Balancing Communications and Computations in Gradient Tracking Algorithms for Distributed Optimization
Additional Numerical Experiments

Albert S. Berahas* Raghu Bollapragada† Shagun Gupta†‡

March 23, 2023

In this document, we present additional numerical experiments that accompany [?]. These experiments illustrate the empirical performance of the methods defined in Table 1 using Python implementations[1]. We present results on two problems: (1) a synthetic strongly convex quadratic problem (Section 1); and, (2) binary classification logistic regression problems (Section 2) over multiple datasets summarised in Table 3. We investigated network structures (different mixing matrix $\mathbf{W}$) with $n = 16$ nodes summarised in Table 2

The methods defined in Table 1 are denoted as `GTA`$-i(n_c, n_g)$, $i = 1, 2, 3$, where n_c and n_g are the number of communication and computation steps, respectively. The performance of the methods was measured in terms of the optimization error ($\|\bar{x}_k - x^*\|_2$) and the consensus error ($\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|_2$). We do not track the consensus error in auxiliary variable $\mathbf{y}_k$ ($\|\mathbf{y}_k - \bar{\mathbf{y}}_k\|_2$) as this measure does not provide any significant additional insights about the performance of the algorithms. We do not present the consensus error for Fully Connected Networks as in this case, it is always zero for `GTA-2` and `GTA-3` while can be brought to zero for `GTA-1` in one communication. The optimal solution x^* for quadratic problem was obtained analytically and for the logistic regression problems was obtained by running gradient descent in the centralized setting to high accuracy, i.e., $\|\nabla f(x^*)\|_2 \leq 10^{-12}$.

*Department of Industrial and Operations Engineering, University of Michigan. (albertberahas@gmail.com)

†Operations Research and Industrial Engineering Program, University of Texas at Austin. (raghu.bollapragada@utexas.edu,shagungupta@utexas.edu)

‡Corresponding author.

[1] Our code will be made publicly available upon publication of the manuscript. Github repository: https://github.com/Shagun-G/Gradient-Tracking-Algorithmic-Framework.

Table 1: Special cases of Gradient Tracking Algorithms (GTA).

Method	Communication Matrices				Algorithms in literature
	$\mathbf{W}_1$	$\mathbf{W}_2$	$\mathbf{W}_3$	$\mathbf{W}_4$	$(n_c = n_g = 1)$
GTA-1	$\mathbf{W}$	I_n	$\mathbf{W}$	I_n	EXTRA[4] [8], DIGing [5]
GTA-2	$\mathbf{W}$	$\mathbf{W}$	$\mathbf{W}$	I_n	SONATA [9], NEXT [2], [7]
GTA-3	$\mathbf{W}$	$\mathbf{W}$	$\mathbf{W}$	$\mathbf{W}$	Aug-DGM [10], ATC-DIGing [6]

Table 2: Graphs Summary

Graph Name	$\lVert \mathbf{W} - \frac{11^T}{n} \rVert$
Full	0
Graph 1	0.125
Graph 2	0.247
Star	0.95
Cyclic	0.992
Line	0.998

Table 3: Summary of Datasets

Dataset	d	$\sum_{i=1}^n n_i$
Mushroom [3]	117	8124
Australian [3]	41	690
Sonar [1]	60	208
Ionosphere [1]	34	351
Phishing [1]	68	11055

1 Quadratic Problems

We first consider quadratic problems

$$f(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{2} x^T Q_i x + b_i^T x,$$

where $Q_i \in \mathbb{R}^{10\times 10}$, $Q_i \succ 0$ and $b_i \in \mathbb{R}^{10}$ is the local information at each node $i \in \{1, 2, .., n\}$, and $n = 16$. Each local problem is strongly convex and was generated using the procedure described in [4], with global condition number, $\kappa \approx 10^4$.

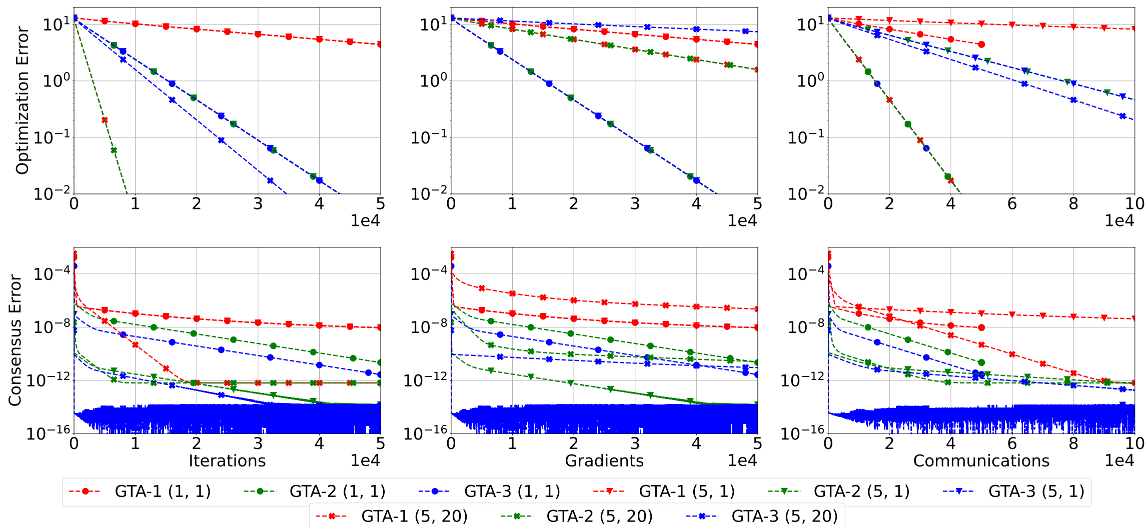

Figure 1: Optimization Error ($\|\bar{x}_k - x^*\|_2$) and Consensus Error ($\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|_2$) of `GTA-1`, `GTA-2` and `GTA-3` with respect to number of iterations, communications and gradient evaluations for the synthetic quadratic problem over the Graph 1 network.

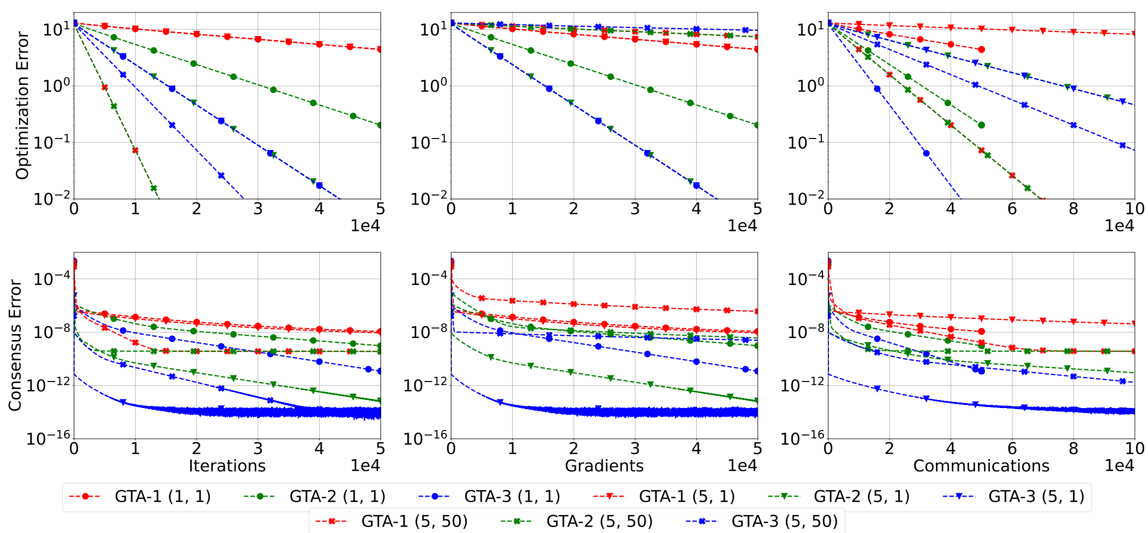

Figure 2: Optimization Error ($\|\bar{x}_k - x^*\|_2$) and Consensus Error ($\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|_2$) of `GTA-1`, `GTA-2` and `GTA-3` with respect to number of iterations, communications and gradient evaluations for the synthetic quadratic problem over the Graph 2 network.

