Differential Privacy with Multiple Selections*

Ashish Goel
Stanford University
ashishg@stanford.edu

Zhihao Jiang
Stanford University
faebdc@stanford.edu

Aleksandra Korolova
Princeton University
korolova@princeton.edu

Kamesh Munagala
Duke University
kamesh@cs.duke.edu

Sahasrajit Sarmasarkar
Stanford University
sahasras@stanford.edu

Abstract

We consider the setting where a user with sensitive features wishes to obtain a recommendation from a server in a differentially private fashion. We propose a "multi-selection" architecture where the server can send back multiple recommendations and the user chooses one from these that matches best with their private features. When the user feature is one-dimensional – on an infinite line – and the accuracy measure is defined w.r.t some increasing function $\mathfrak{h}(.)$ of the distance on the line, we precisely characterize the optimal mechanism that satisfies differential privacy. The specification of the optimal mechanism includes both the distribution of the noise that the user adds to its private value, and the algorithm used by the server to determine the set of results to send back as a response and further show that Laplace is an optimal noise distribution. We further show that this optimal mechanism results in an error that is inversely proportional to the number of results returned when the function $\mathfrak{h}(.)$ is the identity function.

1 Introduction

Consider a user who wants to issue a query to an online server (e.g. to retrieve a search result or an advertisement), but the query itself reveals private information about the use. One commonly studied approach to protect user privacy from the server in this context is for the user to send a perturbed query, satisfying differential privacy under the local trust model [14]. However, since the query itself is changed from the original, the server may not be able to return a result that is very accurate for the original query. Our key observation is that in many situations such as search or content recommendations, the server is free to return many results, and the user can choose the one that is the most appropriate, without revealing the choice to the server. In fact, if the server also returns a model for evaluating the quality of these results for the user, then this choice can be made by a software intermediary such as a client running on the user's device. This software intermediary can also be the one that acts as the user's privacy delegate and is the one ensuring the local differential privacy protection.

We call this, new for the differential privacy (DP) literature system architecture, the "Multi-Selection" approach to privacy, and the key question we ask is: *What is the precise trade-off that can be achieved between the number of returned results and quality under a fixed privacy goal*? Of course,

*Authors are listed in alphabetical order.

had the server simply returned all possible results, there would have been no loss in quality since the client could choose the optimal result. However, this approach is infeasible due to computation and communication costs, as well as due to potential server interest in not revealing proprietary information. We therefore restrict the server to return k results for small k, and study the trade-off between k and the quality when the client sends privacy-preserving queries. Our algorithmic design space consists of choosing the client's algorithm and the server's algorithm, as well as the space of signals they will be sending.

At a high level, in addition to the novel multi-selection framework for differential privacy, our main contributions are two-fold. First, under natural assumptions on the privacy model and the latent space of results and users, we show a tight trade-off between utility and number of returned results via a natural (yet *a priori* non-obvious) algorithm, with the error perceived by a user decreasing as $\Theta(1/k)$ for k results. Secondly, at a technical level, our proof of optimality is via a dual fitting argument and is quite subtle, requiring us to develop a novel duality framework for linear programs over infinite dimensional function spaces, with constraints on both derivatives and integrals of the variables. This framework may be of independent interest for other applications where such linear programs arise.

1.1 System Architecture and Definitions

We denote the set of users by $\mathbb{R}$ and when we refer to a user $u \in \mathbb{R}$, we mean a user with a query value $u \in \mathbb{R}$. Let M denote the set of results and $\text{OPT} : \mathbb{R} \to M$ denotes the function which maps user queries to optimal results. This function OPT is available (known) to the server but is unknown to the users.

1.1.1 Privacy

We adopt a well-studied notion of differential privacy under the local trust model [14]:

Definition 1.1 (adapted from [25, 43])**.** Let $\epsilon > 0$ be a desired level of privacy and let $\mathcal{U}$ be a set of input data and $\mathcal{Y}$ be the set of all possible responses and $\Delta(\mathcal{Y})$ be the set of all probability distributions (over a sufficiently rich σ-algebra of $\mathcal{Y}$ given by $\sigma(\mathcal{Y})$). A mechanism $Q : \mathcal{U} \to \Delta(\mathcal{Y})$ is ϵ-differentially private if for all $S \in \sigma(\mathcal{Y})$ and $u_1, u_2 \in \mathcal{U}$:

$$\mathbb{P}(Qu_1 \in S) \leq e^{\epsilon}\mathbb{P}(Qu_2 \in S).$$

A popular relaxation of differential privacy is geographic differential privacy [4] (GDP), which allows the privacy guarantee to decay with the distance between user values. It reflects the intuition that the user is more interested in protecting the specifics of a medical query they are posing rather than protecting whether they are posing a medical query or an entertainment query, and is thus, appropriate in scenarios such as search. We restate the formal definition from [43] and use it in the rest of the work.

Definition 1.2 (adapted from [43])**.** Let $\epsilon > 0$ be a desired level of privacy and let $\mathcal{U}$ be a set of input data and $\mathcal{Y}$ be the set of all possible responses and $\Delta(\mathcal{Y})$ be the set of all probability distributions (over a sufficiently rich σ-algebra of $\mathcal{Y}$ given by $\sigma(\mathcal{Y})$). A mechanism $Q : \mathcal{U} \to \Delta(\mathcal{Y})$ is ϵ-geographic differentially private if for all $S \in \sigma(\mathcal{Y})$ and $u_1, u_2 \in \mathcal{U}$:

$$\mathbb{P}(Qu_1 \in S) \leq e^{\epsilon|u_1 - u_2|}\mathbb{P}(Qu_2 \in S).$$

1.1.2 "Multi-Selection" Architecture

Our "multi-selection" system architecture (shown in Figure 1) relies on the server returning a small set of results in response to the privatized user input, with the on-device software intermediary deciding, unbeknownst to the server, which of these server responses to use.

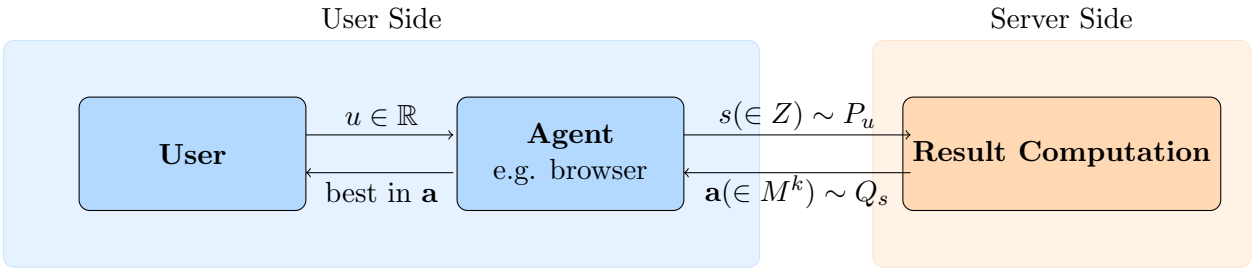


Figure 1: Overall architecture for multi-selection.

The space of mechanisms we consider in this new architecture consists of a triplet $(Z, \mathbf{P}, \mathbf{Q})$:

1. A set of signals Z that can be sent by users.
2. The actions of users, $\mathbf{P}$, which involves a user sampling a signal from a distribution over signals. We use P_u for $u \in \mathbb{R}$ to denote the distribution of the signals sent by user u which is supported on Z. The set of actions over all users is given by $\mathbf{P} = \{P_u\}_{u \in \mathbb{R}}$.
3. The distribution over actions of the server, $\mathbf{Q}$. When the server receives a signal $s \in Z$, it responds with Q_s, which characterizes the distribution of the k results that the server sends (it is supported in M^k). We denote the set of all such server actions by $\mathbf{Q} = \{Q_s\}_{s \in Z}$. [1]

Our central question is to find the triplet over $(Z, \mathbf{P}, \mathbf{Q})$ that satisfies ϵ-geographic differential privacy constraints on $\mathbf{P}$ while ensuring the best-possible utility or the smallest-possible disutility.

1.1.3 The disutility model: Measuring the cost of privacy

We now define the disutility of a user $u \in \mathbb{R}$ from a result $m \in M$. One natural definition would be to look at the difference between (or the ratio) of the cost of the optimum result for u and the cost of the result m returned by a privacy-preserving algorithm. However, we are looking for a general framework, and do not want to presume that this cost measure is known to the algorithm designer, or indeed, that it even exists. Hence, we will instead define the disutility of u as the amount by which u would have to be perturbed for the returned result m to become optimum; this only requires a distance measure in the space in which u resides, which is needed for the definition of the privacy guarantees anyway. For additional generality, we will also allow the disutility to be any increasing function of this perturbation, as defined below.

Definition 1.3. The dis-utility of a user $u \in \mathbb{R}$ from a result $m \in M$ w.r.t some continuously increasing function $\mathfrak{h}(.)$ is given by [2]

$$\text{Dis-util}^{\mathfrak{h}(.)}(u, m) := \inf_{u' \in \mathbb{R}: OPT(u') = m} \mathfrak{h}(|u - u'|) \tag{1}$$

[1] We treat this distribution to supported on U^k instead of k-sized subset of U for ease of mathematical typesetting.
[2] If no such u' exists then the dis-utility is ∞ as infimum of a null set is ∞.

We allow any function $\mathfrak{h}(.)$ that satisfies the following conditions:

$$\mathfrak{h}(.) \text{ is a continuously increasing function satisfying } \mathfrak{h}(0) = 0. \tag{2}$$

$$\text{There exists } \mathcal{B} \in \mathbb{R}^+ \text{ s.t. } \log \mathfrak{h}(.) \text{ is Lipschitz continuous in } [\mathcal{B}, \infty) \tag{3}$$

The first condition (2) captures the intuition that dis-utility for the optimal result is zero. The second condition (3) which bounds the growth of $\mathfrak{h}(.)$ by an exponential function is (a not very restrictive condition) required for our mathematical analysis.

Quite surprisingly, to show that our multi-selection framework provides a good trade-off in the above model for every $\mathfrak{h}$ as defined above, we only need to consider the case where the $\mathfrak{h}$ is the identity function. The following example further motivates our choice of the disutility measure:

Example 1. Suppose, one assumes that the result set M and the user set $\mathbb{R}$ are embedded in the same metric space $(d, M \cup \mathbb{R})$. This setup is similar to the framework studied in the examination of metric distortion of ordinal rankings in social choice [5, 20, 38, 42]. Using triangle inequality, one may argue that $d(u, m') - d(u, m) \leq 2d(u, u')$ where m is the optimal result for user u (i.e. $m = \arg\min_{m \in M} d(u, m)$) and m' is the optimal result for user u'. [3] Thus, $2d(u, u')$ gives an upper bound on the disutility of user u from result $\text{OPT}(u')$.

1.1.4 Restricting users and results to the same set

For ease of exposition, we study a simplified setup restricting the users and results to the same set $\mathbb{R}$. Specifically, since $\text{Dis-util}^{\mathfrak{h}(.)}(u, \text{OPT}(u')) \leq \mathfrak{h}(|u - u'|)$, our simplified setup restricts the users and results to the same set $\mathbb{R}$ where the dis-utility of user $u \in \mathbb{R}$ from a result $a \in \mathbb{R}$ is given by $\mathfrak{h}(|u - a|)$. Our results will extend to the model where the users and results lie in different sets, and we refer the reader to Appendix A.5 for the treatment. We note that even in the simplified setup, while we use $a \in \mathbb{R}$ to denote the result, what we mean is that the server sends back $\text{OPT}(a) \in M$.

1.1.5 Definition of the cost function in the simplified setup

We use $\texttt{Set}(\mathbf{a})$ to convert a vector $\mathbf{a} = (a_1, a_2, \ldots, a_k)^T \in \mathbb{R}^k$ to a set of at most k elements, formally $\texttt{Set}(\mathbf{a}) = \{a_i : i \in [k]\}$.

Recall from Section 1.1.4, the dis-utility of user $u \in \mathbb{R}$ from a result $a \in \mathbb{R}$ in the simplified setup may be written as

$$\text{Dis-util}^{\mathfrak{h}(.)}_{sim}(u, a) = \mathfrak{h}(|u - a|) \tag{4}$$

Since we restrict the users and results to the same set, Q_s is supported on $\mathbb{R}^k$ for every $s \in Z$. Thus, the cost for a user u from the mechanism $(Z, \mathbf{P}, \mathbf{Q})$ is given by

$$\text{cost}^{\mathfrak{h}(.)}(u, (Z, \mathbf{P}, \mathbf{Q})) = \mathop{\mathbb{E}}_{s \sim P_u} \left[\mathop{\mathbb{E}}_{\mathbf{a} \sim Q_s} \left[\min_{a \in \texttt{Set}(\mathbf{a})} \text{Dis-util}^{\mathfrak{h}(.)}_{sim}(u, a) \right] \right] \tag{5}$$

$$= \mathop{\mathbb{E}}_{s \sim P_u} \left[\mathop{\mathbb{E}}_{\mathbf{a} \sim Q_s} \left[\min_{a \in \texttt{Set}(\mathbf{a})} \mathfrak{h}(|u - a|) \right] \right], \tag{6}$$

where the expectation is taken over the randomness in the action of user and the server.

We now define the cost of a mechanism by supremizing[4] over all users $u \in \mathbb{R}$. Since, we refrain from making any distributional assumptions over the users, supremization rather than mean over

[3] This follows since $d(u, m') - d(u, u') \leq d(u', m') \leq d(u', m) \leq d(u, u') + d(u, m)$

[4] We write supremum instead of maximum as the maximum over an infinite set may not always be defined.

the users is the logical choice.

$$\text{cost}^{\mathfrak{h}(.)}(Z,\mathbf{P},\mathbf{Q}) := \sup_{u\in\mathbb{R}} \text{cost}^{\mathfrak{h}(.)}(u,(Z,\mathbf{P},\mathbf{Q})) = \sup_{u\in\mathbb{R}} \mathop{\mathbb{E}}_{s\sim P_u}\left[\mathop{\mathbb{E}}_{\mathbf{a}\sim Q_s}\left[\min_{a\in\texttt{Set}(\mathbf{a})} \mathfrak{h}(|u-a|)\right]\right] \tag{7}$$

We use $\mathbb{1}(.)$ to denote the identity function, i.e. $\mathbb{1}(x) = x$ for every $x \in \mathbb{R}$ and thus define the cost function when $\mathfrak{h}(.)$ is an identity function as follows:

$$\text{cost}^{\mathbb{1}(.)}(Z,\mathbf{P},\mathbf{Q}) := \sup_{u\in\mathbb{R}} \text{cost}^{\mathbb{1}(.)}(u,(Z,\mathbf{P},\mathbf{Q})) = \sup_{u\in\mathbb{R}} \mathop{\mathbb{E}}_{s\sim P_u}\left[\mathop{\mathbb{E}}_{\mathbf{a}\sim Q_s}\left[\min_{a\in\texttt{Set}(\mathbf{a})} |u-a|\right]\right] \tag{8}$$

Recall our central question is to find the triplet over $(Z,\mathbf{P},\mathbf{Q})$ that ensures the smallest possible disutility / cost while ensuring that $\mathbf{P}$ satisfies ϵ-geographic differential privacy. We denote the value of this cost by $f^{\mathfrak{h}(.)}(\epsilon,k)$ and it is formally defined as

$$\begin{aligned} f^{\mathfrak{h}(.)}(\epsilon,k) :=& \inf_{Z} \inf_{\mathbf{P}\in\mathcal{P}_Z^{(\epsilon)}} \inf_{\mathbf{Q}\in\mathcal{Q}_Z} \left(\text{cost}^{\mathfrak{h}(.)}(Z,\mathbf{P},\mathbf{Q})\right) \\ =& \inf_{Z} \inf_{\mathbf{P}\in\mathcal{P}_Z^{(\epsilon)}} \inf_{\mathbf{Q}\in\mathcal{Q}_Z} \left(\sup_{u\in\mathbb{R}} \mathop{\mathbb{E}}_{s\sim P_u}\left[\mathop{\mathbb{E}}_{\mathbf{a}\sim Q_s}\left[\min_{a\in\texttt{Set}(\mathbf{a})} \mathfrak{h}(|u-a|)\right]\right]\right) \text{ where} \end{aligned} \tag{9}$$

$\mathcal{P}_Z^{(\epsilon)} := \{\mathbf{P} | \forall u \in \mathbb{R}, P_u \text{ is a distribution on } Z, \text{ and } \mathbf{P} \text{ satisfies } \epsilon\text{-geographic differential privacy.}\}$.
$\mathcal{Q}_Z := \{\mathbf{Q} | \forall s \in Z, Q_s \text{ is a distribution on } \mathbb{R}^k\}$.

1.2 Our results and key technical contributions

For any $\mathfrak{h}(.)$ satisfying equations (2) and (3) when the privacy goal is ϵ-geographic DP our main results are:

- The optimal action P_u for a user u, is to add Laplace noise[5] of scale $\frac{1}{\epsilon}$ to its value u (result stated in Theorem 2.13 and proof sketch described in Sections 2.1, 2.2 and 2.3). Further, we emphasize that the optimality of adding Laplace noise is far from obvious[6]. For instance, when users and results are located on a ring, Laplace noise is *not optimal* (see Appendix A.4 for an analysis when $k = 2$).
- The optimal server response $\mathbf{Q}$ could be different based on different $\mathfrak{h}$. We give a recursive construction of $\mathbf{Q}$ for a general $\mathfrak{h}$ (Section 2.4). Furthermore, when $\mathfrak{h}(t) = t$, we give an exact construction of $\mathbf{Q}$ (sketched in Fig. 2 for $k = 5$) and show that $f^{\mathbb{1}(.)}(\epsilon,k) = O(\frac{1}{\epsilon k})$ in Section 2.4 and Appendix B.5.

Formally, our main results can be stated as:

Theorem 1.4 (corresponds to Theorem 2.13 and Theorem B.3)**.** *For ϵ-geographic differential privacy, adding Laplace noise, that is, user u sends a signal drawn from distribution $\mathcal{L}_\epsilon(u)$, is one of the optimal choices of $\mathcal{P}_Z^{(\epsilon)}$ for users. Further, when $\mathfrak{h}(t) = t$, we have $f^{\mathbb{1}(.)}(\epsilon,k) = O(\frac{1}{\epsilon k})$ and the optimal mechanism $(Z,\mathbf{P},\mathbf{Q})$ (choice of actions of users and server) itself can be computed in closed form. For a generic $\mathfrak{h}(.)$, the optimal server action $\mathbf{Q}$ may be computed recursively.*

[5]We use $\mathcal{L}_\epsilon(u)$ to denote a Laplace distribution of scale $\frac{1}{\epsilon}$ centred at u. Formally, a distribution $X \sim \mathcal{L}_\epsilon(u)$ has its probability density function given by $f_X(x) = \frac{\epsilon}{2}e^{-\epsilon|x-u|}$.

[6]In fact, only a few optimal DP mechanisms are known [37, 40, 41, 24], and it is known that for certain scenarios, universally optimal mechanisms do not exist [18].

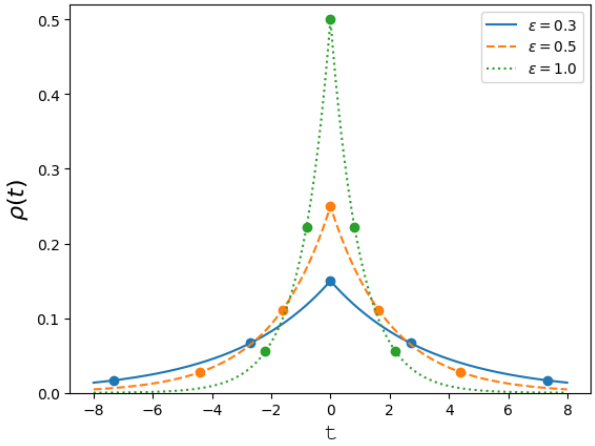


Figure 2: Optimal mechanisms in geographic differential privacy setting when $k = 5$ and $\epsilon \in \{0.3, 0.5, 1.0\}$. Suppose the user has a private value u. Then the user sends a signal s drawn from distribution $\mathcal{L}_\epsilon(u)$ to the server, meaning the user sends $s = v + x$ where x is drawn from the density function $\rho(t)$ in this figure. Suppose the server receives s. Then the server responds $\{s + a_1, ..., s + a_5\}$, where the values of $a_1, a_2, ..., a_5$ are the t-axis values of dots on the density functions.

In addition to our overall framework and the tightness of the above theorem, a key contribution of our work is in the techniques developed. At a high level, our proof proceeds via constructing an infinite dimensional linear program to encode the optimal algorithm under differential privacy constraints. We then use dual fitting to show the optimality of Laplace noise. Finally, the optimal set of results being computable by a simple dynamic program given such noise.

The technical hurdles arise because the linear program for encoding the optimal mechanism is over infinite-dimensional function spaces with linear constraints on both derivatives and integrals, since the privacy constraint translates to constraints on the derivative of the density encoding the optimal mechanism, while capturing the density itself requires an integral. We call it Differential Integral Linear Program (DILP); see Section 2.2. However, there is no weak duality theory for such linear programs in infinite dimensional function spaces, such results only existing for linear programs with integral constraints [3].

We therefore develop a weak duality theory for DILPs (see Section 2.2 with a detailed proof in Appendix B.6), which to the best of our knowledge is novel. The proof of this result is quite technical and involves a careful application of Fatou's lemma [49] and the monotone convergence theorem to interchange integrals with limits, and integration by parts to translate the derivative constraints on the primal variables to derivatives constraint on the dual variables.

We believe our weak duality framework is of independent interest and has broader implications beyond differential privacy; see Appendix A.8 for two such applications.

1.3 Related Work

1.3.1 Differential Privacy

The notion of differential privacy in the trusted curator model is introduced in [26]; see [28] for a survey of foundational results in this model. The idea of local differential privacy dates back to [53], and the algorithms for satisfying it have been studied extensively following the deployment of DP in this model by Google [30] and Apple [6]; see, e.g. [19, 21, 52, 12] and Bebensee [14] for a survey.

Geographic differential privacy was introduced by Andrés *et al.* [4] and has gained widespread adoption for anonymizing location data. Our use of geographic DP utilizes the definition of [4] with the trust assumptions of the local model, and is thus, only a slight relaxation of the traditional local model of differential privacy.

1.3.2 Multi-Selection

An architecture for multi-selection, particularly with the goal of privacy-preserving advertising, was already introduced in *Adnostic* by [50]. Their proposal was to have a browser extension that would run the targeting and ad selection on the user's behalf, reporting to the server only click information using cryptographic techniques. Similarly, *Privad* by [39] propose to use an anonymizing proxy that operates between the client that sends broad interest categories to the proxy and the advertising broker, that transmits all ads matching the broad categories, with the client making appropriate selections from those ads locally. Although both Adnostic and Privad reason about the privacy properties of their proposed systems, unlike our work, neither provides DP guarantees.

Two lines of work in the DP literature can be seen as somewhat related to the multi-selection paradigm – the exponential mechanism (see e.g. [46, 17, 45]) and amplification by shuffling (see e.g. [29, 22, 32]). The exponential mechanism focuses on high-utility private selection from multiple alternatives and is usually deployed in the TCM model. Amplification by shuffling analyzes the improvement in the DP guarantees that can be achieved if the locally privatized data is shuffled by an entity before being shared with the server. As far as we are aware, neither of the results from these lines of work can be directly applied to our version of multi-selection, although combining them is an interesting avenue for future work.

More broadly, several additional directions within DP research can be viewed as exploring novel system architectures in order to improve privacy-utility trade-offs, e.g., using public data to augment private training [47, 11], combining data from TCM and LM models [8, 7, 15], and others. Our proposed architecture is distinct from all of these. Finally, our work is different from how privacy is applied in federated learning [36] – there, the goal is for a centralized entity to be able to build a machine learning model based on distributed data; whereas our goal is to enable personalized, privacy-preserving retrieval from a global ML model.

1.3.3 Optimal DP mechanisms

To some extent, much of the work in DP can be viewed as searching for the optimal DP mechanism, i.e. one that would achieve the best possible utility given a fixed desired DP level. Only a few optimal mechanisms are known [37, 40, 41, 24], and it is known that for certain scenarios, universally optimal mechanisms do not exist [18].

Most closely related to our work is the foundational work of [37] that derives the optimal mechanism for counting queries via a linear programming formulation; the optimal mechanism turns out to be the discrete version of the Laplace mechanism. Its continuous version was studied in [33], where the Laplace mechanism was shown to be optimal. These works focused on the trusted curator model of differential privacy unlike the local trust model which we study.

In the local model, [43] show Laplace noise to be optimal for ϵ-geographic DP. Their proof relies on formulating an infinite dimensional linear program over noise distributions and looking at its dual. Although their proof technique bears a slight resemblance to ours, our proof is different and intricate since it involves the minimisation over the set of returned results in the cost function.

A variation of local DP is considered in [35], in which DP constraints are imposed only when the distance between two users is smaller than a threshold. For that setting, the optimal noise is

piece-wise constant, which is a similar outcome to our optimal Laplace noise distribution. However, our setting of choosing from multiple options makes the problems very different.

1.3.4 Related work on duality theory for infinite dimensional LPs

Strong duality is known to hold for finite dimensional linear programs [1]. However, for infinite dimensional linear programs, strong duality may not always hold (see [3] for a survey). Sufficient conditions for strong duality are presented and discussed in [51, 13] for generalized vector spaces. A class of linear programs (called SCLPs) over bounded measurable function spaces have been studied in [48, 16] with integral constraints on the functions. However, these works do not consider the case with derivative constraints on the functional variables. In [43, Equation 7] a linear program with derivative and integral constraints (DILPs) is formulated to show the optimality of Laplace noise for geographic differential privacy. However, their duality result does not directly generalize to our case since our objective function and constraints are far more complex as it involves minimisation over a set of results.

We thus need to reprove the weak-duality theorem for our DILPs, the proof of which is technical and involves a careful application of integration by parts to translate the derivative constraint on the primal variable to a derivative constraint on the dual variable. Further, we require a careful application of Fatou's lemma [49] and monotone convergence theorem to exchange limits and integrals. Our weak duality result generalizes beyond our specific example and is applicable to a broader class of DILPs. Furthermore, we discuss two problems (one from job scheduling [2] and one from control theory [31]) in Appendix A.8 which may be formulated as a DILP.

2 Characterizing the Optimal Mechanism: Proof Sketch of Theorem 1.4

We now present a sketch of the proof of Theorem 1.4; the full proof involving the many technical details is presented in the Appendix. We first show that for the sake of analysis, the server can be removed by making the signal set coincide with result sets (Section 2.1) assuming that the function OPT is publicly known both to the user and server.[7] Then in Section 2.2 we construct a primal linear program $\mathcal{O}$ for encoding the optimal mechanism, and show that it falls in a class of infinite dimensional linear programs that we call DILPs, as defined below.

Definition 2.1. Differential-integral linear program (DILP) is a linear program over Riemann integrable function spaces involving constraints on both derivatives and integrals.

A simple example is given in Equation (10). Observe that in equation (10), we define $\mathcal{C}_1$ to be a continuously differentiable function.

$$\tilde{\mathcal{O}} = \begin{cases} \inf\limits_{g(.):\mathcal{C}^1(\mathbb{R}\to\mathbb{R}^+)} \int_{\mathbb{R}} |v| g(v) dv \\ \quad \text{s.t.} \quad \int_{\mathbb{R}} g(v) dv = 1 \\ \qquad -\epsilon g(v) \le g'(v) \le \epsilon g(v) \ \forall v \in \mathbb{R} \end{cases} \tag{10}$$

[7] One should note that this removal is just for analysis and the server is needed since the OPT function is unknown to the user.

We next construct a dual DILP formulation $\mathcal{E}$ in Section 2.2, and show that the formulation satisfies weak duality. As mentioned before, this is the most technically intricate result since we need to develop a duality theory for DILPs. We relegate the details of the proof here to the Appendix.

Next, in Section 2.3, we show the optimality of the Laplace noise mechanism via dual-fitting, *i.e.*, by constructing a feasible solution to DILP $\mathcal{E}$ with objective identical to that of the Laplace noise mechanism. Finally, in Section 2.4, we show how to find the optimal set of k results given Laplace noise. We give a construction for general functions $\mathfrak{h}(.)$ as well as a closed form for the canonical case of $\mathfrak{h}(t) = t$. This establishes the error bound and concludes the proof of Theorem 1.4.

2.1 Restricting the signal set Z to $\mathbb{R}^k$

We first show that it is sufficient to consider a more simplified setup where the user sends a signal set in $\mathbb{R}^k$ and the server sends back the results corresponding to the signal set. Since we assumed users and ads lie in the same set, for the purpose of analysis, this removes the server from the picture. To see this, note that for the setting discussed in Section 1.1.4, the optimal result for user u is the result u itself, where when we refer to "result u", we refer to the result $\text{OPT}(u) \in M$.

Thus, this approach is used only for a simplified analysis as the OPT function is not known to the user and our final mechanism will actually split the computation between the user and the server.

Therefore a user can draw a result set directly from the distribution over the server's action and send the set as the signal. The server returns the received signal, hence removing it from the picture. In other words, it is sufficient to consider mechanisms in $\mathcal{P}^{(\epsilon)}_{\mathbb{R}^k}$, which are in the following form (corresponding to Theorem 2.2).

1. User $v \in \mathbb{R}$ reports s that is drawn from a distribution P_v over $\mathbb{R}^k$.
2. The server receives s and returns s.

We give an example to illustrate this statement below.

Example 2. Fix a user u and let $Z = \{s_1, s_2\}$ and $\{\mathbf{a}^{(1)}, \mathbf{a}^{(2)}\} \subseteq \mathbb{R}^k$ and consider a mechanism $\mathcal{M}_1$ where user u sends s_1 and s_2 with equal probability. The server returns $\mathbf{a} \in \mathbb{R}^k$ on receiving signal s, with the following probability.

$$\mathbb{P}\left(\mathbf{a} = \mathbf{a}^{(1)} | s = s_1\right) = 0.2, \quad \mathbb{P}\left(\mathbf{a} = \mathbf{a}^{(2)} | s = s_1\right) = 0.8,$$
$$\mathbb{P}\left(\mathbf{a} = \mathbf{a}^{(1)} | s = s_2\right) = 0.4, \quad \mathbb{P}\left(\mathbf{a} = \mathbf{a}^{(2)} | s = s_2\right) = 0.6.$$

Then the probability that u receives $\mathbf{a}^{(1)}$ is 0.3 and it receives $\mathbf{a}^{(2)}$ is 0.7. Now consider another mechanism $\mathcal{M}_2$ with the same cost satisfying differential privacy constraints, where $Z = \{\mathbf{a}^{(1)}, \mathbf{a}^{(2)}\}$, with user u sending signal $\mathbf{a}^{(1)}$ and $\mathbf{a}^{(2)}$ with probabilities 0.3 and 0.7. When the server receives $\mathbf{a} \in \mathbb{R}^k$, it returns $\mathbf{a}$.

