proof Given,

$$\frac{\delta t_\rho^{\star n}}{\beta^\rho_{t_\rho^{n-1}+\delta t_\rho^{\star n}} I(Y_{\delta t_\rho^{\star n}};\rho)} \geq \frac{8K^2}{\log(1+K\sigma^{-2})\epsilon_\rho^2}$$

$$\implies K\sqrt{\frac{8\beta^{\rho}_{t_\rho^{n-1}+\delta t_\rho^{\star n}} I(Y_{\delta t_\rho^{\star n}};\rho)}{\delta t_\rho^{\star n}\log(1+K\sigma^{-2})}} \leq \epsilon_\rho$$

$$\frac{\sum_{t_\rho^{n-1}+1}^{t_\rho^{n-1}+\delta t_\rho^{\star n}} w_t}{\delta t_\rho^{\star n}} \leq \epsilon_\rho \qquad \text{(Using Eq. (58) in Lem. 14)}$$

$$\implies \min_{t\in[t_\rho^{n-1}+1, t_\rho^{n-1}+\delta t_\rho^{\star n}]} w_t \leq \epsilon_\rho$$

Hence there exists $\delta t_\rho^n < \delta t_\rho^{\star n}$, such that $w_{t_\rho^{n-1}+\delta t_\rho^n+1} \leq \epsilon_\rho$. ■

Lemma 16 *Let* $\delta \in (0,1)$ *and* $\beta_t^{\rho 1/2} = B_\rho + 4\sigma_\rho\sqrt{\gamma_{Kt}^\rho + 1 + \ln(1/\delta)}$ *and* $t_\rho^\star$ *is the smallest integer such that* $\frac{t_\rho^\star}{\beta^\rho_{t_\rho^\star}\gamma^\rho_{Kt_\rho^\star}} \geq \frac{8K^2}{\log(1+K\sigma^{-2})\epsilon_\rho^2}$*, then for any* $n \geq 1$, $t_\rho^{n-1} + \delta t_\rho^n < t_\rho^\star$.

Proof

$$t_\rho^{n-1} + \delta t_\rho^n < \frac{8K^2\beta^\rho_{t_\rho^{n-1}} I(Y_{x^{g,1:K}_{1:t_\rho^{n-1}}};\rho)}{\log(1+K\sigma^{-2})\epsilon_\rho^2} + \frac{8K^2\beta^\rho_{t_\rho^{n-1}+\delta t_\rho^n} I(Y_{\delta t_\rho^n};\rho)}{\log(1+K\sigma^{-2})\epsilon_\rho^2}$$

$$\text{(using Eq. (57), since } t_\rho^1 < t_\rho^{\star 1})$$

$$< \frac{8K^2\beta^\rho_{t_\rho^{n-1}+\delta t_\rho^n}}{\log(1+K\sigma^{-2})\epsilon_\rho^2}(I(Y_{x^{g,1:K}_{1:t_\rho^{n-1}}};\rho) + I(Y_{\delta t_\rho^n};\rho)$$

$$\text{(Since, } \beta_t^\rho \text{ is non decreasing function)}$$

$$= \frac{8K^2\beta^\rho_{t_\rho^{n-1}+\delta t_\rho^n} I(Y_{x^{g,1:K}_{1:t_\rho^{n-1}+\delta t_\rho^n}};\rho)}{\log(1+K\sigma^{-2})\epsilon_\rho^2} \qquad \text{(Since mutual info is additive)}$$

$$< \frac{8K^2\beta^\rho_{t_\rho^{n-1}+\delta t_\rho^n}\gamma^\rho_{K(t_\rho^{n-1}+\delta t_\rho^n)}}{\log(1+K\sigma^{-2})\epsilon_\rho^2} \tag{59}$$

Using Eq. (59) and since, $t_\rho^\star \geq \frac{8K^2\beta^\rho_{t_\rho^\star}\gamma^\rho_{Kt_\rho^\star}}{\log(1+K\sigma^{-2})\epsilon_\rho^2}$, we get $t_\rho^{n-1} + \delta t_\rho^n < t_\rho^\star$. ■