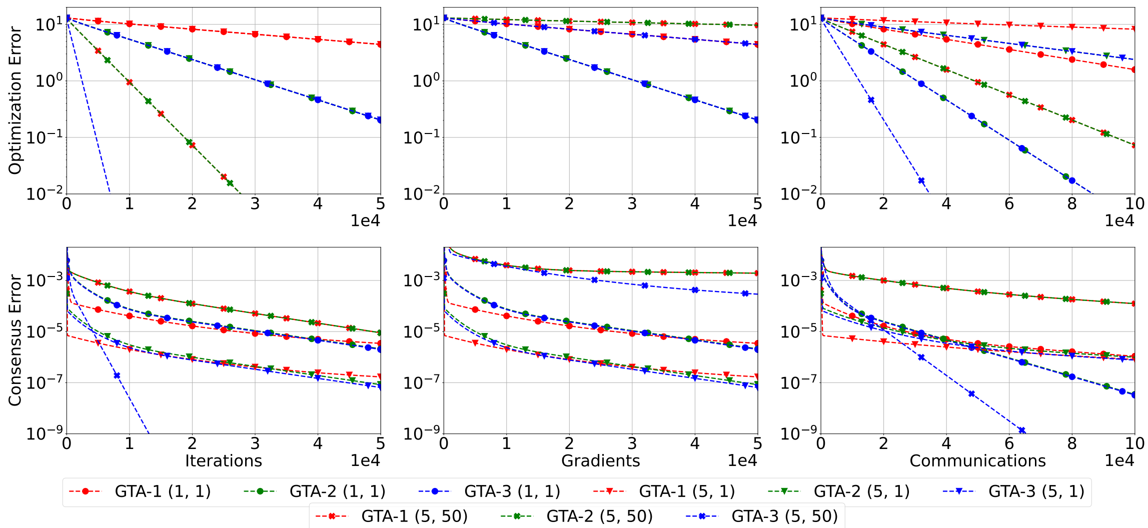

Figure 3: Optimization Error ($\|\bar{x}_k - x^*\|_2$) and Consensus Error ($\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|_2$) of `GTA-1`, `GTA-2` and `GTA-3` with respect to number of iterations, communications and gradient evaluations for the synthetic quadratic problem over the star network.

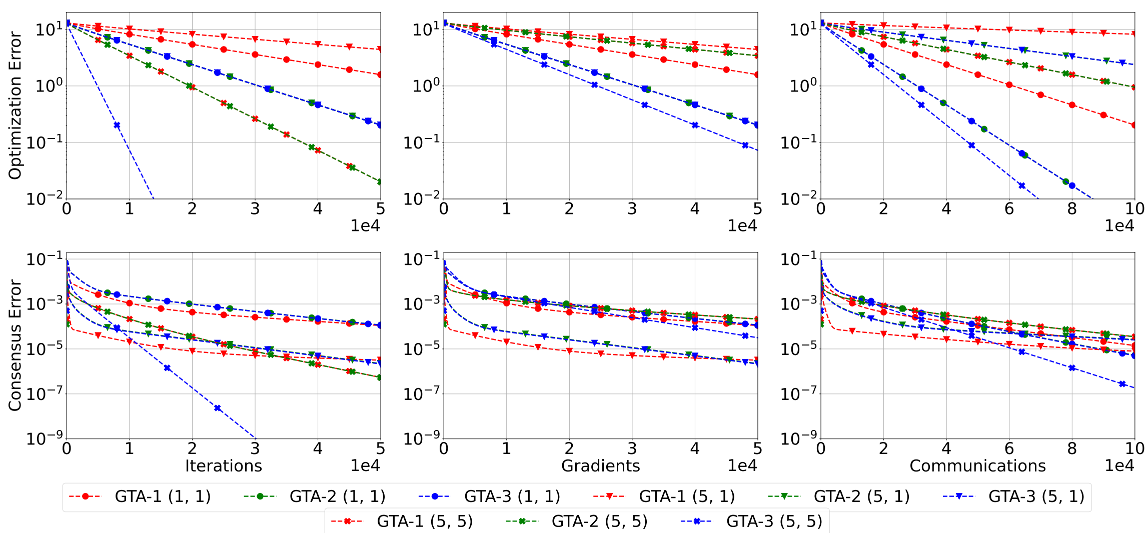

Figure 4: Optimization Error ($\|\bar{x}_k - x^*\|_2$) and Consensus Error ($\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|_2$) of `GTA-1`, `GTA-2` and `GTA-3` with respect to number of iterations, communications and gradient evaluations for the synthetic quadratic problem over the cyclic network.

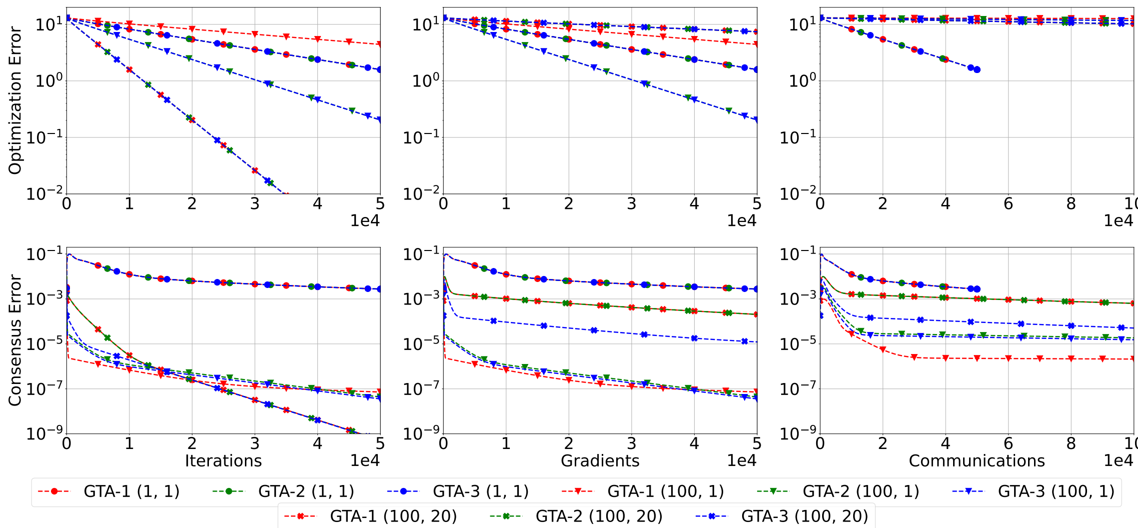

Figure 5: Optimization Error ($\|\bar{x}_k - x^*\|_2$) and Consensus Error ($\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|_2$) of `GTA-1`, `GTA-2` and `GTA-3` with respect to number of iterations, communications and gradient evaluations for the synthetic quadratic problem over the line network.

2 Binary Classification Logistic Regression

Next, we consider ℓ_2-regularized binary classification logistic regression problems of the form

$$f(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{n_i}\log(1+e^{-b_i^T A_i x}) + \frac{1}{n_i}\|x\|_2^2,$$

where each node $i \in \{1, 2, .., n\}$ has a portion of data samples $A_i \in \mathbb{R}^{n_i \times d}$ and corresponding labels $b_i \in \{0, 1\}^{n_i}$. Experiments were performed over the datasets summarised in Table 3.

2.1 Australian Dataset

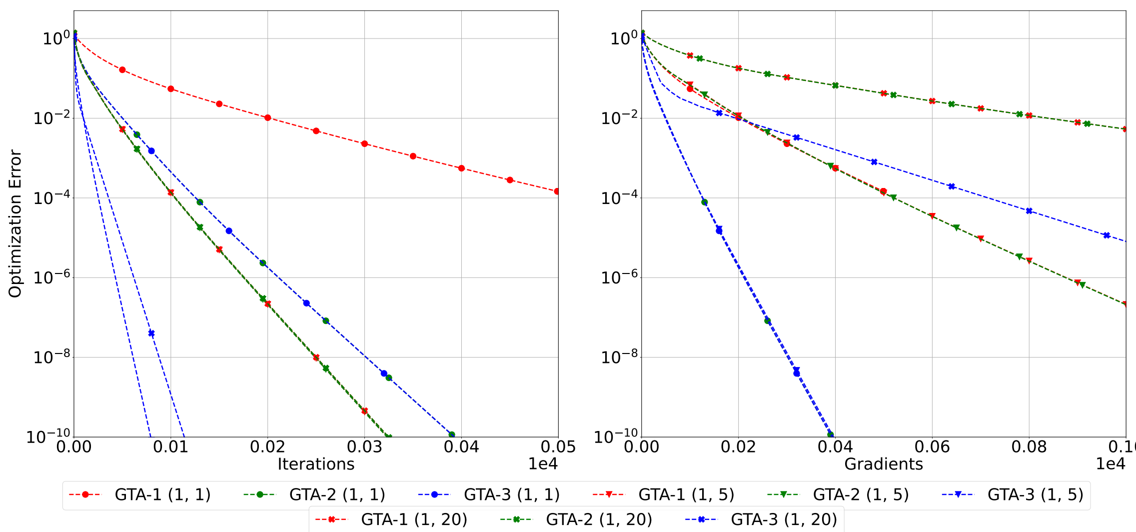

Figure 6: Optimization Error ($\|\bar{x}_k - x^*\|_2$) of `GTA-1`, `GTA-2` and `GTA-3` with respect to number of iterations, communications and gradient evaluations for binary logistic regression on Australian dataset over the Full network.