We show the new mechanism $\mathcal{M}_2$ satisfies differential privacy assuming the original mechanism $\mathcal{M}_1$ satisfies it. As a result, we can assume $Z = \mathbb{R}^k$ when finding the optimal mechanism.

The following theorem states that it is sufficient to study a setup removing the server from the picture and consider mechanisms in set of probability distributions supported on $\mathbb{R}^k$ satisfying ϵ-geographic differential privacy ($\mathcal{P}^{(\epsilon)}_{\mathbb{R}^k}$ as defined in Section 1.1.5).

Theorem 2.2 (detailed proof in Appendix B.1)**.** *It is sufficient to remove the server* ($\mathbf{Q}$) *from the cost function* $f^{\mathfrak{h}(.)}(\epsilon, k)$ *and pretend the user has full-information. Mathematically, it maybe stated as follows.*

$$f^{\mathfrak{h}(.)}(\epsilon, k) = \inf_{\mathbf{P} \in \mathcal{P}^{(\epsilon)}_{\mathbb{R}^k}} \sup_{u \in \mathbb{R}} \mathop{\mathbb{E}}_{\boldsymbol{a} \sim P_u} \left[\min_{a \in \mathit{Set}(\boldsymbol{a})} \mathfrak{h}(|u - a|) \right]. \tag{11}$$

Proof Sketch. Fix $Z, \mathbf{P} \in \mathcal{P}^{(\epsilon)}_Z, \mathbf{Q} \in \mathcal{Q}_Z$. For $u \in \mathbb{R}$ and $S \subseteq \mathbb{R}^k$, let $\tilde{P}_u(S)$ be the probability that the server returns a set in S to user u. Then for any $u_1, u_2 \in \mathbb{R}, S \subseteq \mathbb{R}^k$, we can show that $\tilde{P}_{u_1}(S) \leq e^{\epsilon \cdot |u_1 - u_2|} \tilde{P}_{u_2}(S)$ using post-processing theorem, and thus $\tilde{\mathbf{P}} = \{\tilde{P}_u\}_{u \in \mathbb{R}} \in \mathcal{P}^{(\epsilon)}_{\mathbb{R}^k}$ because $\tilde{P}_u$ is a distribution on $\mathbb{R}^k$ for any $u \in \mathbb{R}$. At the same time,

$$\mathop{\mathbb{E}}_{s \sim \mathbf{P}_u} \left[\mathop{\mathbb{E}}_{\mathbf{a} \sim \mathbf{Q}_s} \left[\min_{a \in \mathtt{Set}(\mathbf{a})} \mathfrak{h}(|u - a|) \right] \right] = \mathop{\mathbb{E}}_{\mathbf{a} \sim \tilde{P}_u} \left[\min_{a \in \mathtt{Set}(\mathbf{a})} \mathfrak{h}(|u - a|)) \right], \text{ so we have}$$

$$f^{\mathfrak{h}(.)}(\epsilon, k) = \inf_{\mathbf{P} \in \mathcal{P}^{(\epsilon)}_{\mathbb{R}^k}} \sup_{u \in \mathbb{R}} \mathop{\mathbb{E}}_{\mathbf{a} \sim P_u} \left[\min_{a \in \mathtt{Set}(\mathbf{a})} \mathfrak{h}(|u - a|) \right],$$

□

2.2 Differential integral linear programs to represent $f(\epsilon, k)$ and a weak duality result

Recall the definition of DILP from Definition 2.1. In this section, we construct an infimizing DILP $\mathcal{O}$ to represent the constraints and the objective in the cost function $f(\epsilon, k)$. We then construct a dual DILP $\mathcal{E}$, and provide some intuition for this formulation. The proof of weak duality is our main technical result, and its proof is defered to the Appendix.

2.2.1 Construction of DILP $\mathcal{O}$ to represent cost function $f(\epsilon, k)$

We now define the cost of a mechanism $\mathbf{P}$ which overloads the cost definition in Equation 7

Definition 2.3. Cost of mechanism $\mathbf{P} \in \mathcal{P}^{(\epsilon)}_{\mathbb{R}^k}$: We define the cost of mechanism $\mathbf{P}$ as

$$\mathit{cost}(\mathbf{P}) := \sup_{u \in \mathbb{R}} \mathop{\mathbb{E}}_{\mathbf{a} \sim P_u} \left[\min_{a \in \mathtt{Set}(\mathbf{a})} \mathfrak{h}(|u - a|) \right] \tag{12}$$

Observe that in Definition 2.3 we just use $\mathbf{P}$ instead of the tuple $(Z, \mathbf{P}, \mathbf{Q})$ as in Equation (7). Observe that it is sufficient to consider $\mathbf{P}$ in the cost since $\mathbf{P}$ simulates the entire combined action of the user and the server as shown in Theorem 11 in Section 2.1. We now define the notion of approximation using cost of mechanism by a sequence of mechanisms which is used in the construction of DILP $\mathcal{O}$.

Definition 2.4. Arbitrary cost approximation: We call mechanisms $\mathbf{P}^{(\eta)} \in \mathcal{P}^{(\epsilon)}_{\mathbb{R}^k}$ an arbitrary cost approximation of mechanisms $\mathbf{P} \in \mathcal{P}^{(\epsilon)}_{\mathbb{R}^k}$ if $\lim_{\eta \to 0} \mathit{cost}(\mathbf{P}^{(\eta)}) = \mathit{cost}(\mathbf{P})$

Now we define the DILP $\mathcal{O}$ to characterise $f(\epsilon, k)$ in Equation (11). In this formulation, the variables are $g(.,.) : \mathcal{I}^B(\mathbb{R} \times \mathbb{R}^k \to \mathbb{R}^+)$, which we assume are non-negative *Riemann integrable* bounded functions. These variables capture the density function P_u.

$$\mathcal{O} = \begin{cases} \inf\limits_{g(.,.):\mathcal{I}^B(\mathbb{R}\times\mathbb{R}^k\to\mathbb{R}^+),\kappa\in\mathbb{R}} & \kappa \\ \text{s.t.} & \kappa - \int\limits_{\mathbf{x}\in\mathbb{R}^k} \left[\min\limits_{a\in\texttt{Set}(\mathbf{x})} \mathfrak{h}(|u-a|)\right] g(u,\mathbf{x}) d\left(\prod_{i=1}^k x_i\right) \geq 0 \ \forall u \in \mathbb{R} \\ & \int\limits_{\mathbf{x}\in\mathbb{R}^k} g(u,\mathbf{x}) d\left(\prod_{i=1}^k x_i\right) = 1 \ \forall u \in \mathbb{R} \\ & \epsilon g(u,\mathbf{x}) + \underline{g}_u(u,\mathbf{x}) \geq 0; \ \forall u \in \mathbb{R}; \mathbf{x} \in \mathbb{R}^k \\ & \epsilon g(u,\mathbf{x}) - \overline{g}_u(u,\mathbf{x}) \geq 0; \ \forall u \in \mathbb{R}; \mathbf{x} \in \mathbb{R}^k \end{cases} \tag{13}$$

In DILP $\mathcal{O}$, we define $\underline{g}_u(u,\mathbf{x})$ and $\overline{g}_u(u,\mathbf{x})$ to be the lower and upper partial derivative of $g(u,\mathbf{x})$ at u. Now observe that, we use lower and upper derivatives instead of directly using derivatives as the derivatives of a probability density function may not always be defined (for example, the left and right derivatives are unequal in the Laplace distribution at origin).

Note that the DILP $\mathcal{O}$ involves integrals and thus requires mechanisms to have a valid probability density function, however not every distribution is continuous, and, as a result, may not have a density function (e.g. point mass distributions like $\hat{P}_u^{\mathcal{L}_\epsilon}$ defined in Definition 2.9). Using ideas from mollifier theory [34] we construct mechanisms $\mathbf{P}^{(\eta)}$ with a valid probability density function that are an arbitrary good approximation of mechanism $\mathbf{P}$ in Lemma 2.7, hence showing that it suffices to define $\mathcal{O}$ over bounded, non-negative Reimann integrable functions g.

We now prove that the DILP constructed above captures the optimal mechanism, in other words, $\text{opt}(\mathcal{O}) = f(\epsilon, k)$.

Lemma 2.5. *Let opt($\mathcal{O}$) denote the optimal value of DILP* (13)*, then $f(\epsilon,k) = opt(\mathcal{O})$.*

To prove this lemma, we show Lemma 2.6, which relates the last two constraints of the DILP $\mathcal{O}$ to ϵ-geographic differential privacy, and Lemma 2.7, which shows that an arbitrary cost approximation of mechanism $\mathbf{P}$ can be constructed with valid probability density functions.

Lemma 2.6. *Let P_u have a probability density function given by $g(u,.) : \mathbb{R}^k \to \mathbb{R}$ for every $u \in \mathbb{R}$. Then, $\mathbf{P}$ satisfies ϵ-geographic differential privacy iff* $\max(|\overline{g}_u(u,\boldsymbol{x})|, |\underline{g}_u(u,\boldsymbol{x})|) \leq \epsilon g(u,\boldsymbol{x}) \ \forall u \in \mathbb{R}; \forall \boldsymbol{x} \in \mathbb{R}^k$ [8]

The proof of this lemma (Appendix B.2.1) proceeds by showing that ϵ-geographic differential privacy is equivalent to Lipschitz continuity of $\log g(u,\mathbf{x})$ in u.[9]

Lemma 2.7. *(Proven in Appendix B.2.3) Given any mechanism $\boldsymbol{P} \in \mathcal{P}^{(\epsilon)}_{\boldsymbol{R}^k}$ (satisfying ϵ-geographic differential privacy), we can construct a sequence of mechanisms $\boldsymbol{P}^{(\eta)}$ with bounded probability density functions such that $\boldsymbol{P}^{(\eta)}$ is an arbitrary cost approximation of mechanism $\boldsymbol{P} \in \mathcal{P}^{(\epsilon)}_{\boldsymbol{R}^k}$.*

Using Lemmas 2.7 and 2.6, we give the proof of Lemma 2.5.

Proof of Lemma 2.5. Consider any $\zeta > 0$. As established in Lemma 2.7, it follows for every mechanism $\mathbf{P} \in \mathcal{P}^{(\epsilon)}_{\mathbf{R}^k}$, we can construct another mechanism $\mathbf{P}^{(\eta)}$ with bounded probability density functions whose cost is a ζ approximation of the cost of mechanism $\mathbf{P}$. Thus, we can use Lemma 2.6 to conclude that the optimum value of DILP $\mathcal{O}$ is precisely $f(\epsilon,k)$. □

[8] $\underline{g}_u(u,\mathbf{x})$, $\overline{g}_u(u,\mathbf{x})$ denote the lower and upper partial derivative w.r.t. u

[9] We handle the case when the log is not defined as the density is zero at some point separately in the proof.

2.2.2 Dual DILP $\mathcal{E}$ and statement of weak duality theorem

Now, we write the *dual* of the DILP $\mathcal{O}$ as the DILP $\mathcal{E}$ in Equation (14). Observe that, we have the constraint that $\delta(.)$ and $\lambda(.)$ is non-negative, $\mathcal{C}^0$ (continuous) and $\nu(.,.)$ is a $\mathcal{C}^1$ function i.e. $\nu(r, \mathbf{v})$ is *continuously differentiable* in r and *continuous* in $\mathbf{v}$. Thus, we may rewrite the equations as

$$\mathcal{E} = \begin{cases} \sup\limits_{\substack{\delta(.),\lambda(.):\mathcal{C}^0(\mathbb{R}\to\mathbb{R}^+);\\ \nu(.,.):\mathcal{C}^1(\mathbb{R}\times(\mathbb{R})^k\to\mathbb{R})}} \displaystyle\int_{r\in\mathbb{R}} \lambda(r)dr \\ \quad \text{s.t.} \quad \displaystyle\int_{r\in\mathbb{R}} \delta(r)dr \leq 1 \\ \quad -\left[\min\limits_{a\in\texttt{Set}(\mathbf{v})} \mathfrak{h}(|r-a|)\right]\delta(r) + \lambda(r) + \nu_r(r,\mathbf{v}) + \epsilon|\nu(r,\mathbf{v})| \leq 0 \ \forall r \in \mathbb{R}; \mathbf{v} \in (\mathbb{R})^k \\ \quad\quad \exists U : \mathcal{C}^0(\mathbb{R}^k \to \mathbb{R}) \text{ s.t. } \nu(r,\mathbf{v}) \geq 0 \ \forall r \geq U(\mathbf{v}) \ \forall \mathbf{v} \in (\mathbb{R})^k \\ \quad\quad \exists L : \mathcal{C}^0(\mathbb{R}^k \to \mathbb{R}) \text{ s.t. } \nu(r,\mathbf{v}) \leq 0 \ \forall r \leq L(\mathbf{v}) \ \forall \mathbf{v} \in (\mathbb{R})^k \end{cases} \tag{14}$$

To get intuition behind the construction of our dual DILP $\mathcal{E}$, relate the equations in DILP $\mathcal{O}$ to the dual variables of DILP $\mathcal{E}$ as follows. The first equation denoted by $\{\delta(r)\}_{r\in\mathbb{R}}$, second equation denoted by $\{\lambda(r)\}_{r\in\mathbb{R}}$ and the last two equations are *jointly* denoted by $\{\nu(r,\mathbf{v})\}_{r\in\mathbb{R};\mathbf{v}\in(\mathbb{R})^k}$ [10]. The last two terms in the second constraint of DILP $\mathcal{E}$ are a consequence of the last two equations on DILP $\mathcal{O}$ and observe that it involves a derivative of the dual variable $\nu(u,\mathbf{v})$. The linear constraint on the derivative of the primal variable translates to a derivative constraint on the dual variable by a careful application of integration by parts, discussed in detail in Appendix B.6.

Observe that in our framework we have to prove the weak-duality result as, to the best of our knowledge, existing duality of linear programs in infinite dimensional spaces work for cases involving just integrals. The proof of this Theorem 2.8 is technical and we defer the details to Appendix B.6.

Theorem 2.8. *opt*$(\mathcal{O}) \geq$ *opt*$(\mathcal{E})$.

2.3 Dual fitting to show the optimality of Laplace noise addition

Before starting this section, we first define a function $\hat{f}(\epsilon, k)$ which characterises the optimal placement of k points in $\mathbb{R}$ to minimise the expected minimum dis-utility among these k points measured with respect to some user u sampled from a Laplace distribution. As we shall prove in Theorem 2.10 that it bounds the cost of the Laplace noise addition mechanism.

$$\hat{f}(\epsilon, k) = \min_{\mathbf{a}\in\mathbb{R}^k} \mathop{\mathbb{E}}_{y\sim\mathcal{L}_\epsilon(0)} \left[\min_{a\in\texttt{Set}(\mathbf{a})} \mathfrak{h}(|y-a|)\right] \tag{15}$$

In this section, we first define a mechanism in Definition 2.9 which simulates the action of the server corresponding to the Laplace noise addition mechanism in Section 2.3.1 and show that the cost of Laplace noise addition mechanism is $\hat{f}(\epsilon, k)$. We finally show the optimality of Laplace noise addition mechanism via dual fitting i.e. constructing a feasible solution to the dual DILP $\mathcal{E}$ with an objective function $\hat{f}(\epsilon, k)$ in Section 2.3.2.

[10] Note that the variable $\nu(r,\mathbf{v})$ is constructed from the difference of two non-negative variables corresponding to third and fourth equations, respectively. The detailed proof is in Appendix B.6,

2.3.1 Bounding cost function $f(\epsilon, k)$ by the cost of Laplace noise adding mechanism

We now define the mechanism $\hat{\mathbf{P}}^{\mathcal{L}_\epsilon} = \{\hat{P}_u^{\mathcal{L}_\epsilon}\}_{u\in\mathbb{R}}$ which corresponds to simulating the action of the server on receiving signal $S_u \sim \mathcal{L}_\epsilon(u)$ from user u. We often call this in short as the Laplace noise addition mechanism.

Definition 2.9. The distribution $\hat{P}_u^{\mathcal{L}_\epsilon}$ is defined as follows for every $u \in \mathbb{R}$.

$$\hat{\mathbf{a}} \sim \hat{P}_u^{\mathcal{L}_\epsilon} \iff \hat{\mathbf{a}} = \arg\min_{\mathbf{a}\in\mathbb{R}^k} \mathop{\mathbb{E}}_{y\sim\mathcal{L}_\epsilon(S_u)} \left[\min_{a\in\texttt{Set}(\mathbf{a})} \mathfrak{h}(|y-a|)\right] \text{ where } S_u \sim \mathcal{L}_\epsilon(u) \tag{16}$$

$$\overset{(a)}{=} \arg\min_{\mathbf{a}\in\mathbb{R}^k} \mathop{\mathbb{E}}_{y\sim\mathcal{L}_\epsilon(0)} \left[\min_{a\in\texttt{Set}(\mathbf{a})} \mathfrak{h}(|y-a|)\right] + S_u \text{[11]where } S_u \sim \mathcal{L}_\epsilon(u) \tag{17}$$

Equality (a) follows from the fact that $y \sim \mathcal{L}_\epsilon(z) \implies y - z \sim \mathcal{L}_\epsilon(0)$ for every $z \in \mathbb{R}$.

Observe that the server responds with set of points $\texttt{Set}(\mathbf{a})$ for some $\mathbf{a} \in \mathbb{R}^k$ so as to minimise the expected cost with respect to some user sampled from a Laplace distribution centred at S_u.

We show that the following lemma which states that $\hat{P}^{\mathcal{L}_\epsilon}$ satisfies ϵ-geographic differential privacy constraints and bound $f(\epsilon, k)$ by $\hat{f}(\epsilon, k)$.

Lemma 2.10 (detailed proof in Appendix B.3)**.** *$\hat{\boldsymbol{P}}^{\mathcal{L}_\epsilon}$ satisfies ϵ-geographic differential-privacy constraints i.e. $\hat{\boldsymbol{P}}^{\mathcal{L}_\epsilon} \in \mathcal{P}_{\mathbb{R}^k}^{(\epsilon)}$ and thus, we have $f(\epsilon, k) \le cost(\hat{P}^{\mathcal{L}_\epsilon}) = \hat{f}(\epsilon, k)$*

Proof Sketch. Observe that $\hat{P}^{\mathcal{L}_\epsilon} \in \mathcal{P}_{\mathbb{R}^k}^{(\epsilon)}$ from the post processing theorem, refer to [27] since $S_u \sim \mathcal{L}_\epsilon(u)$ satisfies ϵ-geographic differential privacy constraints.[12]. Thus, we prove $f(\epsilon, k) \le \text{cost}(\hat{P}^{\mathcal{L}_\epsilon})$. The equality is fully proven in Appendix B.3.

□

2.3.2 Obtaining a feasible solution to DILP $\mathcal{E}$

We now construct feasible solutions to DILP $\mathcal{E}$. For some $\zeta > 0$ and $\hat{\lambda} > 0$, we define

$$\delta^{(c)}(r) = (\zeta/2)e^{-\zeta|r|} \text{ and } \lambda^{(c)}(r) = \hat{\lambda} \cdot (\zeta/2)e^{-\zeta|r|} \ \forall r \in \mathbb{R} \tag{18}$$

Now define $v_{med} = \text{Median}(\texttt{Set}(\mathbf{v}))$ and for every $\mathbf{v} \in \mathbb{R}^k$, we consider the following Differential Equation (19) in $\hat{\nu}(.)$.

$$-\left[\min_{a\in\texttt{Set}(\mathbf{v})} \mathfrak{h}(|r-a|)\right] \delta^{(c)}(r) + \lambda^{(c)}(r) + \frac{d\hat{\nu}(r)}{dr} + \epsilon|\hat{\nu}(r)| = 0; \text{ with } \hat{\nu}(v_{med}) = 0 \tag{19}$$

Observe that this equation precisely corresponds to the second constraint of DILP $\mathcal{E}$ (inequality replaced by equality) with an initial value. We now show that a solution $\hat{\nu}(.)$ to differential equation (19) exists such that $\hat{\nu}(r)$ is non-negative for sufficiently large r and non-positive for sufficiently small r to satisfy the last two constraints of DILP $\mathcal{E}$ in Lemma 2.11.

Observe that the structure of our differential equation is similar to that in [43, Equation 19]. However, our differential equation has significantly more complexity since we are minimising over a set of points $\mathbf{v} \in \mathbb{R}^k$ and also our equation has to be solved for every $\mathbf{v} \in \mathbb{R}^k$ making it more complex.

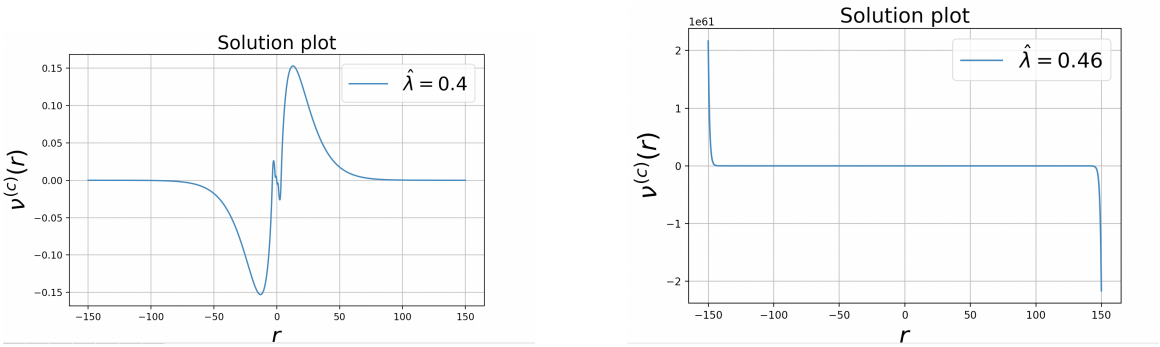
[11] Observe that we choose a deterministic tie-breaking rule amongst all vectors minimising this objective.

[12] Post processing theorem can be proven even for ϵ-geographic differential privacy similarly

Lemma 2.11 (Proof in Appendix B.4.2)**.** *Choose $\zeta < \epsilon$ and $0 < \hat{\lambda} \leq \frac{\epsilon-\zeta}{\epsilon+\zeta}\hat{f}(\epsilon+\zeta, k)$, then equation* (19) *has a unique $\mathcal{C}^1$ solution $\nu^{(c)}(.)$ and there exists $U, L \in \mathbb{R}$ satisfying $\nu^{(c)}(r) \geq 0\ \forall r \geq U$ and $\nu^{(c)}(r) \leq 0\ \forall r \leq L$.*

Intuitive explanation. We just give an intuition for this proof for the case where $\hat{\lambda}$ exceeds $\frac{\epsilon-\zeta}{\epsilon+\zeta}\hat{f}(\epsilon, k)$ by showing two plots in Figure 3a and 3b for the two cases where $\hat{\lambda} < \frac{\epsilon-\zeta}{\epsilon+\zeta}\hat{f}(\epsilon, k)$ and $\hat{\lambda} > \frac{\epsilon-\zeta}{\epsilon+\zeta}\hat{f}(\epsilon, k)$ respectively. In the first case, $\nu^{(c)}(r)$ is positive for sufficiently large r and in second case, it goes negative for large r demonstrating the requirement of the bound $\frac{\epsilon-\zeta}{\epsilon+\zeta}\hat{f}(\epsilon, k)$ on $\hat{\lambda}$.

The two plots are for the case when $\epsilon = 1$, $\zeta = 0.1$, $\mathfrak{h}(z) = z$ and thus $\frac{\epsilon-\zeta}{\epsilon+\zeta}\hat{f}(\epsilon, k)$ may be approximately by $\frac{9}{11} \times \frac{1}{2} = 0.44$ as shown in Section 2.4. For the purpose of the plots, we choose $\mathbf{v} = [-\log 4;\ 0;\ \log 4]^T$[13] and demonstrate the point in the Lemma.



(a) Solution for $\hat{\lambda} = 0.40$

(b) Solution for $\hat{\lambda} = 0.46$

Figure 3: Solutions for Differential Equation (19) for $\mathbf{v} = [-\log 4;\ 0;\ \log 4]^T$

These spikes in the solution may be observed due to the selection of $\mathbf{v} \in \mathbb{R}^3$ due to the term $\left[\min_{a \in \texttt{Set}(\mathbf{v})} \mathfrak{h}(|r-a|)\right]$ in the differential equation. □

Lemma 2.12 (detailed proof in Appendix B.4.2)**.** *opt*$(\mathcal{E}) \geq \hat{f}(\epsilon, k)$.

We present a proof sketch where we do not explicitly show the continuity of the bounds $U(.)$ and $L(.)$. In Appendix B.7, we prove a claim showing the existence of such continuous bounds.

Proof Sketch. Recall the functions $\lambda^{(c)}(.), \delta^{(c)}(.)$ defined in (18). Also for every $\mathbf{v} \in \mathbb{R}^k$, we obtain a function $\nu^{(c)}(., \mathbf{v})$ [solution of Equation (19)] with bounds $U(\mathbf{v})$ and $L(\mathbf{v})$ satisfying $\nu^{(c)}(r, \mathbf{v}) \geq 0\ \forall u \geq U(\mathbf{v})$ and $\nu^{(c)}(r, \mathbf{v}) \leq 0\ \forall u \leq L(\mathbf{v})$ and this solution is feasible.

The objective value of this feasible solution is $\hat{\lambda}$ and the constructed solution is feasible for any $\hat{\lambda} \leq \frac{\epsilon-\zeta}{\epsilon+\zeta}\hat{f}(\epsilon+\zeta, k)$ and $\zeta > 0$. Now, since $\hat{f}(\epsilon, k)$ is continuous in ϵ, choosing ζ to be arbitrarily small enables us to obtain the objective value of the solution arbitrarily close to $\hat{f}(\epsilon, k)$ and thus, opt$(\mathcal{E}) \geq \hat{f}(\epsilon, k)$. □

Observe that although we defined the Laplace noise addition mechanism ($\hat{\mathbf{P}}^{\mathcal{L}_\epsilon}$) (see Definition 2.9) entirely in terms of the user's action, we can consider an alternate mechanism splitting ($\hat{\mathbf{P}}^{\mathcal{L}_\epsilon}$) into user's action and server's response attaining the same cost:

[13] We choose this vector since it minimises equation (15) and a detailed calculation is given in Section 2.4.

- User u sends $S_u \sim \mathcal{L}_\epsilon(u)$ to the server.
- The server on receiving S_u responds with a vector $\mathbf{a} = \arg\min_{\mathbf{a}\in\mathbb{R}^k} \mathop{\mathbb{E}}_{y\sim\mathcal{L}_\epsilon(S_u)} \left[\min_{a\in\texttt{Set}(\mathbf{a})} \mathfrak{h}(|y-a|)\right]$.

Theorem 2.13. *For ϵ-geographic differential privacy, sending Laplace noise, that is, user u sends a signal drawn from distribution $\mathcal{L}_\epsilon(u)$ is one of the optimal choices of $\mathcal{P}_Z^{(\epsilon)}$ for users, and in this case $f^{\mathfrak{h}(.)}(\epsilon,k) = \hat{f}(\epsilon,k)$.*

Proof. Combining the results in Lemmas 2.10, 2.5, 2.12 and Theorems 2.8 and Theorem 2.2, we obtain $\hat{f}(\epsilon,k) \le \text{opt}(\mathcal{E}) \le \text{opt}(\mathcal{O}) \le \hat{f}(\epsilon,k)$ where $\hat{f}(\epsilon,k)$ denotes the cost of Laplace noise addition mechanism $\hat{P}^{\mathcal{L}_\epsilon}$ i.e. $cost(\hat{P}^{\mathcal{L}_\epsilon}) = \hat{f}(\epsilon,k)$.

□

2.4 Server response given the user sends Laplace Noise

Recall that we proved in Theorem 2.13 that the Laplace noise addition mechanism is an optimal action for the users. We now focus on the construction of an optimal server action on receiving the signal s from an user.

1. User with value $v \in \mathbb{R}$ reports s after adding Laplace noise of scale $\frac{1}{\epsilon}$. More formally, s is drawn from Laplace distribution $\mathcal{L}_\epsilon(v)$.
2. The server receives s and respond $(s+a_1, \ldots, s+a_k)$, where $a_1, \ldots, a_k$ are fixed real values.

For the case of $\mathfrak{h}(t) = t$, the optimal mechanism is simple enough that the values $a_1, a_2, \ldots, a_k$ can be computed by dynamic programming, as we sketch in the following subsection, and this concludes the proof of Theorem 1.4. For other general increasing functions, the optimal solution for $\{a_i\}_{i=1}^k$ may not always be written in closed form, however we can always write a recursive expression to compute the points.

2.4.1 Sketch of server response for odd k

We now show the optimal choice of A to optimize cost function $\hat{f}(\epsilon,k)$ [in Equation (15)]. Specifically, we assume odd k in this section. The solution for even k (refer Theorem B.3) can be constructed using a similar induction where the base case for $k=2$ can be directly optimized.

Assuming the symmetry of A, let $A = \{-y_{b-1}, \ldots - y_1, 0, y_1, \ldots, y_{b-1}\}$, where $y_1, \ldots, y_{b-1}$ are positive numbers in increasing order. We will construct the set $y_1, \ldots, y_{b-1}$ inductively.Let x be a random variable drawn from Laplace distribution $\mathcal{L}_\epsilon(0)$ with parameter ϵ, and the goal is to minimize $D_b = \mathbb{E}_{x\sim\mathcal{L}_\epsilon(0)} \left[\min_{a\in A} |x-a|\right]$. Since the density function of $\mathcal{L}_\epsilon(0)$ satisfies $\rho_{\mathcal{L}_\epsilon(0)}(x) = \rho_{\mathcal{L}_\epsilon(0)}(-x)$, we have

$$D_b = \mathbb{E}_{x\sim\mathcal{L}_\epsilon(0)} \left[\min_{a\in A} \mathfrak{h}(|x-a|) \middle| x\right] > 0,$$

i.e. the user has a positive private value. Under this conditioning, the variable x is an exponential random variable of mean 1. In this case, the search result being used by the server will be one of $y_0, y_1, \ldots, y_{b-1}$. Clearly, $D_1 = 1$. To compute D_{b+1}, let $s = y_1$.

