Evaluating and Incentivizing Diverse Data Contributions in Collaborative Learning

Anonymous Authors[1]

Abstract

For a federated learning model to perform well, it is crucial to have a diverse and representative dataset. However, the data contributors may only be concerned with the performance on a specific subset of the population, which may not reflect the diversity of the wider population. This creates a tension between the principal (the FL platform designer) who cares about global performance and the agents (the data collectors) who care about local performance. In this work, we formulate this tension as a game between the principal and multiple agents, and focus on the linear experiment design problem to formally study their interaction. We show that the statistical criterion used to quantify the diversity of the data, as well as the choice of the federated learning algorithm used, has a significant effect on the resulting equilibrium. We leverage this to design simple optimal federated learning mechanisms that encourage data collectors to contribute data representative of the global population, thereby maximizing global performance.

1. Introduction

Collaborative learning can be viewed as a transactional process where participants collectively receive a reduction in uncertainty in return for sharing their data (Karimireddy et al., 2022). However, participants may be concerned with uncertainty in different sub-populations. Thus a reduction in uncertainty on the global population may not necessarily translate to an improvement for every participant.

Consider a collaborative learning project between multiple countries to study rare cancers (Moncada-Torres et al., 2020; Geleijnse et al., 2020). Different countries operate cancer registries with the goal of collecting comprehensive data on rare cancer cases within their jurisdictions. These registries collaborate to pool their data and resources. However, each registry has the responsibility to prioritize the benefit to their own population while minimizing the risks associated with data collection and sharing. Thus, the global performance needs to be balanced with the specific needs and goals of each registry.

The need to balance local and global interests becomes even more critical when collecting data from marginalized communities. Issues of equity and autonomy underpin indigenous critiques of genetic research and the sharing of genomic data (Hudson et al., 2020; Chediak, 2020). Such communities have historically faced exploitation and mis/under-representation in research studies (Graham, 2015; Albain et al., 2009; Nana-Sinkam et al., 2021). Therefore, it is essential to carefully consider the costs incurred by and benefits provided to them individually.

We formalize this as a game between a principal (the platform designer), and multiple agents (participants) whose needs and agency should be respected–see Fig. 1. Together, they wish to determine a statistical model between responses and variables. Each agent has access to a set of experiment conditions relevant to specific demographic groups within their population. They autonomously decide how many (as well as which) samples to collect and share. The platform then employs federated learning to train a model on the collective data, which is then shared back to the agents. Notably, each agent wishes to minimize the data costs incurred while maximizing uncertainity reduction.

This can be seen as a "multi-agent" version of the classic *optimal experiment design* problem (Wald, 1943; Kiefer, 1958; Kiefer & Wolfowitz, 1960; Karlin & Studden, 1966; Atwood, 1969; Fedorov & Malyutov, 1972), where the final allocation of samples among experiment conditions results from decisions made by multiple agents. Unlike classical theories, we introduce game-theoretic subtleties since each agent is primarily concerned with the validity of the model for its specific demographic group. Hence, we must account for the strategic behaviors that emerge due to both data diversity and cost heterogeneity. In this context, two fundamental questions arise.

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

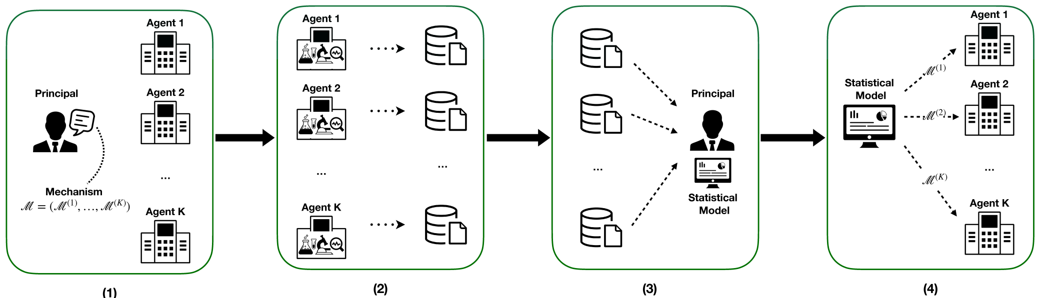

Figure 1. Principal-agents experiment design. (1) The principal publishes a mechanism incentivizing participation. (2) Each agent strategically selects how many and which of their available experiment conditions to collect and share. (3) The agents engage in collaborative learning with the principal, who utilizes the collected data points to train a statistical model. (4) The principal applies the mechanism to the trained model for each agent and subsequently distributes the models to them.

First is efficiency–it is crucial to allocate resources such as time, money, and materials in the most efficient and informative manner. This was the main concern of classic theory of optimal experiment design which proposed different optimality (efficiency) criterion. In our version, we ask:

When is it in the agents' best interest to follow the globally efficient optimal experiment design?

Second is maximizing the amount of information collected. Conventional data sharing mechanisms like federated learning face the critical issue of free-riding (Baumol, 2004; Choi & Robertson, 2019; Sarikaya & Ercetin, 2019; Lin et al., 2019; Ding et al., 2020; Sim et al., 2020; Xu et al., 2021), where self-interested agents may contribute minimal or no data but still benefit from an improved model. The principal may instead want to maximize the information generated by data contributions from all agents (without regards to efficiency), raising the following challenge:

Can the principal design mechanisms to incentivize strategic agents to contribute their fair share of data, thereby maximizing the information produced?

The quality and diversity of data play a vital role in this context. In this study, we address the aforementioned challenges by specifically focusing on linear experiment design (Pukelsheim, 2006; Silvey, 2013), where diversity is characterized by the Fisher information matrix and quality is assessed using optimality criteria.

2. Principal-agents experiment design

This section introduces the problem of principal-agents experiment design which studies the game-theoretic notions and properties of experiment designs in a principal-agents framework (Laffont & Martimort, 2009). Our framework is summarized in Algorithm 1 with detailed discussion below.

We model the interaction between multiple agents and a coordinating principal in linear experiment design. Consider K self-interested agents. Each agent k has a local design space $\mathcal{X}_k = \{x_i\}_{i \in G_k} \subset \mathbb{R}^d$ and the global design space is then $\mathcal{X} = \cup_{k \in [K]} \mathcal{X}_k = \{x_1, \ldots, x_n\}$. For ease of presentation, assume that the indices are sorted such that $G_1, \ldots, G_K$ form consecutive partitions of $[n]$. Then, $w = (w_{G_1}, \ldots, w_{G_K})$ is the global design measure with each agent k controlling w_{G_k}. The data contribution of agent k can thus be summarized by w_{G_k}.

Mechanism definition. The principal is given access to the entire data contributions i.e. the global design measure w. Then the principal sets up a *mechanism* to assign a subset of this contribution to each agent k. Thus, we can define a mechanism $\mathcal{M}$ as follows:

$$\mathcal{M} := (\mathcal{M}^{(k)} : \mathbb{R}^n_+ \to \mathbb{R}^n_+)_{k \in [K]} \ \text{ s.t. } \ \mathcal{M}^{(k)}(w) \leq w\,.$$

The inequality $\mathcal{M}^{(k)}(w) \leq w$ applies element-wise for all $i \in [n]$. Thus, a mechanism represents a re-allocation of the design measure (and hence data) to each of the clients. In Appendix B, we discuss implementation of mechanisms.

Agent utility. Each agent wants to minimize uncertainty (on their design space) while also minimizing the costs of data collection and sharing. While Eq. (9) defined the information matrix for the global design space $\mathcal{X}$, agent k only cares about $\mathcal{X}_k$, which is of rank r_k (that may be less than d). We thus need to consider a *local information matrix* (Sibson, 1974; Silvey, 1978) representing uncertainty along direction only in $\mathcal{X}_k$:

$$\mathbf{M}^{(k)}(w) := {A^{(k)}}^\top \mathbf{M}(w) A^{(k)}\,. \tag{1}$$

Here, $A^{(k)} \in \mathbb{R}^{n \times r_k}$ such that $A^{(k)} {A^{(k)}}^\top$ is a projection matrix onto $\operatorname{span}(\mathcal{X}_k)$. The value of a design strategy w for agent k on $\mathcal{X}_k$ using criterion $f^{(k)}(\cdot)$ can then be written as $f^{(k)}\left(\mathbf{M}^{(k)}(w)^{-1}\right)$. Here, $f^{(k)}$ is an optimality criterion (Pukelsheim, 2006; Silvey, 2013) that depends on k since it may implicitly depend on the design space $\mathcal{X}_k$. In particular, the G-criterion takes a max over the design space $\mathcal{X}_k$, and the V-criterion takes an average.

Next, note that agent k only has control over w_i for $i \in G_k$ i.e. it can only decide the sampling strategy over $\mathcal{X}_k$. For convenience, define $w_{G_k} = (w_i)_{i \in G_k}$. Suppose that the cost of collecting one data point is $c^{(k)}$. This can represent both the actual cost incurred in collecting and storing the data as well as potential privacy risks associated with storing and sharing it. Then, the total cost incurred by agent k is $c^{(k)} \sum_{i \in G_k} w_i$. Putting all of this together, the utility enjoyed by client k under mechanism $\mathcal{M}$ can be written as

$$\left(u^{(k)} \circ \mathcal{M}^{(k)}\right)(w) := f^{(k)}\left(\left(\mathbf{M}^{(k)}(\mathcal{M}^{(k)}(w))\right)^{-1}\right) - c^{(k)} \sum_{i \in G_k} w_i . \quad (2)$$

Here, $u^{(k)} \circ \mathcal{M}^{(k)}$ represents the composition of the agent's utility function $u^{(k)}$, which depends solely on the agent's personal valuation, and the mechanism $\mathcal{M}^{(k)}$ implemented by the principal.

Finally, note that the cost incurred by agent k is completely independent of $w_{G_k^c} := (w_j)_{j \notin G_k}$. However, the local information matrix $\mathbf{M}^{(k)}(w)$ depends on the whole w and $\mathcal{X}$. In particular, if the complementary design space $x_j \in \mathcal{X} \setminus \mathcal{X}_k$ is similar to $\mathcal{X}_k$ and $w_j > 0$, then the local information matrix $\mathbf{M}^{(k)}(w)$ as well as $f_k\left(\mathbf{M}^{(k)}(w)^{-1}\right)$ will be larger. Thus, $w_{G_k^c}$ represents free outside information given to agent k, and importantly it affects the optimal choice of w_{G_k}.

Remark 2.1 (Accuracy shaping)**.** *The mechanism Eq. 1 can be understood as shaping the accuracy (Karimireddy et al., 2022) of the model that is sent to each agent. For example, the standard federate learning mechanism $\mathcal{M}_{\text{fed}}^{(k)}$ would distribute the global model to all the agents. This corresponds to setting $\mathcal{M}_{\text{fed}}^{(k)}(w) = w$ for all agents. To incentivize agents to contribute high-quality data, the mechanism may adjust the accuracy of the model depending on the data quality generated by each agent.*

Strategic behavior of agents. In principal-agents experiment design problems, theoretically interesting and practically relevant scenario arises for strategic agents, who make rational decisions on the design w_{G_k} depending on the actions of other agents, knowing the design spaces $\mathcal{X}_j$ and the costs $c^{(j)}$ for all $j \in [K]$. We characterize the behaviors of strategic agents through the following definition.

Definition 2.2 (Strategic responses)**.** We say the designs $w^* = (w^*_{G_1}, w^*_{G_2}, \ldots, w^*_{G_K})$ is a strategic response to the mechanism $\mathcal{M} = (\mathcal{M}^{(1)}, \mathcal{M}^{(2)}, \ldots, \mathcal{M}^{(k)})$ if:

- Individual Rationality: For any $k \in [k]$, if $\sum_{i \in G_k} w^*_i > 0$ then
$$\left(u^{(k)} \circ \mathcal{M}^{(k)}\right)(w^*) \geq v_*^{(k)} \quad (3)$$
where $v_*^{(k)}$ is the maximum possible utility agent k can achieve if she opts out of the collaborated learning and trains a model using her own data, i.e.
$$v_*^{(k)} := \max_{w_{G_k}} f^{(k)}\left(\left(\sum_{i \in G_k} w_i \cdot (A^{(k)})^\top x_i x_i^\top A^{(k)}\right)^{-1}\right) - c^{(k)} \sum_{i \in G_k} w_i. \quad (4)$$
- Pure Nash Equilibrium: $(w^*_{G_1}, w^*_{G_2}, \ldots, w^*_{G_K})$ is the pure Nash equilibrium of the game defined by concave utilities $(u^{(k)} \circ \mathcal{M}^{(k)})_{k \in [K]}$ and actions $(w_{G_k})_{k \in [K]}$, i.e. it satisfies $\forall\, k \in [K], \forall\, w_{G_k} \in \mathbb{R}_+^{|G_k|}$
$$\left(u^{(k)} \circ \mathcal{M}^{(k)}\right)(w^*) \geq \left(u^{(k)} \circ \mathcal{M}^{(k)}\right)(w_{G_k}, w^*_{G_k^c}).$$
Here, $(w_{G_k}, w^*_{G_k^c})$ denotes concatenation i.e. $(w^*_{G_1}, \ldots, w^*_{G_{k-1}}, w_{G_k}, w^*_{G_{k+1}}, \ldots, w^*_{G_K})$.

The first condition indicates that an agent will never choose an action that results in a worse outcome than v^*, the status quo that agent k can obtain no matter she takes part in the collaboration learning or not. The second condition asserts that an agent can not obtain higher utility by unilaterally changing her action. Thus, the strategic response w^* represents a stable fixed point to the game from which no agent has an incentive to deviate from their chosen action.

3. Efficiency under federated learning

The first question of interest about principal-agents experiment design is the efficiency of the mechanism. By classic optimal experiment design, a design measure $w = (w_{G_1}, \ldots, w_{G_K})$ is efficient for optimality criterion f if w is proportional to the optimal design measure $\pi^* = \arg\max f(\mathbf{M}(\pi)^{-1}),\ \text{s.t.}\ \pi \in \Delta(\mathcal{X})$.

In this section, we explore the conditions under which the standard federated learning mechanism which always sets $\mathcal{M}_{\text{fed}}^{k}(w) = w$ is efficient. We establish that the D-criterion is the only criterion among common criteria for which the federated learning mechanism is efficient.

Definition 3.1 (Incentive-compatibly efficient)**.** A mechanism $\mathcal{M}$ is *incentive-compatibly efficient* for a criterion f, if for any choice of design spaces $(\mathcal{X}_k)_{k \in [K]}$, all strategic responses w^* are efficient designs for criterion f that satisfy $w^* \propto \pi^*$.

Proposition 3.2. *Suppose $c^{(1)} = \cdots = c^{(K)} = c \in \mathbb{R}_+$. Then, among all optimality criteria, the federated learning mechanism ($\mathcal{M}_{\text{fed}}^{(k)}(w) = w$) is efficient only for D-criterion. More precisely,*

1. *When all agents k use criterion $f_D^{(k)}$, the agent's strategic response is the design given by $(\frac{d}{c} \cdot \pi^*_{G_k})_{k \in [K]}$, where $\pi^* \in \arg\max_{\pi \in \Delta(\mathcal{X})} f_D(\mathbf{M}(\pi)^{-1})$.*
2. *For every other standard criteria (E, A, V, or G), there exists a design space $\mathcal{X}$ such that federated learning mechanism is not efficient.*

When each agent incurs the same marginal cost for sampling data, differences in efficiency can be attributed to the agents' data generation capacities rather than variations in data acquisition costs. This setup allows for a fair comparison among agents and serves as the natural framework for studying efficiency. Our result implies that D-criterion is the only criterion that aligns the interest of each agent with the statistical efficiency of the multi-agent system. Therefore, it is the most suitable for experiment design problems involving multiple agents.

Remark 3.3 (Efficiency of D-optimality)**.** *That D-optimality uniquely satisfies incentive-compatible efficiency is remarkable. Numerous reviews and textbooks compare and contrast the different criteria but fail to identify a single-best one (Chaloner & Verdinelli, 1995; Fedorov & Hackl, 1997; Pukelsheim, 2006; Atkinson et al., 2007; Goos & Jones, 2011). In fact, the popularity of D-optimality stemmed from its perceived equivalence to G-optimality, while being easier to optimization. The multi-agent perspective provides a novel lens with which to distinguish them and recommend D-criterion over the rest. However, a note of caution is warranted–these results hold with our specific linear cost model. With different cost functions, it is possible that the conclusions differ.*

The above results leaves the question of efficiency under heterogeneous costs. The standard federated learning does not suffice anymore, and we instead require non-trivial mechanisms. We deferred this discussion to Appendix G.

4. Information maximization

This section addresses the second question posed in the introduction. Following Proposition 3.2, we assume that every agent k uses the D-criterion $f_D^{(k)}$. It is worth noting that this choice is compatible with our information-theoretic considerations as maximizing the D-criterion is equivalent to minimizing the differential entropy of $\hat{\theta}$.

To achieve information maximization, we need to first understand what is the maximum information that could be possibly generated by strategic agents. We have the following result on the maximum achievable information.

Proposition 4.1. *For any mechanism $\mathcal{M}$ and any strategic response $\widetilde{w}$ under $\mathcal{M}$, we have $\log\det \mathbf{M}(\widetilde{w}) \leq \log\det \mathbf{M}(w_{\max})$, where we define $w_{\max}$ as*

$$\arg\max_{w\in\mathbb{R}_+^n} \log\det \mathbf{M}(w), \ \textit{s.t.}\ u^{(k)}(w) \geq v_*^{(k)}. \tag{5}$$

The above proposition states that the maximum achievable information is attained when the data collection is allocated according to $w_{\max}$. However, abundant evidence in Appendix E show that free-riding (Baumol, 2004; Choi & Robertson, 2019; Sarikaya & Ercetin, 2019; Lin et al., 2019; Ding et al., 2020; Sim et al., 2020; Xu et al., 2021) will occur under the federated learning mechanism.