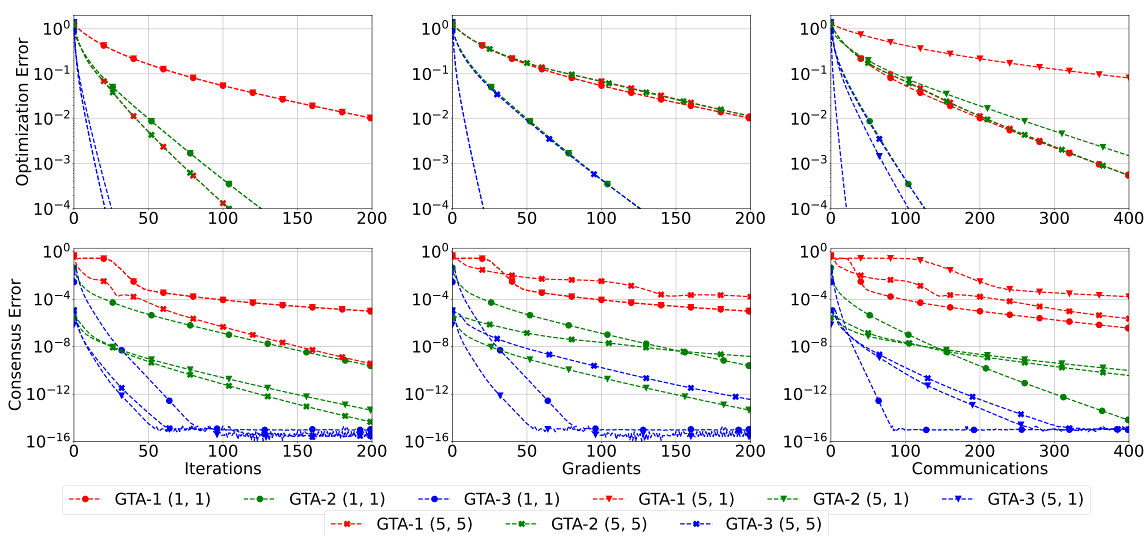

Figure 7: Optimization Error ($\|\bar{x}_k - x^*\|_2$) and Consensus Error ($\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|_2$) of `GTA-1`, `GTA-2` and `GTA-3` with respect to number of iterations, communications and gradient evaluations for binary logistic regression on Australian dataset over the Graph 1 network.

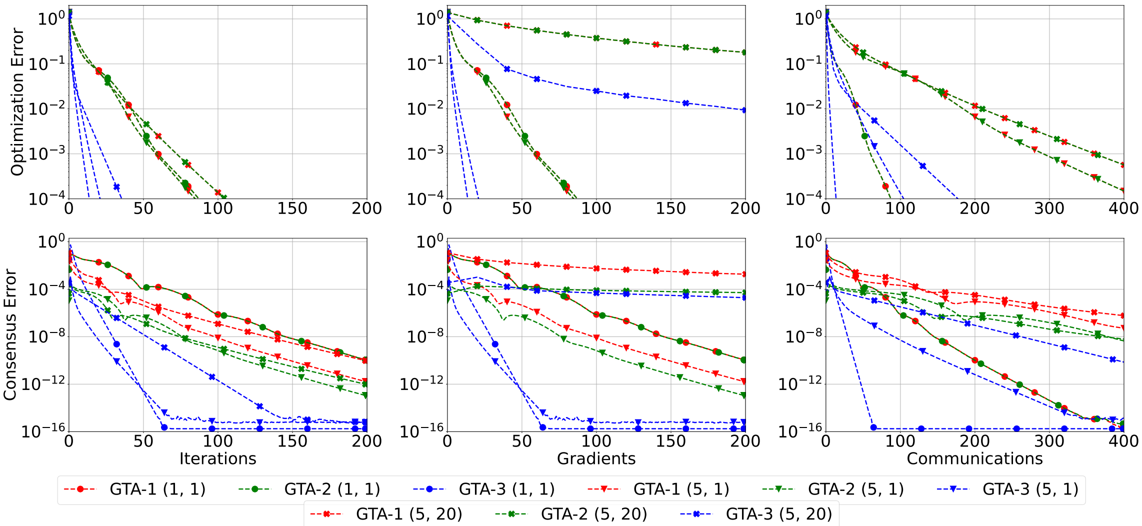

Figure 8: Optimization Error ($\|\bar{x}_k - x^*\|_2$) and Consensus Error ($\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|_2$) of `GTA-1`, `GTA-2` and `GTA-3` with respect to number of iterations, communications and gradient evaluations for binary logistic regression on Australian dataset over the Graph 2 network.

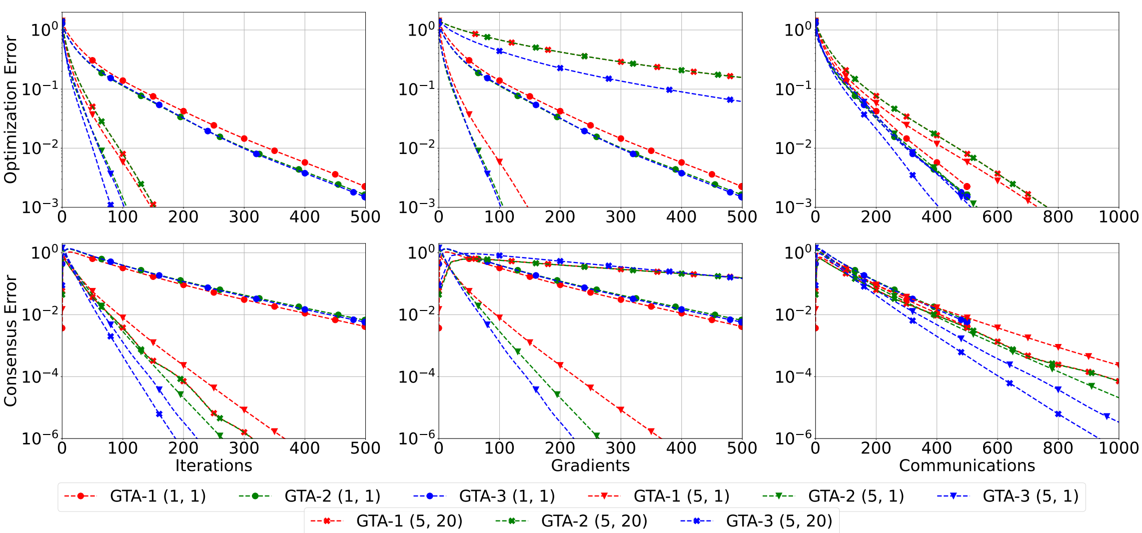

Figure 9: Optimization Error ($\|\bar{x}_k - x^*\|_2$) and Consensus Error ($\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|_2$) of `GTA-1`, `GTA-2` and `GTA-3` with respect to number of iterations, communications and gradient evaluations ffor binary logistic regression on Australian dataset over the star network.

2.2 Mushroom Dataset

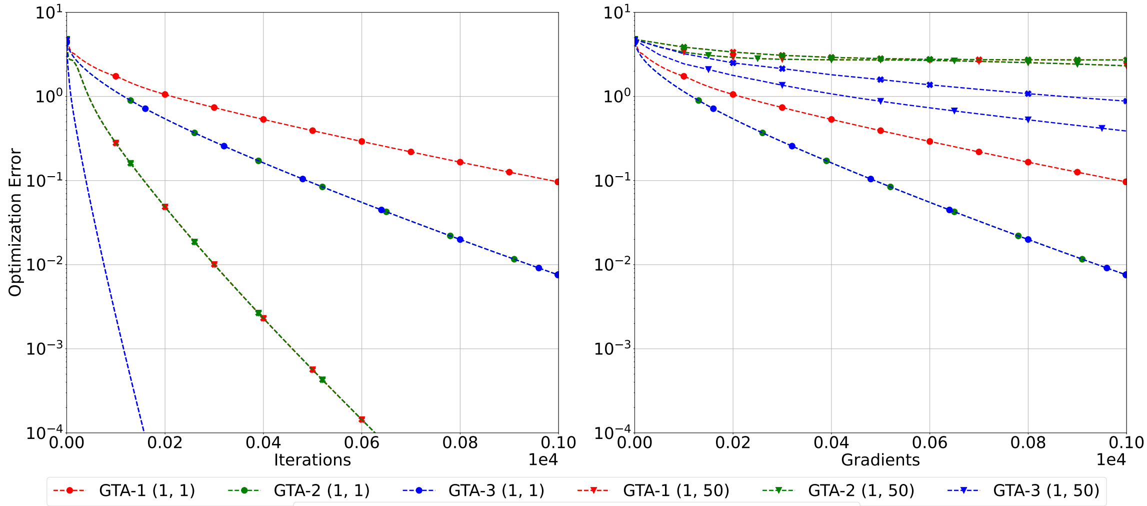

Figure 10: Optimization Error ($\|\bar{x}_k - x^*\|_2$) of `GTA-1`, `GTA-2` and `GTA-3` with respect to number of iterations, communications and gradient evaluations for binary logistic regression on Mushroom dataset over the Full network.