Then using the memorylessness property of exponential random variables, we get the recurrence

$$D_{b+1} = \int_{t=0}^{s} \min\{\mathfrak{h}(t), \mathfrak{h}(s-t)\} e^{-t} dt + e^{-s} D_b$$

$$= \int_{t=0}^{s/2} \mathfrak{h}(t)e^{-t}dt + s\int_{t=s/2}^{s} e^{-t}dt - \int_{t=s/2}^{s} \mathfrak{h}(t)e^{-t}dt + e^{-s}D_b.$$

The optimal D_{b+1} given D_b can be computed by minimising over all $s \in \mathbb{R}$. However, for the case where $\mathfrak{h}(.)$ is an identity function, we may give a closed form expression below.

$$\begin{aligned}
D_{b+1} &= \int_{t=0}^{s/2} te^{-t}dt + s\int_{t=s/2}^{s} e^{-t}dt - \int_{t=s/2}^{s} te^{-t}dt + e^{-s}D_b \\
&= \left(1 - (s/2)e^{-s/2} - e^{-s/2}\right) + s\left(e^{-s/2} - e^{-s}\right) - \left((s/2)e^{-s/2} + e^{-s/2} - se^{-s} - e^{-s}\right) + e^{-s}D_b \\
&= 1 - 2e^{-s/2} + e^{-s} + e^{-s}D_b \\
&= \left(1 - e^{-s/2}\right)^2 + \left(e^{-s/2}\right)^2 D_b
\end{aligned}$$

Setting $\gamma = e^{-s/2}$, and minimizing by taking derivatives, we get $-2(1-\gamma) + 2\gamma D_b = 0$ which in turn gives

$$\gamma = \frac{1}{D_b + 1} \qquad \text{and} \qquad D_{b+1} = \frac{D_b}{D_b + 1}.$$

Plugging in the inductive hypothesis of $D_b = 1/b$, we get $D_{b+1} = 1/(b+1)$. Further, we get $s = 2\ln(1 + 1/b)$. Thus, by returning $k = 2b - 1$ results, the expected "cost of privacy" can be reduced by a factor of b. To obtain the actual positions $y_1, .., y_{b-1}$ we have to unroll the induction. For $i = 1, \ldots, b-1$, the position y_i is given by:

$$y_i = y_{i-1} + 2\ln(1 + 1/(b-i)).$$

Based on the above arguments in the four sections, we have the main theorem 1.4.

3 Further extensions

We describe some additional results below.

- When the user is not able to perform the optimal action, we show in Appendix A.6 that $\text{cost}^{\mathbb{1}(.)}(Z, \mathbf{P}, \mathbf{Q}) = O(\frac{\log k}{k\epsilon})$ for an appropriate server response $\mathbf{Q}$[14] if the user's action $\mathbf{P}$ consists of adding symmetric noise whose distribution satisfies log-concave property[15]. Observe that this property is satisfied by most natural distributions like Exponential and Gaussian.

- In practice, the set of users may not belong to $\mathbb{R}$ but in many cases may have a feature vector embedding in $\mathbb{R}^d$. In this scenario, a server could employ dimensionality reduction techniques such as Principal Component Analysis (PCA) to create a small number d' of dimensions which have the strongest correlation to the dis-utility of a hypothetical user with features identical to the received signal. The server may project the received signal only along these dimensions to select the set of k results. In this case, we show that $\text{cost}^{\mathbb{1}(.)}(Z, \mathbf{P}, \mathbf{Q}) = O\left(\frac{1}{\epsilon k^{1/d'}}\right)$ under some assumptions as discussed Appendix A.7 when the user's action $\mathbf{P}$ consists of adding independent Gaussian noise to every feature.

[14] The server's action $\mathbf{Q}$ involves sampling from the posterior of the noise distribution.

[15] If the random noise with log-concave distribution g is given by Y, then we have $\mathbb{E}[Y^+ \cup \{0\}] = \frac{1}{\epsilon}$ and $g(y) = g(-y)$.

- We further show that Laplace noise continues to be an optimal noise distribution for the user even under a relaxed definition of geographic differential privacy (defined in Definition 3.2) in Section 3.1. This definition captures cases when privacy guarantees are imposed only when the distance between users is below some threshold (recall from Section 1.3.3 that such a setup was studied in [35]).

3.1 A generalization of Geographic differential privacy

Here we consider a generalization of ϵ-geographic differential privacy and define $\mathfrak{g}(.)$-geographic differential privacy for an increasing convex function $\mathfrak{g}(.)$ satisfying Assumption 3.1.

Assumption 3.1. $\mathfrak{g}(.)$ is a increasing convex function satisfying $\mathfrak{g}(0) = 0$ and $\mathfrak{g}(.)$ is differentiable at 0 with $\mathfrak{g}'(0) \neq 0$.

Definition 3.2 (alternate definition of geo-DP)**.** Let $\epsilon > 0$ be a desired level of privacy and let $\mathcal{U}$ be a set of input data and $\mathcal{Y}$ be the set of all possible responses and $\Delta(\mathcal{Y})$ be the set of all probability distributions (over a sufficiently rich σ-algebra of $\mathcal{Y}$ given by $\sigma(\mathcal{Y})$). For any $\mathfrak{g}(.)$ satisfying Assumption 3.1 a mechanism $Q : u \to \Delta(\mathcal{Y})$ is $\mathfrak{g}(.)$-geographic differentially private if for all $S \in \sigma(\mathcal{Y})$ and $u_1, u_2 \in \mathcal{U}$:

$$\mathbb{P}(Qu_1 \in S) \leq e^{\mathfrak{g}(|u_1 - u_2|)} \mathbb{P}(Qu_2 \in S).$$

Since this definition allows the privacy guarantee to decay non-linearly with the distance between the user values, it is a relaxation of ϵ-geographic DP as defined in Definition 1.2. Observe that this definition captures cases where the privacy guarantees exist only when the distance between users is below some threshold by defining $\mathfrak{g}(t)$ to be ∞ if $t > T_0$ for some threshold T_0.

Under this notion of differential privacy, we may redefine cost function $f^{\text{alt},\mathfrak{h}(.)}(\epsilon, k)$ as follows.

$$f^{\text{alt},\mathfrak{h}(.)}(\mathfrak{g}(.), k) := \inf_{Z} \inf_{\mathbf{P} \in \mathcal{P}_Z^{\mathfrak{g}(.)}} \inf_{\mathbf{Q} \in \mathcal{Q}_Z} \sup_{u \in \mathbb{R}} \mathop{\mathbb{E}}_{s \sim \mathbf{P}_u} \left[\mathop{\mathbb{E}}_{\mathbf{a} \sim \mathbf{Q}_s} \left[\min_{a \in \texttt{Set}(\mathbf{a})} \mathfrak{h}\left(|u - a|\right) \right] \right],$$

where $\mathcal{P}_Z^{\mathfrak{g}(.)} := \{\mathbf{P} | \forall u \in \mathbb{R}, P_u$ is a distribution on Z, and $\mathfrak{g}(.)$-geographic differential privacy is satisfied$\}$. The definition of $\mathcal{Q}_Z$ are similar to that in Section 1.1.2.

We now show that adding Laplace noise continues to remain an optimal action for the users even under this relaxed model of geographic differential privacy.

Theorem 3.3. *For $\mathfrak{g}(.)$-geographic differential privacy, adding Laplace noise, whose density function is $\rho(x) = \frac{\mathfrak{g}'(0)}{2} \cdot e^{-\mathfrak{g}'(0)|x|}$, is one of the optimal choices of $\mathcal{P}_Z^{\mathfrak{g}(.)}$ for users. Further, when $\mathfrak{h}(z) = z$, we have $f^{alt,\mathfrak{h}(.)}(\mathfrak{g}(.), k) = O\left(\frac{1}{\mathfrak{g}'(0)k}\right)$ and the optimal mechanism (choice of actions of users and server) itself can be computed in closed form.*

Proof. The proof of this theorem follows identically to that of Theorem 1.4. However, we require a slight modification of Lemma 2.6 to prove it as stated and proven in Lemma 3.4. □

Lemma 3.4. *Suppose, P_u has a probability density function given by $g(u, .) : \mathbb{R}^k \to \mathbb{R}$ for every $u \in \mathbb{R}$. Then, P satisfies $\mathfrak{g}(.)$-geographic differential privacy iff $\max(|\overline{g}_u(u, \boldsymbol{x})|, |\underline{g}_u(u, \boldsymbol{x})|) \leq \mathfrak{g}'(0) g(u, \boldsymbol{x})$ $\forall u \in \mathbb{R}; \forall x \in \mathbb{R}^k$ whenever $\mathfrak{g}(.)$ satisfies Assumption 3.1.*

The proof of this Lemma is very similar to that of Lemma 2.6 and proven in Section B.2.2.

4 Conclusion

We have defined a new architecture for differential privacy with a small number of multiple selections, and shown in a stylized model that significant improvements in the privacy-accuracy tradeoffs are indeed possible. Our model ignores some practical considerations, namely, the client's request lives in a high dimensional feature space (and not in one-dimension), and the server has a machine learning model to evaluate the quality of a result that it needs to convey to the client in some compressed form. Addressing these issues while preserving the privacy-accuracy trade-off either theoretically or empirically, will be the focus of future work.

References

[1] Duality in linear programming. `https://sites.math.washington.edu//~burke/crs/407/notes/section4.pdf`. Accessed: 2010-09-30.

[2] EJ Anderson. A new continuous model for job-shop scheduling. *International journal of systems science*, 12(12):1469–1475, 1981.

[3] E.J. Anderson and P. Nash. *Linear Programming in Infinite-dimensional Spaces: Theory and Applications*. A Wiley-Interscience publication. Wiley, 1987. URL: `https://books.google.com/books?id=O2VRAAAAMAAJ`.

[4] Miguel E Andrés, Nicolás E Bordenabe, Konstantinos Chatzikokolakis, and Catuscia Palamidessi. Geo-indistinguishability: Differential privacy for location-based systems. In *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*, pages 901–914, 2013.

[5] Elliot Anshelevich and John Postl. Randomized social choice functions under metric preferences, 2016. `arXiv:1512.07590`.

[6] Apple. Learning with privacy at scale. *Apple Machine Learning Journal*, 1, 2017. `https://machinelearning.apple.com/2017/12/06/learning-with-privacy-at-scale.html`.

[7] Brendan Avent, Yatharth Dubey, and Aleksandra Korolova. The power of the hybrid model for mean estimation. *The 20th Privacy Enhancing Technologies Symposium (PETS)*, 2020.

[8] Brendan Avent, Aleksandra Korolova, David Zeber, Torgeir Hovden, and Benjamin Livshits. BLENDER: Enabling local search with a hybrid differential privacy model. In *26th USENIX Security Symposium (USENIX Security 17); Journal of Privacy and Confidentiality*, 2019.

[9] M. Bagnoli and T. Bergstrom. Log-Concave Probability And Its Applications. Papers 89-23, Michigan - Center for Research on Economic & Social Theory, 1989. URL: `https://ideas.repec.org/p/fth/michet/89-23.html`.

[10] Kenneth R Baker. Introduction to sequencing and scheduling. *(No Title)*, 1974.

[11] Raef Bassily, Albert Cheu, Shay Moran, Aleksandar Nikolov, Jonathan Ullman, and Steven Wu. Private query release assisted by public data. In *International Conference on Machine Learning*, pages 695–703. PMLR, 2020.

[12] Raef Bassily, Kobbi Nissim, Uri Stemmer, and Abhradeep Guha Thakurta. Practical locally private heavy hitters. *Advances in Neural Information Processing Systems*, 30, 2017.

[13] Amitabh Basu, Kipp Martin, and Christopher Thomas Ryan. Strong duality and sensitivity analysis in semi-infinite linear programming, 2015. arXiv:1510.07531.

[14] Björn Bebensee. Local differential privacy: a tutorial. *arXiv preprint arXiv:1907.11908*, 2019.

[15] Amos Beimel, Aleksandra Korolova, Kobbi Nissim, Or Sheffet, and Uri Stemmer. The power of synergy in differential privacy: Combining a small curator with local randomizers. In *Information-Theoretic Cryptography (ITC)*, 2020.

[16] R. Bellman. *Dynamic Programming*. Princeton Landmarks in Mathematics and Physics. Princeton University Press, 2010. URL: https://books.google.co.in/books?id=92aYDwAAQBAJ.

[17] Avrim Blum, Katrina Ligett, and Aaron Roth. A learning theory approach to noninteractive database privacy. *Journal of the ACM (JACM)*, 60(2):1–25, 2013.

[18] Hai Brenner and Kobbi Nissim. Impossibility of differentially private universally optimal mechanisms. *SIAM Journal on Computing*, 43(5):1513–1540, 2014.

[19] Mark Bun, Jelani Nelson, and Uri Stemmer. Heavy hitters and the structure of local privacy. *ACM Transactions on Algorithms (TALG)*, 15(4):1–40, 2019.

[20] Ioannis Caragiannis, Nisarg Shah, and Alexandros A. Voudouris. The metric distortion of multiwinner voting. *Artificial Intelligence*, 313:103802, 2022. URL: https://www.sciencedirect.com/science/article/pii/S0004370222001424, doi:https://doi.org/10.1016/j.artint.2022.103802.

[21] Rui Chen, Haoran Li, A Kai Qin, Shiva Prasad Kasiviswanathan, and Hongxia Jin. Private spatial data aggregation in the local setting. In *2016 IEEE 32nd International Conference on Data Engineering (ICDE)*, pages 289–300. IEEE, 2016.

[22] Albert Cheu, Adam Smith, Jonathan Ullman, David Zeber, and Maxim Zhilyaev. Distributed differential privacy via shuffling. In *Advances in Cryptology–EUROCRYPT 2019: 38th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Darmstadt, Germany, May 19–23, 2019, Proceedings, Part I 38*, pages 375–403. Springer, 2019.

[23] Richard Walter Conway. " theory of scheduling,". *Addison Wesley*, 1967.

[24] Bolin Ding, Janardhan Kulkarni, and Sergey Yekhanin. Collecting telemetry data privately. *Advances in Neural Information Processing Systems*, 30, 2017.

[25] John C. Duchi, Michael I. Jordan, and Martin J. Wainwright. Local privacy and statistical minimax rates. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pages 429–438, 2013. doi:10.1109/FOCS.2013.53.

[26] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography: Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006. Proceedings 3*, pages 265–284. Springer, 2006.

[27] Cynthia Dwork and Aaron Roth. 2014. doi:10.1561/0400000042.

[28] Cynthia Dwork, Aaron Roth, et al. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407, 2014.

[29] Úlfar Erlingsson, Vitaly Feldman, Ilya Mironov, Ananth Raghunathan, Kunal Talwar, and Abhradeep Thakurta. Amplification by shuffling: From local to central differential privacy via anonymity. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2468–2479. SIAM, 2019.

[30] Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. RAPPOR: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, CCS '14, pages 1054–1067, New York, NY, USA, 2014. ACM. URL: `http://doi.acm.org/10.1145/2660267.2660348`.

[31] Lawrence C Evans. An introduction to mathematical optimal control theory version 0.2. *Lecture notes available at http://math. berkeley. edu/˜ evans/control. course. pdf*, 1983.

[32] Vitaly Feldman, Audra McMillan, and Kunal Talwar. Hiding among the clones: A simple and nearly optimal analysis of privacy amplification by shuffling. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 954–964. IEEE, 2022.

[33] Natasha Fernandes, Annabelle McIver, and Carroll Morgan. The laplace mechanism has optimal utility for differential privacy over continuous queries. In *2021 36th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*, pages 1–12. IEEE, 2021.

[34] K. O. Friedrichs. The identity of weak and strong extensions of differential operators. *Transactions of the American Mathematical Society*, 55(1):132–151, 1944. URL: `http://www.jstor.org/stable/1990143`.

[35] Quan Geng and Pramod Viswanath. The optimal noise-adding mechanism in differential privacy. *IEEE Transactions on Information Theory*, 62(2):925–951, 2015.

[36] Robin C. Geyer, Tassilo J. Klein, and Moin Nabi. Differentially private federated learning: A client level perspective. arXiv preprint arXiv:1712.07557, 2017.

[37] Arpita Ghosh, Tim Roughgarden, and Mukund Sundararajan. Universally utility-maximizing privacy mechanisms. *SIAM Journal on Computing*, 41(6):1673–1693, 2012.

[38] Vasilis Gkatzelis, Daniel Halpern, and Nisarg Shah. Resolving the optimal metric distortion conjecture. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1427–1438. IEEE, 2020.

[39] Saikat Guha, Bin Cheng, and Paul Francis. Privad: Practical privacy in online advertising. In *USENIX conference on Networked systems design and implementation*, pages 169–182, 2011.

[40] Mangesh Gupte and Mukund Sundararajan. Universally optimal privacy mechanisms for minimax agents. In *Proceedings of the twenty-ninth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 135–146, 2010.

[41] Peter Kairouz, Sewoong Oh, and Pramod Viswanath. Extremal mechanisms for local differential privacy. *Advances in neural information processing systems*, 27, 2014.

[42] Fatih Erdem Kizilkaya and David Kempe. Plurality veto: A simple voting rule achieving optimal metric distortion. *arXiv preprint arXiv:2206.07098*, 2022.

[43] Fragkiskos Koufogiannis, Shuo Han, and George J Pappas. Optimality of the laplace mechanism in differential privacy. *arXiv preprint arXiv:1504.00065*, 2015.

[44] Camillo De Lellis. Cauchy-Lipschitz theorem. https://encyclopediaofmath.org/wiki/Cauchy-Lipschitz_theorem.

[45] Jingcheng Liu and Kunal Talwar. Private selection from private candidates. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 298–309, 2019.

[46] Frank McSherry and Kunal Talwar. Mechanism design via differential privacy. In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)*, pages 94–103. IEEE, 2007.

[47] Nicolas Papernot, Shuang Song, Ilya Mironov, Ananth Raghunathan, Kunal Talwar, and Úlfar Erlingsson. Scalable private learning with pate, 2018. arXiv:1802.08908.

[48] Malcolm C. Pullan. An algorithm for a class of continuous linear programs. *SIAM Journal on Control and Optimization*, 31(6):1558–1577, 1993. arXiv:https://doi.org/10.1137/0331073, doi:10.1137/0331073.

[49] W. Rudin. *Principles of Mathematical Analysis*. International series in pure and applied mathematics. McGraw-Hill, 1976. URL: https://books.google.com/books?id=kwqzPAAACAAJ.

[50] Vincent Toubiana, Arvind Narayanan, Dan Boneh, Helen Nissenbaum, and Solon Barocas. Adnostic: Privacy preserving targeted advertising. *NDSS*, 2010.

[51] N.T. Vinh, D.S. Kim, N.N. Tam, and N.D. Yen. Duality gap function in infinite dimensional linear programming. *Journal of Mathematical Analysis and Applications*, 437(1):1–15, 2016. URL: https://www.sciencedirect.com/science/article/pii/S0022247X15011762, doi:https://doi.org/10.1016/j.jmaa.2015.12.043.

[52] Tianhao Wang, Jeremiah Blocki, Ninghui Li, and Somesh Jha. Locally differentially private protocols for frequency estimation. In *26th USENIX Security Symposium (USENIX Security 17)*, pages 729–745, 2017.

[53] Stanley L Warner. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309):63–69, 1965.

[54] Mariusz Wodzicki. Notes on Reinmann Integral. https://math.berkeley.edu/~wodzicki/H104.F10/Integral.pdf.

A Extensions to the models and more detailed model description

A.1 Model and Overview (With More Details)

A.1.1 Notations

We use $[m]$ to denote the set $\{1, 2, \ldots, m\}$, for any non-negative integer m. $\mathbb{N}$ is the set of positive integers, $\mathbb{Q}$ is the set of rational real numbers, and $\mathbb{R}$ is the set of real numbers. We denote $\mathcal{A} = \mathbb{R}^k$

$\mathbb{E}[V]$ denotes the expectation of real random variable V. $\mathbb{P}(E)$ denotes the probability of an event E. $\mathbb{I}\{E\}$ is the indicator, which has value 1 when the event E happens and has value 0 when E does not happen, and thus we have $\mathbb{P}(E) = \mathbb{E}[\mathbb{I}\{E\}]$.

For any set S, $\mathbb{D}(S)$ is defined as the set of probability measures on S.

We sometimes have integral operations. Since we are optimizing the objective over all possible mechanisms, some functions may not be continuous, and some distributions may not have density functions, so we use the Lebesgue integral.

We use $\mathbb{E}_{X\sim\mu}[V(X)]$ to denote the expectation of $V(X)$ when the probability measure of X is μ, so

$$\mathbb{E}_{X\sim\mu}[V(X)] := \int_x V(x)\mu(dx).$$

Similarly, we use $\mathbb{P}_{X\sim\mu}(E(X))$ to denote the probability that event $E(X)$ happens when the probability measure of X is μ, so

$$\mathbb{P}_{X\sim\mu}(E(X)) := \mathbb{E}_{X\sim\mu}[\mathbb{I}\{E(X)\}] = \int_x \mathbb{I}\{E(x)\}\,\mu(dx).$$

When a probability measure μ has a probability density (or in other words is continuous), we use $\rho_\mu(x)$ to denote its probability density at x and the Lebesgue integral may be replaced by a Reimann integral.

Similarly, we use $\texttt{Set}$ to convert a vector in $\mathbb{R}^k$ to a set. More formally, for any $\mathbf{a}\in\mathbb{R}^k$, $\texttt{Set}(\mathbf{a})$ is denoted by $\{a_i : i\in[k]\}$.

A.2 Problem setting for the restricted setup (results and users lie in the same space)

Users send a signal to the server, and the server sends back k results A which we denote as a vector in $\mathbb{R}^k$. In this subsection, we give a measure-theoretic view of the mechanism $(Z,\mathbf{P},\mathbf{Q})$.

We aim to determine a mechanism with the following ingredients.

1. A set of signals Z.
2. Action of users $P_u(\mathcal{H})$, denoting the probability that user u sends signal $s\in\mathcal{H}$, for $u\in U$ and $\mathcal{H}\in\mathcal{F}_Z$, where $\mathcal{F}_Z$ is the set of all the measurable subsets of Z.
3. Action of server $Q_s(\mathcal{H})$, denoting the probability that server sends back results $A\in\mathcal{H}$ when receiving signal s, for $s\in Z$ and $\mathcal{H}\in\mathcal{F}_\mathcal{A}$, where $\mathcal{F}_Z$ is all the measurable subset of Z.

The system should have geographic differential privacy. For any $\mathcal{H}\in\mathcal{F}_Z$, for any $u_1,u_2\in U$, $P_{u_1}(\mathcal{H})\le e^{\epsilon\cdot|u_1-u_2|}P_{u_2}(\mathcal{H})$,

We want to minimize the distance between the user and the closest ad that the user receives, in the worst case with respect to the distribution of users. Formally, we want to compute

$$f^{\mathfrak{h}(.)}(\epsilon,k) := \inf_Z \inf_{\mathbf{P}\in\mathcal{P}_Z^{(\epsilon)},\mathbf{Q}\in\mathcal{Q}_Z} \sup_{u\in\mathbb{R}} \mathop{\mathbb{E}}_{s\sim P_u}\left[\mathop{\mathbb{E}}_{\mathbf{a}\sim Q_s}\left[\min_{a\in\texttt{Set}(\mathbf{a})}\mathfrak{h}(|u-a|)\right]\right],$$

where

1. Z can be any set.
2. When we are interested in geographic differential privacy on $\mathbb{R}$. We call $\mathop{\mathbb{E}}_{s\sim P_u,\mathbf{a}\sim Q_s}\left[\min_{a\in\texttt{Set}(\mathbf{a})}\mathfrak{h}(|u-a|)\right]$ the cost of user u from mechanism $(Z,\mathbf{P},\mathbf{Q})$
3. $\mathcal{P}_Z^{(\epsilon)}$ is the set of mechanism satisfying geographic differential privacy.
4. $\mathcal{Q}_Z := \{\mathbf{Q}|\forall s\in Z, Q_s(\cdot)\in\mathbb{D}(\mathcal{A})\}$, which is the set of available actions of the server.

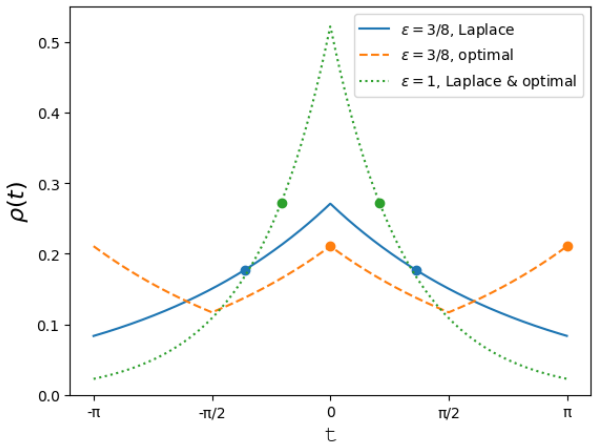


Figure 4: Geographic differential privacy setting when users and results are located on a unit ring, for $k = 2$ and $\epsilon \in \{3/8, 1\}$, showing the stark difference between Laplace noise and the optimal noise. Suppose the user has a private value u. Then the user sends $u + x$ to the server, where x is drawn from a noise distribution with density $\rho(t)$, depicted here for both Laplace noise and the optimal noise. Suppose the server receives s. Then the server's optimal response is $s + a_1$ and $s + a_2$, where the values of a_1, a_2 are the t-axis values of dots on the density functions, again assuming both Laplace noise and the optimal noise. Laplace is not optimal when $\epsilon = 3/8$, while Laplace is optimal when $\epsilon = 1$.

A.3 Summary of Results

In either model, one of the optimal mechanisms satisfies $H = \mathcal{A}$, which means the signals sent by users and servers can be drawn from the same set. In fact, the server will directly return the signal it receives. Furthermore, we do not need to consider all the distribution over users, we only care about the user that has the largest error.

Theorem A.1. *For ϵ-geographic differential privacy adding Laplace noise i.e. user u draws a signal from $\mathcal{L}_\epsilon(u)$ is one of the optimal choices for users, and in this case $f^{\mathfrak{h}(.)}(\epsilon, k)$ is $O(1/(\epsilon k))$ when $\mathfrak{h}(t) = t$. Furthermore, when $\mathfrak{h}(t) = t$, the optimal mechanism can be computed in closed form (Theorem B.3).*

A.4 Calculation of optimal mechanism on a ring for the case of $k = 2$

We calculate the optimal mechanism in geographic differential privacy setting, on a unit ring, when $\epsilon = 3/8$, and the number of results is $k = 2$. In this section, we define $d(u, a) = \langle u, a \rangle$

We use real numbers in $[-\pi, \pi)$ to denote users and results on a unit ring, and $\langle x, a \rangle$ denotes $|x - a|$. Figure 4 illustrates the optimal mechanism under geographic DP for $k = 2$. This mechanism uses noise that is a piece-wise composition of Laplace noises; we obtain a cost of 0.72 whereas Laplace noise gives a cost of 0.75.

To find the optimal mechanism for the case of the ring, we solve the DILP $\mathcal{O}$ using a linear program solver and obtain the plot shown in Figure 4 with cost of 0.72. However, when the user sends Laplace noise, the server on receiving signal z responds with two points $z + a_1$ and $z + a_2$ which maybe calculated by the following problem.

$$\min_{a_1 \in [-\pi,\pi), a_2 \in [-\pi,\pi)} \int_{-\pi}^{\pi} \min\{\langle x, a_1 \rangle, \langle x, a_2 \rangle\} \rho(x) dx,$$

where $\rho(x)$ is a density function for the Laplace distribution,

A.5 Restricted and Unrestricted Setup of the Multi-Selection model

Recall the setup in Section 1 where the users and results belonged to different sets $\mathbb{R}$ and M with the definition of dis-utility in Definition 1.3. In section 1.1.4, we considered an alternate setup where the users and results belonged to the same set $\mathbb{R}$ and the optimal result for an user u was the result u itself. In this section, we call these setups unrestricted and restricted respectively and define our "multi-selection" model separately for both these setups. Finally, we bound the cost function in the unrestricted setup by the cost function in the restricted setup in Theorem A.2 thus, showing that it is sufficient to consider the cost function in the restricted setup.

A.5.1 Unrestricted setup

Recall that results and users are located in sets M and $\mathbb{R}$ respectively and function $\text{OPT} : \mathbb{R} \to M$ maps every user to its optimal result(ad). Recall that the dis-utility of an user u from a result m is defined in Definition 1.3.

A.5.2 Restricted setup

This setup is very similar to the setup described except the fact that users and results(ads) lie on the same set $\mathbb{R}$. Recall from the description in Section 1.1.4, the dis-utility of an user $u \in \mathbb{R}$ from a result $a \in \mathbb{R}$ is given by $\mathfrak{h}(|u - a|)$ for some function $\mathfrak{h}(.)$ satisfying equations (2) and (3).

A.5.3 The space of server/user actions

Recall that the goal is to determine a mechanism that has the following ingredients:

1. A set of signals Z.
2. The action of users, which involves choosing a signal from a distribution over signals. We use P_u for $u \in \mathbb{R}$ to denote the distribution of the signals sent by user u. This distribution is supported on Z.
3. The distribution over actions of the server, Q_s when it receives $s \in Z$. This distribution denoting the distribution of the result set returned by the server given signal s may be supported on either $\mathbb{R}^k$ or M^k, for the restricted setup and unrestricted setup respectively.

The optimal mechanism is computed by jointly optimizing over the tuple $(Z, \mathbf{P}, \mathbf{Q})$.

And thus, we define the set of server responses by $\mathcal{Q}_{\text{unrestricted},Z}$ and $\mathcal{Q}_{\text{restricted},Z}$ for unrestricted and restricted setup respectively.

- $\mathcal{Q}_{\text{unrestricted},Z} := \{\mathbf{Q} | \forall s \in Z, Q_s \text{ is a distribution on } M^k\}$.
- $\mathcal{Q}_{\text{restricted},Z} := \{\mathbf{Q} | \forall s \in Z, Q_s \text{ is a distribution on } \mathbb{R}^k\}$.

In any *feasible geographic DP mechanism*, the user behavior should satisfy ϵ-geographic differential privacy: for any $u_1, u_2 \in \mathbb{R}$, it should hold that $P_{u_1}(S) \leq P_{u_2}(S)e^{\epsilon|u_1-u_2|}$ where S is any measurable subset of Z. For any fixed response size k, in order to maximize utility while ensuring the specified level of privacy, the goal is to minimize the disutility of the user from the result that the gives the user minimum dis-utility where the minimisation is the worst case user u in $\mathbb{R}$.