Motivated by these results, we design mechanisms $(\mathcal{M}_{\max}^{(k)})_{k\in[K]}$ to incentive agents to contribute $w_{\max}$ amount of data. Let $\mathcal{M}_{\max}^{(k)}$ simply scale the design by a constant $\gamma_k \leq 1$

$$\mathcal{M}_{\max}^{(k)}(w) := \gamma_k w\,, \tag{6}$$

$$\text{for } \gamma_k^{-1} := \exp\left(\frac{c^{(k)}}{r_k} \cdot \textstyle\sum_{i\in G_k} (w_{\max,i} - w_i)_+\right).$$

Here, $(x)_+ := \max\{x, 0\}$. In this mechanism, agents are penalized for contributing less data than required for information maximization ($w_{\max}$). The k-th agent's utility $\big(u^{(k)} \circ \mathcal{M}_{\max}^{(k)}\big)(w)$ is then given by

$$-\log\det\big((A^{(k)})^\top \mathbf{M}(w)^\dagger A^{(k)}\big) - c^{(k)} \textstyle\sum_{i\in G_k} w_i$$
$$- c^{(k)} \textstyle\sum_{i\in G_k} (w_{\max,i} - w_i)_+ \,.$$

We have the following proposition that establishes information maximization as the unique strategic response of the information mechanism $\mathcal{M}_{\max}$.

Proposition 4.2 (Information maximization)**.** *The information maximization design $(w_{\max,G_k})_{k\in[K]}$ Eq. (5) is the unique startegic response of the agents to the information mechanism $\mathcal{M}_{\max}$ in Eq. (6).*

Proposition 4.2 establishes that information maximization design, represented by the $(w_{\max,G_k})_{k\in[K]}$, is the unique strategic response to the mechanism $\mathcal{M}_{\max}$. This thus addresses the second question posed in the introduction and offers a tentative resolution for the federated learning community to maximizes data creation from multiple autonomous parties (Graham, 2015; Albain et al., 2009; Hudson et al., 2020; Chediak, 2020; Zhan et al., 2021; Shi et al., 2021), thereby generating positive societal impact. We discuss fairness and price of anarchy (Koutsoupias & Papadimitriou, 1999) of this mechanism in Appendix D.

5. Conclusion

In this paper, we formulated the problem of principal-agents experiment design to capture the game theoretic tensions between the principal and strategic agents in collaborative learning. We showed that under standard federated learning, strategic agents will adopt the optimal design strategy if and only if the D-optimality criterion is used. Additionally, we have highlighted that strategic agents often exhibit free-riding behavior, driven by factors such as data diversity and cost heterogeneity. This observation has motivated us to develop a mechanism that incentivizes strategic agents to maximize the overall information. The proposed mechanism has significant societal implications as it promotes autonomy and equity in clinical trials, collaborative cancer research, etc.

Our results come with some limitations, while opening new avenues for future research. Firstly, our framework does not analyze concrete algorithms with realistic considerations such as unknown design spaces $\mathcal{X}_k$. Overcoming this is an important direction of future work. Furthermore, it would be intriguing to generalize our results to mixed effect models or nonlinear models, which would broaden the scope of our analysis and uncover additional nuances in the principal-agents experiment design problem. Finally, our theoretical analysis uses a somewhat stylized model for the behavior of agents. Translating insights gained in our work to the real world is challenging but necessary.

References

Albain, K. S., Unger, J. M., Crowley, J. J., Coltman, C. A., and Hershman, D. L. Racial disparities in cancer survival among randomized clinical trials patients of the southwest oncology group. *JNCI: Journal of the National Cancer Institute*, 101(14):984–992, 2009.

Atkinson, A., Donev, A., and Tobias, R. *Optimum experimental designs, with SAS*, volume 34. OUP Oxford, 2007.

Atwood, C. L. Optimal and efficient designs of experiments. *The Annals of Mathematical Statistics*, pp. 1570–1602, 1969.

Barak, B. Hopes, fears, and software obfuscation. *Communications of the ACM*, 59(3):88–96, 2016.

Baumol, W. J. *Welfare Economics and the Theory of the State*. Springer, 2004.

Bolton, P. and Dewatripont, M. *Contract Theory*. MIT Press, 2004.

Chaloner, K. and Verdinelli, I. Bayesian experimental design: A review. *Statistical science*, pp. 273–304, 1995.

Chediak, L. Sharing sensitive health data in a federated data consortium model: An eight step guide. In *World Economic Forum. http://www3. weforum. org/docs/WEF_Sharing_Sensitive_Health_Data_2020. pdf*, 2020.

Cheng, C.-S. Optimality of certain asymmetrical experimental designs. *The Annals of Statistics*, 6(6):1239–1261, 1978.

Choi, T. and Robertson, P. J. Contributors and free-riders in collaborative governance: A computational exploration of social motivation and its effects. *Journal of Public Administration Research and Theory*, 29(3):394–413, 2019.

Ding, N., Fang, Z., and Huang, J. Incentive mechanism design for federated learning with multi-dimensional private information. In *2020 18th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOPT)*, pp. 1–8. IEEE, 2020.

Donahue, K. and Kleinberg, J. Optimality and stability in federated learning: A game-theoretic approach. *Advances in Neural Information Processing Systems*, 34: 1287–1298, 2021.

Donahue, K. and Kleinberg, J. Fairness in model-sharing games. In *Proceedings of the ACM Web Conference 2023*, pp. 3775–3783, 2023.

Eccleston, J. and Hedayat, A. On the theory of connected designs: characterization and optimality. *The annals of statistics*, pp. 1238–1255, 1974.

Fedorov, V. and Malyutov, M. Optimal design in regression experiments. *Math Operationsforsch Stat*, 14:237–324, 1972.

Fedorov, V. V. and Hackl, P. *Model-oriented design of experiments*, volume 125. Springer Science & Business Media, 1997.

Fraboni, Y., Vidal, R., and Lorenzi, M. Free-rider attacks on model aggregation in federated learning. In *International Conference on Artificial Intelligence and Statistics*, pp. 1846–1854. PMLR, 2021.

Geleijnse, G., Chiang, R. C.-J., Sieswerda, M., Schuurman, M., Lee, K., van Soest, J., Dekker, A., Lee, W.-C., and Verbeek, X. A. Prognostic factors analysis for oral cavity cancer survival in the netherlands and taiwan using a privacy-preserving federated infrastructure. *Scientific reports*, 10(1):1–9, 2020.

Ghorbani, A. and Zou, J. Data shapley: Equitable valuation of data for machine learning. In *International Conference on Machine Learning*, pp. 2242–2251. PMLR, 2019.

Goos, P. and Jones, B. *Optimal design of experiments: a case study approach*. John Wiley & Sons, 2011.

Graham, G. Disparities in cardiovascular disease risk in the united states. *Current cardiology reviews*, 11(3):238–245, 2015.

Huang, J., Talbi, R., Zhao, Z., Boucchenak, S., Chen, L. Y., and Roos, S. An exploratory analysis on users contributions in federated learning. In *2020 Second IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA)*, pp. 20–29. IEEE, 2020.

Hudson, M., Garrison, N., Sterling, R., Caron, N. R., Fox, K., Yracheta, J., Anderson, J., Wilcox, P., Arbour, L., Brown, A., et al. Rights, interests and expectations: Indigenous perspectives on unrestricted access to genomic data. *Nature Reviews Genetics*, 21(6):377–384, 2020.

Jia, R., Dao, D., Wang, B., Hubis, F. A., Hynes, N., Gürel, N. M., Li, B., Zhang, C., Song, D., and Spanos, C. J. Towards efficient data valuation based on the Shapley value. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pp. 1167–1176. PMLR, 2019.

Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., et al. Advances and open problems in federated learning. *Foundations and Trends® in Machine Learning*, 14(1–2):1–210, 2021.

Karimireddy, S. P., Guo, W., and Jordan, M. I. Mechanisms that incentivize data sharing in federated learning. *arXiv preprint arXiv:2207.04557*, 2022.

Karlin, S. and Studden, W. J. Optimal experimental designs. *The Annals of Mathematical Statistics*, 37(4):783–815, 1966.

Kiefer, J. On the nonrandomized optimality and randomized nonoptimality of symmetrical designs. *The Annals of Mathematical Statistics*, 29(3):675–699, 1958.

Kiefer, J. Construction and optimality of generalized youden designs ii. *A Survey of statistical designs and linear models*, pp. 333–353, 1975.

Kiefer, J. and Wolfowitz, J. The equivalence of two extremum problems. *Canadian Journal of Mathematics*, 12:363–366, 1960.

Kim, H., Park, J., Bennis, M., and Kim, S.-L. Blockchained on-device federated learning. *IEEE Communications Letters*, 24(6):1279–1283, 2019.

Konečnỳ, J., McMahan, H. B., Ramage, D., and Richtárik, P. Federated optimization: Distributed machine learning for on-device intelligence. *arXiv preprint arXiv:1610.02527*, 2016.

Koutsoupias, E. and Papadimitriou, C. Worst-case equilibria. In *Stacs*, volume 99, pp. 404–413. Springer, 1999.

Laffont, J.-J. and Martimort, D. The theory of incentives. In *The Theory of Incentives*. Princeton University Press, 2009.

Li, L., Fan, Y., Tse, M., and Lin, K.-Y. A review of applications in federated learning. *Computers & Industrial Engineering*, 149:106854, 2020.

Lin, J., Du, M., and Liu, J. Free-riders in federated learning: Attacks and defenses. *arXiv preprint arXiv:1911.12560*, 2019.

Mancini, J. Data portability, interoperability and digital platform competition: Oecd background paper. *Interoperability and Digital Platform Competition: OECD Background Paper (June 8, 2021)*, 2021.

McMahan, B., Moore, E., Ramage, D., Hampson, S., and y Arcas, B. A. Communication-efficient learning of deep networks from decentralized data. In *Artificial Intelligence and Statistics*, pp. 1273–1282. PMLR, 2017.

Mohri, M., Sivek, G., and Suresh, A. T. Agnostic federated learning. In *International Conference on Machine Learning*, pp. 4615–4625. PMLR, 2019.

Moncada-Torres, A., Martin, F., Sieswerda, M., Van Soest, J., and Geleijnse, G. Vantage6: an open source privacy preserving federated learning infrastructure for secure insight exchange. In *AMIA Annual Symposium Proceedings*, volume 2020, pp. 870. American Medical Informatics Association, 2020.

Nana-Sinkam, P., Kraschnewski, J., Sacco, R., Chavez, J., Fouad, M., Gal, T., AuYoung, M., Namoos, A., Winn, R., Sheppard, V., et al. Health disparities and equity in the era of covid-19. *Journal of clinical and translational science*, 5(1):e99, 2021.

Pukelsheim, F. *Optimal design of experiments*. SIAM, 2006.

Richardson, A., Filos-Ratsikas, A., and Faltings, B. Budget-bounded incentives for federated learning. In *Federated Learning*, pp. 176–188. Springer, 2020.

Sabt, M., Achemlal, M., and Bouabdallah, A. Trusted execution environment: what it is, and what it is not. In *2015 IEEE Trustcom/BigDataSE/Ispa*, volume 1, pp. 57–64. IEEE, 2015.

Sarikaya, Y. and Ercetin, O. Motivating workers in federated learning: A stackelberg game perspective. *IEEE Networking Letters*, 2(1):23–27, 2019.

Shah, K. R., Sinha, B. K., Shah, K. R., and Sinha, B. K. Optimality criteria in design of experiments. *Theory of Optimal Designs*, pp. 1–16, 1989.

Shi, Y., Yu, H., and Leung, C. A survey of fairness-aware federated learning. *arXiv preprint arXiv:2111.01872*, 2021.

Sibson, R. D_A-optimality and duality. *Progress in Statistics*, 9:677–692, 1974.

Silvey, S. Optimal design measures with singular information matrices. *Biometrika*, 65(3):553–559, 1978.

Silvey, S. *Optimal design: an introduction to the theory for parameter estimation*, volume 1. Springer Science & Business Media, 2013.

Sim, R. H. L., Zhang, Y., Chan, M. C., and Low, B. K. H. Collaborative machine learning with incentive-aware model rewards. In *International Conference on Machine Learning*, pp. 8927–8936. PMLR, 2020.

Smith, K. On the standard deviations of adjusted and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distribution of observations. *Biometrika*, 12(1/2):1–85, 1918.

Smith, S. A. *Contract Theory*. OUP Oxford, 2004.

Tian, M., Chen, Y., Liu, Y., Xiong, Z., Leung, C., and Miao, C. A contract theory based incentive mechanism for federated learning. *arXiv preprint arXiv:2108.05568*, 2021.

Wald, A. On the efficient design of statistical investigations. *The annals of mathematical statistics*, 14(2):134–140, 1943.

Wang, T., Rausch, J., Zhang, C., Jia, R., and Song, D. A principled approach to data valuation for federated learning. In *Federated Learning*, pp. 153–167. Springer, 2020.

Xu, X., Lyu, L., Ma, X., Miao, C., Foo, C. S., and Low, B. K. H. Gradient driven rewards to guarantee fairness in collaborative machine learning. *Advances in Neural Information Processing Systems*, 34, 2021.

Yu, H., Liu, Z., Liu, Y., Chen, T., Cong, M., Weng, X., Niyato, D., and Yang, Q. A fairness-aware incentive scheme for federated learning. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, pp. 393–399, 2020.

Zhan, Y., Zhang, J., Hong, Z., Wu, L., Li, P., and Guo, S. A survey of incentive mechanism design for federated learning. *IEEE Transactions on Emerging Topics in Computing*, 2021.

Zhang, N., Ma, Q., and Chen, X. Enabling long-term cooperation in cross-silo federated learning: A repeated game perspective. *IEEE Transactions on Mobile Computing*, 2022.

A. Backgrounds on experiment design

Many scientific problems involve determining the underlying parameter θ in relationships of type

$$y = \theta^\top x + e\,, \tag{7}$$

where $x \in \mathbb{R}^d$ represents certain experiment condition on which the data is collected, and e is standard Gaussian noise i.e. zero meaned and unit variance. Given a set of observations $(y_i, x_i)_{i=1,\ldots,m}$, Ordinary Least Squares (OLS) yields the estimator $\hat{\theta} = (X^\top X)^\dagger X^\top Y$ where $X = (x_1, \cdots, x_m)^\top \in \mathbb{R}^{m\times d}$ and $Y = (y_1, \ldots, y_m)^\top \in \mathbb{R}^m$ is the experiments/data and responses. The variability of $\hat{\theta}$ is governed by the Fisher information matrix $\mathcal{I}(X;\theta) = \sum_{i=1}^m x_i x_i^\top$. Specifically, the expected error on a prediction for x, conditioned on X is

$$\mathcal{E}(x; X) = \mathbb{E}\big(\hat{\theta}^\top x - \mathbb{E}(y|x)\big)^2 = x^\top \mathcal{I}^\dagger x\,. \tag{8}$$

Thus, the problem of optimal experiment design is to select the training data X which would "maximize" the information matrix $\mathcal{I}$, thereby minimizing uncertainty. More precisely, given a *design space* $\mathcal{X} \subset \mathbb{R}^d$ containing all possible data points which could be collected, the designer makes choice of a sampling strategy π (called *design measure*), which is a measures over $\mathcal{X}$. For technical simplicity, we assume that the design space is finite with $\mathcal{X} = \{x_1, \ldots, x_n\}$. The *information matrix* is defined as a function of π:

$$\mathbf{M}(\pi) := \textstyle\sum_{i=1}^n \pi_i x_i x_i^\top\,. \tag{9}$$

From (8), it can be seen that $\mathbf{M}(\pi)^{-1}$ is a matrix representing uncertainty along different directions under a sampling strategy π. To reduce this to a single scalar that can serve as an objective, it is convenient using an *optimality criterion* f[1] which is a function from the set of symmetric matrices in $\mathbb{R}^{d\times d} \to \mathbb{R}$:

$$\max_{\pi} f(\mathbf{M}(\pi)^{-1}), \text{ s.t. } \pi \in \Delta(\mathcal{X})\,. \tag{10}$$

Some popular choices of optimality criterion are as follows:

- E-criterion: $f_E(M^{-1}) = -\|M^{-1}\|_2$
- A-criterion: $f_A(M^{-1}) = -\mathrm{tr}\left(M^{-1}\right)$
- V-criterion: $f_V(M^{-1}) = -\frac{1}{|\mathcal{X}|}\sum_{x\in\mathcal{X}}\left(x^\top M^{-1} x\right)$.
- D-criterion: $f_D(M^{-1}) = \log\det M$
- G-criterion: $f_G(M^{-1}) = -\max_{x\in\mathcal{X}}\left(x^\top M^{-1} x\right)$

Notations For any vector $w \in \mathbb{R}^n$ and index set $G \subset [n]$, let $w_G \in \mathbb{R}^{|G|}$ denote the vector formed by the coordinates of w in the index set G (preserving the order), and let G^c denote the complement of G. We use $dF(u,v)$ to denote the Gateaux derivative of F at u in the direction v. Let $\mathbb{R}^d_+$ denote the nonnegative orthant in $\mathbb{R}^d$, i.e. $\mathbb{R}^d_+ = \{x \in \mathbb{R}^d : x_i \geq 0, \forall i \in [d]\}$. Denote $\langle A, B\rangle := \mathrm{tr}[A^\top B]$ for $A, B \in \mathbb{R}^{d\times d}$. We will use $\mathrm{supp}(\cdot)$ to denote support of a distribution or vector, i.e., the set consisting of all indices corresponding to nonzero entries. Let $\mathbb{S}^d_+$ denote the set of symmetric positive semidefinite matrices in $\mathbb{R}^{d\times d}$, matrix Loewner order $\preceq$ is a partial ordering on $\mathbb{S}^d_+$, such that $A \preceq B$ iff $B - A \in \mathbb{S}^d_+$. Furthermore, $A \prec B$ if $B - A$ is positive definite. We overload this notation and say $x \preceq y$ for two vectors $x, y \in \mathbb{R}^d$ iff $x_i \leq y_i$ for all $i = 1, \ldots, d$. Let e_i denote the one-hot vector (whose dimension will be specified in the context) with the i-th coordinate being 1 and the rest coordinate being zero. We also define $\mathbb{1}(A) = \begin{cases} 1, & A \\ 0, & \neg A \end{cases}$. Finally, $M^\dagger$ represents the Moore-Penrose inverse of matrix M.