Lemma 17 *When* SafeMaC *converges, i.e,* $U \coloneqq \{S_t^{o,\epsilon_q,i} \backslash S_t^{p,i}) \cap D_t^i, \forall i \in [K]\} = \emptyset$*, then the following inequality holds,*

$$\sum_{B\in\mathcal{B}_t} F(X_t^B;\rho, S_t^{p,B}) = \sum_{B\in\mathcal{B}_t} F(X_t^B;\rho, S_t^{u,B})$$

Proof Since, $U = \emptyset$,

$$\{(S_t^{o,\epsilon_q,i} \backslash S_t^{p,i}) \cap D_t^i, \forall i \in [K]\} = \emptyset$$

$$\implies (S_t^{o,\epsilon_q,i} \cap D_t^i) \subseteq (S_t^{p,i} \cap D_t^i) \quad \forall i \in [K]$$

$$= (S_t^{u,i} \cap D_t^i) \ \ \forall i \in [K] \qquad \text{(Since } S_t^{u,i} := S_t^{p,i} \cup S_t^{o,\epsilon_q,i}\text{)}$$

Based on the last equality, it directly follows,

$$\sum_{B \in \mathcal{B}_t} F(X_t^B; \rho, S_t^{p,B}) = \sum_{B \in \mathcal{B}_t} F(X_t^B; \rho, S_t^{u,B}).$$

■

F.1 Intermediate recommendation is near-optimal at SafeMaC's convergence

Lemma 18 *Let $\delta \in (0,1)$ and β_t^ρ as in [16], i.e., $\beta_t^{\rho 1/2} = B_\rho + 4\sigma_\rho \sqrt{\gamma_{Kt}^\rho + 1 + \ln(1/\delta)}$ and $t_\rho^\star$ be the smallest integer such that $\frac{t_\rho^\star}{\beta_{t_\rho^\star} \gamma_{Kt_\rho^\star}} \geq \frac{8K^2}{\log(1+K\sigma^{-2})\epsilon_\rho^2}$. Let β_t^q and $t_q^\star$ be defined analogously. Then, there exists $t < t_q^\star + t_\rho^\star$, such that with probability at least $1-\delta$*

$$\sum_{B \in \mathcal{B}_T} F(X_T^B; \rho, \bar{R}_0(X_0^B)) \geq (1 - \frac{1}{e}) \sum_{B \in \mathcal{B}} F(X_\star^B; \rho, \bar{R}_{\epsilon_q}(X_0^B)) - \epsilon_\rho \tag{60}$$

where,

$$X_T = \underset{X_T, X_T^l, T \in [1,t]}{\arg\max} \Big\{ \sum_{B \in \mathcal{B}_T^p} F(X_T^B; l_{T-1}^\rho, S_T^{p,B}), \sum_{B \in \mathcal{B}_T^p} F(X_T^{l,B}; l_{T-1}^\rho, S_T^{p,B}) \Big\} \ s.t. X_T \in S_T^p \tag{61}$$

and $X_t^{l,B}$, i.e., the greedy solution w.r.t. the worst-case objective, $F(\cdot; l_{t-1}^\rho, S_t^{p,B}) \, \forall B \in \mathcal{B}_t^p$.

Proof We prove the lemma in two parts. First, we prove the near optimality of SafeMaC's solution X_t but evaluated using l_{t-1}^ρ instead of ρ. This will imply the near optimality at convergence of the 1^{st} term ($\sum_{B \in \mathcal{B}_T^p} F(X_T^B; l_{T-1}^\rho, S_T^{p,B})$) in the above recommendation rule. Secondly, due to the arg max operator, the near optimality of the 1^{st} term is sufficient to establish the optimality of the recommendation rule in Eq. (61).
Notations. $X_t = \cup_{B \in \mathcal{B}_t} X_t^B$, $\Delta(\cdot; \rho, V)$ as defined in Eq. (17).
Given. From Theorem 2, for $t < t_q^\star + t_\rho^\star$ with probability at least $1-\delta$,

$$\sum_{B \in \mathcal{B}_t} F(X_t^B; \rho, \bar{R}_0(X_0^B)) \geq (1 - \frac{1}{e}) \sum_{B \in \mathcal{B}} F(X_\star^B; \rho, \bar{R}_{\epsilon_q}(X_0^B)) - \epsilon_\rho \tag{62}$$

and

$$\sum_{B \in \mathcal{B}_t} \sum_{i \in B} 2\sqrt{\beta_t^\rho} \max_{v \in D_t^{i-}} \sigma_{t-1}^\rho(v) \leq \epsilon_\rho$$