A.5.4 Cost functions in both the setups

For the unrestricted and restricted setups, we define the cost functions $f^{\mathfrak{h}(.)}_{\text{unrestricted}}(\epsilon, k)$ and $f^{\mathfrak{h}(.)}_{\text{unrestricted}}(\epsilon, k)$ respectively below. Recall that Z may denote any set.

$$f^{\mathfrak{h}(.)}_{\text{unrestricted}}(\epsilon, k) := \inf_{Z} \inf_{\substack{\mathbf{P}\in\mathcal{P}^{(\epsilon)}_Z \\ \mathbf{Q}\in\mathcal{Q}_{\text{unrestricted},Z}}} \sup_{u\in\mathbb{R}} \mathop{\mathbb{E}}_{s\sim P_u, \mathbf{a}\sim Q_s} \left[\min_{a\in\texttt{Set}(\mathbf{a})} \left(\text{Dis-util}^{\mathfrak{h}(.)}(u,a)\right)\right],$$

$$f^{\mathfrak{h}(.)}_{\text{restricted}}(\epsilon, k) := \inf_{Z} \inf_{\substack{\mathbf{P}\in\mathcal{P}^{(\epsilon)}_Z \\ \mathbf{Q}\in\mathcal{Q}_{\text{restricted},Z}}} \sup_{u\in\mathbb{R}} \mathop{\mathbb{E}}_{s\sim P_u, \mathbf{a}\sim Q_s} \left[\min_{a\in\texttt{Set}(\mathbf{a})} \mathfrak{h}\left(|u-a|\right)\right], \text{ where}$$

$\mathcal{P}^{(\epsilon)}_Z := \{\mathbf{P} | \forall u \in \mathbb{R}, P_u$ is a distribution on Z, and ϵ-geographic differential privacy is satisfied$\}$.
We state a theorem upper bounding $f^{\mathfrak{h}(.)}_{\text{unrestricted}}(\epsilon, k)$ by $f^{\mathfrak{h}(.)}_{\text{restricted}}(\epsilon, k)$.

Theorem A.2. *For any $\mathfrak{h}(.)$ satisfying equation (2), we have $f^{\mathfrak{h}(.)}_{unrestricted}(\epsilon, k) \leq f^{\mathfrak{h}(.)}_{restricted}(\epsilon, k)$.*

Proof. Recall that

$$f^{\mathfrak{h}(.)}_{\text{unrestricted}}(\epsilon, k) := \inf_{Z} \inf_{\substack{\mathbf{P}\in\mathcal{P}^{(\epsilon)}_Z \\ \mathbf{Q}\in\mathcal{Q}_{\text{unrestricted},Z}}} \sup_{u\in\mathbb{R}} \mathop{\mathbb{E}}_{s\sim P_u, \mathbf{a}\sim Q_s} \left[\min_{a\in\texttt{Set}(\mathbf{a})} \left(\text{Dis-util}^{\mathfrak{h}(.)}(u,a)\right)\right],$$

$$f^{\mathfrak{h}(.)}(\epsilon, k) := \inf_{Z} \inf_{\substack{\mathbf{P}\in\mathcal{P}^{(\epsilon)}_Z \\ \mathbf{Q}\in\mathcal{Q}_{\text{restricted},Z}}} \sup_{u\in\mathbb{R}} \mathop{\mathbb{E}}_{s\sim P_u, \mathbf{a}\sim Q_s} \left[\min_{a\in\texttt{Set}(\mathbf{a})} \mathfrak{h}\left(|u-a|\right)\right],$$

and

$$\text{Dis-util}^{\mathfrak{h}(.)}(u, m) := \inf_{u':OPT(u')=m} \mathfrak{h}(|u-u'|).$$

So we need to prove

$$\inf_{Z} \inf_{\substack{\mathbf{P}\in\mathcal{P}^{(\epsilon)}_Z \\ \mathbf{Q}\in\mathcal{Q}_{\text{unrestricted},Z}}} \sup_{u\in\mathbb{R}} \mathop{\mathbb{E}}_{s\sim P_u, \mathbf{a}\sim Q_s} \left[\min_{a\in\texttt{Set}(\mathbf{a})} \left(\inf_{u':OPT(u')=a} \mathfrak{h}(|u-u'|)\right)\right]$$
$$\leq \inf_{Z} \inf_{\substack{\mathbf{P}\in\mathcal{P}^{(\epsilon)}_Z \\ \mathbf{Q}\in\mathcal{Q}_{\text{restricted},Z}}} \sup_{u\in\mathbb{R}} \mathop{\mathbb{E}}_{s\sim P_u, \mathbf{a}\sim Q_s} \left[\min_{a\in\texttt{Set}(\mathbf{a})} \mathfrak{h}\left(|u-a|\right)\right]$$

It is sufficient to show, for any Z and $\mathbf{P} \in \mathcal{P}^{(\epsilon)}_Z$,

$$\inf_{\mathbf{Q}\in\mathcal{Q}_{\text{unrestricted},Z}} \sup_{u\in\mathbb{R}} \mathop{\mathbb{E}}_{s\sim P_u, \mathbf{a}\sim Q_s} \left[\min_{a\in\texttt{Set}(\mathbf{a})} \left(\inf_{u':OPT(u')=a} \mathfrak{h}(|u-u'|)\right)\right]$$

$$\leq \inf_{\mathbf{Q}\in\mathcal{Q}_{\text{restricted},Z}} \sup_{u\in\mathbb{R}} \mathop{\mathbb{E}}_{s\sim P_u,\mathbf{a}\sim Q_s} \left[\min_{a\in\texttt{Set}(\mathbf{a})} \mathfrak{h}\left(|u-a|\right)\right] \tag{20}$$

For $\mathbf{Q} \in \mathcal{Q}_{\text{restricted},Z}$, since $\text{Dis-util}^{\mathfrak{h}(.)}(u,m) := \inf_{u':OPT(u')=m} \mathfrak{h}(|u-u'|)$, we have

$$\mathop{\mathbb{E}}_{s\sim P_u,\mathbf{a}\sim Q_s} \left[\min_{a\in\texttt{Set}(\mathbf{a})} \mathfrak{h}\left(|u-a|\right)\right] \geq \mathop{\mathbb{E}}_{s\sim P_u,\mathbf{a}\sim Q_s} \left[\min_{a\in\texttt{Set}(\mathbf{a})} \text{Dis-util}^{\mathfrak{h}(.)}(u,\text{OPT}(a))\right].$$

$$\implies \sup_{u\in\mathbb{R}} \mathop{\mathbb{E}}_{s\sim P_u,\mathbf{a}\sim Q_s} \left[\min_{a\in\texttt{Set}(\mathbf{a})} \mathfrak{h}\left(|u-a|\right)\right] \geq \sup_{u\in\mathbb{R}} \mathop{\mathbb{E}}_{s\sim P_u,\mathbf{a}\sim Q_s} \left[\min_{a\in\texttt{Set}(\mathbf{a})} \text{Dis-util}^{\mathfrak{h}(.)}(u,\text{OPT}(a))\right]. \tag{21}$$

Given $\mathbf{Q} \in \mathcal{Q}_{\text{restricted},Z}$, we draw $\mathbf{a}$ from $\mathbf{Q}$, and let $\mathbf{b} = [\text{OPT}(a_1), \text{OPT}(a_2), \ldots, \text{OPT}(a_k)]^T$. Suppose the distribution of $\mathbf{b}$ is $\mathbf{Q}'$, and we have

$$\mathop{\mathbb{E}}_{s\sim P_u,\mathbf{a}\sim Q_s} \left[\min_{a\in\texttt{Set}(\mathbf{a})} \text{Dis-util}^{\mathfrak{h}(.)}(u,\text{OPT}(a))\right] = \mathop{\mathbb{E}}_{s\sim P_u,\mathbf{b}\sim Q'_s} \left[\min_{\mathbf{b}\in\texttt{Set}(\mathbf{b})} \text{Dis-util}^{\mathfrak{h}(.)}(u,b)\right]. \tag{22}$$

Note that $\mathbf{Q}' \in \mathcal{Q}_{\text{unrestricted},Z}$, so combining Inequality 21 and Equality 22, we have Equation 20, which finishes the proof.

□

And thus, it is sufficient to study $f^{\mathfrak{h}(.)}_{\text{restricted}}(\epsilon,k)$ (defined as $f^{\mathfrak{h}(.)}(\epsilon,k)$ in Section 1.1.2).

A.6 Noise satisfying Monotone Hazard Rate property

Let Y denote the random noise with density g. We assume Y is symmetric about the origin, and let $X = Y^+ \cup \{0\}$. Let f denote the density function of X (so that $f(x) = 2g(x)$ for $x \geq 0$), and let $F(x) = \mathbb{P}(X \geq x)$. We assume that $\mathbb{E}[X] = \frac{1}{\epsilon}$.[16] We assume f is continuously differentiable and log-concave. By [9], we have F is also log-concave. Note that several natural distributions such as Exponential (Laplace noise) and Gaussian are log-concave.

We are interested in choosing $2K-1$ values $S = \{-a_{K-1}, -a_{K-2}, \ldots, -a_1, 0, a_1, \ldots, a_{K-1}\}$ such that for a random draw $y \sim Y$, the expected error in approximating y by its closest point in S is small. For $i = 0, 1, 2, \ldots, K-1$, we will choose $a_i = F^{-1}\left(1 - \frac{i}{K}\right)$. Let $\phi = \mathbb{E}_{x\sim X}\left[\min_{v\in S}|v-x|\right]$. Note that the error of Y with respect to S is exactly ϕ.

Our main result is the following theorem:

Theorem A.3. $\phi = O\left(\frac{\log K}{K\epsilon}\right)$.

Proof. Let $G(z) = F^{-1}(z)$ for $z \in [0,1]$. For upper bounding ϕ, we map each $x \sim X$ to the immediately smaller value in S. If we draw $z \in [0,1]$ uniformly at random, the error is upper bounded as:

$$\phi \leq \int_0^1 G(z)dz - \sum_{i=1}^{K} \frac{1}{K} G\left(\frac{i}{K}\right) \leq \int_0^{1/K} \left(G(z) - G(1/K)\right) dz + \frac{1}{K} G(1/K).$$

Let $q = G(1/K)$. Then the above can be rewritten as:

$$\phi \leq \frac{q}{K} + \int_q^{\infty} F(x)dx.$$

[16] This implies that a large ϵ is equivalent to the magnitude of noise being smaller and vice-versa. Although this distribution does not satisfy ϵ geographic differential privacy, this follows a similar trend w.r.t ϵ.

Next, it follows from [9] that if F is log-concave, then so is $\int_r^\infty F(x)dx$. This means the function

$$\ell(r) = \frac{\int_r^\infty F(x)dx}{F(r)}$$

is non-increasing in r. Therefore,

$$\int_q^\infty F(x)dx \leq F(q)\int_0^\infty F(x)dx = \frac{1}{K}\mathbb{E}[X] = \frac{1}{\epsilon K}.$$

Let $h(x) = -\log F(x)$. Then, h is convex and increasing. Further, $h(q) = \log K$. Let $s = F^{-1}(1/e)$ so that $h(s) = 1$. Since

$$h(q) - h(s) \geq (q-s)h'(s),$$

we have

$$q - s \leq \frac{\log K}{h'(s)}.$$

Further, $h(s) \leq sh'(s)$ so that

$$\frac{1}{h'(s)} \leq s.$$

Since $F(s) = 1/e$ and $\mathbb{E}[X] = \frac{1}{\epsilon} \geq \int_0^s F(x)dx$, we have $s \leq \frac{e}{\epsilon}$. Therefore, $h'(s) \geq 1/e\epsilon$. Putting this together,

$$q \leq \frac{s}{\epsilon} + \frac{\log K}{h'(s)} \leq \frac{e}{\epsilon}(1 + \log K) = O(\frac{\log K}{\epsilon}).$$

Therefore,

$$\phi \leq \frac{q}{K} + \int_q^\infty F(x)dx = O\left(\frac{\log K}{K\epsilon}\right) + \frac{1}{K\epsilon}$$

completing the proof. □

A.7 User Features lying on some high dimensional space

In this subsection, we relax the assumption that users lying on an infinite dimensional line, rather for every user u belongs to some set U and every movie m belongs to some set M and consider some function $\kappa : U \to \mathbb{R}^d$ mapping every user to some user feature vector and denote the dis-utility of an user u from a movie m by $d(\kappa(u), m)$. Observe that this could be any complex function in high dimensions which could be captured by some machine learning model. We now make the following assumption for movies lying in some vicinity of user u. Informally, this means that the the user does not get dis-utility from every possible feature but only gets a dis-utility only from some d' directions.

Assumption A.4. Fix any user $u \in U$. There exists a subset of movies $M_u \subseteq M$ (call it movies in neighbourhood of user u), a matrix $P_u \in \mathbb{R}^{d' \times d}$ function $\lambda_u : M_u \to \mathbb{R}^{d'}$ satisfying the following properties

- The disutility of users $u' \in U$ lying in a neighbourhood of u i.e. $|\kappa(u') - \kappa(u)|_1 < \delta$ from movies $m \in M_u$ may be approximated by $d(u', m) \approx ||P_u \cdot \kappa(u') - \lambda_u(m)||_1$.
- The function λ_u is invertible i.e. the set of movies M_u is densely populated and movies exist along most directions.

These assumptions imply that the functions described above hold not just for user u but is also true for users u' with $\kappa(u')$ lying in a neighbourhood of $\kappa(u)$. The function λ_u denotes a map from the movie space to the feature space for an user $u \in U$. P_u effectively captures the linear combination of the features that play a role in the dis-utility of user u.

Note that some of these assumptions a bit tight and may not be true in reality but in this section we analyse the server response and user action under this mechanism under the Gaussian noise addition mechanism. We analyse the Gaussian mechanism as it is easier to analyse since additions of Gaussian is also a Gaussian. Also we make an assumption that noise parameter $1/\epsilon << \delta$.

Notations: For some $Y \sim \mathcal{N}(0, \sigma^2)$ and suppose $X = |Y|$. Now denote $F_\sigma(x) = \Pr(X \geq x)$ for some $x > 0$ and observe that F_σ is invertible from $(0, 1] \to [0, \infty)$. We also extend its definition to $(0, 2)$ by defining $F_\sigma^{-1}(2 - x) = -F_\sigma^{-1}(x)$ for $x > 0$

Let us denote the sum of all squares of all entries in row i of matrix P is given by $P^{(2)}[i, :]$ and denote all integers from a to b by $[a \dots b]$ and $[a \dots b]^k$ denotes the set of all k dimensional vectors in with each component taking an integral value from a to b.

User's Action:

An user $u \in U$ adds Gaussian noise with parameter ϵ i.e. $\mathcal{N}(0, \frac{1}{\epsilon^2} I)$ to its feature $\kappa(u)$ and sends it to the server, let us call it distribution $\mathcal{F}_u$.

Server's Action:

Suppose the server receives a signal $\hat{f}$. The signal sends back $k = (2k' + 1)^{d'}$ along each of the the d' dimensions which we describe below. Further consider some user u' whose feature vector $\kappa(u')$ lies in $1/\epsilon$ vicinity of signal $\hat{f}$. Let us index the movies by $\mathcal{M}_{\hat{f}} = [\mathfrak{m}_{\mathbf{i}}]_{\mathbf{i} \in [-k' \dots k']^{d'}}$.

For every $\mathbf{i} \in [-k' \dots k']^{d'}$, define $\mathfrak{m}_{\mathbf{i}}$ as follows:

$$m_{\mathbf{i}} := \lambda_{u'}^{-1}(P_{u'} \hat{f} + \mathbf{v}) \text{ for } \mathbf{v} \in \mathbb{R}^{d'} \tag{23}$$

$$\text{where } v_j := F^{-1}_{\frac{P^{(2)}_{u'}[j,:]}{\epsilon^2}}(1 - i_j/k') \ \forall j \in [1, \dots, d'] \tag{24}$$

Lemma A.5. $\mathbb{E}_{\hat{f} \sim \mathcal{F}_u}\left[\min_{m \in \mathcal{M}_{\hat{f}}} d(u, m)\right] \overset{(a)}{\approx} O\left(d' \frac{\log k'}{\epsilon k'}\right) = O\left(\frac{\log k}{\epsilon k^{1/d'}}\right)$

Proof. This follows since f' is a Gaussian Random vector and thus j^{th} component of $P_{u'} \hat{f}$ is given by Gaussian random variable with variance $\frac{P^{(2)}_{u'}[j,:]}{\epsilon^2}$. The first point in Assumption A.4 implies the disutility may be approximated by $||P_{u'} \cdot \kappa(u) - \lambda_{u'}(m)||_1$ since the

Now apply Theorem A.3 and since the placement of points is identical to that in Section A.6, we get the desired equality in (a). □

A.8 Additional problems that may be be formulated as a DILP

We now discuss two scenarios where a problem maybe solved by formulation as an optimization problem over function spaces involving constraints on both derivatives and integrals.

A.8.1 A problem from job scheduling

Job scheduling has been a well-studied problem and has attracted a lot of attention in theory of operations research starting from [23, 10]. In [2], a new continuous time model was proposed for the same described formally in [2, Section 2] and the optimization variable functions are given by $\{x_i(.)\}_{i=1}^N : \mathbb{R} \to \mathbb{R}^+, \{y_{i,j}(.), u_{i,j}(.)\}_{j=1,i=1}^{N_i,N} : \mathbb{R} \to \mathbb{R}^+$. In this model, we need to produce multiple

items with each item goes through multiple job shops for an operation to be performed. and there is a continuous demand for each item. We aim to minimise the cumulative backlog for items in continuous time.

We now introduce some notations for the same by restating the problem in [2].

- N : number of parts
- n_i: number of operations required by part i
- M: number of machines
- $m_{i,j}$: machine on which the j^{th} operation is performed for part i
- $r_{i,j}$: Maximum rate of procession on $m_{i,j}$ of the j^{th} operation on part i
- $s_{i,j}$ Initial stock of part i between the j^{th} and $(j+1)^{th}$ operation.
- b_i: Initial backlog of part i
- $d_i(t)$: Rate of demand of part i at time t
- k_i: cost per unit time of unit backlog of part i
- T time horizon

Observe that the optimization variables in this problem are $x_i(t) : [0,1] \to \mathbb{R}^+$ for $i \in [n]$ and $y_{i,j}(t) : [0,1] \to \mathbb{R}^+$ for $j \in [n_i], i \in [n]$

As discussed in [2], we now formulate our optimization problem below.

$$\begin{cases} \min \int_{t\in 0}^{T} \sum_{i=1}^{n} k_i(t) x_i(t) du \\ \text{s.t.} \quad x_i'(t) = d_i(t) - r_{i,n_i} u_{i,n_i}(t) \ \forall j \in [n_i]; i \in [n]; \\ \quad\quad y_{i,j}'(t) = r_{i,j} u_{i,j}(t) - r_{i,j+1} u_{i,j+1}(t) \ \forall j \in [n_i]; i \in [n]; \\ \quad\quad x_i(t) \geq 0; y_{i,j}(t) \geq 0; u_{i,j}(t) \geq 0 \ \forall j \in [n_i]; i \in [n]; t \in [0,T] \\ \quad\quad \sum_{i,j:m_{i,j}=k} u_{i,j}(t) \leq 1 \ \forall k \in [M] \\ \quad\quad x_i(0) = b_i; y_{i,j}(0) = s_{i,j}; \forall j \in [n_i]; i \in [n]; \end{cases} \tag{25}$$

Observe that since it involves constraints on both the derivatives and integrals, this maybe modelled as a differential integral linear program (DILP) and our weak duality framework may be a useful tool in analysing the same.

A.8.2 A problem from control theory

In control theory [31], we often have a state $\mathbf{x}(t)$ whose evolution is given by a differential equation. We redefine the following equation from [31, Chapter 4] below to characterise its evolution.

$$\begin{cases} \mathbf{x}(t) = \mathbf{f}(\mathbf{x}(t), \alpha(t)) \\ \mathbf{x}(0) = x^0 \end{cases} \tag{26}$$

Here $x^0 \in \mathbb{R}^n$, $A \in \mathbb{R}^m$, $\mathbf{f} : \mathbb{R}^n \times A \to \mathbb{R}, \alpha : [0,\infty) \to A$ is the control, and $\mathbf{x} : [0,\infty) \to \mathbb{R}^n$ is the response of the system. Denote the set of admissible controls by $\mathcal{A} = \{\alpha(.) : [0,\infty) \to A | \alpha(.) \text{ is measurable}\}$

We also introduce the payoff functional (P)

$$P[\alpha(.)] = \int_{t=0}^{T} r(x(t), \alpha(t))dt + g(x(T))$$

,

where the terminal time $T > 0$, running payoff $r : \mathbb{R}^n \times A \to \mathbb{R}$ and terminal payoff $g : \mathbb{R}^n \to \mathbb{R}$ are given. Observe that if the function $f(.)$ is linear and the running payoff is linear, we may represent it as a differential integral linear program.

It may also be typical to have constraints on the the control input α, for example there could be some constraints on its rate of growth or some constraints on its Lipschitz continuity.

B Proofs of skipped lemmas and theorems

B.1 Simulating the server: Detailed proof of Theorem 2.2

We first show it is sufficient to consider mechanisms in which servers directly return the received signal from the user, thus removing the server from the picture.

Recall that $\mathcal{P}_Z^{(\epsilon)}$ is the set of differential private mechanisms that adopt signal set Z. Then $\mathcal{P}_{\mathcal{A}}^{(\epsilon)}$ is the set of differential private mechanisms in which users pick signals from $\mathcal{A} = \mathbb{R}^k$.

Theorem (Restatement of Theorem 2.2). We have

$$f^{\mathfrak{h}(.)}(\epsilon, k) = \inf_{\mathbf{P} \in \mathcal{P}_{\mathbb{R}^k}^{(\epsilon)}} \sup_{u \in \mathbb{R}} \mathbb{E}_{\mathbf{a} \sim P_u} \left[\min_{a \in \mathtt{Set}(\mathbf{a})} \mathfrak{h}(|u - a|) \right].$$

Proof. Fix $Z, \mathbf{P} \in \mathcal{P}_Z^{(\epsilon)}, \mathbf{Q} \in \mathcal{Q}_Z$ and recall that that $\mathcal{A} = \mathbb{R}^k$ from Appendix A.1.1. For any $\mathcal{H} \in \mathcal{F}_{\mathcal{A}}$, let

$$\tilde{P}_u(\mathcal{H}) = \mathbb{P}_{s \sim P_u, \mathbf{a} \sim Q_s}(\mathbf{a} \in \mathcal{H}),$$

which is the probability that user u receives a result set in $\mathcal{H}$.

Then for any $u_1, u_2 \in \mathbb{R}, \mathcal{H} \in \mathcal{F}_{\mathcal{A}}$, we have

$$\begin{aligned}
\tilde{P}_{u_1}(\mathcal{H}) = \mathbb{P}_{s \sim P_{u_1}, \mathbf{a} \sim Q_s}(\mathbf{a} \in \mathcal{H}) &= \mathbb{E}_{s \sim P_{u_1}}\left[\mathbb{P}_{\mathbf{a} \sim Q_s}(\mathbf{a} \in \mathcal{H})\right] \\
&= \int_{s \in Z} \left(\mathbb{P}_{\mathbf{a} \sim Q_s}(\mathbf{a} \in \mathcal{H}) \cdot P_{u_1}(ds)\right) \\
&\leq e^{\epsilon \cdot |u_1 - u_2|} \int_{s \in Z} \left(\mathbb{P}_{\mathbf{a} \sim Q_s}(\mathbf{a} \in \mathcal{H}) \cdot P_{u_2}(ds)\right) \\
&= e^{\epsilon \cdot |u_1 - u_2|}\, \mathbb{E}_{s \sim P_{u_2}}\left[\mathbb{P}_{\mathbf{a} \sim Q_s}(\mathbf{a} \in \mathcal{H})\right] \\
&= e^{\epsilon \cdot |u_1 - u_2|}\, \mathbb{E}_{s \sim P_{u_2}, \mathbf{a} \sim Q_s}\left[\mathbb{I}\{\mathbf{a} \in \mathcal{H}\}\right] \\
&= e^{\epsilon \cdot |u_1 - u_2|} \tilde{P}_{u_2}(\mathcal{H}),
\end{aligned}$$

so $\tilde{\mathbf{P}} \in \mathcal{P}_{\mathcal{A}}^{(\epsilon)}$ because $\tilde{P}_u(\cdot) \in \mathbb{D}(\mathcal{A})$ for any $u \in \mathbb{R}$.

At the same time,

$$\mathbb{E}_{s\sim P_u,\mathbf{a}\sim Q_s}\left[\min_{a\in\texttt{Set}(\mathbf{a})}\mathfrak{h}(|u-a|)\right]=\mathbb{E}_{\mathbf{a}\sim\tilde{P}_u}\left[\min_{a\in\texttt{Set}(\mathbf{a})}\mathfrak{h}(|u-a|)\right] \text{ so we have,}$$

$$f^{\mathfrak{h}(.)}(\epsilon,k)=\inf_{\mathbf{P}\in\mathcal{P}_{\mathcal{A}}^{(\epsilon)}}\sup_{u\in\mathbb{R}}\mathbb{E}_{\mathbf{a}\sim P_u}\left[\min_{a\in\texttt{Set}(\mathbf{a})}\mathfrak{h}(|u-a|)\right].$$

□

It is sufficient to assume the server directly sends back the received signal, and we care about the user with largest cost

B.2 Dual and primal DILPs: Proof of Lemmas 2.6, Lemma 3.4 and 2.7

B.2.1 Proof of Lemma 2.6

Lemma. Suppose, P_u has a probability density function given by $g(u,.):\mathbb{R}^k\to\mathbb{R}$ for every $u\in\mathbb{R}$. Then, P satisfies ϵ-geographic differential privacy iff $\max(|\overline{g}_u(u,\mathbf{x})|,|\underline{g}_u(u,\mathbf{x})|)\leq\epsilon g(u,\mathbf{x})\ \forall u\in\mathbb{R};\forall x\in\mathbb{R}^k$ [17]

Proof. **We first prove the only if statement.**

Here assuming ϵ- geographic differential privacy, we prove the desired constraint on $g(u,.)$.

Observe that since, $\{P_u\}_{u\in\mathbb{R}}$ satisfies ϵ - geographic-differential privacy, we must have $g(u_1,\mathbf{x})\leq e^{\epsilon|u_1-u_2|}g(u_2,\mathbf{x})$.

Thus, applying log on both sides we get [assuming $g(u,x)\neq 0$],

$$\begin{aligned}&\log g(u+\delta u,\mathbf{x})-\log g(u,\mathbf{x})\leq\epsilon|\delta u|\\ \implies&\frac{\max(|\overline{g}_u(u,\mathbf{x})|,|\underline{g}_u(u,\mathbf{x})|)}{g(u,\mathbf{x})}\leq\epsilon\end{aligned}$$

The last implication follows on limiting δu towards zero and applying the lower and upper limits respectively and thus we prove the only if statement.

However, if $g(u,x)=0$, we get the following

$$\begin{aligned}&g(u+\delta u,\mathbf{x})-g(u,x)\leq 0\\ \implies&\max(|\overline{g}_u(u,\mathbf{x})|,|\underline{g}_u(u,\mathbf{x})|)\leq 0\end{aligned}$$

The last inequality follows on applying δu tending to 0 and applying upper and lower limits respectively.

We now prove the if statement.

Here assuming, $\max(|\overline{g}_u(u,\mathbf{x})|,|\underline{g}_u(u,\mathbf{x})|)\leq\epsilon g(u,\mathbf{x})$ we prove ϵ-geographic differential privacy.

We first show that $g(u,\mathbf{x})$ is continuous in u. Observe that $\liminf_{\delta u\to 0}g(u+\delta u)-g(u)=\lim_{\delta u\to 0}\delta u\times\liminf_{\delta u\to 0}\frac{g(u+\delta u)-g(u)}{\delta u}=0$ as $|\underline{g}_u(u,\mathbf{x})|=|\liminf_{\delta u\to 0}\frac{g(u+\delta u)-g(u)}{\delta u}|\leq\epsilon g(u,\mathbf{x})$. Similarly, using the bound

[17] $\underline{g}_u(u,\mathbf{x})$, $\overline{g}_u(u,\mathbf{x})$ denote the lower and upper partial derivative w.r.t. u

on upper derivative, we show that $\limsup_{\delta u \to 0} g(u+\delta u) - g(u) = 0$ and thus, we have continuity of $g(u, \mathbf{x})$ at u for every $u \in \mathbb{R}^k$.

We first prove assuming $g(u, x) \neq 0 \forall u \in \mathbb{R}$. Now consider $u_1 > u_2$ for some $u_1, u_2 \in \mathbb{R}$.

$$\log g(u_1, \mathbf{x}) - \log g(u_2, \mathbf{x}) \leq \int_{u=u_2}^{u_1} \frac{\overline{g}_u(u, \mathbf{x})}{g(u, \mathbf{x})} \leq \epsilon(u_1 - u_2)$$

The first inequality follows since we use an upper derivative.

$$\log g(u_1, \mathbf{x}) - \log g(u_2, \mathbf{x}) \geq \int_{u=u_2}^{u_1} \frac{\underline{g}_u(u, \mathbf{x})}{g(u, \mathbf{x})} \geq -\epsilon(u_1 - u_2)$$

Thus, we get $g(u_1, \mathbf{x}) \leq e^{\epsilon|u_1 - u_2|} g(u_2, \mathbf{x})$ on taking exponent on both sides

We now consider the other case that $g(u, \mathbf{x}) = 0$ for some $u \in \mathbb{R}$. We now show that this implies $g(u, \mathbf{x}) = 0 \ \forall u \in \mathbb{R}$. Suppose not and there exists some u_1 s.t. $g(u_1, \mathbf{x}) > 0$ and consider the largest $u < u_1$ s.t. $g(u, \mathbf{x}) = 0$ (call it u_0)

Now, observe the following observe that we define the limits over extended reals i.e. $\mathbb{R} \cup \{-\infty, \infty\}$. Observe that

$$\begin{aligned}
\lim_{u \to u_1} g(u, \mathbf{x}) &= \lim_{u \to u_0} g(u, \mathbf{x}) + \lim_{\substack{d^l \to u_0 \\ d^u \to u_1}} (\log g(d^u, \mathbf{x}) - \log g(d^l, \mathbf{x})) \\
&\leq \lim_{u \to u_0} g(u, \mathbf{x}) + \int_{u=u_0}^{u_1} \frac{\overline{g}_u(u, \mathbf{x})}{g(u, \mathbf{x})} \\
&\leq \lim_{u \to u_0} g(u, \mathbf{x}) + \epsilon(u_1 - u_0) \\
&\overset{(a)}{=} -\infty + \epsilon(u_1 - u_0) = -\infty
\end{aligned}$$

(a) follows from the fact that $g(u, \mathbf{x})$ is continuous in u and $g(u_0, \mathbf{x}) = 0$.

Since, $g(u, \mathbf{x})$ is a continuous function, this implies that $g(u_1, \mathbf{x}) = 0$. Since we can make this argument for every $u_1 \in \mathbb{R}$, we argue that $g(u, \mathbf{x}) = 0 \ \forall u \in \mathbb{R}$ and thus satisfy the condition of ϵ-geographic DP.