B. Framework of principal-agents experiment design

Implementing the mechanism. The mechanism needs to return a $\hat{\theta}^{(k)}$ to agent k using data $\mathcal{M}^{(k)}(w)$. This is equivalent to requiring that $\hat{\theta}^{(k)}$ satisfy the following

$$\hat{\theta}^{(k)} \sim \mathcal{N}\Big(\theta\,,\ \mathbf{M}\big(\mathcal{M}^{(k)}(w)\big)^{-1}\Big)\,. \tag{11}$$

Thus, $\hat{\theta}^{(k)}$ needs to be an unbiased estimator of the ground truth with covariance $\mathbf{M}\big(\mathcal{M}^{(k)}(w)\big)^{-1}$.

A straightforward method of achieving this would be to run K parallel federated learning algorithms. Each of these would train a model $\hat{\theta}^{(k)}$ for agent k using only a subset of the data points as dictated by $\mathcal{M}^{(k)}(w)$.

[1] W.l.o.g. we assume $f(M^{-1}) = -\infty$ when M is singular, and so restrict ourselves to non-singular $\mathbf{M}(w)$.

Algorithm 1 Principal-Agents Collaborative Experiment Design

1: The principal selects and publishes a mechanism $\mathcal{M}$ satisfying

$$\mathcal{M} := (\mathcal{M}^{(k)} : \mathbb{R}_+^{\mathcal{X}} \to \mathbb{R}_+^{\mathcal{X}})_{k\in[K]} \text{ satisfying } \mathcal{M}^{(k)}(w) \leq w\,.$$

2: Each agent k decides whether to join the collaborative learning depending on Eq. (3,4).

3: If joining, agent k chooses a *design* $w_{G_k} \in \mathbb{R}_+^{|G_K|}$ which maximizes her utility:

$$\left(u^{(k)} \circ \mathcal{M}^{(k)}\right)(w) := f^{(k)}\left(\left(\mathbf{M}^{(k)}(\mathcal{M}^{(k)}(w))\right)^{-1}\right) - c^{(k)} \textstyle\sum_{i\in G_k} w_i\,.$$

If the agent is strategic, then w_{G_k} will correspond to the Nash equilibrium

$$\left(u^{(k)} \circ \mathcal{M}^{(k)}\right)(w^*) \geq \left(u^{(k)} \circ \mathcal{M}^{(k)}\right)(w_{G_k}, w^*_{G_k^c}),\ \forall\, w_{G_k} \in \mathbb{R}_+^{|G_k|}\,.$$

4: For all $i \in G_k$, she collects w_i independent samples from x_i, incurring a cost $c^{(k)}$ per unit.

5: The agents commit all the collected data to a collaborative learning procedure coordinated by the principal. Based on this aggregated data, the principal computes the OLS estimator $\hat{\theta}$.

6: Then, to each agent k, the principal sends back a possibly degraded $\hat{\theta}^{(k)}$ in accordance with the published $\mathcal{M}^{(k)}$ Eq. (11).

Alternatively, the principal could present a degraded model by adding noise. After a global model $\hat{\theta}$ is trained using federated learning on the combined data w, we add carefully tuned noise to determine $\hat{\theta}^{(k)}$ as:

$$\hat{\theta}^{(k)} \sim \mathcal{N}\Big(\hat{\theta}\,,\ \mathbf{M}\big(\mathcal{M}^{(k)}(w)\big)^{-1} - \mathbf{M}(w)^{-1}\Big)\,.$$

While perhaps conceptually simpler, computing $\mathbf{M}\big(\mathcal{M}^{(k)}(w)\big)^{-1}$ may be equally cumbersome as running K parallel federated learning algorithms.

Remark B.1 (Computational burden)**.** *While implementing the full mechanism may seem computationally burdensome, note that we only incur this burden if $\mathcal{M}^{(k)}(w) \neq w$. As we will see in the next sections, under equilibrium conditions we will always expect to see $\mathcal{M}^{(k)}(w) = w$ and so no additional computation is required. The mechanism is merely a deterrent.*

Finally, our framework assumes that an agent only has access to the final output of the mechanism, but not to any intermediaries. This is important since if we are learning $\hat{\theta}$ using FL, the agents may utilize the intermediary estimates (which may be of better quality), instead of $\hat{\theta}^{(k)}$. Here, we can appeal to security and cryptographic solutions.

Remark B.2 (Hiding intermediaries)**.** *The entire mechanism can be implemented in an encrypted/obfuscated software (Barak, 2016), or in a trusted execution environment (TEE) (Sabt et al., 2015). These solutions ensure that only the final output of the mechanism can be accessed and all intermediary computations remain hidden. Thus, the agents are prevented from cheating and follow our mechanism.*

Remark B.3 (Implementing via early stopping)**.** *The information maximizing mechanism (Eq. 6) is quite simple: it scales the design measure by a scalar $\gamma_k \in [0,1]$. This corresponds to randomly sub-sampling a γ_k fraction of the data to train $\hat{\theta}^{(k)}$. Instead of implementing this via K parallel sub-samplings and federated learning runs, a more convenient approximation may be achieved using early stopping. Intuitively, early stopping also effectively subsamples data. During training of the global model $\hat{\theta}$ for a total of T rounds, the model at the $\gamma_k T$ round is returned to agent k as its $\hat{\theta}^{(k)}$.*

C. Related works

In recent years, Federated Learning (FL) (Konečnỳ et al., 2016; McMahan et al., 2017; Kim et al., 2019; Kairouz et al., 2021; Li et al., 2020; Mancini, 2021) has become an emerging machine learning paradigm that allows multiple distributed clients to train a central statistical model under the orchestration of a principal. The wide application raises several ethical concerns such as free-riding and fairness (Baumol, 2004; Fraboni et al., 2021; Mohri et al., 2019; Huang et al., 2020;

Shi et al., 2021). To deal with these concerns, a number of works (Richardson et al., 2020; Sarikaya & Ercetin, 2019; Lin et al., 2019; Fraboni et al., 2021; Ding et al., 2020; Zhang et al., 2022) investigate free-rider attacks and develop methods for detection. Another line of works (Ghorbani & Zou, 2019; Jia et al., 2019; Wang et al., 2020) designs metrics for quantifying the contribution of each agent. More related are the recent works that apply the theory of contracts and incentives (Smith, 2004; Laffont & Martimort, 2009; Bolton & Dewatripont, 2004) to FL. Among them, (Tian et al., 2021) proposed a mechanism to achieve improved generalization accuracy by eliciting the private type information; (Sim et al., 2020; Xu et al., 2021) propose mechanisms based on notions from the cooperative game theory literature to incentivize agents through the model quality; (Karimireddy et al., 2022) introduces mechanisms based on accuracy-shaping to maximize the number of data points generated by each agent.

Our work is different from the above works in that (i) we study an autonomous data generation process in which each agent can strategically choose what experiment condition to collect data from; (ii) we explicitly model the utility and costs of self-interested agents. These challenges motivate us to formulate our problem based on the statistical problem of experiment design (Pukelsheim, 2006; Silvey, 2013). There is a long line of works (Smith, 1918; Wald, 1943; Kiefer, 1958; Eccleston & Hedayat, 1974; Kiefer, 1975; Cheng, 1978) studying optimally in linear experiment design. Among them, D-criterion is the most widely used optimality criteria (Shah et al., 1989) and our work is most relevant to (Kiefer & Wolfowitz, 1960; Sibson, 1974). To our knowledge, there is no existing study of multi-agent systems of experiment design.

D. Further discussions of $\mathcal{M}_{\max}$

D.1. Fairness

The $\mathcal{M}_{\max}$ mechanism ensures that the principal obtains the maximum possible information while preventing any agent from free-riding. Meanwhile, it is also crucial to examine notions of fairness to ensure that none of the participating agents are exploited. In order to address this, we analyze the utility of agents under the mechanism $\mathcal{M}_{\max}$ and present the following result.

Corollary D.1 (Incentive Compatibility). *Under mechanism $\mathcal{M}_{\max}$, the strategic response $w_{\max}$ satisfies $\left(u^{(k)} \circ \mathcal{M}_{\max}^{(k)}\right)(w_{\max}) = v_*^{(k)}$ for all $k \in [K]$.*

The above corollary is straightforward from the optimization problem in Eq. (5). Nevertheless, it carries two important implications. First, this corollary implies that the utility obtained by agent k through strategic participation in the collaborative learning, given by $\left(u^{(k)} \circ \mathcal{M}_{\max}^{(k)}\right)(w_{\max})$, is equal to the maximum utility $v_*^{(k)}$ that the agent can obtain by training individually. Therefore, all participating agents benefit equally from the collaborative learning process. In fact, the surplus generated by agents is directed towards enhancing the value of the statistical model, ultimately benefiting the social welfare. This highlights the equitable distribution of benefits and the collective progress achieved through collaboration. Secondly, Corollary D.1 implies that the utility of agent k under the mechanism depends solely on the resources and capacities of agent k itself, represented by $\mathcal{X}_k, A^{(k)}, f^{(k)}, c^{(k)}$, and is independent of other agents. Consequently, any improvements or innovations made by agent k to enhance experimental conditions or reduce marginal costs will be fully exploited within the mechanism. This incentivizes participating agents to enhance their own capacities and resources, promoting an environment of continuous improvement. Thus, the mechanism $\mathcal{M}_{\max}$ exhibits incentive compatibility, fostering agents' motivation to optimize their contributions.

The issue of fairness in the principal-agent experiment design problem is particularly relevant in the exchangeable data setting, where all data points have the same value (Karimireddy et al., 2022). In this scenario, there are no inherent distinctions between the resources and targets of different agents, therefore demanding the mechanism to avoid introducing extrinsic unfairness among the agents. Fortunately, our proposed mechanism satisfies a monotonic notion of fairness.

Proposition D.2 (Fairness under exchangeable data regime). *In the exchangeable data regime (i.e., X_i's are the same), the information maximization mechanism $\mathcal{M}_{\max}$ is fair in the sense that any strategic response $\bar{w} = (\bar{w}_{G_1}, \bar{w}_{G_2}, \ldots, \bar{w}_{G_K})$ satisfies that for all $k \in [K]$*

$$\left(u^{(k)} \circ \mathcal{M}_{\max}^{(k)}\right)(\bar{w}) \geq \left(u^{(k')} \circ \mathcal{M}_{\max}^{(k')}\right)(\bar{w}) \implies \|\bar{w}_{G_k}\|_1 \geq \|\bar{w}_{G'_k}\|_1.$$

This proposition states that in the exchangeable data regime, an agent must contribute more data in order to achieve a higher utility, which aligns with existing notions of fairness in the federated learning literature (Yu et al., 2020; Donahue & Kleinberg, 2021; 2023). When the data points are not exchangeable, fairness becomes more challenging to define due to the inherent heterogeneity of learning targets and resources. We leave the discussions regarding fairness in such scenarios the subject of future research. We leave the relevant discussions to future works.

D.2. Price of Anarchy

In this section we discuss price of anarchy (Koutsoupias & Papadimitriou, 1999) of the information maximization mechanism $\mathcal{M}_{\max}$.

Definition D.3 (Price of Anarchy)**.** We define the social good as

$$\mathrm{SG}(w) = \sum_{k=1}^{K} \left(u^{(k)} \circ \mathcal{M}_{\max}^{(k)} \right)(w).$$

and price of anarchy by the ratio between the maximal social good and the social good at strategic response, i.e.

$$\mathrm{POA} := \max_{w} \frac{\mathrm{SG}(w)}{\mathrm{SG}(w_{\max})}.$$

Price of anarchy measures the inefficiency and suboptimality resulting from strategic behaviors in principal-agents experiment design. The numerator is the optimal 'centralized' social good that can be achieved from the strategy spaces, and the denominator captures the social welfare obtained under selfish behaviors of each agent. To characterize price of anarchy of $\mathcal{M}_{\max}$, we introduce the following concept.

Definition D.4 (Benefit from Collaboration)**.** Define the Benefit from Collaboration of client k as

$$\begin{aligned}\Delta^{(k)} = &\max_{\pi \in \Delta([n])} - \log\det \left((A^{(k)})^{\top} \left(\sum_{i=1}^{n} \pi_i x_i x_i^{\top} \right)^{-1} A^{(k)} \right) \\ &- \max_{\pi \in \Delta(G_k)} \log\det \left((A^{(k)})^{\top} \left(\sum_{i \in G_k} \pi_i x_i x_i^{\top} \right) A^{(k)} \right).\end{aligned}$$

Intuitively, $\Delta^{(k)}$ describes the maximum achievable increase of information for agent k by joining the collaborative learning. We will show that the price of anarchy is bounded by the benefit from collaboration.

Proposition D.5. *Define* $k_0 = \arg\min_{k \in [K]} c^{(k)}$*. then* POA *can be upper bounded by*

$$\frac{\sum_{k=1}^{K} \Delta^{(k)}}{\sum_{k=1}^{K} \left(\theta^{(k)} + r_k \log \frac{r_k}{c^{(k)}} - r_k \right)} + \frac{\sum_{k=1}^{K} \left(r_k \log \frac{c^{(k)} \sum_{k=1}^{K} r_k}{r_k c^{(k_0)}} - (c^{(k)} - c^{(k_0)}) \cdot \|w_{\max, G_k}\|_1 \right)}{\sum_{k=1}^{K} \left(\theta^{(k)} + r_k \log \frac{r_k}{c^{(k)}} - r_k \right)} + 1.$$

To interpret the bound, we notice that the first term, $\frac{\sum_{k=1}^{K} \Delta^{(k)}}{\sum_{k=1}^{K} \left(\theta^{(k)} + r_k \log \frac{r_k}{c^{(k)}} - r_k \right)}$, represents the price of anarchy resulting from data diversity. It captures the extent to which each agent, denoted by k, benefits from a more diverse collection of data points contributed by other agents, which has the potential to improve the low-rank model of agent k. The second term, $\frac{\sum_{k=1}^{K} \left(r_k \log \frac{c^{(k)} \sum_{k=1}^{K} r_k}{r_k c^{(k_0)}} - (c^{(k)} - c^{(k_0)}) \cdot \|w_{\max, G_k}\|_1 \right)}{\sum_{k=1}^{K} \left(\theta^{(k)} + r_k \log \frac{r_k}{c^{(k)}} - r_k \right)}$, captures the price of anarchy resulting from cost heterogeneity and shared representation. It accounts for the potential exploitation of lower costs by the system in a centralized setting and the benefits of utilizing data collected from design spaces of rank r_k to improve the model across all rank $r_{k'}$ spaces for $k \in [K]$. Notice that the $\frac{\sum_{k=1}^{K} \left(-(c^{(k)} - c^{(k_0)}) \cdot \|w_{\max, G_k}\|_1 \right)}{\sum_{k=1}^{K} \left(\theta^{(k)} + r_k \log \frac{r_k}{c^{(k)}} - r_k \right)}$ is a negative term that demonstrates the cost heterogeneity mitigated by the information maximization mechanism $\mathcal{M}_{\max}$.

E. Examples

In this section, we present several illustrative examples that highlight the strategic behaviors of self-interested agents in different scenarios.

E.1. Free-riding behavior

Although the federated learning mechanism $\mathcal{M}_{\text{fed}}$ achieves efficiency in the multi-agent system, it can lead to unfair Nash equilibria in which some agents contribute much fewer data than others. This phenomenon, known as free-riding, is highly undesirable in federated learning (Baumol, 2004; Choi & Robertson, 2019; Sarikaya & Ercetin, 2019; Lin et al., 2019; Ding et al., 2020; Sim et al., 2020; Xu et al., 2021).

The following proposition shows that, in general, $w_{\max}$ cannot be achieved under the common federated learning mechanism $\mathcal{M}_{\text{fed}}$, as there exists at least one agent who can achieve higher utility by contributing fewer samples.

Proposition E.1 (Partial free-riding under federated learning mechanism)**.** *Unless $\sum_{k=1}^{K} r_k = d$, $w_{\max}$ is not the Nash equilibrium of the utility functions $\left(\left(u^{(k)} \circ \mathcal{M}_{\text{fed}}^{(k)}\right)\right)_{k\in[K]}$. More precisely, there exists $k \in [K]$ and $\widetilde{w}_{G_k}$ such that $\widetilde{w}_i \leq w_{\max,i}, \forall i \in G_k$, $\widetilde{w}_i < w_{\max,i}, \exists i \in G_k$, and*

$$\left(u^{(k)} \circ \mathcal{M}_{\text{fed}}^{(k)}\right)\left((\widetilde{w}_{G_k}, w_{\max,G_k^c})\right) > \left(u^{(k)} \circ \mathcal{M}_{\text{fed}}^{(k)}\right)(w_{\max})$$

where $(\widetilde{w}_{G_k}, w_{\max,G_k^c})$ denotes the concatenation of $\widetilde{w}_{G_k}$ and $w_{\max,G_k^c}$.

The condition $\sum_{k=1}^{K} r_k = d$ implies that the covariates in $\mathcal{X}_k$ for each agent k form independent subspaces, and each agent cannot benefit from the data from other agents. Thus, this condition is unlikely to be encountered in the study of collaborative learning.

We illustrate two possible cases where completely free-riding by contributing no data is the optimal strategy for an agent.

Example E.2 (Free-riding due to data diversity)**.** Consider a principal-agents experiment design problem where one agent possesses a data set with high diversity, such that her design space covers the design space of the other agent. In such cases, it can be demonstrated that the second agent will engage in free-riding behavior at pure Nash equilibrium. We establish the following result to formalize this scenario.

Proposition E.3. *Suppose agent k's design space $\mathcal{X}_k$ and agent l's design space $\mathcal{X}_l$ satisfy that $\{x_i x_i^\top : i \in G_l\} \subset \left\{\sum_{i\in G_k} \alpha_i x_i x_i^\top : \alpha \in \mathbb{R}_+^{|G_k|}, \sum_{i\in G_k} \alpha_i < 1, \right\}$ then in any pure Nash equilibrium, $w_{G_l} = 0$.*

Example E.4 (Free-riding due to cost heterogeneity)**.** Consider another scenario where strategic agents with higher marginal costs may engage in free-riding behavior. Intuitively, in equilibrium, an agent with a lower marginal cost experiences a higher marginal increase in utility by sampling more data. If another agent possesses the same experimental capacity but at a higher cost, she is expected to engage in free-riding at pure Nash equilibrium. More precisely, we have the following result.