Near-optimality of SafeMaC's X_t evaluated using l_{t-1}^ρ.

$$\Delta(\tilde{x}_t^i | X_t^{1:i-1}; \rho, S_t^{u,B}) - \Delta(x_t^i | X_t^{1:i-1}; l_{t-1}^\rho, S_t^{u,B})$$

$$= \Big(\sum_{v \in \tilde{D}_t^{i-}} \rho(v) - \sum_{v \in D_t^{i-}} l_{t-1}^{\rho}(v) \Big)/N \qquad \text{(Note } D_t^{i-} \text{ and } \tilde{D}_t^{i-}\text{)}$$

$$\leq \Big(\sum_{v \in D_t^{i-}} \big(\mu_{t-1}^{\rho}(v) + \sqrt{\beta_t^{\rho}} \sigma_{t-1}^{\rho}(v)\big) - \big(\mu_{t-1}^{\rho}(v) - \sqrt{\beta_t^{\rho}} \sigma_{t-1}^{\rho}(v)\big) \Big)/N$$

$$\text{(Using Eq. (55) and definition of } l_{t-1}^{\rho}(v)\text{)}$$

$$= 2\sqrt{\beta_t^{\rho}} \sum_{v \in D_t^{i-}} \sigma_{t-1}^{\rho}(v)/N$$

$$\leq 2\sqrt{\beta_t^{\rho}} \max_{v \in D_t^{i-}} \sigma_{t-1}^{\rho}(v) \tag{63}$$

$$\sum_{B \in \mathcal{B}_t} \sum_{i \in B} \Delta(\tilde{x}_t^i | X_t^{1:i-1}; \rho, S_t^{u,B}) - \Delta(x_t^i | X_t^{1:i-1}; l_{t-1}^{\rho}, S_t^{u,B}) \leq \sum_{B \in \mathcal{B}_t} \sum_{i \in B} 2\sqrt{\beta_t^{\rho}} \max_{v \in D_t^{i-}} \sigma_{t-1}^{\rho}(v)$$

$$\leq \epsilon_{\rho}$$

Using the following two statements,

- $(1 - \frac{1}{e}) F(X_{\star}; \rho, S_t^{u,B}) \leq \sum_{i \in B} \Delta(\tilde{x}^i | X_t^{1:i-1}; \rho, S_t^{u,B})$ from Eq. (39)
- $\bigcup_{i \in B} \bar{R}_{\epsilon_q}(\{x_0^i\}) \subseteq S_t^{u,B} \implies \sum_{B \in \mathcal{B}} F(X_{\star}; \rho, \bar{R}_{\epsilon_q}(X_0)) \leq \sum_{B \in \mathcal{B}_t} F(X_{\star}; \rho, S_t^{u,B})$

we get,

$$\implies \sum_{B \in \mathcal{B}_t} F(X_t^B; l_{t-1}^{\rho}, S_t^{u,B}) \geq (1 - \frac{1}{e}) \sum_{B \in \mathcal{B}} F(X_{\star}^B; \rho, \bar{R}_{\epsilon_q}(X_0^B)) - \epsilon_{\rho} \tag{64}$$

Near-optimality of recommendation as per Eq. (61).
Let's consider the following recommendation rule,

$$X_T = \underset{X_T, T \in [1,t]}{\arg\max} \Big\{ \sum_{B \in \mathcal{B}_T^p} F(X_T^B; l_{T-1}^{\rho}, S_T^{p,B}) \Big\} \; s.t. X_T \in S_T^p \tag{65}$$