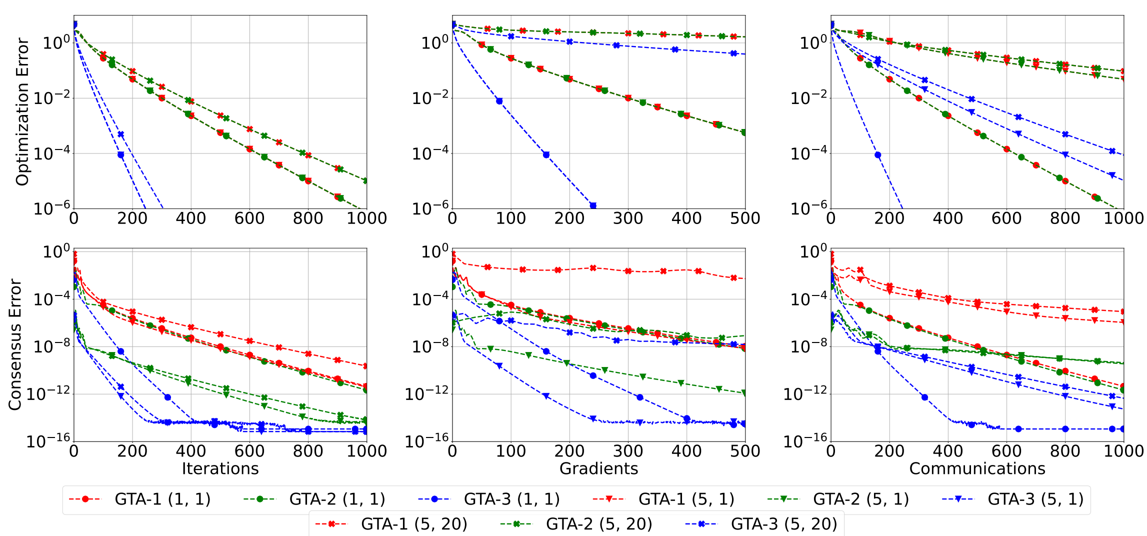

Figure 11: Optimization Error ($\|\bar{x}_k - x^*\|_2$) and Consensus Error ($\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|_2$) of `GTA-1`, `GTA-2` and `GTA-3` with respect to number of iterations, communications and gradient evaluations for binary logistic regression on Mushroom dataset over the Graph 1 network.

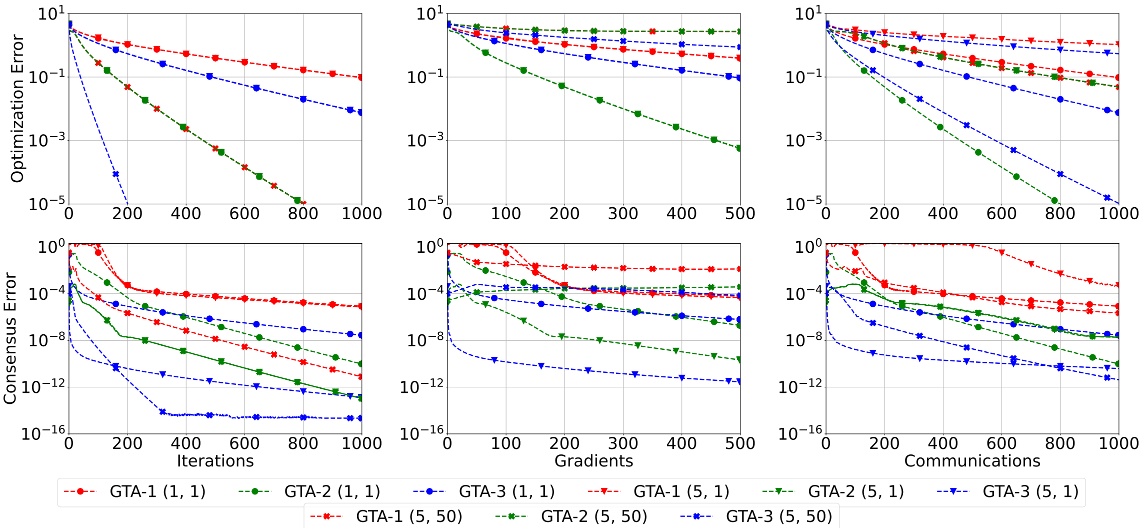

Figure 12: Optimization Error ($\|\bar{x}_k - x^*\|_2$) and Consensus Error ($\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|_2$) of `GTA-1`, `GTA-2` and `GTA-3` with respect to number of iterations, communications and gradient evaluations for binary logistic regression on Mushroom dataset over the Graph 2 network.

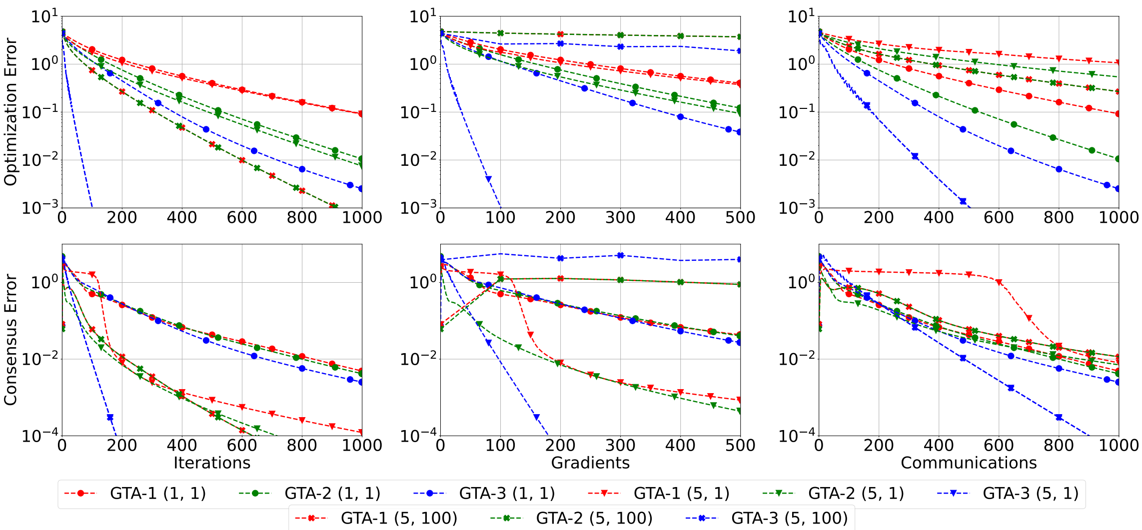

Figure 13: Optimization Error ($\|\bar{x}_k - x^*\|_2$) and Consensus Error ($\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|_2$) of `GTA-1`, `GTA-2` and `GTA-3` with respect to number of iterations, communications and gradient evaluations ffor binary logistic regression on Mushroom dataset over the star network.

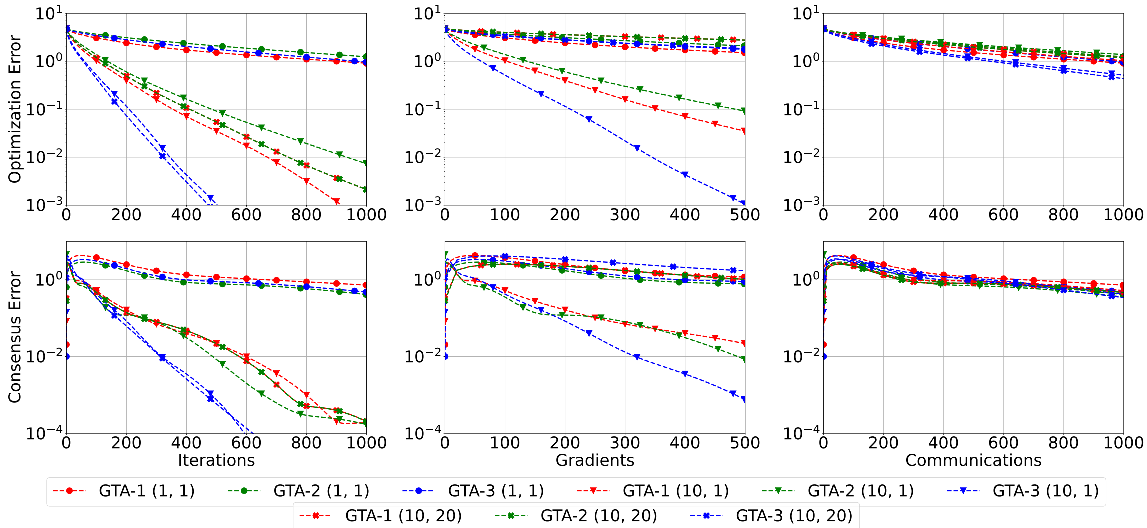

Figure 14: Optimization Error ($\|\bar{x}_k - x^*\|_2$) and Consensus Error ($\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|_2$) of `GTA-1`, `GTA-2` and `GTA-3` with respect to number of iterations, communications and gradient evaluations ffor binary logistic regression on Mushroom dataset over the cyclic network.

2.3 Phishing Dataset

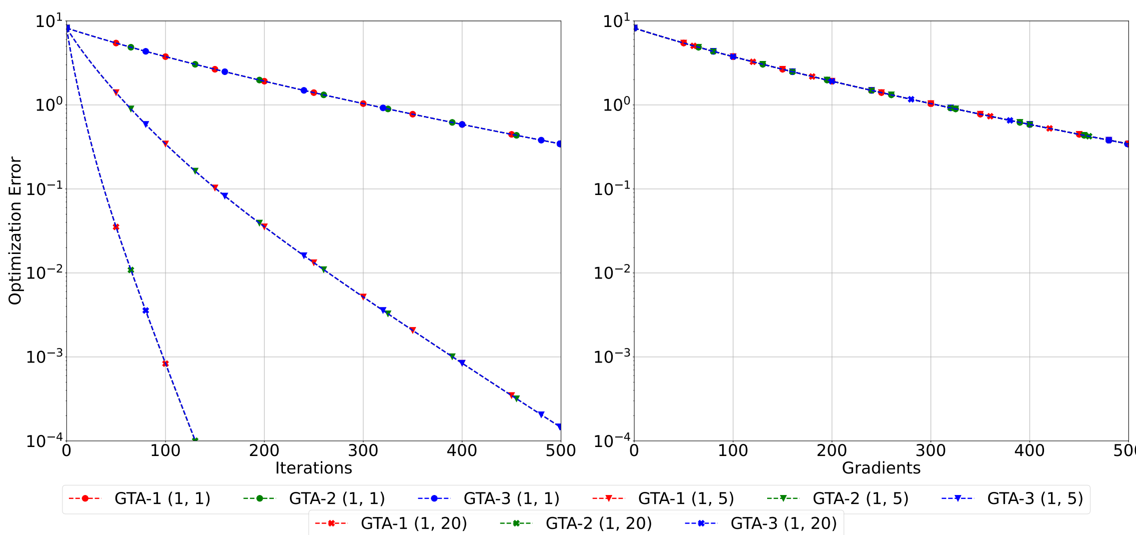

Figure 15: Optimization Error ($\|\bar{x}_k - x^*\|_2$) of `GTA-1`, `GTA-2` and `GTA-3` with respect to number of iterations, communications and gradient evaluations for binary logistic regression on Phishing dataset over the Full network.