Proposition E.5. *Suppose $\mathcal{X}_k = \mathcal{X}_l$ for some $k \neq l$ and $c^{(k)} < c^{(l)}$, then in any pure Nash equilibrium, $w_{G_l} = 0$.*

These examples highlight situations where agents have incentives to free-ride due to factors such as data diversity or cost disparities. Such behaviors can undermine the fairness and collaboration within the multi-agent system. In the subsequent sections, we delve into the analysis of free-riding behaviors and propose mechanisms to mitigate these issues.

E.2. Toy examples and simulations

Example E.6 (Free riding)**.** Consider $\mathcal{X} = \{x_1 = (1,0,0)^\top, x_2 = (0,1,0)^\top, x_3 = (0,0,1)^\top, x_4 = (0,1,1)^\top\}$ and the index sets given by $G_i = \{i\}$ for $i = 1,2,3,4$, i.e. four agents each holding a rank-1 set of experiment condition. Now, let's assume that the cost for the agents are $c^{(1)} = c^{(2)} = c^{(3)} = c \leq 0.5c^{(4)}$.

In this setup, we can observe that the Nash equilibrium for the standard federated learning mechanism is achieved when $w_1 = w_2 = w_3 = \frac{1}{c}$ and $w_4 = 0$. However, in this Nash equilibrium, agent x_4 contributes nothing to the collaborative learning process while benefiting from the information provided by the second and the third agents. This behavior, where agents exploit the contributions of others without contributing themselves, is known as free riding in federated learning.

Example E.7 (Selfish allocation)**.** Consider $u, v > 0$ and $\mathcal{X} = \{x_{2i+1} = u \cdot e_{i+1} \in \mathbb{R}^n, x_{2i+2} = e_1 \in \mathbb{R}^n \ (i = 1,2,\ldots,n-1), x_{2n+1} = v \cdot e_1 \in \mathbb{R}^n\}$ and the index sets given by $G_i = \{2i+1, 2i+2\}$ $(i = 1,2,\ldots,n-1)$, $G_n = \{2n+1\}$. That is, there are n agents; the first $n-1$ agents each holds a rank-2 set $\{u \cdot e_{i+1}, e_1\}$ where e_1 can be seen as a shared feature and e_{i+1} can be seen as the unique feature; the n-th agent holds $\{v \cdot e_1\}$.

In this setup, each agent has a distinct feature and a shared feature. The first $n-1$ agents may selfishly conduct experiments only on their unique feature $(u \cdot e_{i+1})$ while hoping that other agents would experiment on the shared feature (e_1). This results in a selfish allocation of experiments, which can be highly inefficient.

For example, when $c^{(1)} = \cdots = c^{(n-1)} \geq c^{(n)}/v^2$, it is clear that $w_{2i+1} = \frac{1}{c^{(i)}}$ $(i = 1, 2, \ldots, n), w_{2i+2} = 0$ $(i = 1, 2, \ldots, n-1)$ is a Nash equilibrium for the standard federated learning mechanism. In this Nash equilibrium, the first $n-1$ agents only experiment on $u \cdot e_{i+1}$ and in the end only the n-th agent samples from $x_{2n+1} = v \cdot e_1$. However, when $v \ll 1$, x_{2i+2} gives more information and lies in the support of optimal experiment design instead of x_{2n+1}. This is an example of selfish allocation in federated learning.

Example E.8 (Case study of substitutable, orthogonal, and complimentary data)**.** Consider the following design space: $\mathcal{X} = \{x_1 = (\cos\theta, \sin\theta)^\top, x_2 = (1, 0)^\top, x_3 = (0, 1)^\top\}$, and groups $G_1 = 1$ and $G_2 = 2, 3$. Notably, x_2 and x_3 are orthogonal, while the parameter θ governs the degree of complementarity between x_1 and x_2. We investigate the strategic behaviors of the federated learning and information maximization mechanism by varying θ and $c^{(2)}/c^{(1)}$. By direct computation, the strategic response is given by

$$w_1 = \begin{cases} 0, & c^{(2)} < c^{(1)} \\ \frac{1}{c^{(1)}}, & c^{(1)} + c^{(2)}(\sin^2\theta - \cos^2\theta) < 0 \\ \frac{1}{c^{(1)}}, & c^{(1)} - c^{(2)}(\sin^2\theta - \cos^2\theta) < 0 \\ \frac{c^{(2)} - c^{(1)}}{c^{(2)}c^{(1)} - (c^{(1)})^2 - (c^{(2)})^2(\sin^2\theta - \cos^2\theta)^2}, & \text{else} \end{cases}$$

$$w_2 = \begin{cases} \frac{1}{c^{(2)}}, & c^{(2)} < c^{(1)} \\ 0, & c^{(1)} + c^{(2)}(\sin^2\theta - \cos^2\theta) < 0 \\ \frac{1}{c^{(2)}}, & c^{(1)} - c^{(2)}(\sin^2\theta - \cos^2\theta) < 0 \\ \frac{c^{(1)} + c^{(2)}(\sin^2\theta - \cos^2\theta)}{c^{(2)}c^{(1)} - (c^{(1)})^2 - (c^{(2)})^2(\sin^2\theta - \cos^2\theta)^2}, & \text{else.} \end{cases}$$

$$w_3 = \begin{cases} \frac{1}{c^{(2)}}, & c^{(2)} < c^{(1)} \\ \frac{1}{c^{(2)}}, & c^{(1)} + c^{(2)}(\sin^2\theta - \cos^2\theta) < 0 \\ 0, & c^{(1)} - c^{(2)}(\sin^2\theta - \cos^2\theta) < 0 \\ \frac{c^{(1)} - c^{(2)}(\sin^2\theta - \cos^2\theta)}{c^{(2)}c^{(1)} - (c^{(1)})^2 - (c^{(2)})^2(\sin^2\theta - \cos^2\theta)^2}, & \text{else.} \end{cases}.$$

This leads to sub-optimal principal utility compared to the information maximization regime. We show the data contribution w_1, w_2, w_3 and the total information for varying $c^{(2)}$ and θ in Figure 2 and Figure 3. These visualizations demonstrate how cost heterogeneity and data diversity influence the strategic response across various mechanisms. Remarkably, the information maximization strategy $\mathcal{M}_{\max}$ yields improved data contribution and information while exhibiting a more stable behavior.

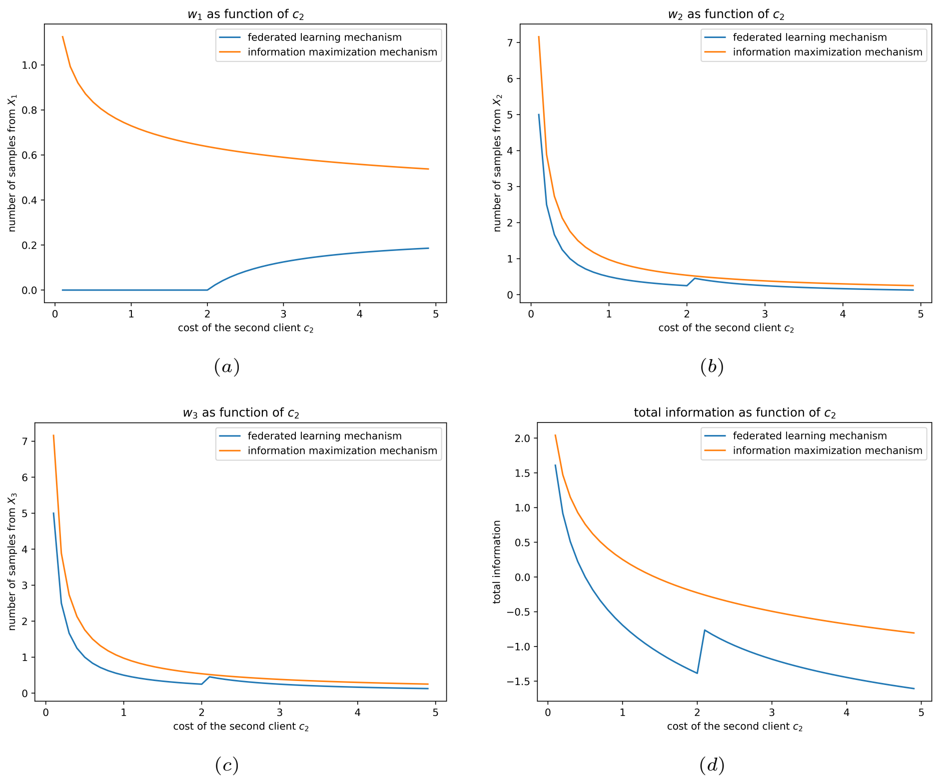

Figure 2. Comparison between federated learning mechanism and information maximization mechanism for different $c^{(2)}$ with fixed $c^{(1)} = 2, \theta = \pi/4$.

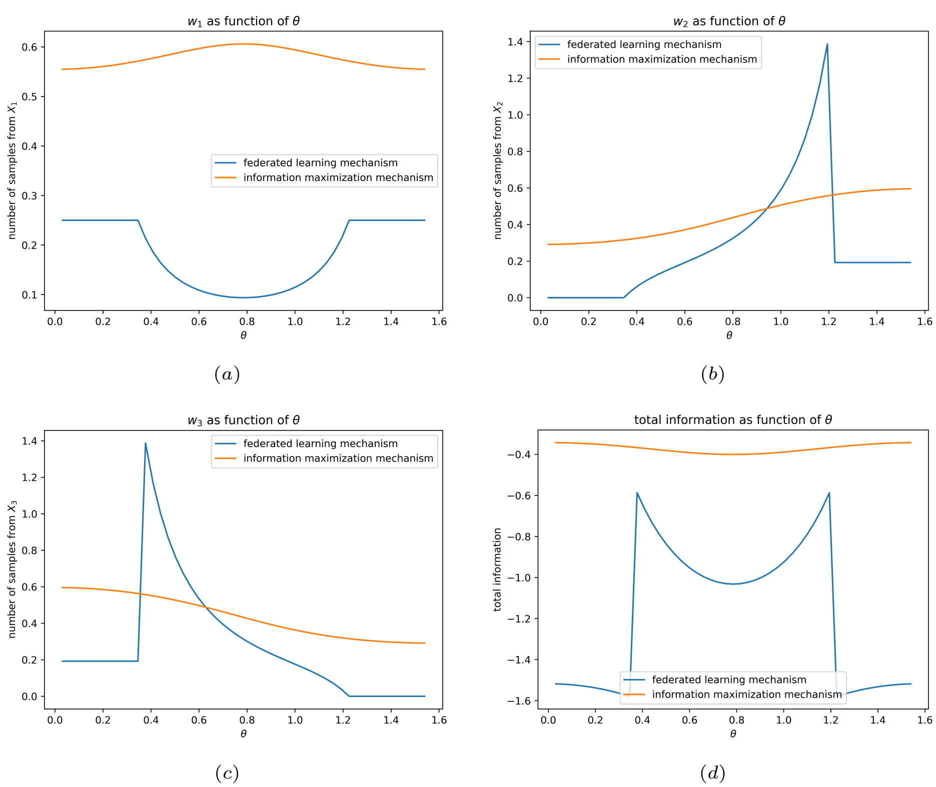

Figure 3. Comparison between federated learning mechanism and information maximization mechanism for different θ with fixed $c^{(1)} = 2, c^{(2)} = 2.5$.

F. Omitted proofs

In this and the following sections, we will use $\text{supp}(\cdot)$ to denote support of a distribution or vector, i.e., the set consisting of all indices corresponding to nonzero entries. Let $\mathbb{S}_+^d$ denote the set of symmetric positive semidefinite matrices in $\mathbb{R}^{d\times d}$, matrix Loewner order $\preceq$ is a partial ordering on $\mathbb{S}_+^d$, such that $A \preceq B$ iff $B - A \in \mathbb{S}_+^d$. Furthermore, $A \prec B$ if $B - A$ is positive definite. We overload this notation and say $x \preceq y$ for two vectors $x, y \in \mathbb{R}^d$ iff $x_i \le y_i$ for all $i = 1, \ldots, d$. Let e_i denote the one-hot vector (whose dimension will be specified in the context) with the i-th coordinate being 1 and the rest coordinate being zero. We also define $\mathbb{1}(A) = \begin{cases} 1, & A \\ 0, & \neg A \end{cases}$.

F.1. Proof of Proposition 3.2

Proof. We first show that when $c^{(1)} = \cdots c^{(K)} = c$, $\widetilde{w} = \frac{d}{c} \cdot \pi^*$ is a pure Nash equilibrium. Consider the following function of w_{G_k}

$$\bar{u}_k(w_{G_k}) = -\log\det\left((A^{(k)})^\top \mathbf{M}\left((w_{G_k}, (\widetilde{w}_{G_k^c}))^{-1} A^{(k)}\right) - c \cdot \sum_{i\in G_k} w_i\right.$$

Indeed, applying Lemma H.1 and Theorem H.5,

$$\begin{aligned} d\bar{u}_k(\widetilde{w}_{G_k}, \Delta w_{G_k}) &= \sum_{i\in G_k} \Delta w_i \langle x_i x_i^\top, \mathbf{M}(\widetilde{w})^{-1}\rangle - c \cdot \sum_{i\in G_k} \Delta w_i \\ &\le \left(\frac{d}{\|\widetilde{w}\|_1} - c\right) \cdot \sum_{i\in G_k} \Delta w_i \\ &= 0,\ \forall \Delta w_{G_k}. \end{aligned}$$

By concavity of $\bar{u}_k(\widetilde{w}_{G_k})$, $\widetilde{w}_{G_k}$ is the maximizer of $\bar{u}_k$.

To show IR, define

$$w^*_{G_k} = \arg\max_{w_{G_k}} f_D^{(k)}\left(\left(\sum_{i\in G_k} w_i \cdot (A^{(k)})^\top x_i x_i^\top A^{(k)}\right)^{-1}\right) - c \sum_{i\in G_k} w_i,$$

we have

$$\begin{aligned} \left(u^{(k)} \circ \mathcal{M}_{\text{fed}}^{(k)}\right)(\widetilde{w}) &\ge \bar{u}_k(w^*_{G_k}) \\ &= -\log\det\left((A^{(k)})^\top \mathbf{M}\left((w^*_{G_k}, \widetilde{w}_{G_k^c})\right)^{-1} A^{(k)}\right) - c \cdot \sum_{i\in G_k} w_i^* \\ &\ge -\log\det\left((A^{(k)})^\top (\sum_{i\in G_k} w_i^* x_i x_i^\top) A^{(k)}\right)^{-1} - c \cdot \sum_{i\in G_k} w_i^* \\ &= v_*^{(k)} \end{aligned}$$

where the first inequality follows from $\widetilde{w}_{G_k} \in \arg\max \bar{u}_k$; the second inequality comes from Lemma H.2.

Next, we show the uniqueness. Suppose $\widetilde{w}$ is a Nash equilibrium of the tuple of utility functions $\left((u^{(k)} \circ \mathcal{M}_{\text{fed}})\right)_{k\in[K]}$. It follows from first-order optimality that for any Δw_{G_k} such that $\text{supp}(\Delta w_{G_k}) \subset \text{supp}(\widetilde{w}_{G_k})$

$$\begin{aligned} 0 &= d\bar{u}_k(\widetilde{w}_{G_k}, \Delta w_{G_k}) \\ &= \sum_{i\in G_k} \Delta w_i \langle x_i x_i^\top, \mathbf{M}(\widetilde{w})^{-1}\rangle - c^{(k)} \cdot \sum_{i\in G_k} \Delta w_i \\ &= \sum_{i\in G_k} \left(\langle x_i x_i^\top, \mathbf{M}(\widetilde{w})^{-1}\rangle - c\right) \cdot \Delta w_i. \end{aligned}$$

Therefore $\langle x_i x_i^\top, \mathbf{M}(\widetilde{w})^{-1}\rangle = c$ holds for all $i \in \operatorname{supp}(\widetilde{w})$. Notice that

$$d = \left\langle \sum_{i=1}^n \widetilde{w}_i x_i x_i^\top, \mathbf{M}(\widetilde{w})^{-1} \right\rangle$$
$$= c \sum_{i=1}^n \widetilde{w}_i.$$

We thus have $\sum_{i=1}^n \widetilde{w}_i = \frac{d}{c}$, and as a result, $\langle x_i x_i^\top, \mathbf{M}(\widetilde{w}/\|\widetilde{w}\|_1)^{-1}\rangle = d$ holds for all $i \in \operatorname{supp}(\widetilde{w})$. Furthermore, for any $i \notin \operatorname{supp}(\widetilde{w})$ first-order optimality implies $\langle x_i x_i^\top, \mathbf{M}(\widetilde{w}/\|\widetilde{w}\|_1)^{-1}\rangle \le d$. Applying Theorem H.5, we know that $\widetilde{w}/\|\widetilde{w}\|_1$ is a D-optimal design. This confirms that any strategic response follows the D-optimal design measure.

Finally, we show that federated learning is not efficient in any other criteria.

V-criterion.

Consider principal-agents experiment design with the accuracy function given by

$$f^{(k)}(w) = -\mathbb{E}_{x\sim p^{(k)}}\left[x^\top \mathbf{M}(w)^{-1} x\right].$$

where $p^{(k)}$ represents the distribution of client k's data and is supported on G_k.