□

B.2.2 Proof of Lemma 3.4

Lemma. Suppose, P_u has a probability density function given by $g(u, .) : \mathbb{R}^k \to \mathbb{R}$ for every $u \in \mathbb{R}$. Then, P satisfies $\mathfrak{g}(.)$-geographic differential privacy iff $\max(|\overline{g}_u(u, \mathbf{x})|, |\underline{g}_u(u, \mathbf{x})|) \leq \mathfrak{g}'(0) g(u, \mathbf{x}) \ \forall u \in \mathbb{R}; \forall x \in \mathbb{R}^k$ [18] whenever $\mathfrak{g}(.)$ satisfies Assumption 3.1.

Proof. **We first prove the only if statement.**

Assuming $\mathfrak{g}(.)$ geographic differential privacy we prove the desired constraint on $g(u, .)$.

Observe that since, $\{P_u\}_{u \in \mathbb{R}}$ satisfies $\mathfrak{g}(.)$ geographic-differential privacy, we must have $g(u_1, \mathbf{x}) \leq e^{\mathfrak{g}(|u_1 - u_2|)} g(u_2, \mathbf{x})$.

[18] $\underline{g}_u(u, \mathbf{x})$, $\overline{g}_u(u, \mathbf{x})$ denote the lower and upper partial derivative w.r.t. u

Thus, applying log on both sides we get,

$$\begin{aligned}&\log g(u+\delta u,\mathbf{x})-\log g(u,\mathbf{x})\leq \mathfrak{g}(|\delta u|)\\ \implies &\frac{\max(|\overline{g}_u(u,\mathbf{x})|,|\underline{g}_u(u,\mathbf{x})|)}{g(u,\mathbf{x})}\leq \mathfrak{g}'(0)\end{aligned}$$

The last implication follows on limiting δu towards zero and $\mathfrak{g}(0)=0$ thus, applying the lower and upper limits respectively and thus we prove the only if statement.

However, if $g(u,x)=0$, we get the following

$$\begin{aligned}&g(u+\delta u,\mathbf{x})-g(u,x)\leq 0\\ \implies &\max(|\overline{g}_u(u,\mathbf{x})|,|\underline{g}_u(u,\mathbf{x})|)\leq 0\end{aligned}$$

The last inequality follows on applying δu tending to 0 and applying upper and lower limits respectively.

We now prove the if statement.

Here assuming $\max(|\overline{g}_u(u,\mathbf{x})|,|\underline{g}_u(u,\mathbf{x})|)\leq \mathfrak{g}'(0)g(u,\mathbf{x})$, we prove $\mathfrak{g}(.)$-geographic differential privacy.

We first show that $g(u,\mathbf{x})$ is continuous in u. Observe that $\liminf_{\delta u\to 0} g(u+\delta u)-g(u)=\lim_{\delta u\to 0}\delta u\times \liminf_{\delta u\to 0}\frac{g(u+\delta u)-g(u)}{\delta u}=0$ as $|\underline{g}_u(u,\mathbf{x})|=|\liminf_{\delta u\to 0}\frac{g(u+\delta u)-g(u)}{\delta u}|\leq \epsilon g(u,\mathbf{x})$. Similarly, using the bound on upper derivative, we show that $\limsup_{\delta u\to 0} g(u+\delta u)-g(u)=0$ and thus, we have continuity of $g(u,\mathbf{x})$ at u for every $u\in\mathbb{R}^k$.

We first prove assuming $g(u,x)\neq 0 \forall u\in\mathbb{R}$. Now consider $u_1>u_2$ for some $u_1,u_2\in\mathbb{R}$.

$$\begin{aligned}\log g(u_1,\mathbf{x})-\log g(u_2,\mathbf{x})\overset{(a)}{\leq}&\int_{u=u_2}^{u_1}\frac{\overline{g}_u(u,\mathbf{x})}{g(u,\mathbf{x})}\leq \mathfrak{g}'(0)(u_1-u_2)\\ &\overset{(b)}{\leq}\mathfrak{g}(u_1-u_2)\end{aligned}$$

The first inequality (a) follows since we use an upper derivative and the inequality (b) follows from the fact that $\mathfrak{g}(.)$ is a convex function and $\mathfrak{g}(0)=0$ as defined in Section 3.1.

$$\begin{aligned}\log g(u_1,\mathbf{x})-\log g(u_2,\mathbf{x})\geq&\int_{u=u_2}^{u_1}\frac{\underline{g}_u(u,\mathbf{x})}{g(u,\mathbf{x})}\geq -\mathfrak{g}'(0)(u_1-u_2)\\ &\overset{(c)}{\geq}-\mathfrak{g}(u_1-u_2)\end{aligned}$$

Inequality (c) follows from the fact that $\mathfrak{g}(.)$ is a convex function. Thus, we get $g(u_1,\mathbf{x})\leq e^{-\mathfrak{g}(|u_1-u_2|)}g(u_2,\mathbf{x})$ on taking exponent on both sides.

We now consider the other case that $g(u,\mathbf{x})=0$ for some $u\in\mathbb{R}$. We now show that this implies $g(u,\mathbf{x})=0\ \forall u\in\mathbb{R}$. Suppose not and there exists some u_1 s.t. $g(u_1,\mathbf{x})>0$ and consider the largest $u<u_1$ s.t. $g(u,\mathbf{x})=0$ (call it u_0)

Now, observe the following observe that we define the limits over extended reals i.e. $\mathbb{R} \cup \{-\infty, \infty\}$. Observe that

$$\begin{aligned}
\lim_{u \to u_1} g(u, \mathbf{x}) &= \lim_{u \to u_0} g(u, \mathbf{x}) + \lim_{\substack{d^l \to u_0 \\ d^u \to u_1}} \left(\log g(d^u, \mathbf{x}) - \log g(d^l, \mathbf{x})\right) \\
&\leq \lim_{u \to u_0} g(u, \mathbf{x}) + \int_{u=u_0}^{u_1} \frac{\overline{g}_u(u, \mathbf{x})}{g(u, \mathbf{x})} \\
&\leq \lim_{u \to u_0} g(u, \mathbf{x}) + \epsilon(u_1 - u_0) \\
&\overset{(a)}{=} -\infty + \epsilon(u_1 - u_0) = -\infty
\end{aligned}$$

(a) follows from the fact that $g(u, \mathbf{x})$ is continuous in u and $g(u_0, \mathbf{x}) = 0$.

Since, $g(u, \mathbf{x})$ is a continuous function, this implies that $g(u_1, \mathbf{x}) = 0$. Since we can make this argument for every $u_1 \in \mathbb{R}$, we argue that $g(u, \mathbf{x}) = 0 \ \forall u \in \mathbb{R}$ and thus satisfy the condition of $\mathfrak{g}(.)$-geographic DP.

□

B.2.3 Proof of Lemma 2.7

To prove this lemma, we first show another lemma showing a necessary and sufficient condition for ϵ-geographic differential privacy of distributions $\{P_u\}_{u \in \mathbb{R}}$ assuming that it has a probability density function.

We now restate and prove Lemma 2.5

Lemma. Given any mechanism $\mathbf{P} \in \mathcal{P}^{(\epsilon)}_{\mathbf{R}^k}$ (satisfying ϵ-geographic differential privacy), we can construct a sequence of mechanisms $\mathbf{P}^{(\eta)} \in \mathcal{P}^{(\epsilon)}_{\mathbf{R}^k}$ with a bounded probability distribution function for every $u \in \mathbb{R}$ such that $\mathbf{P}^{(\eta)}$ is an arbitrary cost approximation of mechanism $\mathbf{P} \in \mathcal{P}^{(\epsilon)}_{\mathbf{R}^k}$.

Proof. For any mechanism $\mathbf{P}$ belonging to the set $\mathcal{P}^{(\epsilon)}_{\mathbb{R}^k}$, we can construct alternative mechanisms $\mathbf{P}^{(\eta)}$ within the same set, aiming to provide an arbitrary cost approximation of the original mechanism.

For this, we borrow ideas from theory of mollifier in [34]. We now give a construction $P^{(\eta)}_u$ with a probability density function $\rho_V(u, .) : \mathbb{R}^k \to \mathbb{R}$ for every $u \in \mathbb{R}$.

For $\eta > 0$ and $\mathbf{a} \in R^k$, define the η-ball of $\mathbf{a}$, denoted by $\mathcal{B}_{\mathbf{a},\eta}$, to be

$$\mathcal{B}_{A,\eta} := \left\{ \mathbf{a}' \in R^k : |\mathbf{a} - \mathbf{a}'\|_\infty < \eta \right\},$$

and use $|\mathcal{B}_{\mathbf{a},\eta}|$ to denote the Lebesgue measure (corresponding to the volume) of $\mathcal{B}_{A,\eta}$. Note that there exists a real number, denoted by $|\mathcal{B}_\eta| \in \mathbb{R}$, satisfying $|\mathcal{B}_{\mathbf{a},\eta}| = |\mathcal{B}_\eta|$ for any $\mathbf{a} \in \mathbb{R}^k$ as this volume is independent of $\mathbf{a}$.

Fix $\eta > 0$, we define another probability measure $P^{(\eta)}_u$ which has density $\rho_V(u, \mathbf{a})$ for all $\mathbf{a} \in \mathbb{R}^k$.

$$\rho_V(u, \mathbf{a}) := \frac{P_u(\mathcal{B}_{\mathbf{a},\eta})}{|\mathcal{B}_\eta|}.$$

$\rho_V(u, .)$ defines a valid probability measure because

$$\int_{\mathbb{R}^k} \rho_V(u, \mathbf{a}) d\mathbf{a} = \int_{\mathbb{R}^k} \frac{P_u(\mathcal{B}_{\mathbf{a},\eta})}{|\mathcal{B}_\eta|} d\mathbf{a}$$

$$
\begin{aligned}
&= \frac{1}{|\mathcal{B}_\eta|} \int_{\mathbf{a}\in\mathbb{R}^k} \int_{\mathbf{b}\in\mathcal{B}_{\mathbf{a},\eta}} P_u(d\mathbf{b}) d\mathbf{a} \\
&= \frac{1}{|\mathcal{B}_\eta|} \int_{\mathbf{b}\in\mathbb{R}^k} P_u(d\mathbf{b}) \int_{\mathbf{a}\in\mathcal{B}_{\mathbf{b},\eta}} d\mathbf{a} \\
&= \frac{1}{|\mathcal{B}_\eta|} \int_{\mathbf{b}\in\mathbb{R}^k} P_u(d\mathbf{b}) \, |\mathcal{B}_\eta| = 1.
\end{aligned}
$$

To show $\mathbf{P}^{(\eta)} \in \mathcal{P}^{(\epsilon)}_{\mathbb{R}^k}$, we now show

$$
\rho_V(u_1, \mathbf{a}) = \frac{P_{u_1}(\mathcal{B}_{\mathbf{a},\eta})}{|\mathcal{B}_\eta|} \leq \frac{P_{u_2}(\mathcal{B}_{\mathbf{a},\eta}) e^{\epsilon|u_1-u_2|}}{|\mathcal{B}_\eta|} = e^{\epsilon|u_1-u_2|} \rho_V(u_2, \mathbf{a})
$$

This implies for any measurable set S in $\mathbb{R}^k$, we have $P^{(\eta)}_{u_1}(S) \leq e^{\epsilon|u_1-u_2|} P^{(\eta)}_{u_2}(S)$ and thus satisfies ϵ-geographic differential privacy and also implies that that $\rho_V(u, \mathbf{a})$ is continuous in u for every $\mathbf{a} \in \mathbb{R}^k$.

Also observe that $\rho_V(u, \mathbf{a})$ is always bounded above by $\frac{1}{|B_\eta|}$ as $P_u(.)$ is a probability measure. Observe that since, we showed that $\rho_V(., \mathbf{a})$ is *Riemann integrable* and is non-negative for every $\mathbf{a} \in \mathbb{R}^k$ and $\rho_V(u, \mathbf{a})$ is continuous in u, we showed that $\rho_V(.,.)$ is Riemann integrable over $\mathbb{R} \times \mathbb{R}^k$. [19]

In the next set of equations (34) and (42), we upper and lower bound $\mathbb{E}_{\mathbf{a}\sim P^{(\eta)}_u}\left[\min_{a\in\texttt{Set}(\mathbf{a})} \mathfrak{h}(|u-a|)\right]$ and then apply limit $\eta \to 0$ on both the upper and lower bounds in Equations (48) and (49) to show that $\mathbf{P}^{(\eta)}$ arbitrarily approximates $\mathbf{P}$ in Equation (50).

$$
\begin{aligned}
\mathbb{E}_{\mathbf{a}\sim P^{(\eta)}_u}\left[\min_{a\in\texttt{Set}(\mathbf{a})} \mathfrak{h}(|u-a|)\right] &= \int_{\mathbf{a}\in\mathbb{R}^k} \min_{a\in\texttt{Set}(\mathbf{a})} \mathfrak{h}(|u-a|) \rho_V(u, \mathbf{a}) d\mathbf{a} && (27)\\
&= \int_{\mathbf{a}\in\mathbb{R}^k} \min_{a\in\texttt{Set}(\mathbf{a})} \mathfrak{h}(|u-a|) \frac{P_u(\mathcal{B}_{\mathbf{a},\eta})}{|\mathcal{B}_\eta|} d\mathbf{a} && (28)\\
&= \int_{\mathbf{a}\in\mathbb{R}^k} \int_{\mathbf{b}\in\mathcal{B}_{\mathbf{a},\eta}} \min_{a\in\texttt{Set}(\mathbf{a})} \mathfrak{h}(|u-a|) \frac{P_u(d\mathbf{b})}{|\mathcal{B}_\eta|} d\mathbf{a} && (29)\\
&= \frac{1}{|\mathcal{B}_\eta|} \int_{\mathbf{b}\in\mathbb{R}^k} \int_{\mathbf{a}\in\mathcal{B}_{\mathbf{b},\eta}} P_u(d\mathbf{b}) \min_{a\in\texttt{Set}(\mathbf{a})} \mathfrak{h}(|u-a|) d\mathbf{a} && (30)\\
&\leq \frac{1}{|\mathcal{B}_\eta|} \int_{\mathbf{b}\in\mathbb{R}^k} \int_{\mathbf{a}\in\mathcal{B}_{\mathbf{b},\eta}} \left(\min_{a\in\texttt{Set}(\mathbf{b})} \mathfrak{h}(|u-a|+\eta)\right) d\mathbf{a} P_u(d\mathbf{b}) && (31)\\
&= \frac{1}{|\mathcal{B}_\eta|} \int_{\mathbf{b}\in\mathbb{R}^k} \int_{\mathbf{a}\in\mathcal{B}_{\mathbf{b},\eta}} \left(\min_{a\in\texttt{Set}(\mathbf{b})} \mathfrak{h}(|u-a|+\eta)\right) d\mathbf{a} P_u(d\mathbf{b}) && (32)\\
&= \int_{\mathbf{b}\in\mathbb{R}^k} \left(\min_{a\in\texttt{Set}(\mathbf{b})} \mathfrak{h}(|u-a|+\eta)\right) P_u(d\mathbf{b}) && (33)\\
&= \mathbb{E}_{\mathbf{a}\sim P_u}\left[\min_{a\in\texttt{Set}(\mathbf{a})} \mathfrak{h}(|u-a|+\eta)\right] && (34)
\end{aligned}
$$

where Inequality 31 holds because for any $\mathbf{b} \in \mathbb{R}^k$ and $\mathbf{a} \in \mathcal{B}_{\mathbf{b},\eta}$, we have

$$
\min_{a\in\texttt{Set}(\mathbf{a})} (|u-a|) \overset{(a)}{\leq} \min_{a\in\texttt{Set}(\mathbf{b})} (|u-a|) + \eta.
$$

[19] Observe that this integral may go to ∞ but is always defined.

$$\overset{(b)}{\implies} \min_{a\in\texttt{Set}(\mathbf{a})} \mathfrak{h}(|u-a|) \leq \min_{a\in\texttt{Set}(\mathbf{b})} \mathfrak{h}(|u-a|+\eta)$$

(a) follows from triangle inequality and the fact that $|a_i - b_i| \leq \eta \ \forall i \in [k]$. (b) follows since $\mathfrak{h}$ is monotonic.

Equality (32) follows from monotone convergence theorem since $\min_{a\in\texttt{Set}(\mathbf{b})} \mathfrak{h}(|u-a|+\eta)$ is monotonic in η as $\mathfrak{h}(.)$ is monotonic.

We now show that $\mathbb{E}_{\mathbf{a}\sim P_u^{(\eta)}}\left[\min_{a\in\texttt{Set}(\mathbf{a})} \mathfrak{h}(|u-a|)\right] \geq \mathbb{E}_{\mathbf{a}\sim P_u}\left[\min_{a\in\texttt{Set}(\mathbf{a})} \mathfrak{h}([|u-a|-\eta]_+)\right]$. Note that in the following expression $[x]_+$ denotes $\max(x,0)$.

$$\begin{aligned}
\mathbb{E}_{\mathbf{a}\sim P_u^{(\eta)}}\left[\min_{a\in\texttt{Set}(\mathbf{a})} \mathfrak{h}(|u-a|)\right] &= \int_{\mathbf{a}\in\mathbb{R}^k} \min_{a\in\texttt{Set}(\mathbf{a})} \mathfrak{h}(|u-a|)\rho_V(u,\mathbf{a})d\mathbf{a} && (35)\\
&= \int_{\mathbf{a}\in\mathbb{R}^k} \min_{a\in\texttt{Set}(\mathbf{a})} \mathfrak{h}(|u-a|)\frac{P_u(\mathcal{B}_{\mathbf{a},\eta})}{|\mathcal{B}_\eta|}d\mathbf{a} && (36)\\
&= \int_{\mathbf{a}\in\mathbb{R}^k}\int_{\mathbf{b}\in\mathcal{B}_{\mathbf{a},\eta}} \min_{a\in\texttt{Set}(\mathbf{a})} \mathfrak{h}(|u-a|)\frac{P_u(d\mathbf{b})}{|\mathcal{B}_\eta|}d\mathbf{a} && (37)\\
&= \frac{1}{|\mathcal{B}_\eta|}\int_{\mathbf{b}\in\mathbb{R}^k}\int_{\mathbf{a}\in\mathcal{B}_{\mathbf{b},\eta}} P_u(d\mathbf{b}) \min_{a\in\texttt{Set}(\mathbf{a})} \mathfrak{h}(|u-a|)d\mathbf{a} && (38)\\
&\geq \frac{1}{|\mathcal{B}_\eta|}\int\limits_{\mathbf{b}\in\mathbb{R}^k}\int\limits_{\mathbf{a}\in\mathcal{B}_{\mathbf{b},\eta}} \min_{a\in\texttt{Set}(\mathbf{b})} \mathfrak{h}\left([|u-a|-\eta]_+\right) d\mathbf{a}P_u(d\mathbf{b}) && (39)\\
&= \frac{1}{|\mathcal{B}_\eta|}\int\limits_{\mathbf{b}\in\mathbb{R}^k}\int\limits_{\mathbf{a}\in\mathcal{B}_{\mathbf{b},\eta}} \min_{a\in\texttt{Set}(\mathbf{b})} \mathfrak{h}\left([|u-a|-\eta]_+\right) d\mathbf{a}P_u(d\mathbf{b}) && (40)\\
&= \int_{\mathbf{b}\in\mathbb{R}^k} \min_{a\in\texttt{Set}(\mathbf{b})} \mathfrak{h}\left([|u-a|-\eta]_+\right) P_u(d\mathbf{b}) && (41)\\
&= \mathbb{E}_{\mathbf{a}\sim P_u}\left[\min_{a\in\texttt{Set}(\mathbf{a})} \mathfrak{h}\left([|u-a|-\eta]_+\right)\right] && (42)
\end{aligned}$$

where Inequality 39 holds because for any $\mathbf{b}\in\mathbb{R}^k$ and $\mathbf{a}\in\mathcal{B}_{\mathbf{b},\eta}$, we have

$$\min_{a\in\texttt{Set}(\mathbf{a})} (|u-a|) \overset{(a)}{\geq} \min_{a\in\texttt{Set}(\mathbf{b})} [(|u-a|)-\eta]_+ .$$
$$\overset{(b)}{\implies} \min_{a\in\texttt{Set}(\mathbf{a})} \mathfrak{h}(|u-a|) \geq \min_{a\in\texttt{Set}(\mathbf{b})} \mathfrak{h}\left([|u-a|-\eta]_+\right)$$

(a) follows from triangle inequality and the fact that $|a_i - b_i| \leq \eta \ \forall i \in [k]$ and the fact (b) follows since $\mathfrak{h}$ is monotonic and $\mathfrak{h}(0) = 0$

Now observe that $\mathbb{E}_{\mathbf{a}\sim P_u}\left[\min_{a\in\texttt{Set}(\mathbf{a})} \mathfrak{h}\left([|u-a|-\eta]_+\right)\right]$ and $\mathbb{E}_{\mathbf{a}\sim P_u}\left[\min_{a\in\texttt{Set}(\mathbf{a})} \mathfrak{h}(|u-a|+\eta)\right]$ is continuous in η for every u. Observe that each of these functions is monotonic in η.

Now we bound $\mathbb{E}_{\mathbf{a}\sim P_u}\left[\min_{a\in\texttt{Set}(\mathbf{a})} \mathfrak{h}(|u-a|+\eta)\right]$ below by some linear function on $\mathbb{E}_{\mathbf{a}\sim P_u}\left[\min_{a\in\texttt{Set}(\mathbf{a})} \mathfrak{h}(|u-a|)\right]$ for every $u\in\mathbb{R}$ and use it to show the convergence of the supremum in (48) and a similar approach for $\mathbb{E}_{\mathbf{a}\sim P_u}\left[\min_{a\in\texttt{Set}(\mathbf{a})} \mathfrak{h}(|u-a|+\eta)\right]$. Let us denote the Lipschitz constant

of $\log \mathfrak{h}(.)$ in $[\mathcal{B}, \infty)$ by M. We now split $\mathbb{R}^k$ into two sets $\mathbf{a}_1(u) = \{\mathbf{a} : \min_{a \in \texttt{Set}(\mathbf{a})} |u - a| > \mathcal{B}\}$ [recall from definition of $\mathcal{B}$ from Property 3] and $\mathbf{a}_2(u) = \{\mathbf{a} : \min_{a \in \texttt{Set}(\mathbf{a})} |u - a| \leq \mathcal{B}\}$. With this, we now define

$$\kappa(\eta) := \sup_{\mathfrak{v} \in [0, \mathcal{B}]} \mathfrak{h}(\mathfrak{v} + \eta) - \mathfrak{h}(\mathfrak{v}) \tag{43}$$

Observe that this supremum is finite for every η since, the interval is closed and bounded and the function is continuous. We can now apply Dini's theorem [49, Theorem 7.13] to conclude $\kappa(\eta)$ goes to 0 as η tends to 0. This argument goes as follows.

- Observe that $\mathfrak{h}(\mathfrak{v} + \eta)$ converges point-wise in $\mathfrak{v}$ monotonically to $\mathfrak{h}(\mathfrak{v})$
- Since, $[0, \mathcal{B}]$ is a compact set(closed and bounded), we can argue that $\mathfrak{h}(\mathfrak{v} + \eta)$ uniformly converges to $\mathfrak{h}(\mathfrak{v})$ for $\mathfrak{v} \in [0, \mathcal{B}]$ implying convergence of $\kappa(\eta)$ to 0 as $\eta \to 0$.

Now observe for every $\mathbf{a} \in \mathbf{a}_1(u)$, we get the following

$$\begin{aligned} \min_{a \in \texttt{Set}(\mathbf{a})} \mathfrak{h}\left(|u - a| + \eta\right) = \mathfrak{h}\left(\min_{a \in \texttt{Set}(\mathbf{a})} \mathfrak{h}(|u - a| + \eta\right) &\overset{(a)}{\leq} \mathfrak{h}\left(\min_{a \in \texttt{Set}(\mathbf{a})} |u - a|\right) + \kappa(\eta) \\ &= \min_{a \in \texttt{Set}(\mathbf{a})} \mathfrak{h}(|u - a|) + \kappa(\eta) \end{aligned} \tag{44}$$

Observe that (a) follows from Definition 43 since $\min_{a \in \texttt{Set}(\mathbf{a})} |u - a| \leq \mathcal{B}$ and the equality follows since $\mathfrak{h}(.)$ is monotonic thus, $\mathfrak{h}(\min(a, b)) = \min\left(\mathfrak{h}(a), \mathfrak{h}(b)\right)$

Similarly, for every $\mathbf{a} \in \mathbf{a}_2(u)$, we get the following

$$\begin{aligned} \min_{a \in \texttt{Set}(\mathbf{a})} \mathfrak{h}(|u - a| + \eta) = \mathfrak{h}\left(\min_{a \in \texttt{Set}(\mathbf{a})} (|u - a| + \eta)\right) &\overset{(b)}{\leq} \mathfrak{h}\left(\min_{a \in \texttt{Set}(\mathbf{a})} |u - a|\right) \cdot e^{M\eta} \\ &= \min_{a \in \texttt{Set}(\mathbf{a})} \mathfrak{h}(|u - a|) \cdot e^{M\eta} \end{aligned} \tag{45}$$

(b) follows since, $\log \mathfrak{h}(z)$ is Lipschitz-continous in $[\mathcal{B}, \infty)$, thus $\mathfrak{h}(z + \eta) \leq \mathfrak{h}(z) e^{M\eta}$ for $z \geq \mathcal{B}$.

Combining equations (44) and (45) coupled with the fact that $\mathbf{a}_1(u) \cup \mathbf{a}_2(u) = \mathbb{R}^k$, we get the following result for every $\mathbf{a} \in \mathbb{R}^k$

$$\min_{a \in \texttt{Set}(\mathbf{a})} \mathfrak{h}(|u - a| + \eta) \leq \max\left(\min_{a \in \texttt{Set}(\mathbf{a})} \mathfrak{h}(|u - a|) \cdot e^{M\eta}, \min_{a \in \texttt{Set}(\mathbf{a})} \mathfrak{h}(|u - a|) + \kappa(\eta)\right) \tag{46}$$

$$\leq \min_{a \in \texttt{Set}(\mathbf{a})} \mathfrak{h}(|u - a|) \cdot e^{M\eta} + \kappa(\eta) \tag{47}$$

With this, we now observe that

$$\begin{aligned} &\sup_{u \in \mathbb{R}} \left(\mathbb{E}_{\mathbf{a} \sim P_u}\left[\min_{a \in \texttt{Set}(\mathbf{a})} \mathfrak{h}(|u - a| + \eta)\right]\right) \leq \sup_{u \in \mathbb{R}} \left(\mathbb{E}_{\mathbf{a} \sim P_u}\left[\min_{a \in \texttt{Set}(\mathbf{a})} \mathfrak{h}(|u - a|)\right]\right) e^{M\eta} + \kappa(\eta) \\ \overset{(c)}{\implies} &\lim_{\eta \to 0} \left(\sup_{u \in \mathbb{R}} \left(\mathbb{E}_{\mathbf{a} \sim P_u}\left[\min_{a \in \texttt{Set}(\mathbf{a})} \mathfrak{h}(|u - a| + \eta)\right]\right)\right) \leq \sup_{u \in \mathbb{R}} \left(\mathbb{E}_{\mathbf{a} \sim P_u}\left[\min_{a \in \texttt{Set}(\mathbf{a})} \mathfrak{h}(|u - a|)\right]\right) \end{aligned} \tag{48}$$

(c) follows since $\lim_{\eta\to 0}\kappa(\eta)=0$

Using a very similar approach, we can show that

$$\lim_{\eta\to 0}\left(\sup_{u\in\mathbb{R}}\left(\mathbb{E}_{\mathbf{a}\sim P_u}\left[\min_{a\in\texttt{Set}(\mathbf{a})}\mathfrak{h}([|u-a|-\eta]_+\right]\right)\right)\geq\sup_{u\in\mathbb{R}}\left(\mathbb{E}_{\mathbf{a}\sim P_u}\left[\min_{a\in\texttt{Set}(\mathbf{a})}\mathfrak{h}(|u-a|)\right]\right)\tag{49}$$

Now, combining equations (48), (49) and bounds in equation (42) and (34), we obtain

$$\lim_{\eta\to 0}\sup_{u\in\mathbb{R}}\left(\mathbb{E}_{\mathbf{a}\sim P_u^{(\eta)}}\left[\min_{a\in\texttt{Set}(\mathbf{a})}\mathfrak{h}(|u-a|)\right]\right)=\sup_{u\in\mathbb{R}}\left(\mathbb{E}_{\mathbf{a}\sim P_u}\left[\min_{a\in\texttt{Set}(\mathbf{a})}\mathfrak{h}(|u-a|)\right]\right)\tag{50}$$

Thus, we show that the distributions $\mathbf{P}^{(\eta)}\in\mathcal{P}^{(\epsilon)}_{\mathbb{R}^k}$ with a valid probability density function defined by $\rho_V(u,.)$ arbitrarily approximates $\mathbf{P}$ for every distribution $\mathbf{P}\in\mathcal{P}^{(\epsilon)}_{\mathbb{R}^k}$.

We now construct the DILP $\mathcal{O}$ in (13) where $g(u,.):\mathbb{R}^k\to\mathbb{R}$ denotes the probability density corresponding to $P_u^{(\eta)}$ for every $u\in\mathbb{R}$. The last 2 conditions in DILP $\mathcal{O}$ follow as a consequence of Lemma 2.6.