At convergence, $S_t^{p,i} \cap D_t^i = S_t^{u,i} \cap D_t^i \implies (S^{o,\epsilon_q} \backslash S_t^{p,i}) \cap D_t^i = \emptyset$, using this SafeMaC recommendation X_t can be written as,

$$\sum_{B \in \mathcal{B}_t} F(X_t^B; l_{t-1}^{\rho}, S_t^{u,B}) = \sum_{i \in [K]} \Delta(x_t^i | X_t^{1:i-1}; l_{t-1}^{\rho}, S_t^{p,i}) = \sum_{B \in \mathcal{B}_t^p} F(X_t^B; l_{t-1}^{\rho}, S_t^{p,B})$$

$$\sum_{B \in \mathcal{B}_T^p} F(X_T^B; l_{T-1}^{\rho}, S_T^{p,B}) \geq \sum_{B \in \mathcal{B}_t^p} F(X_t^B; l_{t-1}^{\rho}, S_t^{p,B})$$

$$\text{(since, } X_T^B = \underset{X_T^B, T \in [1,t]}{\arg\max} \textstyle\sum_{B \in \mathcal{B}_T^p} F(X_T^B; l_{T-1}^{\rho}, S_T^{p,B})\text{)}$$

$$\implies \sum_{B \in \mathcal{B}_T^p} F(X_T^B; l_{T-1}^{\rho}, S_T^{p,B}) \geq \sum_{B \in \mathcal{B}_t} F(X_t^B; l_{t-1}^{\rho}, S_t^{u,B})$$

$$\text{(Combining the above 2 equations)}$$

$$\implies \sum_{B\in\mathcal{B}_T^p} F(X_T^B; l_{T-1}^\rho, S_T^{p,B}) \geq (1-\frac{1}{e})\sum_{B\in\mathcal{B}} F(X_\star^B; \rho, \bar{R}_{\epsilon_q}(X_0^B)) - \epsilon_\rho \qquad \text{(using Eq. (64))}$$

Hence, the recommendation of Eq. (65) evaluated with lower bound is near optimal (at convergence $X_T \in S_T^p$). Further, due to arg max operator Eq. (65) also implies near-optimality of recommendation rule in Eq. (61) evaluated with the lower bound. So now using X_T chosen as per Eq. (61) and at convergence, $\forall i, (S_t^{p,i} \cap D_t^i) \subseteq (\bar{R}_0(\{x_0^i\}) \cap D_t^i)$, we get,

$$\sum_{B\in\mathcal{B}_T^p} F(X_T^B; \rho, S_T^{p,B}) \geq (1-\frac{1}{e})\sum_{B\in\mathcal{B}} F(X_\star^B; \rho, \bar{R}_{\epsilon_q}(X_0^B)) - \epsilon_\rho \qquad (l_{t-1}^\rho(v) \leq \rho(v)\forall v)$$

$$\sum_{B\in\mathcal{B}_T^p} F(X_T^B; \rho, \bar{R}_0(X_0)) \geq (1-\frac{1}{e})\sum_{B\in\mathcal{B}} F(X_\star^B; \rho, \bar{R}_{\epsilon_q}(X_0^B)) - \epsilon_\rho$$

Hence Proved. ■

Appendix G. Multi-agent GoOSE version

This section presents our fundamental lemma for the multi-agent version of goose. The result is built upon Theorem 1 of [11]. Since agents are sharing the observations among all in our case, we first derive a finite time bound for learning constrained function up to ϵ_q-accuracy under the cooperative sharing setting. Later, we present our key Lem. 22, which guarantees complete exploration by each agent under finite time while preserving safety.

Lemma 19 *Let $\delta \in (0,1)$ and let $(\beta_t^q)^{1/2} = B_q + 4\sigma_q\sqrt{\gamma_{Kt}^q + 1 + \ln(1/\delta)}$. Then the following hols with probability at least $1-\delta$,*

$$\sum_t \omega_t^2 \leq C_1 \beta_t^q I(Y_A; q) \leq C_1 \beta_t^q \gamma_{KT}^q,$$

where $C_1 = 8/\log(1+\sigma_q^{-2})$, $\omega_t = 2\sqrt{\beta_t^q}\sigma_{t-1}^q(x_t^i)$, and x_t^i is the location visited by some agent i at time t. A is the set of locations visited by agents to collect constraint observations. $I(Y_A; q)$ is information gain due to these interactions and γ_{KT}^q is the information capacity.