In this case, the utility function under federated learning mechanism is given by

$$\left(u^{(k)} \circ \mathcal{M}_{\text{fed}}^{(k)}\right)(w) = -\mathbb{E}_{x\sim p^{(k)}}\left[x^\top \mathbf{M}(w)^{-1} x\right] - c\sum_{i\in G_k} w_i.$$

The Nash equilibrium $w^* \in \mathbb{R}_+^n$ thus gives the following system:

$$\mathbb{E}_{x\sim p^{(k)}}\left[\langle xx^\top, \mathbf{M}(w^*)^{-1} x_l x_l \mathbf{M}(w^*)^{-1}\rangle\right] \begin{cases} = c, & l \in \operatorname{supp}(w^*) \\ \le c, & l \notin \operatorname{supp}(w^*) \end{cases}, \ \forall l \in G_k, k \in [K]. \tag{12}$$

Consider the efficient allocation over the population p ($\operatorname{supp}(p) \subset \mathcal{X}$)

$$\pi^* = \arg\min_{\pi\in\Delta([n])} \mathbb{E}_{x\sim p}\left[x^\top \mathbf{M}(\pi)^{-1} x\right].$$

The optimal design measure requires the following system:

$$\mathbb{E}_{x\sim p}\left[\langle xx^\top, \mathbf{M}(\pi^*)^{-1} x_l x_l \mathbf{M}(\pi^*)^{-1}\rangle\right] \begin{cases} = \langle \mathbb{E}_{x\sim p}\left[xx^\top\right], \mathbf{M}(\pi^*)^{-1}\rangle, & l \in \operatorname{supp}(\pi^*) \\ \le \langle \mathbb{E}_{x\sim p}\left[xx^\top\right], \mathbf{M}(\pi^*)^{-1}\rangle, & l \notin \operatorname{supp}(\pi^*) \end{cases}, \ \forall l \in [n]. \tag{13}$$

Therefore, if the unique pure Nash equilibrium follows the optimal design measure, then $p, p^{(1)}, \ldots, p^{(K)}$ must satisfy the linear system given by Eq. (12) and Eq. (13). The solution of this linear system is generally a subspace of $\Delta([n])$ that has zero measure.

As a result, there exists a design measure such that federated learning mechanism is not efficient.

G-criterion and E-criterion.

Consider the following design space $\mathcal{X} = \{(1,0,0)^\top, (0,1,0)^\top, (0,0,1)^\top\}$ and let there be two agents with index sets $G_1 = \{1\}, G_2 = \{2,3\}$. It is not to see that in this case,

$$f_G^{(k)}(w) = f_E^{(k)}(w) = -\max_{i\in G_k}\left[x_i^\top \mathbf{M}(w)^{-1} x_i\right] = -\|(A^{(k)})^\top \mathbf{M}(w)^{-1} A^{(k)}\|_2$$
$$= \begin{cases} -w_1^{-1}, & k = 1 \\ \min\{-w_2^{-1}, -w_3^{-1}\}, & k = 2. \end{cases}$$

Therefore the unique pure Nash equilibrium is given by $w_1 = c^{-1/2}, w_2 = w_3 = (2c)^{-1/2}$. This is clearly not proportional to the optimal design measure which is uniform over $\mathcal{X}$.

A-criterion.

Consider the following design space $\mathcal{X} = \{(1,1)^\top, (1,0)^\top, (0,1)^\top\}$ and let there be two agents with index sets $G_1 =$

$\{1, 2\}, G_2 = \{3\}$. It is not to see that in this case,

$$-\text{tr}\left[(A^{(k)})^\top \mathbf{M}(w)^{-1} A^{(k)}\right] = \begin{cases} -\frac{2w_1+w_2+w_3}{w_1w_3+w_2w_3+w_1w_3}, & k=1 \\ -\frac{w_1+w_2}{w_1w_3+w_2w_3+w_1w_3}, & k=2. \end{cases}$$

The pure Nash equilibrium (w_1, w_2, w_3) follows the system:

$$\frac{(w_1+w_2)^2}{w_1w_3+w_2w_3+w_1w_3} - c = 0$$
$$\frac{w_2^2+w_3^2}{w_1w_3+w_2w_3+w_1w_3} - c = 0$$
$$\frac{2w_1^2+w_3^2+2w_1w_3}{w_1w_3+w_2w_3+w_1w_3} - c = 0.$$

The optimal design measure (π_1, π_2, π_3) follows the system:

$$\frac{2\pi_1^2+\pi_2^2+2\pi_2\pi_1}{\pi_1\pi_3+\pi_2\pi_3+\pi_1\pi_3} - \lambda = 0$$
$$\frac{\pi_2^2+\pi_3^2}{\pi_1\pi_3+\pi_2\pi_3+\pi_1\pi_3} - \lambda = 0$$
$$\frac{2\pi_1^2+\pi_3^2+2\pi_1\pi_3}{\pi_1\pi_3+\pi_2\pi_3+\pi_1\pi_3} - \lambda = 0$$

where λ is the Lagrangian multiplier.

If(w_1, w_2, w_3) is proportional to (π_1, π_2, π_3), then comparing these two systems yields $\pi_1 = 0$. This is a contradiction.

□

F.2. Proof of Proposition E.3

Proof. Suppose there exists a pure Nash equilibrium $\widetilde{w}$ such that $\widetilde{w}_j \neq 0$ and $j \in G_l$. Then from the first order optimality, for any Δw_{G_l} such that $\text{supp}(\Delta w_{G_l}) \subset \text{supp}(\widetilde{w}_{G_l})$

$$\begin{aligned} 0 &= d\bar{u}_l(\widetilde{w}_{G_l}, \Delta w_{G_l}) \\ &= \sum_{i\in G_l} \Delta w_i \langle x_i x_i^\top, \mathbf{M}(\widetilde{w})^{-1}\rangle - c \cdot \sum_{i \in G_l} \Delta w_i \\ &= \sum_{i\in G_l} \left(\langle x_i x_i^\top, \mathbf{M}(\widetilde{w})^{-1}\rangle - c\right) \cdot \Delta w_i. \end{aligned}$$

It follows that $\langle x_j x_j^\top, \mathbf{M}(\widetilde{w})^{-1}\rangle = c$. Similarly, $\langle x_i x_i^\top, \mathbf{M}(\widetilde{w})^{-1}\rangle \leq c$ for all $i \in G_k$. From the condition, there exists $\alpha_i > 0$ such that $x_j x_j^\top = \sum_{i\in G_k} \alpha_i x_i x_i^\top$ and $\sum_{i \in G_k} \alpha_i < 1$. It follows that

$$\langle x_j x_j^\top, \mathbf{M}(\widetilde{w})^{-1}\rangle = \sum_{i\in G_k} \alpha_i \langle x_i x_i^\top, \mathbf{M}(\widetilde{w})^{-1}\rangle < c.$$

Contradiction! □

F.3. Proof of Proposition E.5

Proof. Suppose there exists a pure Nash equilibrium $\widetilde{w}$ such that $\widetilde{w}_j \neq 0$ and $j \in G_l$. Then first order optimality yields for any Δw_{G_l} such that $\text{supp}(\Delta w_{G_l}) \subset \text{supp}(\widetilde{w}_{G_l})$

$$0 = \sum_{i \in G_l} \left(\langle x_i x_i^\top, \mathbf{M}(\widetilde{w})^{-1}\rangle - c^{(l)}\right) \cdot \Delta w_i.$$

It follows that $\langle x_j x_j^\top, \mathbf{M}(\widetilde{w})^{-1}\rangle = c^{(l)}$. Similarly, $\langle x_i x_i^\top, \mathbf{M}(\widetilde{w})^{-1}\rangle \leq c^{(k)}$ for all $i \in G_k$. This yields $c^{(k)} \geq \langle x_i x_i^\top, \mathbf{M}(\widetilde{w})^{-1}\rangle = c^{(l)}$ for $i \in \text{supp}(\widetilde{w}_{G_l})$, which contradicts the condition. □

F.4. Proof of Proposition 4.1

Proof. We first notice that the problem in Eq. (5) has compact feasible set and concave objective. Therefore, it has unique maximizer $w_{\max}$. Consider any mechanism $\mathcal{M}$. The maximum possible information that can be achieved under $\mathcal{M}$ is given by

$$\max_{w\in\mathbb{R}_+^n} \log\det \mathbf{M}(w), \text{ s.t. } \left(u^{(k)} \circ \mathcal{M}^{(k)}\right)(w) \geq v_*^{(k)}.$$

Let $\widetilde{w}$ be the maximizer of the above program, then due to $\mathcal{M}^{(k)}(w)_i \leq w_i,\ \forall i$ we have

$$\begin{aligned} u^{(k)}(w) &\geq \left(u^{(k)} \circ \mathcal{M}^{(k)}\right)(w) \\ &\geq v_*^{(k)}. \end{aligned}$$

Therefore $\widetilde{w}$ is in the feasible set of the optimization problem in Eq. (5). It follows from definition of $w_{\max}$ that $\log\det \mathbf{M}(\widetilde{w}) \leq \log\det \mathbf{M}(w_{\max})$. □

F.5. Proof of Proposition E.1

Proof. First, notice that

$$u^{(k)}\left(w_{\max}\right) = v_*^{(k)},\ \forall k \in [K].$$

Indeed, if there exist $k \in [K]$ such that $u^{(k)}\left(w_{\max}\right) > v_*^{(k)}$, then by setting $w'_{\max,i} = \begin{cases} (1+\epsilon)\cdot w_{\max,i}, & \text{if } i \in G_k \\ w_{\max,i}, & \text{if } i \notin G_k \end{cases}$ of sufficiently small $\epsilon > 0$, the constraints in Eq. (5) is still satisfied, but $\log\det \mathbf{M}(w'_{\max}) > \log\det \mathbf{M}(w_{\max})$. This contradicts to the fact that $w_{\max}$ is the maximizer.

Suppose $w_{\max}$ is the Nash equilibrium of $\left(\left(u^{(k)} \circ \mathcal{M}_{\text{fed}}^{(k)}\right)\right)_{k\in[K]}$, we will show that $\sum_{i=1}^K r_k = d$.

Indeed, by defining

$$\bar{u}_k(w_{G_k}) := -\log\det\left((A^{(k)})^\top \mathbf{M}\left((w_{G_k}, w_{\max,G_k^c})\right)^{-1} A^{(k)}\right) - c^{(k)} \cdot \sum_{i\in G_k} w_i,$$

it follows that $w_{\max,G_k}$ is the maximizer of $\bar{u}_k$. First-order optimality condition and Lemma H.1 yields that for any k and $l \in \operatorname{supp}(w_{\max,G_k})$

$$0 = d\bar{u}_k(w_{\max,G_k}, e_l) = \langle x_l x_l^\top, \mathbf{M}(w_{\max})^{-1}\rangle - c^{(k)}.$$

As a result,

$$\begin{aligned} \sum_{k=1}^K c^{(k)} \cdot \sum_{i\in G_k} w_{\max,i} &= \sum_{k=1}^K \sum_{i\in G_k} w_{\max,i}\langle x_i x_i^\top, \mathbf{M}(w_{\max})^{-1}\rangle \\ &= \left\langle \sum_{i=1}^n w_{\max,i} x_i x_i^\top, \mathbf{M}(w_{\max})^{-1} \right\rangle \\ &= d. \end{aligned} \tag{14}$$

Define

$$\begin{aligned} v_k(w_{G_k}) &= u^{(k)}\left(\mathbf{0}, \ldots, \mathbf{0}, w_{G_k}, \mathbf{0}, \ldots, \mathbf{0}\right) \\ &= -\log\det\left((A^{(k)})^\top (\sum_{i\in G_k} w_i x_i x_i^\top) A^{(k)}\right)^{-1} - c^{(k)} \cdot \sum_{i\in G_k} w_i. \end{aligned}$$

Let $w_{G_k}^* \in \arg\max_{w_{G_k}} v_k(w_{G_k})$, We have,

$$\begin{aligned}
\left(u^{(k)} \circ \mathcal{M}_{\text{fed}}^{(k)}\right)(w_{\max}) &\geq \bar{u}_k(w_{G_k}^*) \\
&= -\log\det\left((A^{(k)})^\top \mathbf{M}\left((w_{G_k}^*, w_{\max,G_k^c})\right)^{-1} A^{(k)}\right) - c^{(k)} \cdot \sum_{i \in G_k} w_i^* \\
&\geq -\log\det\left((A^{(k)})^\top (\sum_{i \in G_k} w_i^* x_i x_i^\top) A^{(k)}\right)^{-1} - c^{(k)} \cdot \sum_{i \in G_k} w_i^* \\
&= v_*^{(k)} \\
&= \left(u^{(k)} \circ \mathcal{M}_{\text{fed}}^{(k)}\right)(w_{\max})
\end{aligned}$$

where the second inequality is due to Lemma H.2. Therefore, the above inequalities are all equalities, which implies $w_{G_k}^* \in \arg\max \bar{u}_k(w_{G_k})$ and

$$-\log\det\left((A^{(k)})^\top \mathbf{M}\,(w_{\max})^{-1} A^{(k)}\right) = -\log\det\left((A^{(k)})^\top (\sum_{i \in G_k} w_{\max,i} x_i x_i^\top) A^{(k)}\right)^{-1}.$$

It follows that $w_{\max,G_k} \in \arg\max v_k(w_{G_k})$ and thus $\|w_{\max,G_k}\|_1 = \|w_{G_k}^*\|_1$.

First-order optimality condition and Theorem H.5 yields that for any k and $l \in G_k$

$$0 = d\bar{v}_k(w_{G_k}^*, e_l) = \frac{r_k}{\|w_{G_k}^*\|_1} - c^{(k)}.$$

As a result, $\|w_{\max,G_k}\|_1 = \|w_{G_k}^*\|_1 = \frac{r_k}{c^{(k)}}$. Combining this and Eq. (14), we have

$$d = \sum_{k=1}^K c^{(k)} \cdot \sum_{i \in G_k} w_{\max,i} = \sum_{k=1}^K c^{(k)} \cdot \frac{r_k}{c^{(k)}} = \sum_{k=1}^K r_k.$$

This establishes the first statement.

If $\sum_{k=1}^K r_k > d$, then the above arguments imply that there exist $k \in [K]$ and $i \in \text{supp}(w_{\max,G_k})$ such that $d\bar{u}_k(w_{\max,G_k}, e_i) < 0$. It follows that by letting $\widetilde{w}_{G_k} = w_{\max,G_k} - \epsilon e_i$ for sufficiently small $\epsilon > 0$, we have

$$\left(u^{(k)} \circ \mathcal{M}_{\text{fed}}^{(k)}\right)\left((\widetilde{w}_{G_k}, w_{\max,G_k^c})\right) > \left(u^{(k)} \circ \mathcal{M}_{\text{fed}}^{(k)}\right)(w_{\max}).$$

This completes the proof. □

F.6. Proof of Proposition 4.2

Proof. Fix $k \in [K]$. Define

$$\begin{aligned}
\bar{u}_k(w_{G_k}) := &-\log\det\left((A^{(k)})^\top \mathbf{M}\left((w_{G_k}, w_{\max,G_k^c})\right)^{-1} A^{(k)}\right) \\
&- c^{(k)} \cdot \sum_{i \in G_k} w_i - c^{(k)} \cdot \sum_{i \in G_k} (w_{\max,i} - w_i)_+ .
\end{aligned}$$

To see that $w_{\max}$ is a pure NE, it suffices to show that $w_{\max,G_k} = \arg\max \bar{u}_k(w_{G_k})$. Indeed, if $\widetilde{w}_{G_k} = \arg\max \bar{u}_k(w_{G_k})$ and $\widetilde{w}_{G_k} \neq w_{\max,G_k}$. Consider the following two cases.

Case 1: There exists $i \in G_k$ such that $\widetilde{w}_i < w_{\max,i}$.

Let $\widetilde{w}'_j = \begin{cases} w_{\max,i}, & \text{if } j = i \\ \widetilde{w}_j, & \text{otherwise} \end{cases}$. Then

$$\begin{aligned} \bar{u}_k(\widetilde{w}'_{G_k}) := & -\log\det\left((A^{(k)})^\top \mathbf{M}\left((\widetilde{w}'_{G_k}, w_{\max,G_k^c})\right)^{-1} A^{(k)}\right) \\ & - c^{(k)} \cdot \left(w_{\max,i} + \sum_{j\in G_k/\{i\}} \widetilde{w}_j\right) - c^{(k)} \cdot \sum_{j\in G_k/\{i\}} (w_{\max,j} - \widetilde{w}_j)_+ \\ > & -\log\det\left((A^{(k)})^\top \mathbf{M}\left((\widetilde{w}_{G_k}, w_{\max,G_k^c})\right)^{-1} A^{(k)}\right) \\ & - c^{(k)} \cdot \sum_{j\in G_k} \widetilde{w}_j - c^{(k)} \cdot \sum_{j\in G_k} (w_{\max,j} - \widetilde{w}_j)_+ \\ = & \, \bar{u}_k(\widetilde{w}_{G_k}) \end{aligned}$$

where the first step is due to $w_{\max,i} - \widetilde{w}'_i = 0$ and the second step comes from Lemma H.1 and $w_{\max,i} = \widetilde{w}_i + (w_{\max,i} - \widetilde{w}_i)_+$. This contradicts with $\widetilde{w}_{G_k} = \arg\max \bar{u}_k(w_{G_k})$.

Case 2: $\widetilde{w}_j \geq w_{\max,j},\ \forall j \in G_k$ **and there exists** $i \in G_k$ **such that** $\widetilde{w}_i > w_{\max,i}$**.**

Notice that in this case $\log\det \mathbf{M}\left((\widetilde{w}_{G_k}, w_{\max,G_k^c})\right) > \log\det \mathbf{M}\left(w_{\max}\right)$. Therefore there exists $j \in [K]$ such that $\left(u^{(j)} \circ \mathcal{M}^{(j)}_{\max}\right)\left((\widetilde{w}_{G_j}, w_{\max,G_j^c})\right) < v_*^{(j)}$, and it is obvious that such j's must include k. As a result,

$$\bar{u}_k(\widetilde{w}_{G_k}) = \left(u^{(k)} \circ \mathcal{M}^{(k)}_{\max}\right)\left((\widetilde{w}_{G_k}, w_{\max,G_k^c})\right) < v_*^{(k)} \leq \left(u^{(k)} \circ \mathcal{M}^{(k)}_{\max}\right)(w_{\max}) = \bar{u}_k(w_{\max}).$$

Contradiction! Therefore, we have shown that $(w_{\max,G_k})_{k\in[K]}$ is a pure Nash equilibrium. Since $w_{\max}$ is the solution of Eq. (5), Individual Rationality is satisfied. As a result, $w_{\max}$ is a strategic response of mechanism $\mathcal{M}_{\max}$.