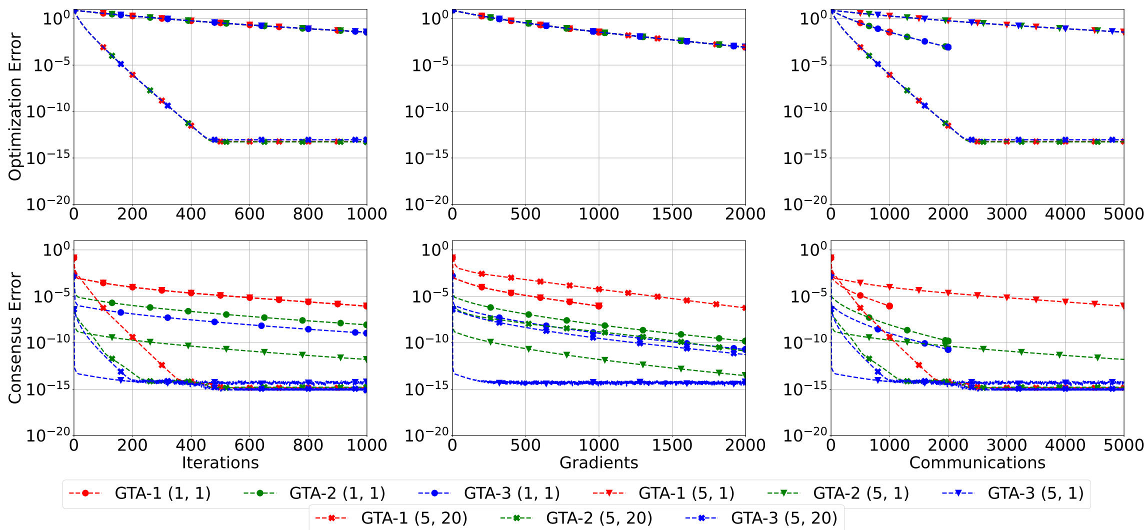

Figure 16: Optimization Error ($\|\bar{x}_k - x^*\|_2$) and Consensus Error ($\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|_2$) of `GTA-1`, `GTA-2` and `GTA-3` with respect to number of iterations, communications and gradient evaluations for binary logistic regression on Phishing dataset over the Graph 1 network.

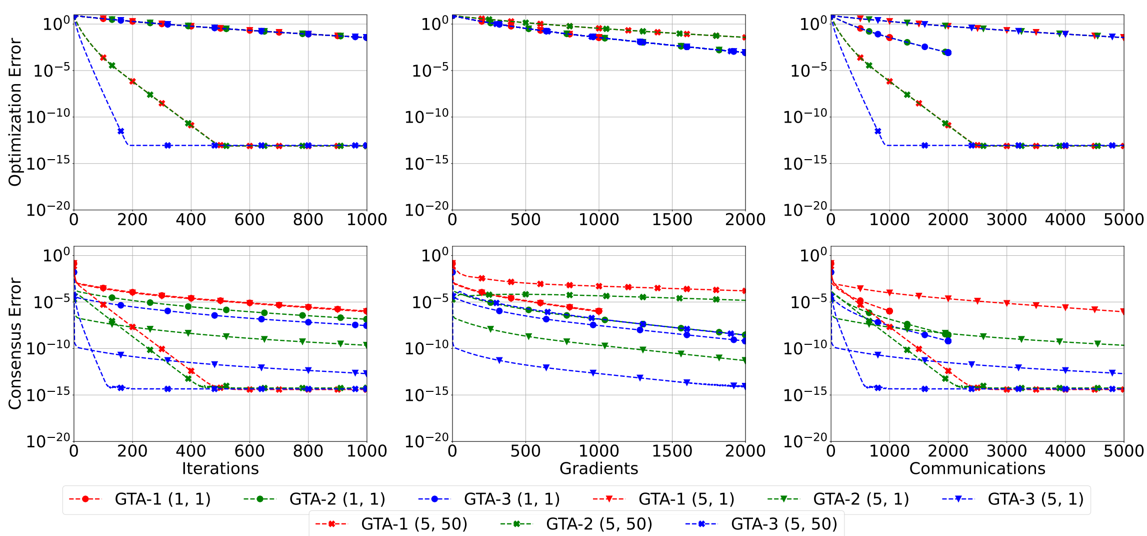

Figure 17: Optimization Error ($\|\bar{x}_k - x^*\|_2$) and Consensus Error ($\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|_2$) of `GTA-1`, `GTA-2` and `GTA-3` with respect to number of iterations, communications and gradient evaluations for binary logistic regression on Phishing dataset over the Graph 2 network.

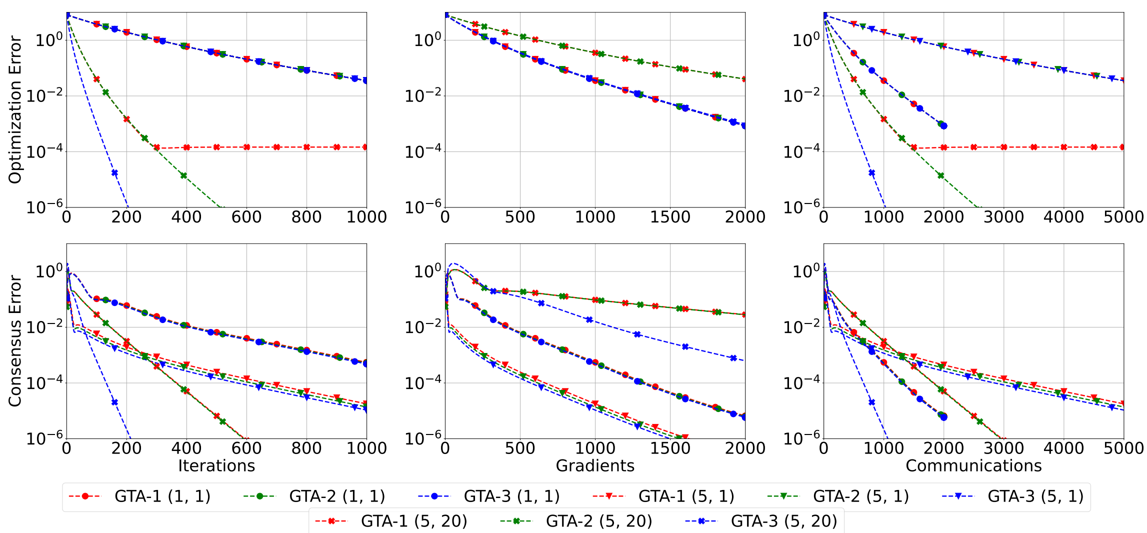

Figure 18: Optimization Error ($\|\bar{x}_k - x^*\|_2$) and Consensus Error ($\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|_2$) of `GTA-1`, `GTA-2` and `GTA-3` with respect to number of iterations, communications and gradient evaluations for binary logistic regression on Phishing dataset over the star network.

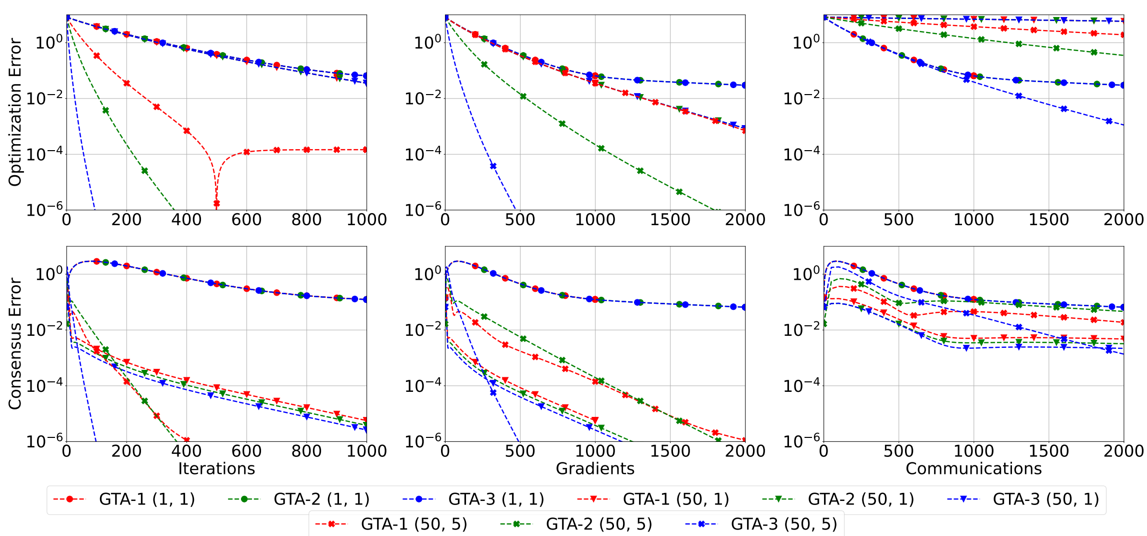

Figure 19: Optimization Error ($\|\bar{x}_k - x^*\|_2$) and Consensus Error ($\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|_2$) of `GTA-1`, `GTA-2` and `GTA-3` with respect to number of iterations, communications and gradient evaluations for binary logistic regression on Phishing dataset over the cyclic network.