Proof Using $\omega_t = 2\sqrt{\beta_t^q}\sigma_{t-1}^q(x_t^i)$,

$$\begin{aligned}
\omega_t^2 &= 4\beta_t^q(\sigma_{t-1}^q(v))^2 = 4\beta_t^q\sigma_q^2\sigma_q^{-2}(\sigma_{t-1}^q(x_t^i))^2 \\
&\leq 4\beta_t^q\sigma_q^2 C_2 \log(1+\sigma_q^{-2}(\sigma_{t-1}^q(x_t^i))^2) \\
&\qquad \text{(Since, } s \leq C_2\log(1+s) \text{ for } s\in[0,\sigma_q^{-2}]\text{, where } C_2 = \sigma_q^{-2}/\log(1+\sigma_q^{-2}) \geq 1) \\
&\qquad \text{(Since, } s = \sigma_q^{-2}\sigma_{t-1}^q(v)^2 \leq \sigma_q^{-2}k^q(v,v) \leq \sigma_q^{-2}, (wlog\ k^q(v,v)\leq 1)) \\
&\leq C_1\beta_t^q \frac{1}{2}\log(1+\sigma_q^{-2}(\sigma_{t-1}^q(x_t^i))^2) && (C_1 = 8\sigma_q^2 C_2) \\
&\leq C_1\beta_t^q \frac{1}{2}\log(1+\sigma_q^{-2}\sum_{i=1}^K(\sigma_{t-1}^q(x_t^i))^2) && \text{(Since, } (\sigma_{t-1}^q(x_t^i))^2 \leq \textstyle\sum_{i=1}^K(\sigma_{t-1}^q(x_t^i))^2) \\
&= C_1\beta_t^q \frac{1}{2}\log(1+\sigma_q^{-2}\sum_i^K \lambda_{i,t}) && (\textstyle\sum_{i=1}^K(\sigma_{t-1}^\rho(x_t^i))^2 = Tr(K^q) = \sum_{i=1}^K\lambda_{i,t}) \\
&\leq C_1\beta_t^q \sum_{i=1}^K \frac{1}{2}\log(1+\sigma_q^{-2}\lambda_{i,t}) \\
&\qquad (\log(1+x_1+x_2) \leq \log(1+x_1)+\log(1+x_2), for\ x_1,x_2\geq 0) \\
&= C_1\beta_t^q I(Y_A;q) && (I(;q) \text{ is defined analogous to } I(;\rho) \text{ in Eq. (44)}) \\
&\leq C_1\beta_t^q\gamma_{KT}^q && (\gamma_{KT}^q = \textstyle\sup_{A\subset V;|A|=KT} I(Y_A;q))
\end{aligned}$$

Hence Proved. ■

Similar to Lem. 8 of Turchetta et al. [11], Let us denote $\mathcal{T}_t^v = \{\tau_1, ..., \tau_j\}$ the set of steps where the constraint q is evaluated at v by step t.

Lemma 20 *For any $t \geq 1$ and for any $v \in V$, it holds that $w_t(v) \leq \sqrt{\frac{C_1\beta_t^q\gamma_{Kt}^q}{|\mathcal{T}_t^v|}}$, with $C_1 = 8/\log(1+\sigma_q^{-2})$.*

Proof

$$
\begin{aligned}
|\mathcal{T}_t^v| w_t^2(v) &\leq \sum_{\tau \in \mathcal{T}_t^v} w_\tau^2(v) \\
&\leq \sum_{\tau \in \mathcal{T}_t^v} 4\beta_\tau^q (\sigma_{t-1}^q(x_t^i))^2 \\
&\leq \sum_{\tau \in t} 4\beta_\tau^q (\sigma_{t-1}^q(x_t^i))^2 \\
&\leq C_1 \beta_t^q \gamma_{Kt}^q
\end{aligned} \tag{66}
$$

Eq. (66), follows due to intersection of confidence interval arguments, Lemma 1 of Turchetta et al. [11] and the inequality follows due to Lem. 19. ■

Let us denote with T_t, the smallest positive integer such that $\frac{T_t}{\beta^q_{t+T_t}\gamma^q_{K,t+T_t}} \geq \frac{C_1}{\epsilon_q^2}$, with $C_1 = 8/\log(1+\sigma_q^{-2})$ and with $t^\star$ the smallest positive integer such that $t^\star \geq |\bar{R}_0(X_0)| T_{t^\star}$.