Next, we display uniqueness. Suppose for the sake of contradiction that there exists a Nash equilibrium $(\widetilde{w}_{G_k})_{k\in[K]} \neq (w_{\max,G_k})_{k\in[K]}$. We follow the above line of arguments and consider the following two cases.

Case 1: There exists $k \in [K]$ **and** $i \in G_k$ **such that** $\widetilde{w}_i < w_{\max,i}$**.**

Let $\widetilde{w}'_j = \begin{cases} w_{\max,i}, & \text{if } j = i \\ \widetilde{w}_j, & \text{otherwise} \end{cases}$. Then

$$\begin{aligned} \left(u^{(k)} \circ \mathcal{M}^{(k)}_{\max}\right)\left((\widetilde{w}'_{G_k}, \widetilde{w}_{G_k^c})\right) := & -\log\det\left((A^{(k)})^\top \mathbf{M}\left((\widetilde{w}'_{G_k}, \widetilde{w}_{G_k^c})\right)^{-1} A^{(k)}\right) \\ & - c^{(k)} \cdot \left(w_{\max,i} + \sum_{j\in G_k/\{i\}} \widetilde{w}_j\right) - c^{(k)} \cdot \sum_{j\in G_k/\{i\}} (w_{\max,j} - \widetilde{w}_j)_+ \\ > & -\log\det\left((A^{(k)})^\top \mathbf{M}\left(\widetilde{w}\right)^{-1} A^{(k)}\right) \\ & - c^{(k)} \cdot \sum_{j\in G_k} \widetilde{w}_j - c^{(k)} \cdot \sum_{j\in G_k} (w_{\max,j} - \widetilde{w}_j)_+ \\ = & \left(u^{(k)} \circ \mathcal{M}^{(k)}_{\max}\right)(\widetilde{w}). \end{aligned}$$

This contradicts with $\widetilde{w}_{G_k} = \arg\max_{w_{G_k}} \left(u^{(k)} \circ \mathcal{M}^{(k)}_{\max}\right)\left((w_{G_k}, \widetilde{w}_{G_k^c})\right)$.

Case 2: $\widetilde{w}_j \geq w_{\max,j},\ \forall j \in [n]$ **and there exists** $k \in [K]$ **and** $i \in G_k$ **such that** $\widetilde{w}_i > w_{\max,i}$**.**

Since $\log\det \mathbf{M}\left(\widetilde{w}\right) > \log\det \mathbf{M}\left(w_{\max}\right)$, there exists $j \in [K]$ such that $\left(u^{(j)} \circ \mathcal{M}^{(j)}_{\max}\right)(\widetilde{w}) < v_*^{(j)}$. Obviously, there exists $i \in G_j$ such that $\widetilde{w}_i > w_{\max,i}$. As a result,

$$\left(u^{(j)} \circ \mathcal{M}^{(j)}_{\max}\right)(\widetilde{w}) < v_*^{(j)} \leq \left(u^{(j)} \circ \mathcal{M}^{(j)}_{\max}\right)(w_{\max}) \leq \left(u^{(j)} \circ \mathcal{M}^{(j)}_{\max}\right)\left((w_{\max,G_j}, \widetilde{w}_{G_j^c})\right).$$

This means

$$\widetilde{w}_{G_j} \notin \arg\max_{w_{G_j} \in \mathbb{R}_+^{|G_j|}} \left(u^{(j)} \circ \mathcal{M}^{(j)}_{\max}\right)\left((w_{G_j}, \widetilde{w}_{G_j^c})\right).$$

Contradiction! ☐

F.7. Proof of Corollary D.1

Proof. Suppose there exist $k \in [K]$ such that $\left(u^{(k)} \circ \mathcal{M}_{\max}^{(k)}\right)(w_{\max}) > v_*^{(k)}$, then the definition of $\mathcal{M}_{\max}$ yields $u^{(k)}(w_{\max}) = \left(u^{(k)} \circ \mathcal{M}_{\max}^{(k)}\right)(w_{\max}) > v_*^{(k)}$. By setting $w_i' = \begin{cases}(1+\epsilon)\cdot w_{\max,i}, & \text{if } i \in G_k \\ w_{\max,i}, & \text{if } i \notin G_k\end{cases}$ of sufficiently small $\epsilon > 0$, we have for any $l \in [K]$,

$$u^{(l)}(w') \geq v_*^{(l)}.$$

Thus the constraints in Eq. (5) is still satisfied, but $\log\det \mathbf{M}(w_{\max}') > \log\det \mathbf{M}(w_{\max})$. This contradicts to the fact that $w_{\max}$ is the optimizer in Eq. 5.

☐

F.8. Proof of Proposition D.2

Proof. Fix $k \neq k'$ and assume $\left(u^{(k)} \circ \mathcal{M}_{\max}^{(k)}\right)(\bar{w}) \geq \left(u^{(k')} \circ \mathcal{M}_{\max}^{(k')}\right)(\bar{w})$. By the data exchangeability and Corollary D.1 we have, modulo constants $2\log\|x\|_2$, that

$$\begin{aligned}\log\left(\|\bar{w}\|_1\right) - c^{(k)} \cdot \|\bar{w}_{G_k}\|_1 &= \left(u^{(k)} \circ \mathcal{M}_{\max}^{(k)}\right)(\bar{w}) \\ &= v_*^{(k)} \\ &= -\log c^{(k)} - 1.\end{aligned}$$

Therefore $c^{(k)} \leq c^{(k')}$ and we have

$$\|\bar{w}_{G_k}\|_1 = \frac{\log\left(\|\bar{w}\|_1\right) + \log c^{(k)} + 1}{c^{(k)}}.$$

Now we define

$$f(c) = \frac{\log\left(\|\bar{w}\|_1\right) + \log c + 1}{c}.$$

Notice that $f'(c) = -\frac{\log(\|\bar{w}\|_1)+\log c}{c^2} < 0$ for any $c \geq \min_{l\in[K]} c^{(l)}$, thus

$$\begin{aligned}\|\bar{w}_{G_k}\|_1 &= \frac{\log\left(\|\bar{w}\|_1\right) + \log c^{(k)} + 1}{c^{(k)}} \\ &\geq \frac{\log\left(\|\bar{w}\|_1\right) + \log c^{(k')} + 1}{c^{(k')}} \\ &= \|\bar{w}_{G_k'}\|_1.\end{aligned}$$

This confirms that $\|\bar{w}_{G_k}\|_1 \geq \|\bar{w}_{G_k'}\|_1$. ☐

F.9. Proof of Proposition D.5

Proof. Define $k_0 = \arg\min_{k\in[K]} c^{(k)}$ and

$$\theta_*^{(k)} := \max_{\pi\in\Delta([n])} -\log\det\left((A^{(k)})^\top \left(\sum_{i=1}^n \pi_i x_i x_i^\top\right)^{-1} A^{(k)}\right)$$

$$\theta^{(k)} := \max_{\pi\in\Delta(G_k)} \log\det\left((A^{(k)})^\top \left(\sum_{i\in G_k} \pi_i x_i x_i^\top\right) A^{(k)}\right).$$

We have

$$
\begin{aligned}
&\mathrm{SG}(w)\\
&=\sum_{k=1}^{K}\left(u^{(k)}\circ\mathcal{M}_{\max}^{(k)}\right)(w)\\
&=\sum_{k=1}^{K}\left(-\log\det\left((A^{(k)})^{\top}\mathbf{M}(w)^{-1}A^{(k)}\right)-c^{(k)}\cdot\sum_{i\in G_k}w_i-c^{(k)}\cdot\sum_{i\in G_k}(w_{\max,i}-w_i)_+\right)\\
&\le\sum_{k=1}^{K}\left(\theta_*^{(k)}+r_k\log\|w\|_1-c^{(k)}\cdot\sum_{i\in G_k}w_i-c^{(k)}\cdot\sum_{i\in G_k}(w_{\max,i}-w_i)_+\right)\\
&\le\sum_{k=1}^{K}\left(\theta_*^{(k)}+r_k\log\frac{\sum_{k=1}^{K}r_k}{c^{(k_0)}}-r_k-(c^{(k)}-c^{(k_0)})\cdot\|w_{\max,G_k}\|_1\right)
\end{aligned}
$$

where the maximizer in the last inequality is given by $\|w_{G_k}\|_1=\begin{cases}\|w_{\max,G_k}\|_1, & k\neq k_0\\ \|w_{\max,G_k}\|_1+\frac{\sum_{k=1}^{K}r_k}{c^{(k_0)}}-\|w_{\max}\|_1, & k=k_0\end{cases}$. Further, notice that

$$
\begin{aligned}
\mathrm{SG}(w_{\max})&=\sum_{k=1}^{K}\left(u^{(k)}\circ\mathcal{M}_{\max}^{(k)}\right)(w_{\max})\\
&=\sum_{k=1}^{K}v_*^{(k)}\\
&=\sum_{k=1}^{K}\left(\theta^{(k)}+r_k\log\frac{r_k}{c^{(k)}}-r_k\right)
\end{aligned}
$$

where the second inequality uses Corollary D.1.

It follows that

$$
\begin{aligned}
&\frac{\mathrm{SG}(w)}{\mathrm{SG}(w_{\max})}\\
&\le\frac{\sum_{k=1}^{K}\left(\theta_*^{(k)}-\theta^{(k)}\right)}{\sum_{k=1}^{K}\left(\theta^{(k)}+r_k\log\frac{r_k}{c^{(k)}}-r_k\right)}+\frac{\sum_{k=1}^{K}\left(r_k\log\frac{c^{(k)}\sum_{k=1}^{K}r_k}{r_kc^{(k_0)}}-(c^{(k)}-c^{(k_0)})\cdot\|w_{\max,G_k}\|_1\right)}{\sum_{k=1}^{K}\left(\theta^{(k)}+r_k\log\frac{r_k}{c^{(k)}}-r_k\right)}+1\\
&=\frac{\sum_{k=1}^{K}\Delta^{(k)}}{\sum_{k=1}^{K}\left(\theta^{(k)}+r_k\log\frac{r_k}{c^{(k)}}-r_k\right)}+\frac{\sum_{k=1}^{K}\left(r_k\log\frac{c^{(k)}\sum_{k=1}^{K}r_k}{r_kc^{(k_0)}}-(c^{(k)}-c^{(k_0)})\cdot\|w_{\max,G_k}\|_1\right)}{\sum_{k=1}^{K}\left(\theta^{(k)}+r_k\log\frac{r_k}{c^{(k)}}-r_k\right)}+1.
\end{aligned}
$$

□

G. Efficiency under heterogeneous costs

In Section 3, we investigated the efficiency of federated learning with homogeneous costs. However, the proof of Proposition 3.2 demonstrates that federated learning is not efficient when costs are heterogeneous. Therefore, in this section, we focus on mechanism designs to incentivize efficient allocation. Specifically, we consider the objective $w_{\text{eff}}=n_{\max}\cdot\pi^*$, where π^* represents an optimal design measure under the D-criterion, and $n_{\max}$ is defined as

$$
\begin{aligned}
n_{\max}&=\max_{n\in\mathbb{R}_+}\ n, \qquad (15)\\
&\text{s.t. } u^{(k)}(n\cdot\pi^*)\ge v_*^{(k)},\forall k\in[K].
\end{aligned}
$$

The objective w_{eff} aims to maximize the total number of data while preserving efficient allocation of experiments. However, the feasibility of the program defined by Eq. (15) is not guaranteed in general. We provide a result that establishes a condition under which w_{eff} is well-defined and lower bounded by $w^*_{G_k}$, where

$$w^*_{G_k} = \arg\max - \log\det\left((A^{(k)})^\top(\sum_{i\in G_k} w_i x_i x_i^\top)A^{(k)}\right)^{-1} - c^{(k)} \cdot \sum_{i\in G_k} w_i.$$

Assumption G.1 (Data compatibility)**.** We assume for any $k, k' \in [K]$,

$$u^{(k')}\left(\frac{\|w^*_{G_k}\|_1}{\|\pi^*_{G_k}\|_1}\cdot\pi^*\right) \geq v_*^{(k')}.$$

This assumption implies that if we scale up the D-optimal design according to $w^*_{G_k}$, the utility for any other agent k' is still no less than the maximum utility that agent k' can achieve if she opts out of the collaborative learning and trains a model using her own data. Therefore, π^* is compatible in the sense that no agent has an incentive to leave the collaborative learning program if they follow π^* and each agent k contribute at least $\|w^*_{G_k}\|_1$ data points. Under this condition, we can derive the following result:

Proposition G.2 (Feasibility and incentivized more contribution)**.** *Suppose Assumption G.1 holds. Then the problem in Eq. (15) is feasible. Furthermore, For all $k \in [K]$ we have $n_{\max}\cdot\sum_{i\in G_k}\pi^*_i \geq \sum_{i\in G_k} w^*_i$.*

Proof. Define $I_k = \left\{n \in \mathbb{R}_+ : u^{(k)}(n\cdot\pi^*) \geq v_*^{(k)}\right\}$. Notice $u^{(k)}$ is concave and

$$\begin{aligned}
&u^{(k)}\left(\frac{\|w^*_{G_k}\|_1}{\|\pi^*_{G_k}\|_1}\cdot\pi^*\right)\\
=&\ -\log\det\left((A^{(k)})^\top\mathbf{M}\left(\frac{\|w^*_{G_k}\|_1}{\|\pi^*_{G_k}\|_1}\cdot\pi^*\right)^{-1}A^{(k)}\right) - c^{(k)}\cdot\|w^*_{G_k}\|_1\\
\geq&\ -\log\det\left((A^{(k)})^\top\mathbf{M}\left(\left(w^*_{G_k}, \left(\frac{\|w^*_{G_k}\|_1}{\|\pi^*_{G_k}\|_1}\cdot\pi^*_j\right)_{j\notin G_k}\right)\right)^{-1}A^{(k)}\right) - c^{(k)}\cdot\|w^*_{G_k}\|_1\\
\geq&\ -\log\det\left((A^{(k)})^\top\left(\sum_{i\in G_k} w^*_i x_i x_i^\top\right)A^{(k)}\right)^{-1} - c^{(k)}\cdot\|w^*_{G_k}\|_1\\
=&\ v_*^{(k)},
\end{aligned}$$

where the second step comes from Applying Lemma H.1 and the fact that u_k is concave wrt w_{G_k}; the third step comes from Lemma H.2. As a result, I_k is a closed interval and $\frac{\|w^*_{G_k}\|_1}{\|\pi^*_{G_k}\|_1} \in I_k$ for any $k \in [K]$. We rewrite $I_k = [a_k, b_k]$ where $a_k \leq \frac{\|w^*_{G_k}\|_1}{\|\pi^*_{G_k}\|_1} \leq b$.

Assumption G.1 implies that $\frac{\|w^*_{G_k}\|_1}{\|\pi^*_{G_k}\|_1} \in I_{k'}$ for any $k' \in [K]$. Therefore, $\cap_{k\in[k]} I_k \neq \emptyset$ and $n_{\max} = \min_{k\in[k]} b_k$. This establishes feasibility and $n_{\max} \geq \frac{\|w^*_{G_k}\|_1}{\|\pi^*_{G_k}\|_1}$. □

G.1. Mechanism design for pure efficient allocation

We begin by considering pure efficient allocation, which best illustrates the nature of the problem. In this subsection, we omit the cost functions and assume $c^{(1)} = \cdots = c^{(K)} = 0$. The goal in this section is to design mechanisms $\mathcal{M}^{(k)}$ such that all Nash equilibrium wrt the utility functions $\left((u^{(k)}\circ\mathcal{M}^{(k)})\right)_{k\in[K]}$ takes the form of $(\lambda\cdot\pi^*_{G_k})_{k\in[K]}$ where $\lambda > 0$, i.e. proportional to the optimal design measure.

We define the following mechanism based on scaling the design by a constant $\eta_k \leq 1$:

$$\mathcal{M}^{(k)}_{\text{pure}}(w) = \eta_k w \text{ where } \eta_k^{-1} = \exp\left(\frac{d}{r_k}\cdot\left(\frac{(\sum_{i\notin G_k}\pi^*_i)(\sum_{i\in G_k} w_i)}{\sum_{i\notin G_k} w_i} - \sum_{i\in G_k}\pi^*_i\right)_+\right). \tag{16}$$

The intuition behind η_k is to introduce competition among agents, penalizing those who contribute proportionally less data than others. In fact, any strategic agent k under this mechanism is incentivized to contribute no less than $\frac{\sum_{i\in G_k}\pi_i^*}{\sum_{i\notin G_k}\pi_i^*}$ times the total amount of data collected by the other agents. Therefore, the mechanism in Eq. (16) ensures that the marginal probability of the aggregated design measure on each agent k, i.e., $(\sum_{i\in G_k} w_i)/(\sum_{i=1}^n w_i)$, aligns with the marginal probability of the optimal design measure, i.e. $(\sum_{i\in G_k}\pi_i^*)/(\sum_{i=1}^n \pi_i^*)$. By leveraging the properties of D-optimal design, we can demonstrate that it further ensures alignment between w and π^* for each coordinate. Besides subsampling, this mechanism can also be efficiently implemented by letting

$$\hat{\theta}^{(k)} = \hat{\theta} + \zeta^{(k)}, \text{ where } \zeta^{(k)} \sim \mathcal{N}\left(0, (\eta_k^{-1} - 1)\cdot \mathbf{M}(w)^{-1}\right).$$

It follows that agent k's utility is given by

$$\left(u^{(k)} \circ \mathcal{M}_{\text{pure}}^{(k)}\right)(w) = -\log\det\left((A^{(k)})^\top \mathbf{M}(w)^{-1} A^{(k)}\right) - d\cdot\left(\frac{(\sum_{i\notin G_k}\pi_i^*)(\sum_{i\in G_k} w_i)}{\sum_{i\notin G_k} w_i} - \sum_{i\in G_k}\pi_i^*\right)_+. \tag{17}$$

Proposition G.3 (Pure Efficient allocation)**.** *For any $\lambda > 0$, $(\lambda\cdot\pi_{G_k}^*)_{k\in[K]}$ is a pure Nash equilibrium of the tuple of utility functions $\left(\left(u^{(k)}\circ\mathcal{M}_{\text{pure}}^{(k)}\right)\right)_{k\in[K]}$. Furthermore, any pure Nash equilibrium takes the form of $(\lambda\cdot\pi_{G_k}^*)_{k\in[K]}$.*

Proof. Fix $k\in\mathbb{Z}_+, \lambda\in\mathbb{R}^+$ and $\bar{w}_i = \lambda\pi_i^*$ for all $i\notin G_k$. For the sake of brevity, we define the following function of w_{G_k}

$$\bar{u}_k(w_{G_k}) = -\log\det\left((A^{(k)})^\top \mathbf{M}((w_{G_k}, (\bar{w}_j)_{j\notin G_k}))^{-1} A^{(k)}\right) - d\cdot\left(\frac{(\sum_{i\notin G_k}\pi_i^*)(\sum_{i\in G_k} w_i)}{\sum_{i\notin G_k}\bar{w}_i} - \sum_{i\in G_k}\pi_i^*\right)_+$$

where $(w_{G_k}, (\bar{w}_j)_{j\notin G_k})$ denotes the concatenation of w_{G_k} and $(\bar{w}_j)_{j\notin G_k}$ such that

$$(w_{G_k}, (\bar{w}_j)_{j\notin G_k})_i = \begin{cases} w_i, & \text{if } i\in G_k \\ \bar{w}_i, & \text{otherwise.}\end{cases}$$

It suffices to show that $\bar{w}_{G_k} := \lambda\pi_{G_k}^*$ is the unique maximizer of $\bar{u}_k(w_{G_k})$.