□

B.3 Dual fitting: Proof of Lemma 2.10

Lemma (detailed proof of Lemma 2.10). $\hat{\mathbf{P}}^{\mathcal{L}_\epsilon}$ satisfies ϵ-geographic differential-privacy constraints i.e. $\hat{P}^{\mathcal{L}_\epsilon}\in\mathcal{P}^{(\epsilon)}_{\mathbb{R}^k}$ and thus, we have $f(\epsilon,k)\leq\text{cost}(\hat{\mathbf{P}}^{\mathcal{L}_\epsilon})=\hat{f}(\epsilon,k)$

Proof. We denote $\mathbf{a}+z$ as $\mathbf{a}+z\mathbb{1}_k$ (i.e.) adding $z\in U$ to every component of $\mathbf{a}\in U^k$

$$\begin{aligned}
f(\epsilon,k)&\overset{(a)}{\leq}\sup_{u\in\mathbb{R}}\mathop{\mathbb{E}}_{\mathbf{a}\sim\hat{P}_u^{\mathcal{L}_\epsilon}}\left[\min_{a\in\texttt{Set}(\mathbf{a})}\mathfrak{h}(|u-a|)\right]\\
&\overset{(b)}{=}\sup_{u\in\mathbb{R}}\mathop{\mathbb{E}}_{S_u\sim\mathcal{L}_\epsilon(u)}\left(\left[\min_{a\in\texttt{Set}(\hat{\mathbf{a}})}\mathfrak{h}(|u-a|)\right]\Bigg|_{\hat{\mathbf{a}}=\underset{\mathbf{a}\in\mathbb{R}^k}{\arg\min}\,\mathbb{E}_{y\sim\mathcal{L}_\epsilon(0)}\left[\min_{a\in\texttt{Set}(\mathbf{a})}\mathfrak{h}(|y-a|)\right]+S_u}\right)\\
&\overset{(c)}{=}\sup_{u\in\mathbb{R}}\mathop{\mathbb{E}}_{S_u\sim\mathcal{L}_\epsilon(u)}\left(\left[\min_{a\in\texttt{Set}(\hat{\mathbf{b}})}\mathfrak{h}(|(u-S_u)-a|)\right]\Bigg|_{\hat{\mathbf{b}}=\underset{\mathbf{a}\in\mathbb{R}^k}{\arg\min}\,\mathbb{E}_{y\sim\mathcal{L}_\epsilon(0)}\left[\min_{a\in\texttt{Set}(\mathbf{a})}\mathfrak{h}(|y-a|)\right]}\right)\\
&\overset{(d)}{=}\sup_{u\in\mathbb{R}}\mathop{\mathbb{E}}_{y\sim\mathcal{L}_\epsilon(0)}\left[\min_{a\in\texttt{Set}(\hat{\mathbf{b}})}\mathfrak{h}(|y-a|)\right]\Bigg|_{\hat{\mathbf{b}}=\underset{\mathbf{a}\in\mathbb{R}^k}{\arg\min}\,\mathbb{E}_{y\sim\mathcal{L}_\epsilon(0)}\left[\min_{a\in\texttt{Set}(\mathbf{a})}\mathfrak{h}(|y-a|)\right]}\\
&\overset{(e)}{=}\min_{\mathbf{a}\in\mathbb{R}^k}\mathbb{E}_{y\sim\mathcal{L}_\epsilon(0)}\left[\min_{a\in\texttt{Set}(\mathbf{a})}\mathfrak{h}(|y-a|)\right]=\hat{f}(\epsilon,k)
\end{aligned}$$

(a) follows from the fact that $\hat{P}^{\mathcal{L}_\epsilon}\in\mathcal{P}^{(\epsilon)}_{\mathbb{R}^k}$ from the post processing theorem, refer to [27] since $S_u\sim\mathcal{L}_\epsilon(u)$ satisfies ϵ-geographic differential privacy constraints.[20]. (b) follows from the definition in Equation (15) and (c) follows on substituting $\hat{\mathbf{b}}=\hat{\mathbf{a}}-S_u$ and (d) follows on substituting $u-S_u$

[20]Post processing theorem can be proven even for ϵ-geographic differential privacy similarly

by y and the fact that $\hat{\mathbf{b}}$ is independent of S_u and (e) follows since $\hat{\mathbf{b}}$ is minimised over the same objective function and independent of u.

□

B.4 Dual fitting: Proofs of Lemmas 2.11 and Lemma 2.12

Recall from Section 2.3 the following assignment to functions $\delta^{(c)}(.)$ and $\lambda^{(c)}(.)$

$$\delta^{(c)}(r) = (\zeta/2)e^{-\zeta|r|} \text{ and } \lambda^{(c)}(r) = \hat{\lambda} \cdot (\zeta/2)e^{-\zeta|r|} \ \forall r \in \mathbb{R}$$

Recall that we considered the following Differential Equation (19) in $\hat{\nu}(.)$.

$$-\left[\min_{a \in \texttt{Set}(\mathbf{v})} \mathfrak{h}(|r-a|)\right] \delta^{(c)}(r) + \lambda^{(c)}(r) + \frac{d\hat{\nu}(r)}{dr} + \epsilon|\hat{\nu}(r)| = 0; \text{ with } \hat{\nu}(v_{med}) = 0$$

We show that there always exists a solution $\hat{\nu}(.)$ to (19) such that $\hat{\nu}(r)$ is non-negative for sufficiently large u and non-positive for sufficiently small r [21] to satisfy the last two constraints of DILP $\mathcal{E}$ in Section below. Observe that the solution could depend on $\mathbf{v}$. In the next subsection, we prove two technical lemmas which is used in the proof of Lemmas and Claims in Section B.4.2.

B.4.1 Technical lemmas used

Lemma B.1. *Consider any vector $\boldsymbol{q} \in \mathbb{R}^k$ and consider any $v_{med} \in \mathbb{R}$ satisfying $v_{med} \geq$ Median($\texttt{Set}(\boldsymbol{q})$). Then the following holds true (recall the definition of $\hat{f}(\epsilon, k)$ from Equation (15))*

$$2\int_{v_{med}}^{\infty} \left[\min_{a \in \texttt{Set}(\boldsymbol{q})} \mathfrak{h}(|t-a|)\right] (\epsilon/2)\, e^{-\epsilon(t-v_{med})} dt \geq \hat{f}(\epsilon, k). \tag{51}$$

Notation: In this proof we use $q_{[j:k]}$ to denote a vector in $\mathbb{R}^{k-j+1}$ constructed from all coordinates of $\mathbf{q}$ starting from j^{th} coordinate to the k^{th} coordinate of $\mathbf{q}$.

Proof. Suppose the median of $\texttt{Set}(\mathbf{q})$ is denoted by q_{med} and construct $\mathbf{v} = \mathbf{q} + (v_{med} - q_{med})\mathbb{1}_k$ [22]. W.L.O.G, we assume that components in $\mathbf{q}$ are sorted in ascending order. Consider the smallest index j in $[k]$ s.t. $q_j > v_{med}$. Clearly, $j > \frac{k+1}{2}$ if k is odd and $j > \frac{k}{2}$ if k is even.

We construct a vector $\tilde{\mathbf{q}} \in \mathbb{R}^k$ by defining $\tilde{q}_{[j:k]} = q_{[j:k]}$ and $\tilde{q}_{[1:k-j+1]} = 2v_{med}\mathbb{1} - q_{[j:k]}$. For all other entries of $\tilde{\mathbf{q}}$, define it to be v_{med}. Observe that we choose a points from $q_{[j:k]}$ and choose other points by symmetrizing around v_{med} and $j > \frac{k}{2} + 1$ implies $v_{med} \in \texttt{Set}(\tilde{\mathbf{q}})$ as $2(k-j+1) < k$. Now consider the following two exhaustive cases.

Case - 1: $j > \frac{k}{2} + 1$.

$$2\int_{v_{med}}^{\infty} \left[\min_{a \in \texttt{Set}(\mathbf{q})} \mathfrak{h}(|t-a|)\right] (\epsilon/2)\, e^{-\epsilon(t-v_{med})} dt \tag{52}$$

$$\overset{(a)}{\geq} 2\int_{v_{med}}^{\infty} \left[\min_{a \in \texttt{Set}(q_{[j:k]}) \cup \{v_{med}\}} \mathfrak{h}(|t-a|)\right] (\epsilon/2)\, e^{-\epsilon(t-v_{med})} dt \tag{53}$$

$$\overset{(b)}{=} \int_{-\infty}^{\infty} \left[\min_{a \in \texttt{Set}(\tilde{\mathbf{q}})} \mathfrak{h}(|t-a|)\right] (\epsilon/2)\, e^{-\epsilon|t-v_{med}|} dt \tag{54}$$

[21] We may still need to prove continuity of the bounds $U(.)$ and $L(.)$ which we discuss in Lemma 2.12

[22] $\mathbb{1}_k$ is a vector in $\mathbb{R}^k$ with all elements unity

$$= \mathbb{E}_{y\sim\mathcal{L}_\epsilon(v_{med})}\left[\min_{a\in\texttt{Set}(\tilde{\mathbf{q}})} \mathfrak{h}(|y-a|)\right] \overset{(c)}{\geq} \hat{f}(\epsilon,k). \tag{55}$$

(a) follows from the fact that q_j is the smallest element in list larger than v_{med}. (b) follows from the fact that $\tilde{\mathbf{q}}$ is constructed from a symmetric extension of $q_{[j:k]}$ about v_{med} and $v_{med} \in \texttt{Set}(\tilde{\mathbf{q}})$. (c) follows from the fact that it is minimised over all collection of k points as defined in Equation (15).

Case -2: $j = \frac{k}{2}+1$. And thus,

$$2\int_{v_{med}}^{\infty}\left[\min_{a\in\texttt{Set}(\mathbf{q})} \mathfrak{h}(|t-a|)\right](\epsilon/2)\, e^{-\epsilon(t-v_{med})}dt \tag{56}$$

$$\overset{(a)}{\geq} 2\int_{v_{med}}^{\infty}\left[\min_{a\in\texttt{Set}(q_{[j:k]})} \mathfrak{h}(|t-a|)\right](\epsilon/2)\, e^{-\epsilon(t-v_{med})}dt \tag{57}$$

$$\overset{(b)}{=} \int_{-\infty}^{\infty}\left[\min_{a\in\texttt{Set}(\tilde{\mathbf{q}})} \mathfrak{h}(|t-a|)\right](\epsilon/2)\, e^{-\epsilon|t-v_{med}|}dt \tag{58}$$

$$= \mathbb{E}_{y\sim\mathcal{L}_\epsilon(v_{med})}\left[\min_{a\in\texttt{Set}(\tilde{\mathbf{q}})} \mathfrak{h}(|y-a|)\right] \overset{(c)}{\geq} \hat{f}(\epsilon,k). \tag{59}$$

Note that (a) follows from the fact that $q_j + q_{j-1} > 2v_{med}$ as Median$(\texttt{Set}(\mathbf{q})) > v_{med}$ and Median$(\texttt{Set}(\mathbf{q})) = \frac{q_j+q_{j-1}}{2}$. (b) follows from the fact that $\tilde{\mathbf{q}}$ is constructed from a symmetric extension of $q_{[j:k]}$ about v_{med}. (c) follows from the fact that it is minimised over all collection of k points as defined in Equation (15).

□

We now state the next lemma which say that we says that there exists a solution to Equation (19). This lemma states that zeros of the solution to Equation (19) with negative derivative is upper bounded and states that zeros with positive derivative is lower bounded.

Lemma B.2. *Choose any $0 < \hat{\lambda} \leq \frac{\epsilon-\zeta}{\epsilon+\zeta}\hat{f}(\epsilon+\zeta,k)$ for some $0 < \zeta < \epsilon$. Then for every $\boldsymbol{v} \in \mathbb{R}^k$, a unique $\mathcal{C}^1$ solution $\nu^{(c)}(.)$ exists to Differential Equation (19) satisfying the following two constraints for some constant C independent of $\boldsymbol{v} \in \mathbb{R}^k$. Note that $v_{\lfloor i \rfloor}$ denotes the i^{th} largest component of $\boldsymbol{v}$ for every $i \in [k]$.*

- $\left\{\mathfrak{r} \in \mathbb{R} : \nu^{(c)}(\mathfrak{r}) = 0 \text{ and } \frac{d\nu^{(c)}(\mathfrak{r})}{d\mathfrak{r}} < 0\right\}$ *is upper bounded by $C + v_{\lfloor k \rfloor}$.*
- $\left\{\mathfrak{r} \in \mathbb{R} : \nu^{(c)}(\mathfrak{r}) = 0 \text{ and } \frac{d\nu^{(c)}(\mathfrak{r})}{d\mathfrak{r}} > 0\right\}$ *is lower bounded by $-C + v_{\lfloor 1 \rfloor}$.*

Proof. Recall Differential Equation from Equation (19), restate it below as follows.

$$-\left[\min_{a\in\texttt{Set}(\mathbf{v})} \mathfrak{h}(|r-a|)\right]\delta^{(c)}(r) + \lambda^{(c)}(r) + \frac{d\hat{\nu}(r)}{dr} + \epsilon|\hat{\nu}(r)| = 0$$

Also recall from Equation (18) that we defined

$$\delta^{(c)}(r) = (\zeta/2)e^{-\zeta|r|} \text{ and } \lambda^{(c)}(r) = \hat{\lambda}\cdot(\zeta/2)e^{-\zeta|r|}\ \forall r \in \mathbb{R}$$

From *global uniqueness and existence* condition of *Cauchy-Lipschitz* theorem in [44, Theorem 2], we can show that there is a *unique* solution $\hat{\nu}^{(c)}(.)$ of Equation (19) satisfying this initial value

condition as *Lipschitz* conditions as described in [44, Theorem 2] is satisfied in Equation (19). Also, observe that this function is *continuously differentiable*.

We now prove the first point in the lemma..

We first prove that $\lim_{u\to\infty} \mathfrak{h}(u) > \hat{\lambda}$. First observe from Claim B.1) (substituting ϵ by $\epsilon+\zeta$) that

$$\int_{v_{med}}^{\infty} \left[\min_{a\in\mathtt{Set(q)}} \mathfrak{h}(|t-a|)\right](\epsilon+\zeta)\, e^{-(\epsilon+\zeta)(t-v_{med})}dt \geq \hat{f}(\epsilon+\zeta,k).$$
$$\implies \int_{v_{med}}^{\infty} \left[\min_{a\in\mathtt{Set(q)}} \mathfrak{h}(|t-a|) - \hat{f}(\epsilon+\zeta,k)\right](\epsilon+\zeta)e^{-(\epsilon+\zeta)(t-v_{med})}dt \geq 0$$

This, implies that for some $t > v_{med}$ s.t. $\left[\min_{a\in\mathtt{Set(q)}} \mathfrak{h}(|t-a|) - \hat{f}(\epsilon+\zeta,k)\right] > 0$ and since, we know $\mathfrak{h}(.)$ is monotonic, we can argue that $\lim_{r\to\infty} \mathfrak{h}(r) > \hat{f}(\epsilon+\zeta,k) > \hat{\lambda}$.[23] Let us denote the the smallest r s.t. $\mathfrak{h}(r) > \hat{\lambda}$ by $\mathfrak{u}^s$. Now observe for $r > (v_{\lfloor k\rfloor}+\mathfrak{r}^s)$, $\left[\min_{a\in\mathtt{Set(v)}} \mathfrak{h}(|r-a|)\right]\delta^{(c)}(r) - \lambda^{(c)}(r)$ is non-negative.

Suppose there exists $r_0 \in [v_{\lfloor k\rfloor}+\mathfrak{r}^s, \infty)$ satisfying, $\nu^{(c)}(r_0) = 0$ and $\nu^{(c)'}(r_0) < 0$ thus, $\nu^{(c)'}(r_0) + \nu^{(c)}(r_0) < 0$, thus contradicting the fact that $\left[\min_{a\in\mathtt{Set(v)}} \mathfrak{h}(|r-a|)\right]\delta^{(c)}(r) - \lambda^{(c)}(r) > 0$ is non-negative in $r \in [v_{\lfloor k\rfloor}+\mathfrak{u}^s, \infty)$. Thus, we show that there does not exist $r_0 > v_{\lfloor k\rfloor}+\mathfrak{r}^s$ satisfying $\nu^{(c)}(u_0) = 0$ and $\nu^{(c)'}(r_0) < 0$ thus proving the desired statement, since $\mathfrak{r}^s$ is a constant independent of $\mathbf{v}$.

The second statement follows from a very similar argument.

□

With these lemmas, we now prove the following lemma 2.11 which shows that $\nu^{(c)}(.)$ is positive for sufficiently large u and negative for sufficiently small u.

B.4.2 Obtaining a feasible solution to DILP $\mathcal{E}$: Proof of Lemma 2.11

Lemma (detailed proof of Lemma 2.11). Choose $\zeta < \epsilon$ and $0 < \hat{\lambda} \leq \frac{\epsilon-\zeta}{\epsilon+\zeta}\hat{f}(\epsilon+\zeta,k)$, then equation (19) has a unique $\mathcal{C}^1$ solution $\nu^{(c)}(.)$ and there exists $U, L \in \mathbb{R}$ satisfying $\nu^{(c)}(r) \geq 0\ \forall r \geq U$ and $\nu^{(c)}(r) \leq 0\ \forall r \leq L$.

Proof. Recall from lemma B.2, that there exists a unique $\mathcal{C}^1$ solution (denoted by $\nu^{(c)}(.)$) to (19). Uniqueness and existence of the solution can be shown from *global uniqueness and existence* condition of *Cauchy-Lipschitz* theorem in [44, Theorem 2] as *Lipschitz* conditions as described in [44, Theorem 2] is satisfied in Equation (19).

We first prove the following Equation (60) in Part - 1 which states that there cannot be an unbounded interval in $[v_{med}, \infty)$ where $\nu^{(c)}(.)$ is non-positive. Next, we use this statement to prove the main lemma in Part - 2.

$$\left\{\mathfrak{r} \in [v_{med}, \infty) \;\middle|\; \nu^{(c)}(r) \leq 0\ \forall r \geq \mathfrak{r}\right\} = \phi \tag{60}$$

Part - 1. We now assume the *contradictory* statement that there exists $r^0 \in \mathbb{R}$ s.t. $\nu^{(c)}(r) \leq 0$ for all $r > r^0$ and define $r^{\max} = \inf\left\{\mathfrak{r} \in [v_{med}, \infty) \;\middle|\; \nu^{(c)}(r) \leq 0\ \forall r \geq \mathfrak{r}\right\}$. Since, $\nu^{(c)}(.)$ is continuous and $\nu^{(c)}(v_{med}) = 0$, we have $\nu^{(c)}(r^{max}) = 0$. Thus, the differential equation (19) can be rewritten in $[r^{\max}, \infty)$ as follows (replacing $|\hat{\nu}(u)|$ by $-\hat{\nu}(u)$)

[23] The limits are defined over extended reals $\mathbb{R} \cup \{-\infty, +\infty\}$

$$-\left[\min_{a\in\texttt{Set}(\mathbf{v})}\mathfrak{h}(|r-a|)\right]\delta^{(c)}(r)+\lambda^{(c)}(r)+\frac{d\hat{\nu}(r)}{dr}-\epsilon\hat{\nu}(r)=0$$

Now, multiplying both sides by $e^{-\epsilon r}$ and applying the initial value condition of $\nu^{(c)}(v_{med})=0$, we have the solution to the Differential Equation in $[r^{\max},\infty)$ is given by

$$\nu^{(c)}(r)=\left(\int_{r^{\max}}^{r}\left[-\lambda^{(c)}(t)+\left[\min_{a\in\texttt{Set}(\mathbf{v})}\mathfrak{h}(|t-a|)\right]\delta^{(c)}(t)\right]e^{-\epsilon t}dt\right)e^{\epsilon r}$$

However, since $\hat{\nu}^{(c)}(r)$ is non-positive in $[r^{\max},\infty)$, we have[24]

$$\int_{r^{\max}}^{\infty}\left[-\lambda^{(c)}(t)+\left[\min_{a\in\texttt{Set}(\mathbf{v})}\mathfrak{h}(|t-a|)\right]\delta^{(c)}(t)\right]e^{-\epsilon t}dt<0 \tag{61}$$

$$\implies\int_{r^{\max}}^{\infty}\left[\min_{a\in\texttt{Set}(\mathbf{v})}\mathfrak{h}(|t-a|)\right]\left(\frac{\zeta}{2}\right)e^{-\zeta|t|}(\epsilon/2)e^{-\epsilon t}dt<\int_{r^{\max}}^{\infty}\lambda^{(c)}(t)(\epsilon/2)e^{-\epsilon t}dt \tag{62}$$

We now consider two exhaustive cases and treat them separately $r^{\max}>0$ and $r^{\max}<0$. We denote $q_i=v_i+v_{med}-r^{\max}\forall i\in[k]$ and $\mathbf{q}=[q_1,q_2,\ldots,q_k]$

Case a: Let us first consider the case where $r^{\max}>0$. We thus get from Equation (62)

$$\overset{(a)}{\implies}e^{-(\epsilon+\zeta)(r^{\max}-v_{med})}.\int_{v_{med}}^{\infty}\left[\min_{a\in\texttt{Set}(\mathbf{q})}\mathfrak{h}(|t-a|)\right]\left(\frac{\zeta}{2}\right)e^{-\zeta t}(\epsilon/2)e^{-\epsilon t}dt<(1/4)\hat{\lambda}e^{-(\epsilon+\zeta)r^{\max}}\left(\frac{\epsilon\zeta}{\epsilon+\zeta}\right) \tag{63}$$

$$\implies 2\int_{v_{med}}^{\infty}\left[\min_{a\in\texttt{Set}(\mathbf{q})}\mathfrak{h}(|t-a|)\right]((\epsilon+\zeta)/2)e^{-(\epsilon+\zeta)t}dt<\hat{\lambda}e^{-(\epsilon+\zeta)v_{med}} \tag{64}$$

$$\implies 2\int_{v_{med}}^{\infty}\left[\min_{a\in\texttt{Set}(\mathbf{q})}\mathfrak{h}(|t-a|)\right]((\epsilon+\zeta)/2)e^{-(\epsilon+\zeta)(t-v_{med})}dt<\hat{\lambda} \tag{65}$$

$$\overset{(b)}{\implies}\hat{f}(\epsilon+\zeta,k)<\hat{\lambda} \tag{66}$$

which is a contradiction.

(a) follows from the fact that $r^{\max}>0$, thus $|t|=t$ and applying change of variables. The argument for (b) follows from Lemma B.1.

Case b: Consider the case when $r^{\max}<0$. We can thus write Equation (62)

$$\implies\int_{r^{\max}}^{0}\left[\min_{a\in\texttt{Set}(\mathbf{v})}\mathfrak{h}(|t-a|)\right]\left(\frac{\epsilon\zeta}{4}\right)e^{\zeta t}e^{-\epsilon t}dt+\int_{0}^{\infty}\left[\min_{a\in\texttt{Set}(\mathbf{v})}\mathfrak{h}(|t-a|)\right]\left(\frac{\zeta\epsilon}{4}\right)e^{-\zeta t}e^{-\epsilon t}dt \tag{67}$$

$$<\int_{r^{\max}}^{0}\hat{\lambda}\left(\epsilon\zeta/4\right)e^{\zeta t}e^{-\epsilon t}dt+\int_{0}^{\infty}\hat{\lambda}\left(\epsilon\zeta/4\right)e^{-\zeta t}e^{-\epsilon t}dt \tag{68}$$

[24]Observe that a strict inequality follows since a contrary assumption would imply that $\nu^{(c)}(r)$ is identically zero in $[r^{med},\infty)$. This would imply that $-\lambda^{(c)}(t)+\left[\min_{a\in\texttt{Set}(\mathbf{v})}\mathfrak{h}(|t-a|)\right]\delta^{(c)}(t)$ is zero in $[r_{max},\infty)$ implying, $\mathfrak{h}(|t-a|)=\hat{\lambda}$ which is a contradiction since, $\lim_{r\to\infty}\mathfrak{h}(r)>\hat{\lambda}$ as shown in proof of Claim B.2

$$\overset{(a)}{\implies} e^{-(\epsilon-\zeta)r^{\max}} \int_{v_{med}}^{v_{med}-r^{\max}} \left[\min_{a\in\texttt{Set}(\mathbf{q})} \mathfrak{h}(|t-a|)\right] e^{-(\epsilon-\zeta)(t-v_{med})}dt \tag{69}$$

$$+e^{-(\epsilon+\zeta)r^{\max}} \int_{v_{med}-r^{\max}}^{0} \left[\min_{a\in\texttt{Set}(\mathbf{q})} \mathfrak{h}(|t-a|)\right] e^{-(\epsilon+\zeta)(t-v_{med})}dt < \hat{\lambda}\left[\frac{1}{(\epsilon+\zeta)} + \frac{e^{-(\epsilon-\zeta)r^{\max}}-1}{(\epsilon-\zeta)}\right] \tag{70}$$

$$\overset{(b)}{\implies} e^{-(\epsilon-\zeta)r^{\max}} \int_{v_{med}}^{\infty} \left[\min_{a\in\texttt{Set}(\mathbf{q})} \mathfrak{h}(|t-a|)\right] e^{-(\epsilon+\zeta)(t-v_{med})}dt < \hat{\lambda}\left[\frac{1}{(\epsilon+\zeta)} + \frac{e^{-(\epsilon-\zeta)r^{\max}}-1}{(\epsilon-\zeta)}\right] \tag{71}$$

$$\implies 2e^{-(\epsilon-\zeta)r^{\max}} \int_{v_{med}}^{\infty} \left[\min_{a\in\texttt{Set}(\mathbf{q})} \mathfrak{h}(|t-a|)\right] \left(\frac{\epsilon+\zeta}{2}\right) e^{-(\epsilon+\zeta)(t-v_{med})}dt < \hat{\lambda}\left[1+\frac{(\epsilon+\zeta)}{(\epsilon-\zeta)}(e^{-(\epsilon-\zeta)r^{\max}}-1)\right] \tag{72}$$

$$\implies \frac{e^{-(\epsilon-\zeta)r^{\max}}2}{\left[1+\frac{(\epsilon+\zeta)}{(\epsilon-\zeta)}(e^{-(\epsilon-\zeta)r^{\max}}-1)\right]} \int_{v_{med}}^{\infty} \left[\min_{a\in\texttt{Set}(\mathbf{q})} \mathfrak{h}(|t-a|)\right] \left(\frac{\epsilon+\zeta}{2}\right) e^{-(\epsilon+\zeta)(t-v_{med})}dt < \hat{\lambda} \tag{73}$$

$$\overset{(c)}{\implies} \left(\frac{\epsilon-\zeta}{\epsilon+\zeta}\right) 2\int_{v_{med}}^{\infty} \left[\min_{a\in\texttt{Set}(\mathbf{q})} \mathfrak{h}(|t-a|)\right] \left(\frac{\epsilon+\zeta}{2}\right) e^{-(\epsilon+\zeta)(t-v_{med})}dt < \hat{\lambda} \tag{74}$$

$$\overset{(d)}{\implies} \left(\frac{\epsilon-\zeta}{\epsilon+\zeta}\right) \hat{f}(\epsilon+\zeta,k) < \hat{\lambda} \tag{75}$$

which is a contradiction.

(b) follows from the fact that $\epsilon-\zeta < \epsilon+\zeta$ and thus, $-(\epsilon-\zeta)r^{\max} < -(\epsilon+\zeta)r^{\max}$ since $r^{\max} < 0$. Also, $-(\epsilon+\zeta)(t-v_{med}) < -(\epsilon-\zeta)(t-v_{med})$ for $t > v_{med}$. (c) follows from by infimising $\frac{e^{-(\epsilon-\zeta)r^{\max}}}{\left[1+\frac{(\epsilon+\zeta)}{(\epsilon-\zeta)}(e^{-(\epsilon-\zeta)r^{\max}}-1)\right]}$ over all $r^{\max} < 0$ (happens as $r^{\max} \to -\infty$) and (d) follows from Lemma B.1.

We thus prove Equation (60) by showing contradiction under both cases.

Part - 2. Observe that Equation (76) follows from Claim B.2 and Equation (60) (restated below) which has been shown above.

$$\left\{\mathfrak{r} \in \mathbb{R} : \nu^{(c)}(\mathfrak{r}) = 0 \text{ and } \frac{d\nu^{(c)}(\mathfrak{r})}{d\mathfrak{r}} < 0\right\} \text{ is upper bounded by } C + v_{\lfloor k \rfloor}. \tag{76}$$

.

$$\left\{\mathfrak{r} \in \mathbb{R} \;\middle|\; \nu^{(c)}(r) \leq 0 \;\forall r > \mathfrak{r}\right\} = \phi$$

Let us denote the set of all the closed intervals in $[v_{med}, \infty)$, where $\nu^{(c)}(.)$ is non-negative by $\mathcal{I}^{pos}$[25]. Now observe, that the Equation (76) implies that *upper bound* of every interval in $\mathcal{I}^{pos}$ must be upper bounded by $C + v_{\lfloor k \rfloor}$. This implies the existence of an unbounded interval where $\nu^{(c)}(.)$ is non-negative or an unbounded interval where $\nu^{(c)}(.)$ is non-positive. However Equation (60) implies that there can not exist an unbounded interval where $\nu^{(c)}(.)$ is non-positive and thus, we show that there exists an unbounded interval where $\nu^{(c)}(.)$ is non-negative. This proves the statement in Claim 2.11 that there exists $U \in \mathbb{R}$ s.t. $\nu^{(c)}(u) \geq 0 \forall u \geq U$.

Using a very similar approach, we can prove that $\nu^{(c)}(u) \leq 0$ for $u \leq L$ for some $L \in \mathbb{R}$.

□

[25]We can construct such intervals as $\nu^{(c)}$ is *continuously differentiable* and there does not exist an interval where $\nu^{(c)}$ is identically zero.

We now prove the main lemma 2.12 which shows that the objective value of $\hat{f}(\epsilon, k)$ is achievable by some solution in the DILP $\mathcal{E}$. Observe that this feasible solution is constructed using $\delta^{(c)}(.), \lambda^{(c)}(.)$ and $\nu^{(c)}(.)$ that we defined to satisfy Equation (19) and ensured positivity and negativity of $\nu^{(c)}(.)$ for "sufficiently" large and small u respectively for every $\mathbf{v} \in \mathbb{R}^k$. However, we also require a technical claim 1 to prove the existence of continuous bounds $U(.)$ and $L(.)$. For sake of brevity, we state and prove it in Appendix B.7.

Lemma (detailed proof of Lemma 2.12). $\text{opt}(\mathcal{E}) \geq \hat{f}(\epsilon, k)$

Proof. Recall the functions $\lambda^{(c)}(.), \delta^{(c)}(.)$ defined in (18). Also for every $\mathbf{v} \in \mathbb{R}^k$, we obtain a function $\nu^{(c)}(., \mathbf{v})$ [solution of Equation (19)] with bounds $U(\mathbf{v})$ and $L(\mathbf{v})$ satisfying $\nu^{(c)}(r, \mathbf{v}) \geq 0 \forall u \geq U(\mathbf{v})$ and $\nu^{(c)}(r, \mathbf{v}) \leq 0 \forall u \leq L(\mathbf{v})$. Now, we argue that this solution is feasible attaining an objective value of $\hat{\lambda}$ which we argue below.

- Observe that $\int_{u \in \mathbb{R}} \lambda^{(c)}(u) = \hat{\lambda}$ and $\int_{u \in \mathbb{R}} \delta^{(c)}(u) = 1$
- The second constraint is satisfied as $\nu^{(c)}(\mathbf{v}, .)$ is a solution of Equation (19).
- Bounds $U(\mathbf{v})$ and $L(\mathbf{v})$ exist from statement in Lemma 2.11. The proof of continuity of $U(.)$ and $L(.)$ is slightly technical and we formally prove this in Claim 1. Observe that all the assumptions in Claim 1 is satisfied due to results from Lemma 2.11 and Lemma B.2.