Lemma 21 *For any $t \leq t^\star$, for any $x \in V$ such that $|\mathcal{T}_t^v| \geq T_{t^\star}$, it holds that $w_t(v) \leq \epsilon_q$.*

Proof Since T_t is an increasing function of t [49], we have $|\mathcal{T}_t^v| \geq T_{t^\star} \geq T_t$. Therefore using Lem. 20 and the definition of T_t, we get,

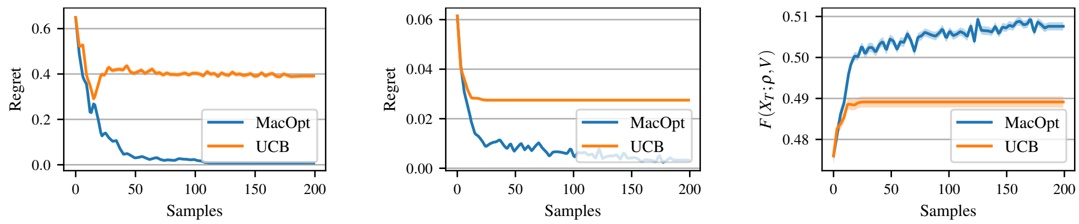
$$
w_t(v) \leq \sqrt{\frac{C_1 \beta_t^q \gamma_{Kt}^q}{T_t}} \leq \sqrt{\frac{C_1 \beta_t^q \gamma_{Kt}^q \epsilon_q^2}{C_1 \gamma^q_{K,t+T_t} \beta^q_{t+T_t}}} \leq \sqrt{\frac{\beta_t^q \gamma_{Kt}^q}{\gamma^q_{K,t+T_t} \beta^q_{t+T_t}}}\epsilon_q \leq \epsilon_q.
$$

The last inequality follows from the fact that both β_t^q and γ_t^q are non-decreasing function of t. ■

Lemma 22 *Assume that $q(\cdot)$ is L_q-Lipschitz continuous w.r.t $d(.,.)$ with $\|q\|_k \leq B_q$, $X_0 \neq \emptyset$, $q(x_0^i) \geq 0$ for all $i \in [K]$. Let $(\beta_t^q)^{1/2} = B_q + 4\sigma_q\sqrt{\gamma_{Kt}^q + 1 + \ln(1/\delta)}$, then, for any heuristic $h_t : V \to \mathbb{R}$, with probability at least $1-\delta$, we have $q(x) \geq 0$, for any x visited by any agent in* SafeMaC. *Moreover, let γ_{Kt}^q denote the information capacity associated with the kernel k^q and let $t_q^\star$ be the smallest integer such that $\frac{t_q^\star}{\beta_{t_q^\star}\gamma_{Kt_q^\star}} \geq \frac{C_1|\bar{R}_0(X_0)|}{\epsilon_q^2}$, with $C_1 = 8/\log(1+\sigma_q^{-2})$, then there exists $t \leq t_q^\star$ such that, with probability at least $1-\delta$, $\bar{R}_{\epsilon_q}(\{x_0^i\}) \subseteq S^{o,\epsilon_q,i} \subseteq S^{p,i} \subseteq \bar{R}_0(\{x_0^i\})$ for all $i \in [K]$.*

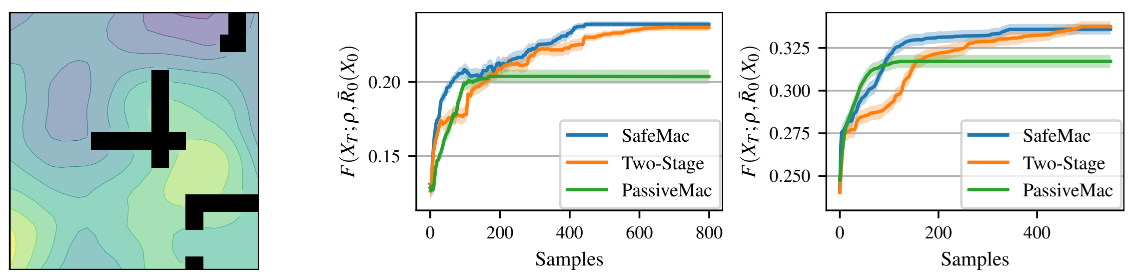
Proof In SafeMaC, each agent have a record for its optimistic and pessimistic set.The lemma is similar to K instances of Theorem 1 of Turchetta et al. [11]; each instance corresponds to per agent case. Safety is a direct consequence of Theorem 2 of Turchetta et al. [51]. Finite time bound while agents are sharing information is consequence of Lem. 19-21. The convergence of the pessimistic and optimistic approximation of the safe sets for each agent is a direct consequence of Lemmas 16-18 of Turchetta et al. [11]. ■

For a detailed discussion, we refer the reader to Appendix D Completeness of Turchetta et al. [11].