For any Δw_{G_k}, Lemma H.1 gives

$$\begin{aligned}
& d\bar{u}_k\left(\bar{w}_{G_k}, \Delta w_{G_k}\right) \\
&= \left\langle \sum_{i\in G_k}\Delta w_i x_i x_i^\top, \mathbf{M}(\bar{w})^{-1}\right\rangle - \frac{d(\sum_{i\notin G_k}\pi_i^*)}{\sum_{i\notin G_k}\bar{w}_i}\left(\sum_{i\in G_k}\Delta w_i\right)_+ \\
&= \left\langle \sum_{i\in G_k}\Delta w_i x_i x_i^\top, \mathbf{M}(\pi^*)^{-1}\right\rangle \cdot \frac{\sum_{i\notin G_k}\pi_i^*}{\sum_{i\notin G_k}\bar{w}_i} - \frac{d(\sum_{i\notin G_k}\pi_i^*)}{\sum_{i\notin G_k} w_i}\left(\sum_{i\in G_k}\Delta w_i\right)_+ \\
&\le \sum_{i\in G_k}\Delta w_i\left(\langle x_i x_i^\top, \mathbf{M}(\pi^*)^{-1}\rangle - d\right)\cdot\frac{\sum_{i\notin G_k}\pi_i^*}{\sum_{i\notin G_k}\bar{w}_i} \\
&\le 0,
\end{aligned}$$

where the second step is due to $\mathbf{M}(\bar{w})^{-1} = \mathbf{M}(\pi^*)^{-1}\cdot\frac{1}{\sum_{i=1}^n \bar{w}_i} = \mathbf{M}(\pi^*)^{-1}\cdot\frac{\sum_{i\notin G_k}\pi_i^*}{\sum_{i\notin G_k}\bar{w}_i}$; the last step uses $\langle x_i x_i^\top, \mathbf{M}(\pi^*)^{-1}\rangle \begin{cases} = d, & i\in\text{supp}(\pi^*) \\ \le d, & i\notin\text{supp}(\pi^*)\end{cases}$ by Theorem H.5. By concavity of $\bar{u}_k$, $\bar{w}_{G_k} := \lambda\pi_{G_k}^*$ is the unique maximizer of $\bar{u}_k(w_{G_k})$.

Therefore for any $\lambda > 0$, $(w_{G_k} = \lambda\cdot\pi_{G_k}^*)_{k\in[K]}$ is a Nash Equilibrium.

In what follows, we show that any pure Nash Equilibrium takes the form of $(w_{G_k} = \lambda\cdot\pi_{G_k}^*)_{k\in[K]}, \lambda\in\mathbb{R}_+$.

Suppose for the sake of contradiction a Nash equilibrium $(\widetilde{w}_{G_k})_{k\in[K]}$ not in the form of $(\lambda\cdot\pi_{G_k}^*)_{k\in[K]}$.

Fix $k \in [K]$. Consider the following utility as a function of w_{G_k}

$$\bar{u}_k(w_{G_k}) = -\log\det\left((A^{(k)})^\top \mathbf{M}\left((w_{G_k}, (\widetilde{w}_{G_k^c})\right)^{-1} A^{(k)}\right) - d\cdot\left(\frac{(\sum_{i\notin G_k}\pi_i^*)(\sum_{i\in G_k} w_i)}{\sum_{i\notin G_k}\widetilde{w}_i} - \sum_{i\in G_k}\pi_i^*\right)_+$$

where $(w_{G_k}, \widetilde{w}_{G_k^c})$ denotes the concatenation of w_{G_k} and $\widetilde{w}_{G_k^c}$ such that

$$(w_{G_k}, \widetilde{w}_{G_k^c})_i = \begin{cases} w_i, & \text{if } i \in G_k \\ \widetilde{w}_i, & \text{otherwise.}\end{cases}$$

We assert that

$$\frac{(\sum_{i\notin G_k}\pi_i^*)(\sum_{i\in G_k}\widetilde{w}_i)}{\sum_{i\notin G_k}\widetilde{w}_i} - \sum_{i\in G_k}\pi_i^* = 0,\ \forall k\in[K]. \tag{18}$$

Indeed, it is obvious that $\frac{(\sum_{i\notin G_k}\pi_i^*)(\sum_{i\in G_k}\widetilde{w}_i)}{\sum_{i\notin G_k}\widetilde{w}_i} - \sum_{i\in G_k}\pi_i^* \geq 0$ for all $k\in[K]$. (If there exists $k\in[K]$ such that $\frac{(\sum_{i\notin G_k}\pi_i^*)(\sum_{i\in G_k}\widetilde{w}_i)}{\sum_{i\notin G_k}\widetilde{w}_i} - \sum_{i\in G_k}\pi_i^* < 0$, then define $\hat{w}_{G_k} := (1+\epsilon)\widetilde{w}_{G_k}$. By applying Lemma H.1,

$$\begin{aligned}\bar{u}_k(\hat{w}_{G_k}) &= -\log\det\left((A^{(k)})^\top \mathbf{M}\left(((1+\epsilon)\widetilde{w}_{G_k}, (\widetilde{w}_{G_k^c})\right)^{-1} A^{(k)}\right)\\ &> -\log\det\left((A^{(k)})^\top \mathbf{M}\left((\widetilde{w}_{G_k}, (\widetilde{w}_{G_k^c})\right)^{-1} A^{(k)}\right)\\ &= \bar{u}_k(\widetilde{w}_{G_k})\end{aligned}$$

holds for sufficiently small $\epsilon > 0$. Contradiction!) Recall that $\pi^* \in \Delta[n]$. Examining $\frac{(\sum_{i\notin G_k}\pi_i^*)(\sum_{i\in G_k}\widetilde{w}_i)}{\sum_{i\notin G_k}\widetilde{w}_i} - \sum_{i\in G_k}\pi_i^* \geq 0$ for all $k\in[K]$ yields Eq. (18).

By Theorem H.5, there must exist $l\in[n]$ such that $\langle x_l x_l^\top, \mathbf{M}(\widetilde{w}/\|\widetilde{w}\|_1)^{-1}\rangle > d$. Suppose $l \in G_k$, then we have

$$\begin{aligned} d\bar{u}_k\left(\widetilde{w}_{G_k}, e_l\right) &\geq \left\langle x_l x_l^\top, \mathbf{M}(\widetilde{w})^{-1}\right\rangle - \frac{d(\sum_{i\notin G_k}\pi_i^*)}{\sum_{i\notin G_k}\widetilde{w}_i}\\ &= \left\langle x_l x_l^\top, \mathbf{M}(\widetilde{w}/\|\widetilde{w}\|_1)^{-1}\right\rangle\cdot\frac{\sum_{i\notin G_k}\pi_i^*}{\sum_{i\notin G_k}\widetilde{w}_i} - \frac{d(\sum_{i\notin G_k}\pi_i^*)}{\sum_{i\notin G_k}\widetilde{w}_i}\\ &> 0\end{aligned}$$

where the first step applies Lemma H.1; the second step uses $\mathbf{M}(\widetilde{w})^{-1} = \mathbf{M}(\widetilde{w}/\|\widetilde{w}\|_1)^{-1}\cdot\frac{1}{\sum_{i=1}^n\widetilde{w}_i} = \mathbf{M}(\widetilde{w}/\|\widetilde{w}\|_1)^{-1}\cdot\frac{\sum_{i\notin G_k}\pi_i^*}{\sum_{i\notin G_k}\widetilde{w}_i}$ due to Eq. (18). It follows that letting $\widetilde{w}' = \epsilon\cdot e_l + \widetilde{w}$ would increase $\bar{u}_k$ for sufficiently small $\epsilon > 0$. Contradiction!

□

G.2. Mechanism design for efficient allocation under different cost parameters

We design the following feasible mechanism to achieve efficient allocation in Eq. (15) when the cost parameters are not the same. Define $\mathcal{M}_{\text{eff}}(k)$ as follows

$$\mathcal{M}_{\text{eff}}^{(k)}(w) = \rho_k w,\ \text{where} \tag{19}$$

$$\rho_k^{-1} = \exp\left(\frac{c^{(k)}}{r_k}\cdot\sum_{i\in G_k}(n_{\max}\cdot\pi_i^* - w_i)_+ + \sum_{i\in G_k}\left(\frac{(\sum_{i\notin G_k}\pi_i^*)w_i}{(\sum_{i\notin G_k}\bar{w}_i)\pi_i^*} - 1\right)_+\cdot\mathbb{1}\left(w_i \geq n_{\max}\cdot\pi_i\right)\right)$$

In ρ_k, the first term is the same as the regularization term in $\mathcal{M}_{\max}$ and functions as incentivizing more data contribution; the second term is similar to the regularization term in $\mathcal{M}_{\text{pure}}$ and serves as incentivizing alignment with optimal design measure. Therefore, although the objective w_{eff} may take complex forms, these two simple terms together create the incentive for each agent to follow the optimal design measure while increasing the total information. Besides subsampling, this mechanism can also be efficiently implemented by letting

$$\hat{\theta}^{(k)} = \hat{\theta} + \zeta^{(k)},\ \text{where } \zeta^{(k)} \sim \mathcal{N}\left(0, (\rho_k^{-1} - 1)\cdot\mathbf{M}(w)^{-1}\right).$$

The k-th agent's utility is then given by

$$\left(u^{(k)} \circ \mathcal{M}_{\text{eff}}^{(k)}\right)(w) = -\log\det\left((A^{(k)})^\top \mathbf{M}(w)^{-1} A^{(k)}\right) - c^{(k)} \cdot \sum_{i \in G_k} (n_{\max} \cdot \pi_i^* - w_i)_+ \\ - r_k \cdot \sum_{i \in G_k} \left(\frac{(\sum_{i \notin G_k} \pi_i^*) w_i}{(\sum_{i \notin G_k} w_i) \pi_i^*} - 1\right)_+ \cdot \mathbb{1}\left(w_i \geq n_{\max} \cdot \pi_i\right) - c^{(k)} \left(\sum_{i \in G_k} w_i\right). \tag{20}$$

Proposition G.4 (Data maximization and efficient allocation)**.** *The efficient allocation design $(w_{\text{eff},G_k})_{k \in [K]}$ is the unique strategic response to the mechanism $\mathcal{M}_{\text{eff}}$.*

Proof. Fix k and $\bar{w}_i = n_{\max} \pi_i^*$ for all $i \notin G_k$. Define $N = \sum_{i \notin G_k} \bar{w}_i x_i x_i^\top$. Define the following function of w_{G_k}

$$\begin{aligned} &\bar{u}_k(w_{G_k}) \\ &= -\log\det\left((A^{(k)})^\top \mathbf{M}((w_{G_k}, \bar{w}_{G_k^c}))^{-1} A^{(k)}\right) - c^{(k)} \cdot \left(\sum_{i \in G_k} w_i\right) - c^{(k)} \cdot \sum_{i \in G_k} (n_{\max} \cdot \pi_i^* - w_i)_+ \\ &\quad - r_k \cdot \sum_{i \in G_k} \left(\frac{(\sum_{i \notin G_k} \pi_i^*) w_i}{(\sum_{i \notin G_k} \bar{w}_i) \pi_i^*} - 1\right)_+ \cdot \mathbb{1}\left(w_i \geq n_{\max} \cdot \pi_i\right) \end{aligned}$$

where $(w_{G_k}, \bar{w}_{G_k^c})$ denotes the concatenation of w_{G_k} and $\bar{w}_{G_k^c}$ such that

$$(w_{G_k}, \bar{w}_{G_k^c})_i = \begin{cases} w_i, & \text{if } i \in G_k \\ \bar{w}_i, & \text{otherwise.} \end{cases}$$

To show that w_{eff} is a pure Nash equilibrium, it suffices to show that $\bar{w}_{G_k} := n_{\max} \pi_{G_k}^*$ is the unique maximizer of $\bar{u}_k(w_{G_k})$.

Indeed, consider any $w_{G_k} \neq \bar{w}_{G_k}$.

Case 1: there exists $i \in G_k$ such that $w_i < n_{\max} \cdot \pi_i^*$.

For all $j \in G_k$ let $\widetilde{w}_j = \begin{cases} w_j, & \text{if } w_j \geq n_{\max} \cdot \pi_j^* \\ n_{\max} \cdot \pi_j^*, & \text{if } w_j < n_{\max} \cdot \pi_j^* \end{cases}$. From Lemma H.1,

$$\begin{aligned} &\bar{u}_k(\widetilde{w}_{G_k}) \\ &= -\log\det\left((A^{(k)})^\top \mathbf{M}((\widetilde{w}_{G_k}, \bar{w}_{G_k^c}))^{-1} A^{(k)}\right) - c^{(k)} \cdot \left(\sum_{i \in G_k} \widetilde{w}_i\right) - c^{(k)} \cdot \sum_{i \in G_k} (n_{\max} \cdot \pi_i^* - \widetilde{w}_i)_+ \\ &\quad - r_k \cdot \sum_{i \in G_k} \left(\frac{(\sum_{i \notin G_k} \pi_i^*) \widetilde{w}_i}{(\sum_{i \notin G_k} \bar{w}_i) \pi_i^*} - 1\right)_+ \cdot \mathbb{1}\left(\widetilde{w}_i \geq n_{\max} \cdot \pi_i\right) \\ &> -\log\det\left((A^{(k)})^\top \mathbf{M}((w_{G_k}, \bar{w}_{G_k^c}))^{-1} A^{(k)}\right) - c^{(k)} \cdot \left(\sum_{i \in G_k} w_i\right) - c^{(k)} \cdot \sum_{i \in G_k} (n_{\max} \cdot \pi_i^* - w_i)_+ \\ &\quad - r_k \cdot \sum_{i \in G_k} \left(\frac{(\sum_{i \notin G_k} \pi_i^*) w_i}{(\sum_{i \notin G_k} \bar{w}_i) \pi_i^*} - 1\right)_+ \cdot \mathbb{1}\left(w_i \geq n_{\max} \cdot \pi_i\right) \\ &= \bar{u}_k(w_{G_k}). \end{aligned}$$

This yields a contradiction.

Case 2: $w_i \geq n_{\max} \cdot \pi_i^*$ for all $i \in G_k$.

Define $\epsilon_j = \frac{w_j}{n_{\max}\cdot\pi_j^*} - 1$ for all $j \in G_k$, then $\min_{i\in G_k}\epsilon_i \geq 0$ and $\max_{i\in G_k}\epsilon_i > 0$. We have

$$\begin{aligned}
&\bar{u}_k(w_{G_k}) \\
&= -\log\det\left((A^{(k)})^\top\left(\sum_{i\in G_k}(1+\epsilon_i)\bar{w}_i x_i x_i^\top + N\right)^{-1} A^{(k)}\right) - c^{(k)}\cdot\sum_{i\in G_k}(n_{\max}\cdot\pi_i^* - w_i)_+ \\
&\quad - c^{(k)}\cdot\left(\sum_{i\in G_k} w_i\right) - r_k\cdot\sum_{i\in G_k}\epsilon_i \\
&\leq -\log\det\left(\left(1+\max_{i\in G_k}\epsilon_i\right)^{-1}(A^{(k)})^\top\left(\sum_{i\in G_k}\bar{w}_i x_i x_i^\top + N\right)^{-1} A^{(k)}\right) \\
&\quad - c^{(k)}\cdot\sum_{i\in G_k}(n_{\max}\cdot\pi_i^* - \bar{w}_i)_+ - c^{(k)}\cdot\left(\sum_{i\in G_k}\bar{w}_i\right) - r_k\cdot\sum_{i\in G_k}\epsilon_i \\
&= \bar{u}_k(\bar{w}_{G_k}) + r_k\cdot\log\left(1+\max_{i\in G_k}\epsilon_i\right) - r_k\cdot\sum_{i\in G_k}\epsilon_i \\
&< \bar{u}_k(\bar{w}_{G_k}).
\end{aligned}$$

It follows that $\bar{u}_k(w_{G_k}) < \bar{u}_k(\bar{w}_{G_k})$, also a contradiction.

Combining the above two cases confirms that $\left(n_{\max}\cdot\pi^*_{G_k}\right)_{k\in[K]}$ is a pure NE. By definition of w_{eff} in Eq. (15), Individual rationality is satisfied. Therefore w_{eff} is a strategic response of $\mathcal{M}_{\text{eff}}$.

In what follows, we show that $n_{\max}\pi^*$ is the unique pure Nash equilibrium. Consider any pure Nash equilibrium $(\widetilde{w}_{G_k})_{k\in[K]}$, we will show in the following three steps that it must be equal to $(n_{\max}\cdot\pi^*_{G_k})_{k\in[K]}$.