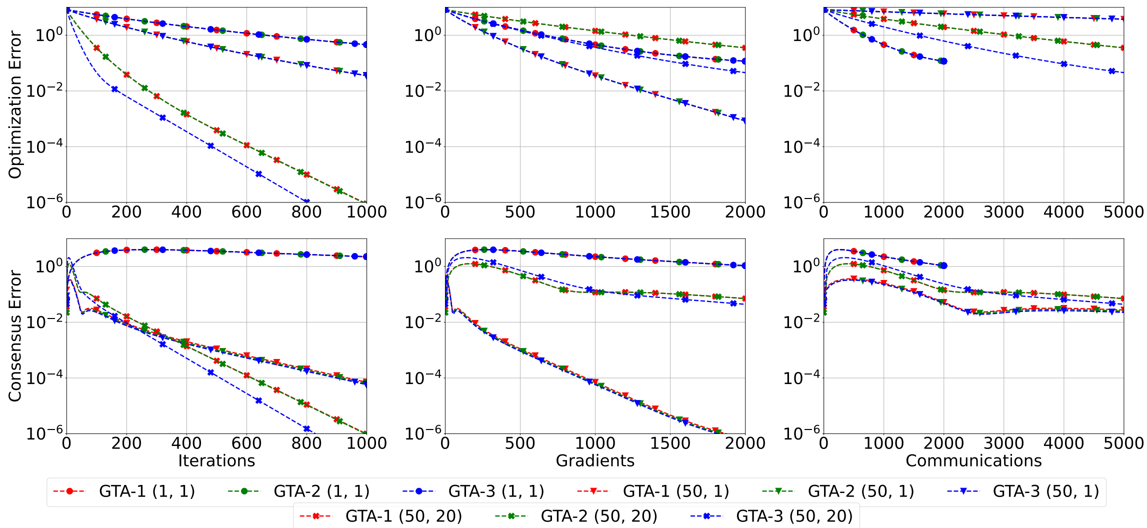

Figure 20: Optimization Error ($\|\bar{x}_k - x^*\|_2$) and Consensus Error ($\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|_2$) of `GTA-1`, `GTA-2` and `GTA-3` with respect to number of iterations, communications and gradient evaluations for binary logistic regression on Phishing dataset over the line network.

2.4 Sonar Dataset

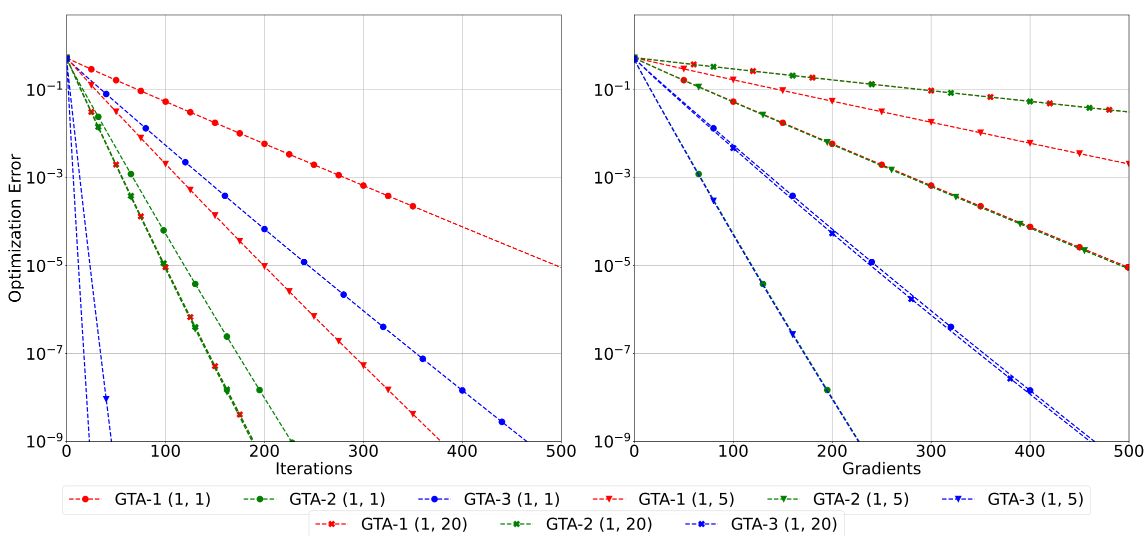

Figure 21: Optimization Error ($\|\bar{x}_k - x^*\|_2$) of `GTA-1`, `GTA-2` and `GTA-3` with respect to number of iterations, communications and gradient evaluations for binary logistic regression on Sonar dataset over the Full network.

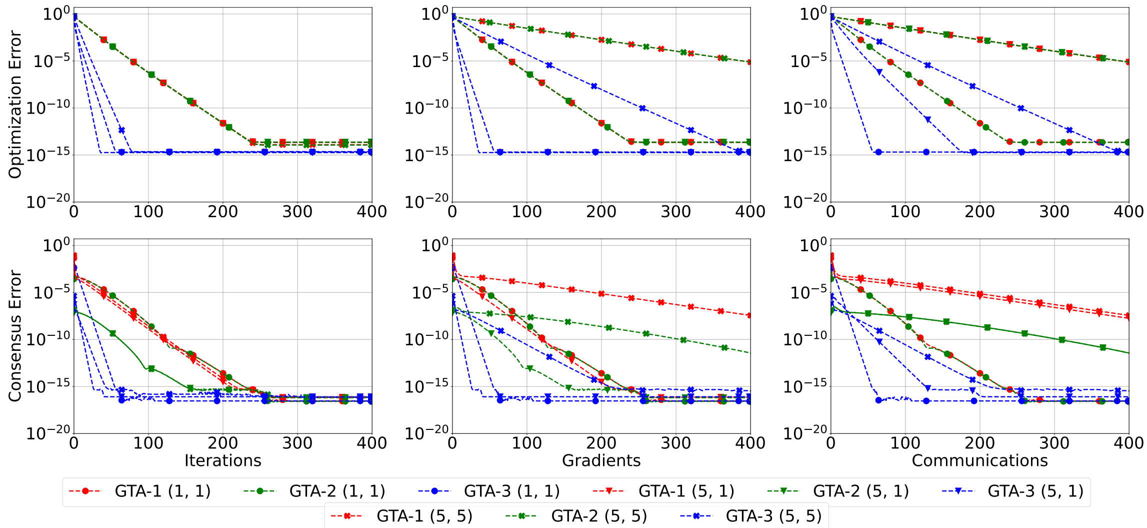

Figure 22: Optimization Error ($\|\bar{x}_k - x^*\|_2$) and Consensus Error ($\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|_2$) of `GTA-1`, `GTA-2` and `GTA-3` with respect to number of iterations, communications and gradient evaluations for binary logistic regression on Sonar dataset over the Graph 1 network.

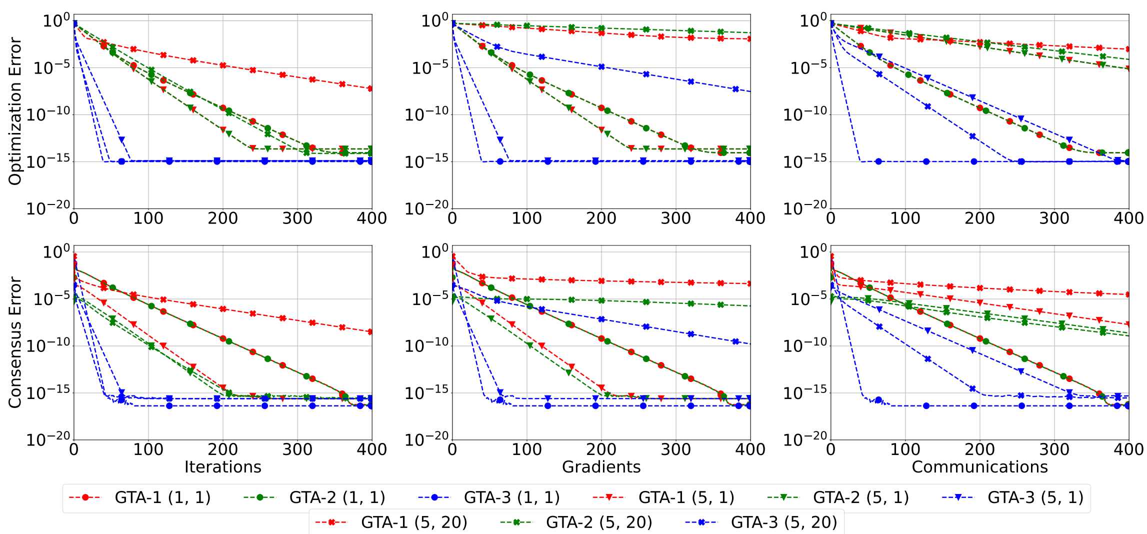

Figure 23: Optimization Error ($\|\bar{x}_k - x^*\|_2$) and Consensus Error ($\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|_2$) of `GTA-1`, `GTA-2` and `GTA-3` with respect to number of iterations, communications and gradient evaluations for binary logistic regression on Sonar dataset over the Graph 2 network.