(a) Gorilla environment (b) GP environment (c) GP environment

Figure 4: Compares MACOPT and UCB on the Gorilla (a) and the GP (b,c) environment. a,b) Compares simple regret r_t Eq. (29) in the unconstrained case (domain V). c) Plots total coverage achieved by both the algorithms.



(a) Obstacles environment (b) Coverage in Obstacle env. (c) Coverage in GP env.

Figure 5: a) The contours show the synthetic density and the obstacles marked by the black blocks. b,c) Comparison of SAFEMAC with PASSIVEMAC and Two-Stage in the Obstacle and the GP environment during optimization

Appendix H. Experiments

Implementation details. We implemented all our algorithms with BoTorch [57] and GPyTorch [58] frameworks, built on top of Pytorch [59]. The code for both the algorithms will be made public along with the competitive baselines. We limit the maximum number of rounds to 300, and with the selected hyperparameters and the given environments, it terminates before that. This roughly takes *10 min* of training for SAFEMAC on a single core CPU. The code is written for running a single instance of the experiment. In practice, we launch nearly 1000 such instances simultaneously on the cluster in parallel to get results about different environments, noise realizations and initializations.

Gorilla Environment. The gorilla environment (Fig. 2a) is defined in a grid of 34×34, with each grid cell being a square of length 0.1. The $K = 3$ agents perform the coverage task, with each having a sensing region defined as a set of locations agents that can travel in 5 steps in the underlying transition graph (Precisely, $D^i = R_5^{\text{reach}}(\{x^i\}), Eq.$ (9)). We considered 10 gorilla environments each differ in the initial location of the agents. The nest density is obtained by fitting a smooth rate function [18] over Gorilla nest site locations which were provided by the Wildlife Conservation Society Takamanda-Mone Landscape project (WCS-TMLP) Funwi-gabga and Mateu [19]. As a proxy for bad weather, we use the cloud coverage data over the Kagwene Gorilla Sanctuary from OpenWeather [55]. The density and the constraint function used are available in our code base. The code for fitting a rate function is available here (https://github.com/Mojusko/sensepy) under the MIT license. We used a lengthscale of 1 for the density and of 2 for the constraint function. The noise variance is set to 10^{-3} and 7×10^{-3} for density and the constraint respectively. However, the performance in the experiments is not sensitive to the hyperparameters and is easily reproducible with other sensible parameters as well.

Obstacles Environment. The obstacle environment (Fig. 5a) is defined on a grid of 30×30, with each grid cell being a square of length 0.1. The sensing region and number of agents are defined similarly to the Gorilla environment. The obstacle is completely defined by the location of its top right corner and the bottom left corner. The obstacle environment is generated by combining a set of such obstacles. The density is directly sampled from a GP with the parameters same as synthetic data. We produced ten instances of environments, each having a different set of obstacles and GP sample and initialization. We used a lengthscale of 2 for both density and the constraint function. The noise variance is set to 10^{-3}. Similar to earlier environments, performance is not sensitive to hyperparameters.

Experiment results.

Unconstrained case Fig. 4a and Fig. 4b plots the simple regret r_t for each round t, precisely, defined as $\sum_{i=1}^{K} \Delta(\tilde{x}|X^{1:i-1}; \rho, V) - \Delta(x_t^i|X^{1:i-1}; \rho, V)$. This quantity upper bounds the actual regret and provides intuition for the convergence rate. We see in the plots that the simple regret goes to zero for MaCOpt, but gets stuck for the UCB algorithm. Due to this, we also observe that MaCOpt can achieve higher coverage value as compared to UCB in Fig. 4c.

Constrained case Fig. 5b and Fig. 5c compares coverage of area attained by SafeMaC, PassiveMaC and the two stage algorithm. Precisely the intermediate locations are recommended as per Eq. (61). We see that SafeMaC finds a comparable solution to two stage more efficiently without exploring the whole environment, where as PassiveMaC gets stuck in the local optimum.