For any $k\in[K]$ define the following utility as a function of w_{G_k}

$$\begin{aligned}
&\bar{u}_k(w_{G_k}) \\
&= -\log\det\left((A^{(k)})^\top\mathbf{M}((w_{G_k},\widetilde{w}_{G_k^c}))^{-1}A^{(k)}\right) - c^{(k)}\cdot\left(\sum_{i\in G_k} w_i\right) - c^{(k)}\cdot\sum_{i\in G_k}(n_{\max}\cdot\pi_i^* - w_i)_+ \\
&\quad - r_k\cdot\sum_{i\in G_k}\left(\frac{(\sum_{i\notin G_k}\pi_i^*)w_i}{(\sum_{i\notin G_k}\widetilde{w}_i)\pi_i^*} - 1\right)_+ \cdot\mathbb{1}\left(w_i \geq n_{\max}\cdot\pi_i\right)
\end{aligned}$$

where $(w_{G_k},\widetilde{w}_{G_k^c})$ denotes the concatenation of w_{G_k} and $\widetilde{w}_{G_k^c}$ such that

$$(w_{G_k},\widetilde{w}_{G_k^c})_i = \begin{cases} w_i, & \text{if } i\in G_k \\ \widetilde{w}_i, & \text{otherwise.}\end{cases}$$

It follows that $\bar{u}_k(\widetilde{w}_{G_k}) = \max_{w_{G_k}}\bar{u}_k(w_{G_k}),\ \forall k\in[K]$.

Step 1. We first show that $\widetilde{w}_i \geq n_{\max}\pi_i^*$ for any $i\in[n]$.

Indeed, if there exists $k\in[K]$ and $i\in G_k$ such that $\widetilde{w}_i < n_{\max}\pi_i^*$. Let $\hat{w}_j = \begin{cases}\widetilde{w}_j, & \text{if } \widetilde{w}_j \geq n_{\max}\cdot\pi_j^* \\ n_{\max}\cdot\pi_j^*, & \text{if } \widetilde{w}_j < n_{\max}\cdot\pi_j^*\end{cases}$. From Lemma H.1,

$$\begin{aligned}
&\bar{u}_k(\hat{w}_{G_k}) \\
&= -\log\det\left((A^{(k)})^\top\mathbf{M}((\hat{w}_{G_k},\widetilde{w}_{G_k^c}))^{-1}A^{(k)}\right) - c^{(k)}\cdot\left(\sum_{i\in G_k}\hat{w}_i\right) - c^{(k)}\cdot\sum_{i\in G_k}(n_{\max}\cdot\pi_i^* - \hat{w}_i)_+ \\
&\quad - r_k\cdot\sum_{i\in G_k}\left(\frac{(\sum_{i\notin G_k}\pi_i^*)\hat{w}_i}{(\sum_{i\notin G_k}\widetilde{w}_i)\pi_i^*} - 1\right)_+ \cdot\mathbb{1}\left(\hat{w}_i \geq n_{\max}\cdot\pi_i\right) \\
&> \bar{u}_k(\widetilde{w}_{G_k}).
\end{aligned}$$

This contradicts with the fact that $\widetilde{w}_{G_k}$ is the maximizer of $\bar{u}_k$.

Step 2. We show that there exists $\lambda \geq n_{\max}$ such that $\widetilde{w} = \lambda \cdot \pi^*$.

Suppose $\exists\ k \in [K]$ and $i \in G_k$ such that $\frac{(\sum_{j\notin G_k} \pi_i^*)\widetilde{w}_i}{\sum_{j\notin G_k} \widetilde{w}_i} > \pi_i^*$, define $\epsilon_i = \frac{(\sum_{j\notin G_k} \pi_j^*)\widetilde{w}_i}{(\sum_{j\notin G_k} \widetilde{w}_j)\pi_i^*} - 1 > 0$. Let $\hat{w}_j = \begin{cases} \widetilde{w}_j, & \text{if } j \neq i \\ \frac{(\sum_{j\notin G_k} \widetilde{w}_j)\pi_i^*}{\sum_{j\notin G_k} \pi_j^*}, & \text{if } j = i \end{cases}$. We have

$$
\begin{aligned}
&\bar{u}_k(\widetilde{w}_{G_k}) \\
&= -\log\det\left((A^{(k)})^\top \left((1+\epsilon_i)\hat{w}_i x_i x_i^\top + \sum_{j\neq i} \widetilde{w}_j x_j x_j^\top\right)^{-1} A^{(k)}\right) - c^{(k)} \cdot \left(\sum_{j\in G_k} \widetilde{w}_j\right) \\
&\quad - c^{(k)} \cdot \sum_{j\in G_k} \left(n_{\max}\cdot\pi_j^* - \widetilde{w}_j\right)_+ - r_k \cdot \sum_{j\in G_k} \left(\frac{(\sum_{l\notin G_k} \pi_l^*)\widetilde{w}_j}{(\sum_{l\notin G_k} \widetilde{w}_l)\pi_j^*} - 1\right)_+ \cdot \mathbb{1}\left(\hat{w}_j \geq n_{\max}\cdot\pi_j\right) \\
&\leq -\log\det\left((1+\epsilon_i)^{-1}(A^{(k)})^\top \left(\sum_{j\in G_k} \hat{w}_j x_j x_j^\top\right)^{-1} A^{(k)}\right) - r_k\cdot\epsilon_i - c^{(k)} \cdot \left(\sum_{j\in G_k} \hat{w}_j\right) \\
&\quad - c^{(k)} \cdot \sum_{j\in G_k} \left(n_{\max}\cdot\pi_j^* - \hat{w}_j\right)_+ - r_k \cdot \sum_{j\in G_k, j\neq i} \left(\frac{(\sum_{l\notin G_k} \pi_l^*)\hat{w}_j}{(\sum_{l\notin G_k} \widetilde{w}_l)\pi_j^*} - 1\right)_+ \cdot \mathbb{1}\left(\hat{w}_j \geq n_{\max}\cdot\pi_j\right) \\
&= \bar{u}_k(\hat{w}_{G_k}) + r_k \cdot \log(1+\epsilon_i) - r_k\cdot\epsilon_i \\
&< \bar{u}_k(\hat{w}_{G_k}).
\end{aligned}
$$

This contradicts with the fact that $\widetilde{w}_{G_k}$ is the maximizer of $\bar{u}_k$.

As a result, $\frac{(\sum_{i\notin G_k} \pi_i^*)\widetilde{w}_i}{\sum_{i\notin G_k} \widetilde{w}_i} \geq \pi_i^*$ holds for any $i \in [n]$. Examining this inequality for all $i \in [n]$ yields that $\widetilde{w} = \lambda\cdot\pi^*$ for some λ which must be greater than or equal to $n_{\max}$.

Step 3. We show that $\lambda = n_{\max}$.

From the definition of $n_{\max}$, for any $\lambda > n_{\max}$ there exists $k \in [K]$ such that

$$
\begin{aligned}
\left(u^{(k)} \circ \mathcal{M}_{\text{eff}}^{(k)}\right)(\lambda\cdot\pi^*) &< v_*^{(k)} \\
&\leq \left(u^{(k)} \circ \mathcal{M}_{\text{eff}}^{(k)}\right)(n_{\max}\cdot\pi^*) \\
&\leq \left(u^{(k)} \circ \mathcal{M}_{\text{eff}}^{(k)}\right)\left((n_{\max}\cdot\pi_{G_k}^*, \lambda\cdot\pi_{G_k^c}^*)\right).
\end{aligned}
$$

This means that $\lambda\cdot\pi_{G_k}^*$ is not a NE, yielding a contradiction. It follows that $\left(n_{\max}\cdot\pi_{G_k}^*\right)_{k\in[K]}$ is the unique pure NE.

□

H. Support lemmata

Lemma H.1. *We abbreviate $A^{(k)}$ and $A^{(k)}(A^{(k)})^\top$ as A and P respectively. Suppose $\mathbf{M}(w) \succ 0$. For any $l \in G_k$, we have*

$$\frac{\partial - \log\det\left(A^\top \mathbf{M}(w)^{-1} A\right)}{\partial w_l} = \left\langle x_l x_l^\top, \mathbf{M}(w)^{-1} \right\rangle > 0.$$

Furthermore, fixing $w_{G_k^c}$, the function $\log\det\left(A^\top \mathbf{M}(w)^{-1} A\right)$ is concave in w_{G_k}.

Proof. Indeed, for $l \in G_k$

$$\begin{aligned}
&\frac{\partial - \log\det\left(A^\top \mathbf{M}(w)^{-1} A\right)}{\partial w_l} \\
&= -\left\langle \left(A^\top \mathbf{M}(w)^{-1} A\right)^{-1}, \frac{\partial A^\top \mathbf{M}(w)^{-1} A}{\partial w_l} \right\rangle \\
&= -\left\langle \left(A^\top \mathbf{M}(w)^{-1} A\right)^{-1}, A^\top \frac{\partial \mathbf{M}(w)^{-1}}{\partial w_l} A \right\rangle \\
&= -\left\langle \left(A^\top \mathbf{M}(w)^{-1} A\right)^{-1}, A^\top \left(\sum_{i,j \in [d]} -\mathbf{M}(w)^{-1} e_i e_j^\top \mathbf{M}(w)^{-1} (x_l x_l^\top)_{i,j} \right) A \right\rangle \\
&= -\left\langle \left(A^\top \mathbf{M}(w)^{-1} A\right)^{-1}, A^\top \left(-\mathbf{M}(w)^{-1} x_l x_l^\top \mathbf{M}(w)^{-1}\right) A \right\rangle \\
&= \left\langle \left(A^\top \mathbf{M}(w)^{-1} A\right)^{-1}, A^\top \mathbf{M}(w)^{-1} A A^\top x_l x_l^\top \mathbf{M}(w)^{-1} A \right\rangle \\
&= \left\langle I, A^\top x_l x_l^\top \mathbf{M}(w)^{-1} A \right\rangle \\
&= \left\langle x_l x_l^\top A A^\top, \mathbf{M}(w)^{-1} \right\rangle \\
&= \left\langle x_l x_l^\top, \mathbf{M}(w)^{-1} \right\rangle,
\end{aligned}$$

where the first step uses the fact that $\frac{\partial}{\partial Y_{i,j}} \log\det Y = [Y^{-1}]_{j,i}$; the third step uses the fact that $\frac{\partial}{\partial Y_{i,j}} Y^{-1} = -Y^{-1} e_i e_j^\top Y^{-1}$; the fourth step comes from

$$\sum_{i,j \in [d]} e_i e_j^\top (x_l x_l^\top)_{i,j} = \sum_{i,j \in [d]} e_i e_i^\top (x_l x_l^\top) e_j e_j^\top = x_l x_l^\top;$$

the fifth and final step use the fact that $AA^\top x_l x_l^\top = x_l x_l^\top$ since $AA^\top = P$ is the projection matrix on $\{x_i\}_{i \in G_k}$.

This establishes the first statement. To show concavity, notice that

$$\begin{aligned}
[-\mathrm{Hess}_f(w)]_{i,j} &= -\frac{\partial \left\langle x_i x_i^\top, \mathbf{M}(w)^{-1} \right\rangle}{\partial w_j} \\
&= \left\langle x_i x_i^\top, \mathbf{M}(w)^{-1} x_j x_j^\top \mathbf{M}(w)^{-1} \right\rangle \\
&= \left(x_i^\top \mathbf{M}(w)^{-1} x_j\right)^2 \ \forall i, j \in G_k.
\end{aligned}$$

Therefore $-\mathrm{Hess}_f(w)$ is Hadamard product of the positive semi-definite matrix $\left(x_i^\top \mathbf{M}(w)^{-1} x_j\right)_{i,j \in G_k}$ and itself. It follows from Schur product theorem that the negative Hessian matrix $-\mathrm{Hess}_f$ is symmetric positive semidefinite everywhere in the domain, which establishes the concavity. □

Lemma H.2. *For any w such that $\sum_{i=1}^n w_i x_i x_i^\top$ is non-singular,*

$$\log\det\left((A^{(k)})^\top \left(\sum_{i \in G_k} w_i x_i x_i^\top\right) A^{(k)}\right) \le \log\det\left((A^{(k)})^\top \left(\sum_{i=1}^n w_i x_i x_i^\top\right)^{-1} A^{(k)}\right)^{-1}.$$

Further, if equality holds, then for any w'_{G_k} we have

$$\left(A^{(k)})^\top (\sum_{i \in G_k} w'_i x_i x_i^\top) A^{(k)}\right)^{-1} = (A^{(k)})^\top \left(\sum_{i \in G_k} w'_i x_i x_i^\top + \sum_{i \notin G_k} w_i x_i x_i^\top\right)^{-1} A^{(k)}.$$

Proof. Fix $k \in [K]$. We use shorthand notations $M_0 = \sum_{i \in G_k} w_i x_i x_i^\top, M = \sum_{i=1}^n w_i x_i x_i^\top, A = A^{(k)}$. Let $M_0 = U \mathrm{diag}(\lambda_1, \ldots, \lambda_{r_k}, 0, \ldots, 0) U^\top$ and $A = U \Lambda V^\top$ denote the Singular Value Decomposition (SVD) decomposition of M_0 and A respectively, where $U \in \mathbb{R}^{d \times d}$ and $V \in \mathbb{R}^{r_k \times r_k}$ are real orthogonal matrices. We write $U(M - M_0)U^\top = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$ where $A_{11} \in \mathbb{R}^{r_k \times r_k}$.

Define $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_{r_k})$. Notice that

$$
\begin{aligned}
A^\top M_0 A &= V \Lambda^\top U^\top U \mathrm{diag}(\lambda_1, \ldots, \lambda_{r_k}, 0, \ldots, 0) U^\top U \Lambda V^\top \\
&= V \begin{pmatrix} I_{r_k \times r_k} & 0_{r_k \times (d - r_k)} \end{pmatrix} \mathrm{diag}(\lambda_1, \ldots, \lambda_{r_k}, 0, \ldots, 0) \begin{pmatrix} I_{r_k \times r_k} \\ 0_{r_k \times (d - r_k)} \end{pmatrix} V^\top \\
&= V D V^\top .
\end{aligned}
$$

We assert that $A^\top M^{-1} A \preceq \left(A^\top M_0 A\right)^{-1}$. Indeed, we have

$$
\begin{aligned}
&\left(A^\top M_0 A\right)^{1/2} A^\top M^{-1} A \left(A^\top M_0 A\right)^{1/2} \\
=& V D^{1/2} V^\top V \Lambda^\top U^\top U \begin{pmatrix} D + A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}^{-1} U^\top U \Lambda V^\top V D^{1/2} V^\top \\
=& V D^{1/2} \begin{pmatrix} I_{r_k \times r_k} & 0_{r_k \times (d - r_k)} \end{pmatrix} \begin{pmatrix} (D + A_{11} - A_{12} A_{22}^{-1} A_{21})^{-1} & \star \\ \star & \star \end{pmatrix} \begin{pmatrix} I_{r_k \times r_k} \\ 0_{r_k \times (d - r_k)} \end{pmatrix} D^{1/2} V^\top \\
=& V D^{1/2} \left(D + A_{11} - A_{12} A_{22}^{-1} A_{21}\right)^{-1} D^{1/2} V^\top \\
\preceq& V V^\top \\
=& I_{r_k \times r_k}
\end{aligned}
$$

where the inequality is due to $A_{11} - A_{12} A_{22}^{-1} A_{21} \succeq 0$ since this is the Schur complement of $M - M_0 \succeq 0$. As a result, $A^\top M^{-1} A \preceq \left(A^\top M_0 A\right)^{-1}$.

Applying the monotonicity of $\log\det(\cdot)$, we have

$$
\log\det\left((A^{(k)})^\top \left(\sum_{i=1}^n w_i x_i x_i^\top\right)^{-1} A^{(k)}\right) \leq \log\det\left((A^{(k)})^\top \left(\sum_{i \in G_k} w_i x_i x_i^\top\right) A^{(k)}\right)^{-1} .
$$

This establishes the inequality. Further, if equality holds then $A_{11} - A_{12} A_{22}^{-1} A_{21} = 0$. As a result,

$$
\begin{aligned}
&A^\top M^{-1} A \\
=& V \Lambda^\top U^\top U \begin{pmatrix} D + A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}^{-1} U^\top U \Lambda V^\top \\
=& V \begin{pmatrix} I_{r_k \times r_k} & 0_{r_k \times (d - r_k)} \end{pmatrix} \begin{pmatrix} (D + A_{11} - A_{12} A_{22}^{-1} A_{21})^{-1} & \star \\ \star & \star \end{pmatrix} \begin{pmatrix} I_{r_k \times r_k} \\ 0_{r_k \times (d - r_k)} \end{pmatrix} V^\top \\
=& V D^{-1} V^\top \\
=& \left(A^\top M_0 A\right)^{-1} .
\end{aligned}
$$

Notice that the validity above argument does not depend on w_{G_k} and in fact holds for any w'_{G_k}. This completes the proof.

□

Claim H.3 (Monotonicity of determinant)**.** Suppose A and B are two symmetric matrices such that $A \succeq B \succ 0$, then $\det A \geq \det B$.

Claim H.4 (Concavity of log-determinant function)**.** Suppose A and B are two symmetric positive semidefinite matrices such that $A \succeq B \succ 0$, then $\log\det(\lambda A + (1 - \lambda) B) \geq \lambda \log\det B + (1 - \lambda) \log\det A$ holds for any $\lambda \in (0, 1)$.

The following important result of J. Kiefer and J. Wolfowitz established the equivalence of the D-optimal design and G-optimal design.

Theorem H.5 (General equivalence theorem of G-optimal design (Kiefer & Wolfowitz, 1960))**.** *Assume* $\mathrm{span}(\mathcal{X}) = \mathbb{R}^d$. *The followings are equivalent:*

- $\pi^* = \arg\max_{\pi \in \Delta(\mathcal{X})} \log\det \mathbf{M}(\pi)$;

- $\pi^* = \arg\min_{\pi \in \Delta(\mathcal{X})} \max_{x \in \mathcal{X}} x^\top \mathbf{M}(\pi)^{-1} x$;

- $\max_{x \in \mathcal{X}} x^\top \mathbf{M}(\pi)^{-1} x = d$.