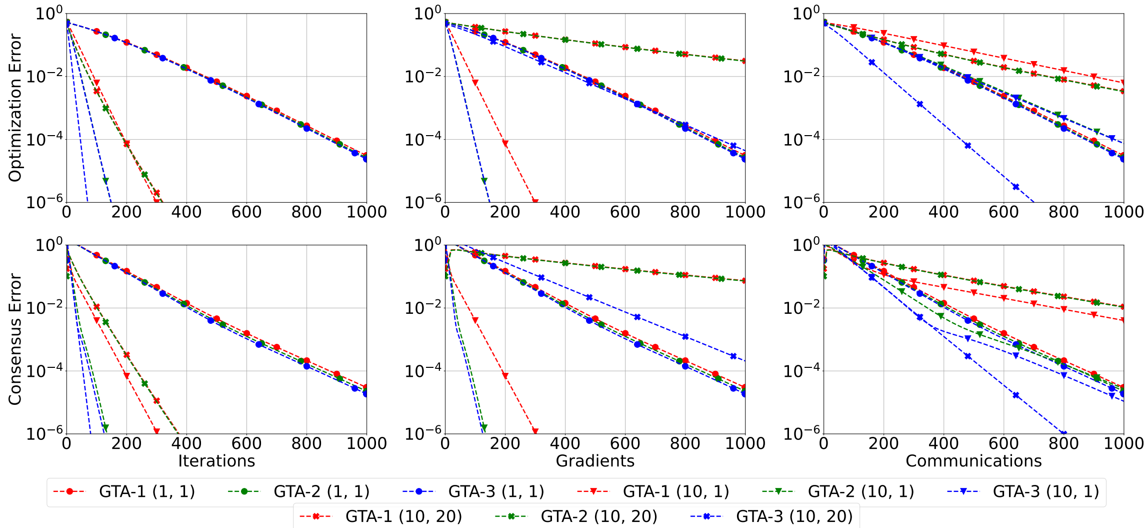

Figure 24: Optimization Error ($\|\bar{x}_k - x^*\|_2$) and Consensus Error ($\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|_2$) of `GTA-1`, `GTA-2` and `GTA-3` with respect to number of iterations, communications and gradient evaluations for binary logistic regression on Sonar dataset over the star network.

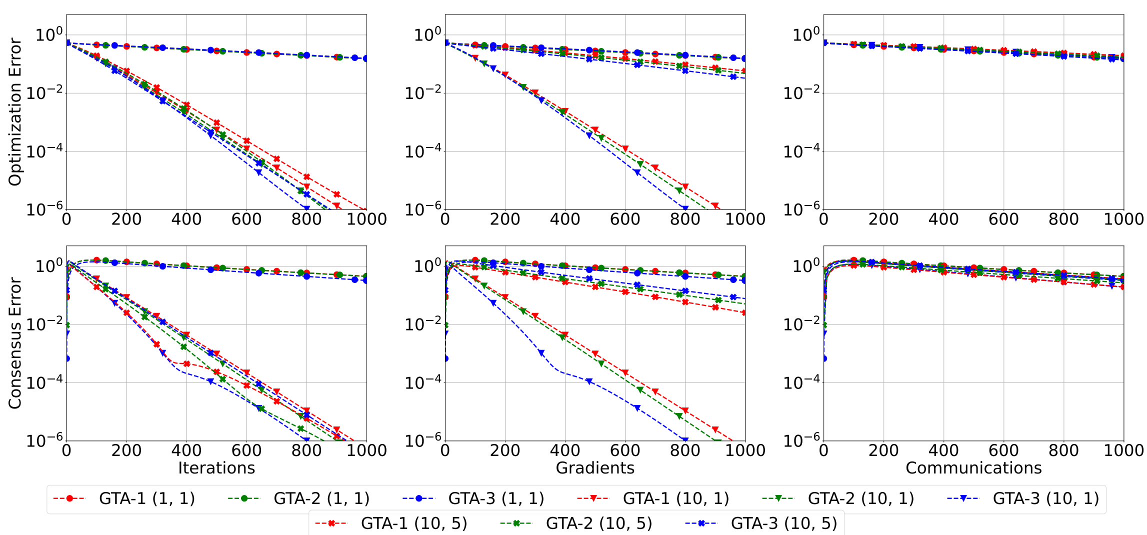

Figure 25: Optimization Error ($\|\bar{x}_k - x^*\|_2$) and Consensus Error ($\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|_2$) of `GTA-1`, `GTA-2` and `GTA-3` with respect to number of iterations, communications and gradient evaluations for binary logistic regression on Sonar dataset over the cyclic network.

2.5 Ionosphere Dataset

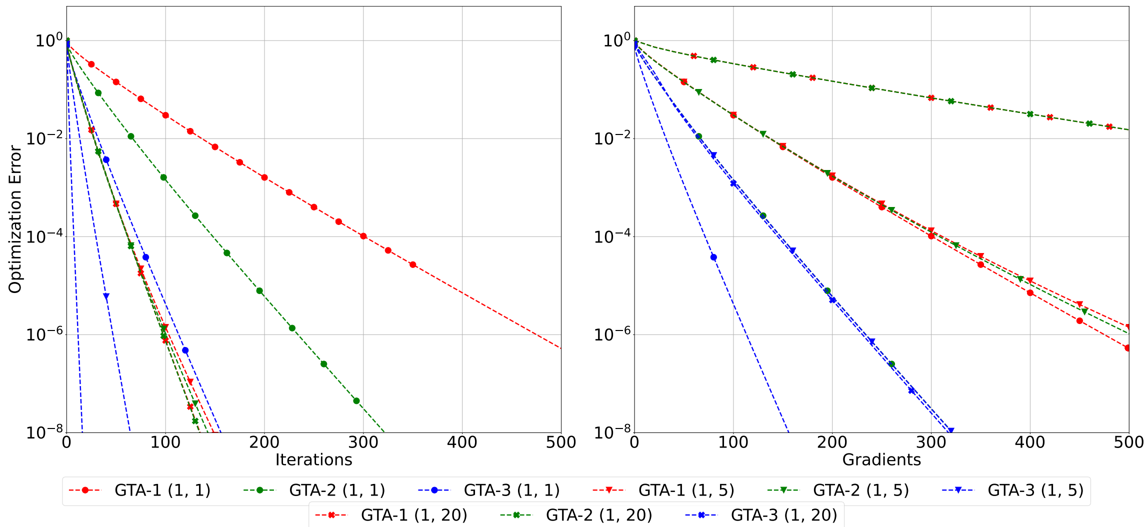

Figure 26: Optimization Error ($\|\bar{x}_k - x^*\|_2$) of `GTA-1`, `GTA-2` and `GTA-3` with respect to number of iterations, communications and gradient evaluations for binary logistic regression on Ionosphere dataset over the Full network.

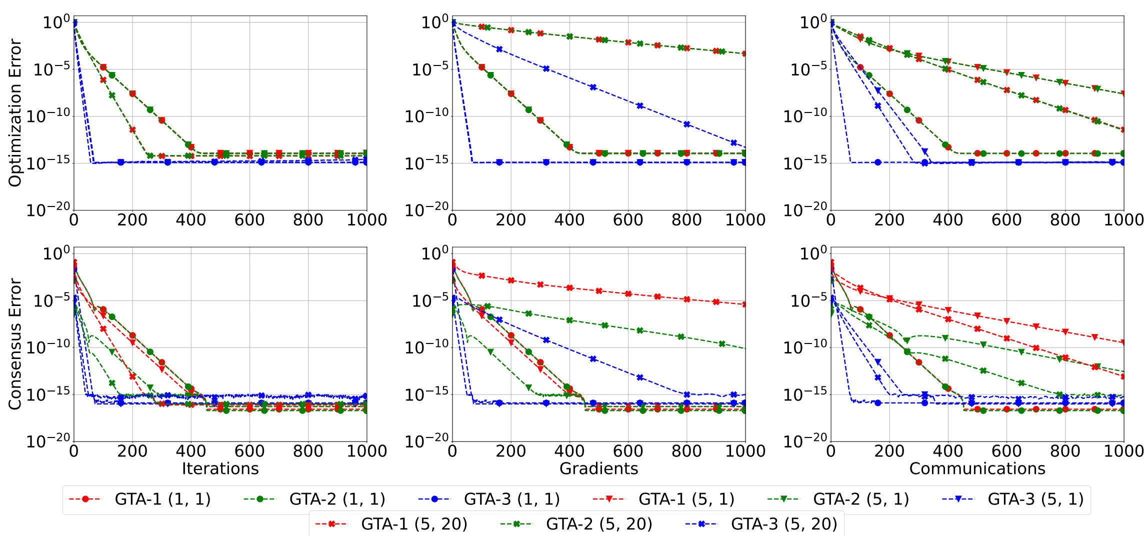

Figure 27: Optimization Error ($\|\bar{x}_k - x^*\|_2$) and Consensus Error ($\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|_2$) of `GTA-1`, `GTA-2` and `GTA-3` with respect to number of iterations, communications and gradient evaluations for binary logistic regression on Ionosphere dataset over the Graph 1 network.