Now observe that the objective value of this feasible solution in $\hat{\lambda}$ and the constructed solution is feasible for any $\hat{\lambda} \leq \frac{\epsilon - \zeta}{\epsilon + \zeta} \hat{f}(\epsilon + \zeta, k)$ and $\zeta > 0$. Now, since $\hat{f}(\epsilon, k)$ is continuous in ϵ, choosing ζ to be arbitrarily small enables us to get obtain the objective value of the solution arbitrarily close to $\hat{f}(\epsilon, k)$ and thus, $\text{opt}(\mathcal{E}) \geq \hat{f}(\epsilon, k)$.

□

B.5 Optimal Result Selection given Laplace noise when $\mathfrak{h}(.)$ is an identity function

Theorem B.3 (corresponds to Theorem 1.4)**.** *Recall the definition of $\hat{f}(\epsilon, k)$ from Equation* (15) *and suppose $A^* = \{x_1, x_2, \ldots, x_k\}$ where $x_1 \leq x_2 \leq \cdots \leq x_k$ minimises the same.*[26] *Then, we have*

1. *When k is odd, $x_{(k+1)/2} = 0$. For $i \in [(k-1)/2]$, $x_i = -2\log((k+3)/2 - i)/\epsilon$. For $i \in [k] \setminus [(k+1)/2]$, $x_i = -x_{k+1-i}$. That is, if $t := (k-1)/2$, then*

$$\{x_1, \ldots, x_k\} = \left\{ 0, \pm\frac{2\log\left(\frac{t+1}{t}\right)}{\epsilon}, \pm\frac{2\log\left(\frac{t+1}{t-1}\right)}{\epsilon}, \ldots, \pm\frac{2\log\left(\frac{t+1}{2}\right)}{\epsilon}, \pm\frac{2\log\left(t+1\right)}{\epsilon} \right\}.$$

2. *When k is even, $x_{k/2} = -\log(1 + 2/k)/\epsilon$. For $i \in [k/2 - 1]$, $x_i = x_{i+1} - (2\log(1 + 1/i))/\epsilon$. For $i \in [k] \setminus [k/2]$, $x_i = -x_{k+1-i}$. That is, if $t := k/2$, then*

$$\{x_1, \ldots, x_k\} = \left\{ \pm\frac{\log\left(\frac{t+1}{t}\right)}{\epsilon}, \pm\frac{\log\left(\frac{t(t+1)}{(t-1)^2}\right)}{\epsilon}, \pm\frac{\log\left(\frac{t(t+1)}{(t-2)^2}\right)}{\epsilon}, \ldots, \pm\frac{\log\left(\frac{t(t+1)}{2^2}\right)}{\epsilon}, \pm\frac{\log\left(t(t+1)\right)}{\epsilon} \right\}.$$

We prove this theorem using the following lemmas.

[26] Here, we denote A^* by a set instead of a vector.

Lemma B.4. *If* $\min_{A\in\mathcal{A}} \mathbb{E}_{x\sim\mathcal{L}_\epsilon(0)}\left[\min_{a\in A}|a-x|\right] = \mathbb{E}_{x\sim\mathcal{L}_\epsilon(0)}\left[\min_{a\in A^*}|a-x|\right]$ *and* $A^* = \{x_1, x_2, \ldots, x_k\}$ *where* $x_1 \le x_2 \le \cdots \le x_k$*, we define* $y_i := (x_i + x_{i+1})/2$ *for* $i \in [k-1]$*,* $y_0 := -\infty$*, and* $y_k := +\infty$*. Then,*

$$\mathbb{P}_{z\sim\mathcal{L}_\epsilon(0)}\left(y_{i-1} < z < x_i\right) = \mathbb{P}_{z\sim\mathcal{L}_\epsilon(0)}\left(x_i < z < y_i\right), \quad \forall \ \ i \in [k]. \tag{77}$$

Proof. Note that for any $i \in [k]$, for any $z \in (y_{i-1}, y_i)$, $\min_{a\in A^*}|a-x| = |x_i - x|$. Then we have

$$\mathbb{E}_{x\sim\mathcal{L}_\epsilon(0)}\left[\min_{a\in A^*}|a-x|\right] = \sum_{i=1}^{k}\mathbb{P}_{z\sim\mathcal{L}_\epsilon(0)}\left(y_{i-1} < z < y_i\right)\mathbb{E}_{z\sim\mathcal{L}_\epsilon(0)}\left[|x_i - z|\,|y_{i-1} < z < y_i\right].$$

If $x_i \neq \mathbb{E}_{z\sim\mathcal{L}_\epsilon(0)}[z|y_{i-1} < z < y_i]$ for a certain $i \in [k]$, we can change the value of x_i so that Equation 77 holds, in which x_i is somehow a median of $\mathcal{L}_\epsilon(0)$ inside the interval (y_{i-1}, y_i), and then $\mathbb{E}_{z\sim\mathcal{L}_\epsilon(0)}\left[|x_i - z|\,|y_{i-1} < z < y_i\right]$ strictly decreases, which implies $\mathbb{E}_{x\sim\mathcal{L}_\epsilon(0)}\left[\min_{a\in A^*}|a-x|\right]$ decreases. However, A^* is the optimal point, so it is a contradiction. As a result, for $i \in [k]$, we have Equations 77.

□

Lemma B.5. *Using the same definition of* x_i *and* y_i *as Lemma B.4, we have*

1. *for* $i \in [k-1]$*, if* $y_i \le 0$*, we have* $x_i = y_i - \frac{\log(1+1/i)}{\epsilon}$*,*
2. *for* $i \in [k] \setminus \{1\}$*, if* $y_{i-1} \ge 0$*, we have* $x_i = y_{i-1} + \frac{\log(1+1/(k-i+1))}{\epsilon}$*.*

Proof. We first show $x_i = y_i - \frac{\log(1+1/i)}{\epsilon}$ when $y_i \le 0$. We prove it by induction. When $i = 1$, by Equation 77, we have $\mathbb{P}_{z\sim\mathcal{L}_\epsilon(0)}\left(z < x_1\right) = \mathbb{P}_{z\sim\mathcal{L}_\epsilon(0)}\left(x_1 < z < y_1\right)$. If $y_1 \le 0$, we have

$$\int_{-\infty}^{x_1} e^{\epsilon z}dz = \int_{x_1}^{y_1} e^{\epsilon z}dz,$$

which is equivalent to

$$e^{\epsilon x_1} = e^{\epsilon y_1} - e^{\epsilon x_1},$$

so $y_1 = x_1 + \log(2)/\epsilon$.

If $y_i \le 0$, assuming $x_{i-1} = y_{i-1} - \frac{\log(1+1/(i-1))}{\epsilon}$, then $x_i = y_{i-1} + \frac{\log(1+1/(i-1))}{\epsilon}$. By Equation 77, we have

$$\int_{y_{i-1}}^{x_i} e^{\epsilon z}dz = \int_{x_i}^{y_i} e^{\epsilon z}dz,$$

which is equivalent to

$$\exp\left(\epsilon x_i\right) - \exp\left(\epsilon x_i - \frac{\log(1+1/(i-1))}{\epsilon}\right) = \exp\left(\epsilon y_i\right) - \exp\left(\epsilon x_i\right),$$

so $y_i = x_i + \frac{\log(1+1/i)}{\epsilon}$.

As a result, for $i \in [k-1]$, if $y_i \le 0$, we have $x_i = y_i - \frac{\log(1+1/i)}{\epsilon}$.

By symmetry, using similar arguments, we know that for $i \in [k] \setminus \{1\}$, if $y_{i-1} \ge 0$, then $x_i = y_{i-1} + \frac{\log(1+1/(k-i+1))}{\epsilon}$.

□

Lemma B.6. *Using the same definition of x_i and y_i as Lemma B.4, if $y_i \le 0 \le y_{i+1}$ where $i \in [k-2]$, then $|i-(k-i)| \le 1$.*

Proof. By Lemma B.5, we have

$$\begin{cases} x_i = y_i - \frac{\log(1+1/i)}{\epsilon}, \\ x_{i+2} = y_{i+1} + \frac{\log(1+1/(k-i-1))}{\epsilon}, \end{cases}$$

which implies

$$y_i + \frac{\log(1+1/i)}{\epsilon} = x_{i+1} = y_{i+1} - \frac{\log(1+1/(k-i-1))}{\epsilon}.$$

Since $y_i \le 0 \le y_{i+1}$, we have $x_{i+1} \in [\beta_1, \beta_2]$, where

$$\beta_1 := -\frac{\log(1+1/(k-i-1))}{\epsilon}, \quad \beta_2 := \frac{\log(1+1/i)}{\epsilon}.$$

Define

$$g_1(x) := \int_{x-\frac{\log(1+1/i)}{\epsilon}}^{x} e^{-\epsilon|z|}dz, \quad \text{and} \quad g_2(x) := \int_{x}^{x+\frac{\log(1+1/(k-i-1))}{\epsilon}} e^{-\epsilon|z|}dz,$$

and define $h(x) := g_1(x)/g_2(x)$. By Lemma B.4, we have $g_1(x_{i+1}) = g_2(x_{i+1})$, which implies $h(x_{i+1}) = 1$.

For $\beta_1 \le y_1 < y_2 \le 0$, we have $g_1(y_1) = e^{y_1-y_2}g_1(y_2)$, and $g_2(y_1) > e^{y_1-y_2}g_2(y_2)$, so $h(y_1) < h(y_2)$. Similarly, for $0 \le y_1 < y_2 \le \beta_2$, we have $g_1(y_1) < e^{y_2-y_1}g_1(y_2)$, and $g_2(y_1) = e^{y_2-y_1}g_2(y_2)$, so $h(y_1) < h(y_2)$. As a result, $h(x)$ is strictly increasing on $[\beta_1, \beta_2]$.

If $i > k-i+1$, $h(\beta_2) < 1$. If $i+1 < k-i$, $h(\beta_1) > 1$. In both cases, for any $x \in [\beta_1, \beta_2]$, $h(x) \ne 1$. It contradicts to $h(x_{i+1}) = 1$ and $x_{i+1} \in [\beta_1, \beta_2]$. As a result, we have $|i-(k-i)| \le 1$. □

Combining Lemmas B.5 and B.6, we have the following corollary B.3.

And thus, Theorem A.1 is proved.

B.6 Proof of weak duality as stated in Theorem 2.8

We now restate the weak duality theorem from Theorem 2.8.

Theorem (proof of Theorem 2.8). $\text{opt}(\mathcal{O}) \ge \text{opt}(\mathcal{E})$.

To prove this lemma, we now redefine the optimization problems $\mathcal{O}$ and $\mathcal{E}$ in (13) and (14).

$$\mathcal{O} = \begin{cases} \inf\limits_{g(.,.):\mathcal{I}^B(\mathbb{R}\times\mathbb{R}^k\to\mathbb{R}^+),\kappa\in\mathbb{R}} & \kappa \\ & \text{s.t.} \quad \kappa - \int\limits_{\mathbf{x}\in\mathbb{R}^k} \left[\min\limits_{a\in\texttt{Set}(\mathbf{x})} \mathfrak{h}(|u-a|)\right] g(u,\mathbf{x}) d\left(\prod_{i=1}^k x_i\right) \ge 0 \ \forall u \in \mathbb{R} \\ & \int\limits_{\mathbf{x}\in\mathbb{R}^k} g(u,\mathbf{x}) d\left(\prod_{i=1}^k x_i\right) = 1 \ \forall u \in \mathbb{R} \\ & \epsilon g(u,\mathbf{x}) + \underline{g}_u(u,\mathbf{x}) \ge 0; \ \forall u \in \mathbb{R}; \mathbf{x} \in \mathbb{R}^k \\ & \epsilon g(u,\mathbf{x}) - \overline{g}_u(u,\mathbf{x}) \ge 0; \ \forall u \in \mathbb{R}; \mathbf{x} \in \mathbb{R}^k \end{cases} \tag{78}$$

$$\mathcal{E} = \begin{cases} \sup\limits_{\substack{\nu(.,.):\mathcal{C}^1(\mathbb{R}\times\mathbb{R}^k\to\mathbb{R}) \\ \lambda(.):\mathcal{C}^0(\mathbb{R}\to\mathbb{R}^+) \\ \delta(.):\mathcal{C}^0(\mathbb{R}\to\mathbb{R}^+)}} \int_{u\in\mathbb{R}} \lambda(u)du \\ \quad \text{s.t.} \quad \int_{u\in\mathbb{R}} \delta(u)du \le 1 \\ \quad -\left[\min\limits_{a\in\texttt{Set}(\mathbf{x})} \mathfrak{h}(|u-a|)\right]\delta(u) + \lambda(u) + \epsilon|\nu(u,\mathbf{x})| + \nu_u(u,\mathbf{x}) \le 0 \ \forall u\in\mathbb{R}, \forall \mathbf{x}\in\mathbb{R}^k \\ \quad\quad \exists U : \mathcal{C}^0(\mathbb{R}^k\to\mathbb{R}) \text{ s.t. } \nu(u,\mathbf{x}) \ge 0 \ \forall u \ge U(\mathbf{x}) \ \forall \mathbf{x}\in(\mathbb{R})^k \\ \quad\quad \exists L : \mathcal{C}^0(\mathbb{R}^k\to\mathbb{R}) \text{ s.t. } \nu(u,\mathbf{x}) \le 0 \ \forall u \le L(\mathbf{x}) \ \forall \mathbf{x}\in(\mathbb{R})^k \end{cases} \tag{79}$$

We now define a DILP $\mathcal{E}^{int}$ which effectively splits the function $\nu(.,.)$ into a negative and a positive part and prove a lemma B.7 bounding the optimal value of $\mathcal{E}$ by $\mathcal{E}^{int}$.

$$\mathcal{E}^{int} = \begin{cases} \sup\limits_{\substack{\delta(.):\mathcal{C}^0(\mathbb{R}\to\mathbb{R}^+),\nu^{(1)}(.,.):\mathcal{C}^1(\mathbb{R}\times\mathbb{R}^k\to\mathbb{R}^+) \\ \nu^{(2)}(.,.):\mathcal{C}^1(\mathbb{R}\times\mathbb{R}^k\to\mathbb{R}^+),\lambda(.):\mathcal{C}^0(\mathbb{R}\to\mathbb{R}^+)}} \int_{u\in\mathbb{R}} \lambda(u)du \\ \quad \text{s.t.} \quad \int_{u\in\mathbb{R}} \delta(u)du \le 1 \\ \quad -\left[\min\limits_{a\in\texttt{Set}(\mathbf{x})} \mathfrak{h}(|u-a|)\right]\delta(u) + \lambda(u) + \epsilon\left(-\nu^{(1)}(u,\mathbf{x}) + \nu^{(2)}(u,\mathbf{x})\right) \\ \quad\quad + \nu_u^{(1)}(u,\mathbf{x}) + \nu_u^{(2)}(u,\mathbf{x}) \le 0 \ \forall u\in\mathbb{R}, \forall \mathbf{x}\in\mathbb{R}^k \\ \quad \exists U : \mathcal{C}^0(\mathbb{R}^k\to\mathbb{R}) \text{ s.t. } -\nu^{(1)}(u,\mathbf{x}) + \nu^{(2)}(u,\mathbf{x}) \ge 0 \ \forall u \ge U(\mathbf{x}) \ \forall \mathbf{x}\in(\mathbb{R})^k \\ \quad \exists L : \mathcal{C}^0(\mathbb{R}^k\to\mathbb{R}) \text{ s.t. } -\nu^{(1)}(u,\mathbf{x}) + \nu^{(2)}(u,\mathbf{x}) \le 0 \ \forall u \le L(\mathbf{x}) \ \forall \mathbf{x}\in(\mathbb{R})^k \end{cases} \tag{80}$$

We now prove the following set of lemmas and denote the optimal value of DILP given by $\mathcal{O}$ as $\text{opt}(\mathcal{O})$

Lemma B.7. *opt*$(\mathcal{E}^{int}) \ge$ *opt*$(\mathcal{E})$

Proof. This proof follows by restricting exactly one of the values $\nu^{(1)}(u,\mathbf{x})$ or $\nu^{(2)}(u,\mathbf{x}) = 0$. In this case, we may define $\nu(u,\mathbf{x}) = -\nu^{(1)}(u,\mathbf{x}) + \nu^{(2)}(u,\mathbf{x})$ and thus, $|\nu(u,\mathbf{x})| = \nu^{(1)}(u,\mathbf{x}) + \nu^{(2)}(u,\mathbf{x})$. Since, we restrict the optimization variables i.e. add an extra constraint, we get $\text{opt}(\mathcal{E}^{int}) \ge \text{opt}(\mathcal{E})$ □

Lemma B.8. *opt*$(\mathcal{O}) \ge$ *opt*$(\mathcal{E}^{int})$.

Before we start the proof, we give some key observations which allow us to prove the weak duality. Observe that the constraint of DILP $\mathcal{O}$ involves a linear constraint of derivative of $g(.,.)$. We use integration by parts in this proof to convert this into a constraint on the derivative of the dual variable $\nu(.)$ in the dual DILP $\mathcal{E}$. We state the equation for this below. [27]

$$\int_{u\in\mathbb{R}} (\nu^{(1)}(u,\mathbf{x}) - \nu^{(2)}(u,\mathbf{x}))g_u(u,\mathbf{x})du \tag{81}$$

[27] For sake of brevity, we give the inequality using derivatives and not upper and lower derivatives. A formal version of this is given in equation (98).

$$= \Big[(\nu^{(1)}(u,\mathbf{x}) - \nu^{(2)}(u,\mathbf{x}))g(u,\mathbf{x})\Big]_{u=-\infty}^{+\infty} - \int_{u\in\mathbb{R}} (\nu_u^{(1)}(u,\mathbf{x}) - \nu_u^{(2)}(u,\mathbf{x}))g(u,\mathbf{x})du \tag{82}$$

$$\leq - \int_{u\in\mathbb{R}} (\nu^{(1)}(u,\mathbf{x}) - \nu^{(2)}(u,\mathbf{x}))g(u,\mathbf{x}) \tag{83}$$

The last step follows from the last two constraints in DILP $\mathcal{E}$ i.e. $\lim_{u\to\infty} \nu^{(1)}(u,\mathbf{x}) - \nu^{(2)}(u,\mathbf{x}) \leq 0$ and $\lim_{u\to-\infty} \nu^{(1)}(u,\mathbf{x}) - \nu^{(2)}(u,\mathbf{x}) \geq 0$

We now move to the proof of the weak duality. Notice that this proof B.6.2 bears resemblance to the weak duality proof, albeit with a clever utilization of integration by parts, Fatou's lemma, and the monotone convergence theorem. This approach allows for the exchange of limits and integrals in a strategic manner.

To get an intuition on the proof steps, we first present an informal proof B.6.1 where we assume functions are always integrable and exchange limits and integrals without giving explicit reasons. To look into its formal treatment refer to Section B.6.2

B.6.1 Informal proof to Theorem 2.8

As discussed above, in this proof we just give an intuition for the steps to gain an understanding without giving formal reasons for exchange of integrals and limits. We further assume that the lower and upper derivatives are integrable in this part. To look into its formal treatment refer to Proof in Section B.6.2.

Informal proof. Now consider any feasible solution $g(.,.)$ and κ in the primal DILP $\mathcal{O}$ and feasible solution $\nu^{(1)}(.,.), \nu^{(2)}(.,.), \lambda(.)$ and $\delta(.)$ in the dual DILP $\mathcal{E}^{int}$.

Now pre-multiply the first constraint in DILP $\mathcal{O}$ by $\delta(u)$ second constraint in DILP by $\lambda(u)$, the third constraint in DILP by $\nu^{(1)}(u,\mathbf{x})$ and fourth constraint in DILP $\mathcal{O}$ by $\nu^{(2)}(u,\mathbf{x})$

$$\int_{u\in\mathbb{R}} \delta(u) \left[\kappa - \int_{\mathbf{x}\in\mathbb{R}^k} \left(\min_{a\in\texttt{Set}(\mathbf{x})} \mathfrak{h}(|u-a|)\right) g(u,\mathbf{x}) d\left(\prod_{i=1}^k x_i\right)\right] du + \int_{u\in\mathbb{R}} \int_{\mathbf{x}\in\mathbb{R}^k} g(u,\mathbf{x})\lambda(u) d\left(\prod_{i=1}^k x_i\right) du$$

$$+ \int_{u\in\mathbb{R}} \int_{\mathbf{x}\in\mathbb{R}^k} (\epsilon g(u,\mathbf{x}) + \underline{g}_u(u,\mathbf{x}))\nu^{(1)}(u,\mathbf{x}) d\left(\prod_{i=1}^k x_i\right) du + \int_{u\in\mathbb{R}} \int_{\mathbf{x}\in\mathbb{R}^k} (\epsilon g(u,\mathbf{x}) - \overline{g}_u(u,\mathbf{x}))\nu^{(2)}(u,\mathbf{x}) d\left(\prod_{i=1}^k x_i\right) du \geq \int_{u\in\mathbb{R}} \lambda(u)du \tag{84}$$

$$\overset{(e)}{\Longrightarrow} \kappa + \int_{\mathbf{x}\in\mathbb{R}^k} \left[\int_{u\in\mathbb{R}} \left(-\left[\min_{a\in\texttt{Set}(\mathbf{x})} \mathfrak{h}(|u-a|)\right]\delta(u) + \lambda(u) + \epsilon\nu^{(2)}(u,\mathbf{x}) + \epsilon\nu^{(1)}(u,\mathbf{x})\right) du\right] g(u,\mathbf{x}) d\left(\prod_{i=1}^k x_i\right)$$

$$+ \int_{\mathbf{x}\in\mathbb{R}^k} \left[\int_{u\in\mathbb{R}} \left(\nu^{(1)}(u,\mathbf{x})(u,\mathbf{x})\underline{g}_u(u,\mathbf{x}) - \nu^{(2)}(u,\mathbf{x})\overline{g}_u(u,\mathbf{x})\right) du\right] d\left(\prod_{i=1}^k x_i\right) \geq \int_{u\in\mathbb{R}} \lambda(u)du \tag{85}$$

(e) follows by exchanging integrals since each term is positive and the constraint $\int_u \delta(u) \leq 1$.

Now, we solve the for the term $\int_{u\in\mathbb{R}} \left[\nu^{(1)}(u,\mathbf{x})\underline{g}_u(u,\mathbf{x})du - \nu^{(2)}(u,\mathbf{x})\overline{g}_u(u,\mathbf{x})\right] du$

$$\int_{u\in\mathbb{R}} \left[\nu^{(1)}(u,\mathbf{x})\underline{g}_u(u,\mathbf{x})du - \nu^{(2)}(u,\mathbf{x})\overline{g}_u(u,\mathbf{x})\right] du \tag{86}$$

$$\overset{(a)}{\leq} \left[\left(\nu^{(1)}(u,\mathbf{x}) - \nu^{(2)}(u,\mathbf{x})\right) g(u,\mathbf{x})\right]_{u=-\infty}^{+\infty} - \left[\int_{u\in\mathbb{R}} (\nu_u^{(1)}(u,\mathbf{x}) - \nu_u^{(2)}(u,\mathbf{x}))g(u,\mathbf{x})du\right] \tag{87}$$

$$\overset{(b)}{\leq} -\left[\int_{u\in\mathbb{R}} \left(\nu_u^{(1)}(u,\mathbf{x}) - \nu_u^{(2)}(u,\mathbf{x})\right) g(u,\mathbf{x})du\right] \tag{88}$$

(a) follows from integration of parts and the inequality from the fact that we take lower derivative and upper derivative of $g(.)$ respectively. (b) follows from the third and fourth constraints in the LP D^{int} on $\nu^{(1)}(u,\mathbf{x}) - \nu^{(2)}(u,\mathbf{x})$ as $u \to \infty$ or $u \to -\infty$.

Combining the inequalities in (97) and (98), we get,

$$\implies \kappa + \int_{\mathbf{x}\in\mathbb{R}^k} \left[\int_{u\in\mathbb{R}} \left(\left[\min_{a\in\texttt{Set}(\mathbf{x})} \mathfrak{h}(|u-a|)\right]\delta(u) + \lambda(u) + \epsilon\nu^{(2)}(u,\mathbf{x}) + \epsilon\nu^{(1)}(u,\mathbf{x})\right) du\right] g(u,\mathbf{x}) d\left(\prod_{i=1}^k x_i\right)$$

$$- \int_{\mathbf{x}\in\mathbb{R}^k} \left[\int_{u\in\mathbb{R}} \left(\nu_u^{(1)}(u,\mathbf{x}) - \nu_u^{(2)}(u,\mathbf{x})\right) g(u,\mathbf{x})du\right] d\left(\prod_{i=1}^k x_i\right) \geq \int_{u\in\mathbb{R}} \lambda(u)du \tag{89}$$

$$\implies \kappa + \int_{\mathbf{x}\in\mathbb{R}^k} \left[\int_{u\in\mathbb{R}} \left(\left(\min_{a\in\texttt{Set}(\mathbf{x})} \mathfrak{h}(|u-a|)\right)\delta(u) + \lambda(u) + \epsilon\nu^{(2)}(u,\mathbf{x}) + \epsilon\nu^{(1)}(u,\mathbf{x})\right.\right. \tag{90}$$

$$\left.\left. - \left(\nu_u^{(1)}(u,\mathbf{x}) - \nu_u^{(2)}(u,\mathbf{x})\right)\right) du\right] g(u,\mathbf{x}) d\left(\prod_{i=1}^k x_i\right) \geq \int_{u\in\mathbb{R}} \lambda(u)du \tag{91}$$

$$\overset{(c)}{\implies} \kappa \geq \int_{u\in\mathbb{R}} \lambda(u)du \tag{92}$$

(c) follows from the first and second constraint in LP $\mathcal{E}^{int}$. Since this inequality is true for every feasible solution in the primal $\mathcal{O}$ and dual $\mathcal{E}^{int}$, we have the proof in the theorem.

□

B.6.2 A formal proof of Theorem 2.8

Proof. Now consider any feasible solution $g(.,.)$ and κ in the primal DILP $\mathcal{O}$ and feasible solution $\nu^{(1)}(.,.), \nu^{(2)}(.,.), \lambda(.)$ and $\delta(.)$ in the dual DILP $\mathcal{E}^{int}$.

Now pre-multiply the first constraint in DILP $\mathcal{O}$ by $\delta(u)$ second constraint in DILP by $\lambda(u)$, the third constraint in DILP by $\nu^{(1)}(u,\mathbf{x})$ and fourth constraint in DILP $\mathcal{O}$ by $\nu^{(2)}(u,\mathbf{x})$ and take the lower Riemann-Darboux integrals for the last two terms and the integral for the first two terms. [28]. This approach is very similar to the use of Lagrange multiplier in weak duality proof in linear programming. Also observe that some limits may take values in the extended reals i.e. $\mathbb{R}\cup\{-\infty,\infty\}$ and thus, we define the limits in extended reals.

.

Observe that $d^l, d^u \in \mathbb{R}$,$\mathbf{c}^l, \mathbf{c}^u \in \mathbb{R}^k$ and c_i^l, c_i^u refers to its i^{th} component of $\mathbf{c}^l$ and $\mathbf{c}^u \in \mathbb{R}^k$ respectively.

[28]The integral may not be defined as $\underline{g}_u$ and $\overline{g}_u$ may not be Riemann integrable, however the first and second terms are integrable which follows since the product of integrable functions is integrable.

Thus, we get[29]

$$\lim_{\substack{d^l\to-\infty\\ d^u\to\infty}} \int_{u=d^l}^{d^u} \delta(u)\left[\kappa - \int_{\mathbf{x}\in\mathbb{R}^k}\left(\min_{a\in\texttt{Set}(\mathbf{x})}\mathfrak{h}(|u-a|)\right) g(u,\mathbf{x})d\left(\prod_{i=1}^k x_i\right)\right]du \tag{93}$$
$$+\lim_{\substack{d^l\to-\infty\\ d^u\to\infty}}\int_{u=d^l}^{d^u}\left(\int_{\mathbf{x}\in\mathbb{R}^k} g(u,\mathbf{x})d\left(\prod_{i=1}^k x_i\right)\right)\lambda(u)du$$

$$+\liminf_{d,\mathbf{c}\to\infty}\underline{\int}_{\substack{u\in[d^l,d^u]\\ \mathbf{x}\in\prod_{i=1}^k[c_i^l,c_i^u]}}\left((\epsilon g(u,\mathbf{x})+\underline{g}_u(u,\mathbf{x}))\nu^{(1)}(u,\mathbf{x})+(\epsilon g(u,\mathbf{x})-\overline{g}_u(u,\mathbf{x}))\nu^{(2)}(u,\mathbf{x})\right)d\left(\prod_{i=1}^k x_i\right)du \geq \int_{u\in\mathbb{R}}\lambda(u)du \tag{94}$$

$$\implies \lim_{\substack{d^l\to-\infty\\ d^u\to\infty}}\int_{u=d^l}^{d^u}\delta(u)\lim_{\substack{\mathbf{c}^u\to\infty\\ \mathbf{c}^l\to-\infty}}\left[\kappa-\int_{\mathbf{x}\in\prod_{i=1}^k[c_i^l,c_i^u]}\left(\min_{a\in\texttt{Set}(\mathbf{x})}\mathfrak{h}(|u-a|)\right)g(u,\mathbf{x})d\left(\prod_{i=1}^k x_i\right)\right]du$$
$$+\lim_{\substack{d^l\to-\infty\\ d^u\to\infty}}\int_{u=d^l}^{d^u}\lim_{\substack{\mathbf{c}^u\to\infty\\ \mathbf{c}^l\to-\infty}}\left[\int_{\mathbf{x}\in\prod_{i=1}^k[c_i^l,c_i^u]} g(u,\mathbf{x})\lambda(u)d\left(\prod_{i=1}^k x_i\right)\right]du$$

$$+\liminf_{d,\mathbf{c}\to\infty}\underline{\int}_{\substack{u\in[d^l,d^u]\\ \mathbf{x}\in\prod_{i=1}^k[c_i^l,c_i^u]}}\left((\epsilon g(u,\mathbf{x})+\underline{g}_u(u,\mathbf{x}))\nu^{(1)}(u,\mathbf{x})+(\epsilon g(u,\mathbf{x})-\overline{g}_u(u,\mathbf{x}))\nu^{(2)}(u,\mathbf{x})\right)d\left(\prod_{i=1}^k x_i\right)du \geq \int_{u\in\mathbb{R}}\lambda(u)du \tag{95}$$

$$\overset{(c)}{\implies}\lim_{d\to\infty}\lim_{\mathbf{c}\to\infty}\int_{u=d^l}^{d^u}\left(\kappa\delta(u)+\int_{\mathbf{x}\in\prod_{i=1}^k[c_i^l,c_i^u]}\left(\left(-\delta(u)\min_{a\in\texttt{Set}(\mathbf{x})}\mathfrak{h}(|u-a|)+\lambda(u)\right)g(u,\mathbf{x})\right)d\left(\prod_{i=1}^k x_i\right)\right)du$$
$$+\liminf_{d,\mathbf{c}\to\infty}\left(\int_{\substack{u\in[d^l,d^u]\\ \mathbf{x}\in\prod_{i=1}^k[c_i^l,c_i^u]}}\left(\epsilon g(u,\mathbf{x})\left(\nu^{(1)}(u,\mathbf{x})+\nu^{(2)}(u,\mathbf{x})\right)\right)d\left(\prod_{i=1}^k x_i\right)du\right.$$
$$\left.+\underline{\int}_{\substack{u\in[d^l,d^u]\\ \mathbf{x}\in\prod_{i=1}^k[c_i^l,c_i^u]}}\left(\underline{g}_u(u,\mathbf{x})\nu^{(1)}(u,\mathbf{x})-\overline{g}_u(u,\mathbf{x})\nu^{(2)}(u,\mathbf{x})\right)d\left(\prod_{i=1}^k x_i\right)du\right)\geq\int_{u\in\mathbb{R}}\lambda(u)du \tag{96}$$

[29] $\underline{\int} f(x)dx$ and $\overline{\int} f(x)dx$ denotes the lower and upper integral respectively.

$$\overset{(d)}{\implies} \kappa + \limsup_{d,\mathbf{c}\to\infty}\left(\int_{\mathbf{x}\in\prod_{i=1}^{k}[c_i^l,c_i^u]}\left[\int_{u=d^l}^{d^u}\left(-\left[\min_{a\in\mathtt{Set}(\mathbf{x})}\mathfrak{h}(|u-a|)\right]\delta(u)+\lambda(u)\right)du\right]g(u,\mathbf{x})d\left(\prod_{i=1}^{k}x_i\right)\right)$$
$$+\liminf_{d,\mathbf{c}\to\infty}\left(\int_{\substack{u\in[d^l,d^u]\\ \mathbf{x}\in\prod_{i=1}^{k}[c_i^l,c_i^u]}}\left(\epsilon g(u,\mathbf{x})\left(\nu^{(1)}(u,\mathbf{x})+\nu^{(2)}(u,\mathbf{x})\right)\right)d\left(\prod_{i=1}^{k}x_i\right)du\right.$$
$$\left.+\underline{\int}_{\mathbf{x}\in\prod_{i=1}^{k}[c_i^l,c_i^u]}\left[\underline{\int}_{u=d^l}^{d^u}\left(\nu^{(1)}(u,\mathbf{x})\underline{g}_u(u,\mathbf{x})-\nu^{(2)}(u,\mathbf{x})\overline{g}_u(u,\mathbf{x})\right)du\right]d\left(\prod_{i=1}^{k}x_i\right)\right)\geq\int_{u\in\mathbb{R}}\lambda(u)du \tag{97}$$

(c) [30] follows from the following observations.