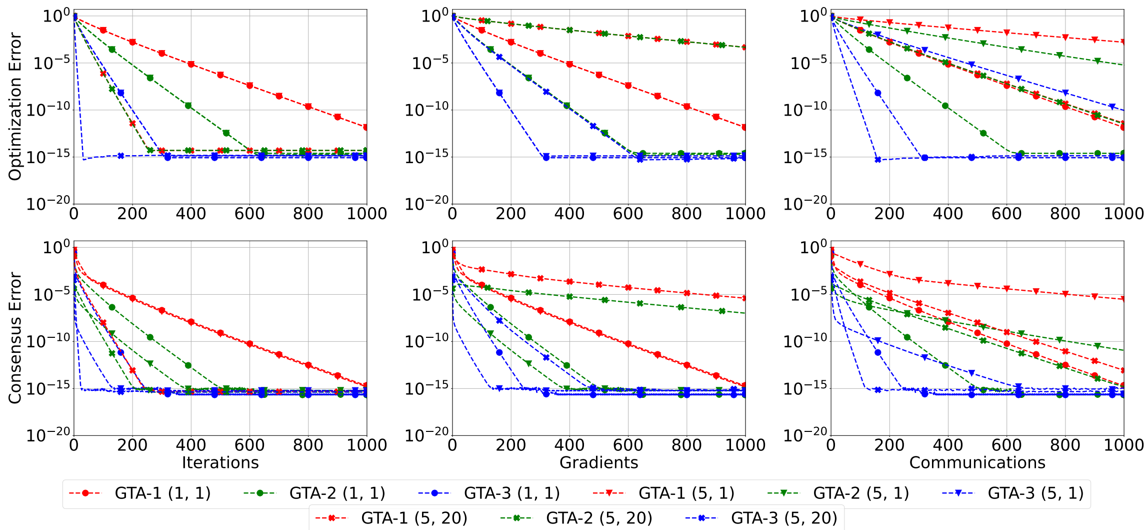

Figure 28: Optimization Error ($\|\bar{x}_k - x^*\|_2$) and Consensus Error ($\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|_2$) of `GTA-1`, `GTA-2` and `GTA-3` with respect to number of iterations, communications and gradient evaluations for binary logistic regression on Ionosphere dataset over the Graph 2 network.

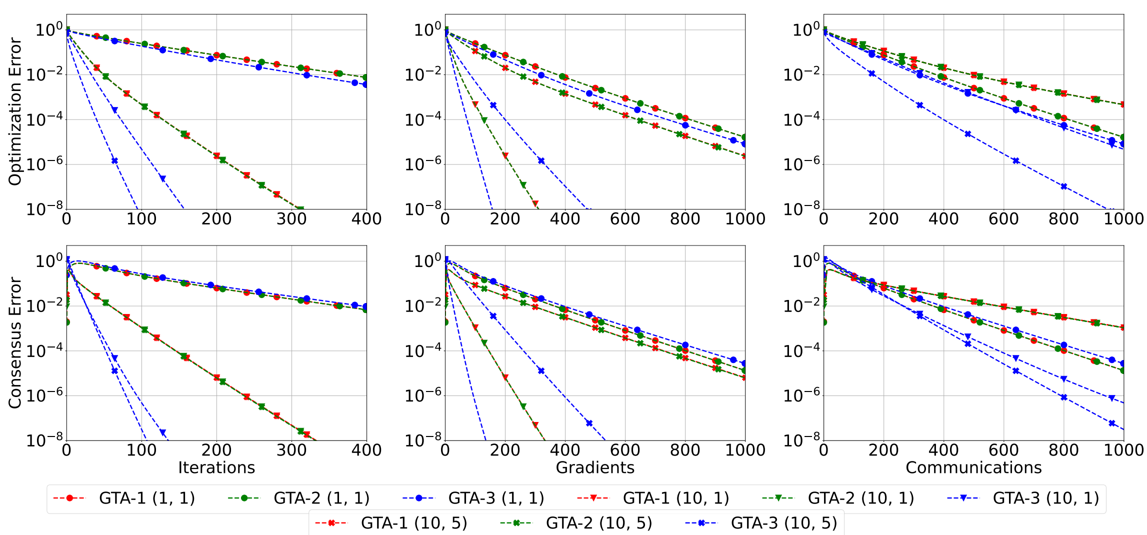

Figure 29: Optimization Error ($\|\bar{x}_k - x^*\|_2$) and Consensus Error ($\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|_2$) of `GTA-1`, `GTA-2` and `GTA-3` with respect to number of iterations, communications and gradient evaluations for binary logistic regression on Ionosphere dataset over the star network.

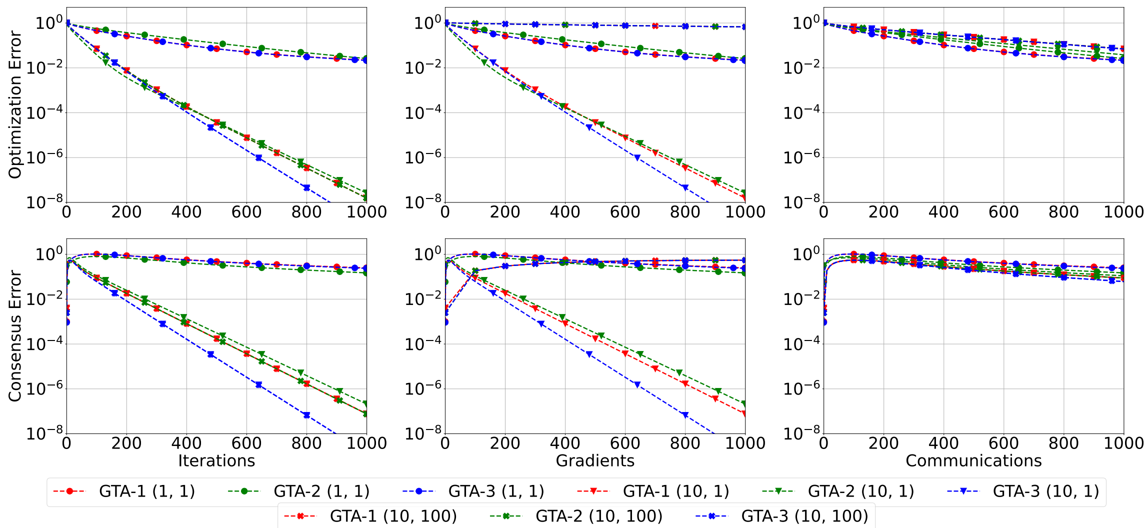

Figure 30: Optimization Error ($\|\bar{x}_k - x^*\|_2$) and Consensus Error ($\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|_2$) of `GTA-1`, `GTA-2` and `GTA-3` with respect to number of iterations, communications and gradient evaluations for binary logistic regression on Ionosphere dataset over the cyclic network.

References

[1] Chih-Chung Chang and Chih-Jen Lin. Libsvm: a library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)*, 2(3):1–27, 2011.

[2] Paolo Di Lorenzo and Gesualdo Scutari. Next: In-network nonconvex optimization. *IEEE Transactions on Signal and Information Processing over Networks*, 2(2):120–136, 2016.

[3] Dheeru Dua and Casey Graff. UCI machine learning repository, 2017.

[4] Aryan Mokhtari, Qing Ling, and Alejandro Ribeiro. Network newton distributed optimization methods. *IEEE Transactions on Signal Processing*, 65(1):146–161, 2016.

[5] Angelia Nedic, Alex Olshevsky, and Wei Shi. Achieving geometric convergence for distributed optimization over time-varying graphs. *SIAM Journal on Optimization*, 27(4):2597–2633, 2017.

[6] Angelia Nedić, Alex Olshevsky, Wei Shi, and César A Uribe. Geometrically convergent distributed optimization with uncoordinated step-sizes. In *2017 American Control Conference (ACC)*, pages 3950–3955. IEEE, 2017.

[7] Shi Pu, Wei Shi, Jinming Xu, and Angelia Nedić. Push–pull gradient methods for distributed optimization in networks. *IEEE Transactions on Automatic Control*, 66(1):1–16, 2020.

[8] Wei Shi, Qing Ling, Gang Wu, and Wotao Yin. Extra: An exact first-order algorithm for decentralized consensus optimization. *SIAM Journal on Optimization*, 25(2):944–966, 2015.

[9] Ying Sun, Gesualdo Scutari, and Amir Daneshmand. Distributed optimization based on gradient tracking revisited: Enhancing convergence rate via surrogation. *SIAM Journal on Optimization*, 32(2):354–385, 2022.

[10] Jinming Xu, Shanying Zhu, Yeng Chai Soh, and Lihua Xie. Augmented distributed gradient methods for multi-agent optimization under uncoordinated constant stepsizes. In *2015 54th IEEE Conference on Decision and Control (CDC)*, pages 2055–2060. IEEE, 2015.