- For the first two integrals, apply monotone convergence theorem (MCT) which allows us to exchange limit and integral. This is possible since the expression in [.] is positive and non-decreasing as $\mathbf{c}^u\to\infty$ and $\mathbf{c}^l\to-\infty$ due to first constraint of DILP $\mathcal{O}$ and the fact that the $g(.,.)$ is non-negative.

(d) follows from the following steps.

- Apply the *interated integral* inequality [54, Proposition 3.9] that $\underline{\int}_{A\times B}g(A,B)dAdB\leq\underline{\int}_{A}\left[\underline{\int}_{B}g(A,B)dB\right]dA$ for the last integral.
- We exchange integrals for the first term using *Fubini's theorem* [54, Theorem 3.10] as the functions are *Riemann integrable* and use the fact that $\int_{u\in\mathbb{R}}\delta(u)\leq 1$.
- We upper bound the iterated limits by a limsup as $\limsup_{m,n}a_{m,n}\geq\lim_m\lim_n a_{m,n}$.

Now, we bound the term $\underline{\int}_{u\in[d^l,d^u]}\left[\nu^{(1)}(u,\mathbf{x})\underline{g}_u(u,\mathbf{x})du-\nu^{(2)}(u,\mathbf{x})\overline{g}_u(u,\mathbf{x})\right]du$.[31]. Observe that,

$$\int_{\underline{u}\in[d^l,d^u]}\left[\nu^{(1)}(u,\mathbf{x})\underline{g}_u(u,\mathbf{x})du-\nu^{(2)}(u,\mathbf{x})\overline{g}_u(u,\mathbf{x})\right]du+\int_{u\in[d^l,d^u]}\left(\nu_u^{(1)}(u,\mathbf{x})-\nu_u^{(2)}(u,\mathbf{x})\right)g(u,\mathbf{x})du$$
$$\overset{(f)}{\leq}\left[\left(\nu^{(1)}(u,\mathbf{x})-\nu^{(2)}(u,\mathbf{x})\right)g(u,\mathbf{x})\right]_{u=d^l}^{d^u} \tag{98}$$

(f) follows from the chain rule of differentiation and the inequality from the fact that we take lower derivative of $-g(.)$ is upper derivative of $g(.)$

Now consider (observe we take the -ve of the last term in Equation (98)).

$$\limsup_{d,\mathbf{c}\to\infty}\left[\int_{\mathbf{x}\in\prod_{i=1}^{k}[c_i^l,c_i^u]}-\left[\left(\nu^{(1)}(u,\mathbf{x})-\nu^{(2)}(u,\mathbf{x})\right)g(u,\mathbf{x})\right]_{u=d^l}^{d^u}d\left(\prod_{i=1}^{k}x_i\right)\right]$$
$$\overset{(i)}{\geq}\limsup_{\mathbf{c}\to\infty}\liminf_{d\to\infty}\left[\int_{\mathbf{x}\in\prod_{i=1}^{k}[c_i^l,c_i^u]}\left(-\left[\left(\nu^{(1)}(u,\mathbf{x})-\nu^{(2)}(u,\mathbf{x})\right)g(u,\mathbf{x})\right]_{u=d^l}^{d^u}\right)d\left(\prod_{i=1}^{k}x_i\right)\right]$$

[30] $d,\to\infty$ is a shorthand notation for $d^u\to\infty$, $d^l\to-\infty$ and $\mathbf{c}\to\infty$ is a shorthand for $\mathbf{c}^u\to\infty$ and $\mathbf{c}^l\to-\infty$

[31] Note that the second integral is defined as the product of *Riemann integrable* functions is also *integrable*

$$\overset{(g)}{\geq} \limsup_{\mathbf{c}\to\infty} \int_{\mathbf{x}\in\prod_{i=1}^{k}[c_i^l,c_i^u]} \liminf_{\substack{d^u\to\infty\\ d^l\to-\infty}} \left[-\left(\nu^{(1)}(u,\mathbf{x})-\nu^{(2)}(u,\mathbf{x})\right) g(u,\mathbf{x})\right]_{u=d^l}^{d^u} d\left(\prod_{i=1}^{k} x_i\right)$$

$$\overset{(h)}{\geq} \limsup_{\mathbf{c}\to\infty} \int_{\mathbf{x}\in\prod_{i=1}^{k}[c_i^l,c_i^u]} \Big(\liminf_{d^u\to\infty}\left[\left(-\nu^{(1)}(d^u,\mathbf{x})+\nu^{(2)}(d^u,\mathbf{x})\right) g(d^u,\mathbf{x})\right] \tag{99}$$

$$+\liminf_{d^l\to-\infty}\left[\left(\nu^{(1)}(d^l,\mathbf{x})-\nu^{(2)}(d^l,\mathbf{x})\right) g(d^l,\mathbf{x})\right]\Big) d\left(\prod_{i=1}^{k} x_i\right)$$

$$\overset{(m)}{\geq} 0 \tag{100}$$

(i) follows from the fact that $\limsup_{m,n} f(m,n) \geq \limsup_m \limsup_n f(m,n) \geq \limsup_m \liminf_n f(m,n)$. (g) follows from the following arguments.

- Observe that $\left[-\left(\nu^{(1)}(u,\mathbf{x})-\nu^{(2)}(u,\mathbf{x})\right) g(u,\mathbf{x})\right]_{u=d^l}^{d^u}$ is non-negative for every $\mathbf{x}\in\prod_{i=1}^{k}[c_i^l,c_i^u]$ whenever, $d_u > \sup\{U(\mathbf{x})|\mathbf{x}\in\prod_{i=1}^{k}[c_i^l,c_i^u]\}$ and $d_l < \inf\{L(\mathbf{x})|\mathbf{x}\in\prod_{i=1}^{k}[c_i^l,c_i^u]\}$ from third and fourth constraint of DILP $\mathcal{E}^{int}$. Note that value is finite as supremeum and infimum of *continuous functions* is finite over a *compact set* since $\prod_{i=1}^{k}[c_i^l,c_i^u]$ is closed and bounded.
- Now apply *Fatou's lemma* to conclude (g).

(h) follows from the fact that $\liminf_n(A_n+B_n) \geq \liminf_n A_n + \liminf_n B_n$ and the last inequality (m) follows from the following 2 statements.

- $\liminf_{d^u\to\infty}\left(-\nu^{(1)}(d^u,\mathbf{x})+\nu^{(2)}(d^u,\mathbf{x})\right) \geq 0$ and $\liminf_{d^l\to-\infty}\left(\nu^{(1)}(d^l,\mathbf{x})-\nu^{(2)}(d^l,\mathbf{x})\right) \geq 0$ follows from third and fourth constraint of DILP $\mathcal{E}^{int}$.
- $g(.,.)$ is a bounded function follows from the constraint in DILP $\mathcal{O}$.

Observe that (99) implies (follows from $\limsup_n A_n = -\liminf_n -A_n$) that

$$\liminf_{d,\mathbf{c}\to\infty}\left[\int_{\mathbf{x}\in\prod_{i=1}^{k}[c_i^l,c_i^u]}\left[\left(\nu^{(1)}(u,\mathbf{x})-\nu^{(2)}(u,\mathbf{x})\right) g(u,\mathbf{x})\right]_{u=d^l}^{d^u} d\left(\prod_{i=1}^{k} x_i\right)\right] \leq 0 \tag{101}$$

Using inequality in (98) and the fact that $\liminf_n(a_n+b_n) \leq \liminf a_n + \limsup_n b_n$, we obtain

$$\liminf_{d,\mathbf{c}\to\infty}\Bigg(\underline{\int}_{\mathbf{x}\in\prod_{i=1}^{k}[c_i^l,c_i^u]}\left[\underline{\int}_{u=d^l}^{d^u}\left(\nu^{(1)}(u,\mathbf{x})\underline{g}_u(u,\mathbf{x})-\nu^{(2)}(u,\mathbf{x})\overline{g}_u(u,\mathbf{x})\right)du\right] d\left(\prod_{i=1}^{k} x_i\right)$$

$$+\int_{\substack{u\in[d^l,d^u]\\ \mathbf{x}\in\prod_{i=1}^{k}[c_i^l,c_i^u]}}\left(\epsilon g(u,\mathbf{x})\left(\nu^{(1)}(u,\mathbf{x})+\nu^{(2)}(u,\mathbf{x})\right)\right) d\left(\prod_{i=1}^{k} x_i\right)\Bigg)du \tag{102}$$

$$\leq \liminf_{d,\mathbf{c}\to\infty}\left[\int_{\mathbf{x}\in\prod_{i=1}^{k}[c_i^l,c_i^u]}\left[\left(\nu^{(1)}(u,\mathbf{x})-\nu^{(2)}(u,\mathbf{x})\right) g(u,\mathbf{x})\right]_{u=d^l}^{d^u} d\left(\prod_{i=1}^{k} x_i\right)\right]$$

$$+\limsup_{d,\mathbf{c}\to\infty}\Bigg(\int_{\mathbf{x}\in\prod_{i=1}^{k}[c_i^l,c_i^u]}\left[-\int_{u\in[d^l,d^u]}\left(\nu_u^{(1)}(u,\mathbf{x})-\nu_u^{(2)}(u,\mathbf{x})\right)g(u,\mathbf{x})du\right]d\left(\prod_{i=1}^{k}x_i\right)$$

$$+\int_{\substack{u\in[d^l,d^u]\\ \mathbf{x}\in\prod_{i=1}^{k}[c_i^l,c_i^u]}}\left(\epsilon g(u,\mathbf{x})\left(\nu^{(1)}(u,\mathbf{x})+\nu^{(2)}(u,\mathbf{x})\right)\right)d\left(\prod_{i=1}^{k}x_i\right)du\Bigg) \quad (103)$$

$$\overset{(k)}{\leq}\limsup_{d,\mathbf{c}\to\infty}\Bigg(\int_{\mathbf{x}\in\prod_{i=1}^{k}[c_i^l,c_i^u]}\left[-\int_{u\in[d^l,d^u]}\left(\nu_u^{(1)}(u,\mathbf{x})-\nu_u^{(2)}(u,\mathbf{x})\right)g(u,\mathbf{x})du\right]d\left(\prod_{i=1}^{k}x_i\right)$$

$$+\int_{\substack{u\in[d^l,d^u]\\ \mathbf{x}\in\prod_{i=1}^{k}[c_i^l,c_i^u]}}\left(\epsilon g(u,\mathbf{x})\left(\nu^{(1)}(u,\mathbf{x})+\nu^{(2)}(u,\mathbf{x})\right)\right)d\left(\prod_{i=1}^{k}x_i\right)du\Bigg) \quad (104)$$

(k) follows on applying Equation (101)

Now combining (104) and (97), we obtain

$$\kappa+\limsup_{d,\mathbf{c}\to\infty}\Bigg(\int_{\mathbf{x}\in\prod_{i=1}^{k}[c_i^l,c_i^u]}\left[\int_{u\in[d^l,d^u]}\left(-\left[\min_{a\in\texttt{Set}(\mathbf{x})}\mathfrak{h}(|u-a|)\right]\delta(u)+\lambda(u)\right)du\right]g(u,\mathbf{x})d\left(\prod_{i=1}^{k}x_i\right)\Bigg)$$

$$+\limsup_{d,\mathbf{c}\to\infty}\Bigg(\int_{\mathbf{x}\in\prod_{i=1}^{k}[c_i^l,c_i^u]}\left[-\int_{u\in[d^l,d^u]}\left(\nu_u^{(1)}(u,\mathbf{x})-\nu_u^{(2)}(u,\mathbf{x})\right)g(u,\mathbf{x})du\right]d\left(\prod_{i=1}^{k}x_i\right)$$

$$+\int_{\substack{u\in[d^l,d^u]\\ \mathbf{x}\in\prod_{i=1}^{k}[c_i^l,c_i^u]}}\left(\epsilon g(u,\mathbf{x})\left(\nu^{(1)}(u,\mathbf{x})+\nu^{(2)}(u,\mathbf{x})\right)\right)d\left(\prod_{i=1}^{k}x_i\right)du\Bigg)\geq\int_{u\in\mathbb{R}}\lambda(u)du \quad (105)$$

$$\overset{(i)}{\implies}\kappa+\lim_{d,\mathbf{c}\to\infty}\Bigg(\int_{\mathbf{x}\in\prod_{i=1}^{k}[c_i^l,c_i^u]}\left[\int_{u\in[d^l,d^u]}\left(-\left[\min_{a\in\texttt{Set}(\mathbf{x})}\mathfrak{h}(|u-a|)\right]\delta(u)+\lambda(u)\right)du\right]g(u,\mathbf{x})d\left(\prod_{i=1}^{k}x_i\right)\Bigg)$$

$$+\limsup_{d,\mathbf{c}\to\infty}\Bigg(\int_{\mathbf{x}\in\prod_{i=1}^{k}[c_i^l,c_i^u]}\left[-\int_{u\in[d^l,d^u]}\left(\nu_u^{(1)}(u,\mathbf{x})-\nu_u^{(2)}(u,\mathbf{x})\right)g(u,\mathbf{x})du\right]d\left(\prod_{i=1}^{k}x_i\right)\Bigg)$$

$$+\int_{\mathbf{x}\in\prod_{i=1}^{k}[c_i^l,c_i^u]}\int_{u\in[d^l,d^u]}\left(\epsilon g(u,\mathbf{x})\left(\nu^{(1)}(u,\mathbf{x})+\nu^{(2)}(u,\mathbf{x})\right)\right)d\left(\prod_{i=1}^{k}x_i\right)du\Bigg)\geq\int_{u\in\mathbb{R}}\lambda(u)du \quad (106)$$

$$\implies\kappa+\limsup_{d,\mathbf{c}\to\infty}\Bigg(\int_{\mathbf{x}\in\prod_{i=1}^{k}[c_i^l,c_i^u]}\left[\int_{u\in[d^l,d^u]}\left(-\left(\min_{a\in\texttt{Set}(\mathbf{x})}\mathfrak{h}(|u-a|)\right)\delta(u)+\lambda(u)+\epsilon\nu^{(2)}(u,\mathbf{x})+\epsilon\nu^{(1)}(u,\mathbf{x})\right)\right. \quad (107)$$

$$\left.-\left(\nu_u^{(1)}(u,\mathbf{x})-\nu_u^{(2)}(u,\mathbf{x})\right)\right)du\Bigg]g(u,\mathbf{x})d\left(\prod_{i=1}^{k}x_i\right)\Bigg)\geq\int_{u\in\mathbb{R}}\lambda(u)du \quad (108)$$

$$\stackrel{(j)}{\implies} \kappa \geq \int_{u\in\mathbb{R}} \lambda(u)du \tag{109}$$

The limsup in first term of (i) can be replaced by a limit as integral (110) is defined in extended reals ($\mathbb{R}\cup\{\infty,-\infty\}$) which follows from the following reasons.[32]

$$\left(\int_{\mathbf{x}\in\mathbb{R}^k}\left[\int_{u\in\mathbb{R}}\left(-\left[\min_{a\in\texttt{Set}(\mathbf{x})}\mathfrak{h}(|u-a|)\right]\delta(u)+\lambda(u)\right)du\right]g(u,\mathbf{x})d\left(\prod_{i=1}^{k}x_i\right)\right) \tag{110}$$

$$=-\int_{\mathbf{x}\in\mathbb{R}^k}\int_{u\in\mathbb{R}}\left[\min_{a\in\texttt{Set}(\mathbf{x})}\mathfrak{h}(|u-a|)\right]\delta(u)g(u,\mathbf{x})du\ d\left(\prod_{i=1}^{k}x_i\right) \tag{111}$$

$$+\int_{\mathbf{x}\in\mathbb{R}^k}\int_{u\in\mathbb{R}}\lambda(u)g(u,\mathbf{x})du\ d\left(\prod_{i=1}^{k}x_i\right) \tag{112}$$

- Observe that $\int_u\delta(u)\leq 1$ and $\int_{\mathbf{x}\in\mathbb{R}^k}\min_{a\in\texttt{Set}(\mathbf{x})}\mathfrak{h}(|u-a|)g(u,\mathbf{x})d\left(\prod_{i=1}^{k}x_i\right)\leq\kappa\ \forall u\in\mathbb{R}$ (first constraint of DILP $\mathcal{O}$) implying that term (111) is lower bounded by $-\kappa$.
- Thus, we may conclude that the following integral (110) is defined in extended reals as the first term (111) is lower bounded by $-\kappa$ and second term (112) is lower bounded by 0.

(j) follows from the first and second constraint in DILP $\mathcal{E}^{int}$. Since this inequality is true for every feasible solution in the primal $\mathcal{O}$ and dual $\mathcal{E}^{int}$, we have the proof in the theorem. □

We now prove Theorem 2.8.

Proof. Combining Claim B.7 and B.8, we obtain that $\text{opt}(\mathcal{O})\geq\text{opt}(\mathcal{E})$. □

B.7 Claim to prove the existence of continuous bounds $U(\mathbf{v})$ and $L(\mathbf{v})$ for the feasible solution in $\mathcal{E}$ in the proof of Lemma 2.12

Claim 1. Consider a function $\nu(.,.):\mathcal{C}^1(\mathbb{R}\times\mathbb{R}^k\to\mathbb{R})$ s.t. zeros of $\nu(.,\mathbf{v})$ i.e. $\{\mathfrak{u}:\nu(\mathfrak{u},\mathbf{v})=0\}$ is a *countable* set for every $\mathbf{v}\in\mathbb{R}^k$.

It also satisfies the following two conditions for some constant C independent of $\mathbf{v}$. Note that $v_{\lfloor i\rfloor}$ denotes the i^{th} largest component of $\mathbf{v}$ for every $i\in[k]$.

- $\left\{\mathfrak{u}\in\mathbb{R}:\nu(\mathfrak{u},\mathbf{v})=0 \text{ and } \nu_u(\mathfrak{u},\mathbf{v})<0\right\}$ is upper bounded by $C+v_{\lfloor k\rfloor}$ for every $\mathbf{v}\in\mathbb{R}^k$.
- $\forall\mathbf{v}\in\mathbb{R}^k\ \exists\ \mathcal{U}$ s.t $\nu(u,\mathbf{v})\geq 0\ \forall u>\mathcal{U}$.

Then there exists $U:\mathcal{C}^0(\mathbb{R}^k\to\mathbb{R})$ s.t. $\nu(u,\mathbf{v})\geq 0\ \forall u\geq U(\mathbf{v})\ \forall\mathbf{v}\in\mathbb{R}^k$.

Proof. We prove this statement as follows. For every $\mathbf{v}\in\mathbb{R}^k$, denote the largest zero of $\nu(.,\mathbf{v})$ by $p(\mathbf{v})$. Now construct $U(\mathbf{v})=\max(p(\mathbf{v}),C+v_{\lfloor k\rfloor})$. Observe that $\nu(u,\mathbf{v})\geq 0\ \forall u\geq U(\mathbf{v})\ \forall\mathbf{v}\in\mathbb{R}^k$ follows from the second assumption.

To prove continuity at $\mathbf{v}\in\mathbb{R}^k$, we aim to show the following statement.

[32] Intuitively it shows that $\infty-\infty$ scenario cannot arrive in this integral

$$\forall \epsilon > 0 \ \exists \delta > 0 \text{ s.t. } \forall \mathbf{z} \in \mathbb{R}^k;\ ||\mathbf{z}-\mathbf{v}||_1 \leq \delta \implies |U(\mathbf{z}) - U(\mathbf{v})| \leq \epsilon. \tag{113}$$

We now show $U(\mathbf{v})$ is continuous for every $\mathbf{v} \in \mathbb{R}^k$. We consider two exhaustive cases below.

Case 1: $p(\mathbf{v}) \geq C + v_{\lfloor k \rfloor}$ and thus, $U(\mathbf{v}) = p(\mathbf{v})$. Recall from the second assumption that $\nu(u, \mathbf{v})$ goes from negative to positive at $u = p(\mathbf{v})$.

We now discuss the construction of δ below for a *sufficiently small* ϵ as follows.

For some ϵ sufficiently small, we now discuss the construction of an interval around $p_{\mathbf{v}}$ as $[\mathcal{A}_{\mathbf{v}}, \mathcal{B}_{\mathbf{v}}] = [p_{\mathbf{v}} - \frac{\epsilon}{2}, p_{\mathbf{v}} + \frac{\epsilon}{2}]$. Since, ϵ is sufficiently small, observe that $\nu(\mathcal{A}_{\mathbf{v}}, \mathbf{v}) = -\zeta_1$ and $\nu(\mathcal{B}_{\mathbf{v}}, \mathbf{v}) = \zeta_2$ for some $\zeta_1, \zeta_2 > 0$.

Observe that $\nu(\mathcal{A}_{\mathbf{v}}, \mathbf{v})$ must be continuous in $\mathbf{v}$ and thus there must exist $\delta_1 > 0$ s.t. $\forall \mathbf{z} \in \mathbb{R}^k;\ ||\mathbf{z}-\mathbf{v}|| \leq \delta_1 \implies |\nu(\mathcal{A}_{\mathbf{v}}, \mathbf{v}) - \nu(\mathcal{A}_{\mathbf{v}}, \mathbf{z})| \leq \frac{\zeta_1}{2}$ and similarly, choose $\delta_2 > 0$ s.t. $\forall \mathbf{z} \in \mathbb{R}^k;\ ||\mathbf{z}-\mathbf{v}_1|| \leq \delta_2 \implies |\nu(\mathcal{B}_{\mathbf{v}}, \mathbf{v}) - \nu(\mathcal{B}_{\mathbf{v}}, \mathbf{z})| \leq \frac{\zeta_2}{2}$.[33]

We choose $\delta = \min(\delta_1, \delta_2, \frac{\epsilon}{2})$ and consider any $\mathbf{z} \in \mathbb{R}^k$ satisfying $||\mathbf{z}-\mathbf{v}||_1 < \delta$ and show (113)

This implies $\nu(\mathcal{A}_{\mathbf{v}}, \mathbf{z}) < -\frac{\zeta_1}{2} < 0$ and $\nu(\mathcal{B}_{\mathbf{v}}, \mathbf{z}) > \frac{\zeta_2}{2} > 0$ since, $\nu(\mathcal{A}_{\mathbf{v}}, \mathbf{v}) = -\zeta_1$ and $\nu(\mathcal{B}_{\mathbf{v}}, \mathbf{v}) = \zeta_2$. Thus, from *intermediate value and LMVT theorem*, there must exist some zero of $\nu(., \mathbf{z})$ (call it u_0) with $u_0 \in [\mathcal{A}_{\mathbf{v}}, \mathcal{B}_{\mathbf{v}}]$ with $\nu(.)$ going from negative to positive i.e. $\nu_u(u_0, \mathbf{z}) \geq 0$ and $\nu(u_0, \mathbf{z}) = 0$

Also observe that,

$$\begin{aligned} \mathcal{B}_{\mathbf{v}} &\overset{(a)}{\geq} C + v_{\lfloor k \rfloor} + \frac{\epsilon}{2} \\ &\overset{(b)}{>} C + z_{\lfloor k \rfloor} + \frac{\epsilon}{2} - \delta \geq C + z_{\lfloor k \rfloor} \end{aligned} \tag{114}$$

(a) follows from $p(\mathbf{v}) \geq C + v_{\lfloor k \rfloor}$ and the definition of $\mathcal{B}_{\mathbf{v}}$. (b) follows from the fact $||\mathbf{v}-\mathbf{z}|| \leq \delta$. Observe that the last inequality follows from the fact that $\delta < \frac{\epsilon}{2}$.

Equation (114) and the first condition in Claim 1implies that $U(u, \mathbf{z})$ cannot have a zero in u with $U(u, \mathbf{z})$ going from positive to negative on the right side of interval $[\mathcal{A}_{\mathbf{v}}, \mathcal{B}_{\mathbf{v}}]$. This implies that the largest zero of $\nu(u, \mathbf{z})$ in interval $[\mathcal{A}_{\mathbf{v}}, \mathcal{B}_{\mathbf{v}}]$ must be the largest zero and thus, it must be $p(\mathbf{z})$. Also, $p(\mathbf{z}) > \mathcal{A}_{\mathbf{v}} > C + z_{\lfloor k \rfloor}$ and thus, $U(\mathbf{z}) = p(\mathbf{z})$

We thus, have $|U(\mathbf{z}) - U(\mathbf{v})| = |p(\mathbf{z}) - p(\mathbf{v})| \leq \epsilon$, since $\max(|\mathcal{A}_{\mathbf{v}} - p(\mathbf{v})|, |\mathcal{B}_{\mathbf{v}} - p(\mathbf{v})|) < \epsilon$ thus proving the desired result in (113)

Case 2: $p(\mathbf{v}) < C + v_{\lfloor k \rfloor}$, thus $U(\mathbf{v}) = C + v_{\lfloor k \rfloor}$.

Since, $p(\mathbf{v})$ was the largest zero of $\nu(., \mathbf{v})$ with $\nu_u(p(\mathbf{v}), \mathbf{v}) > 0$, observe that $\nu(C + v_{\lfloor k \rfloor}, \mathbf{v}) = \zeta > 0$. Now since $\nu(u, \mathbf{v})$ is continuous, there must exist $\delta_1 > 0$ s.t. $\forall \hat{u} \in \mathbb{R}$ satisfying $|\hat{u} - (C + v_{\lfloor k \rfloor})| < \delta_1$, we have $|\nu(\hat{u}, \mathbf{v}) - \nu(\hat{u}, \mathbf{v})| < \frac{\zeta}{4}$. Similarly, there must exist $\delta_2 > 0$ s.t. $\forall \mathbf{z} \in \mathbb{R}$ satisfying $||\mathbf{v}-\mathbf{z}||_1 < \delta_2$, we have $|\nu(C + z_{\lfloor k \rfloor}, \mathbf{v}) - \nu(C + z_{\lfloor k \rfloor}, \mathbf{z})| \leq \frac{\zeta}{4}$. [34]

Now we choose $\delta = \min(\delta_1, \delta_2, \epsilon)$ and consider any $\mathbf{z} \in \mathbb{R}^k$ satisfying $||\mathbf{z}-\mathbf{v}||_1 \leq \delta$ to show (113)**.**

Observe that, since, $|(C + v_{\lfloor k \rfloor}) - (C + z_{\lfloor k \rfloor})| \leq \delta \leq \delta_1$, we must have $\nu(C + z_{\lfloor k \rfloor}, \mathbf{v}) = \nu(C + v_{\lfloor k \rfloor}, \mathbf{v}) + (\nu(C + z_{\lfloor k \rfloor}, \mathbf{v}) - \nu(C + v_{\lfloor k \rfloor}, \mathbf{v})) \geq \zeta - \frac{\zeta}{4} = \frac{3\zeta}{4}$. Similarly, since $||\mathbf{v}-\mathbf{z}|| \leq \delta \leq \delta_2$, we must have $\nu(C + z_{\lfloor k \rfloor}, \mathbf{z}) = \nu(C + z_{\lfloor k \rfloor}, \mathbf{v}) + (\nu(C + z_{\lfloor k \rfloor}, \mathbf{v}) - \nu(C + z_{\lfloor k \rfloor}, \mathbf{v})) \geq \frac{3\zeta}{4} - \frac{\zeta}{4} = \frac{\zeta}{2} > 0$.

Since, $\nu(C + z_{\lfloor k \rfloor}, \mathbf{z}) > 0$ and since the second condition in lemma 1 says that there can be no zero of $\nu(u, \mathbf{z})$ in u with $\nu(u, \mathbf{z})$ going from positive to negative beyond $C + z_{\lfloor k \rfloor}$ we have the largest zero of $\nu(., \mathbf{v})$ should be smaller than $C + z_{\lfloor k \rfloor}$ and thus $U(\mathbf{z}) = C + z_{\lfloor k \rfloor}$.

[33] Note that we can do this since, the continuity result holds true for every $\epsilon > 0$ and we can choose any ϵ we want.

[34] This holds since continuity result holds for every $\epsilon > 0$.

Now observe that $|U(\mathbf{z}) - U(\mathbf{y})| = |(C + z_{\lfloor k \rfloor}) - (C + y_{\lfloor k \rfloor})| \leq \delta \leq \epsilon$, thus proving the desired statement in (113).

□