Differential Privacy via Distributionally Robust Optimization

Aras Selvi*[1], Huikang Liu[2] and Wolfram Wiesemann[1]

[1]Imperial College Business School, Imperial College London, United Kingdom
[2]Research Institute for Interdisciplinary Sciences, School of Information Management and Engineering, Shanghai University of Finance and Economics, China

April 25, 2023

Abstract

In recent years, differential privacy has emerged as the *de facto* standard for sharing statistics of datasets while limiting the disclosure of private information about the involved individuals. This is achieved by randomly perturbing the statistics to be published, which in turn leads to a privacy-accuracy trade-off: larger perturbations provide stronger privacy guarantees, but they result in less accurate statistics that offer lower utility to the recipients. Of particular interest are therefore optimal mechanisms that provide the highest accuracy for a pre-selected level of privacy. To date, work in this area has focused on specifying families of perturbations *a priori* and subsequently proving their asymptotic and/or best-in-class optimality.

In this paper, we develop a class of mechanisms that enjoy non-asymptotic and unconditional optimality guarantees. To this end, we formulate the mechanism design problem as an infinite-dimensional distributionally robust optimization problem. We show that the problem affords a strong dual, and we exploit this duality to develop converging hierarchies of finite-dimensional upper and lower bounding problems. Our upper (primal) bounds correspond to implementable perturbations whose suboptimality can be bounded by our lower (dual) bounds. Both bounding problems can be solved within seconds via cutting plane techniques that exploit the inherent problem structure. Our numerical experiments demonstrate that our perturbations can outperform the previously best results from the literature on artificial as well as standard benchmark problems.

Keywords: Differential Privacy; Privacy-Accuracy Trade-Off; Distributionally Robust Optimization.

*Corresponding author: a.selvi19@imperial.ac.uk

1 Introduction

When organizations collect personal data about individuals, it is their responsibility to protect that data when they share information about it with third parties. Data anonymization, which aims to alter the data so that individuals are no longer identifiable, is often insufficient in that regard as it is prone to, among others, reconstruction attacks (Dinur and Nissim 2003) and de-identification attacks (Sweeney 1997, Heffetz and Ligett 2014). To address this issue, manifold definitions of data privacy have been proposed, including k-map (Sweeney 2001, §4.3), k-anonymity (Sweeney 2002), ℓ-diversity (Machanavajjhala et al. 2007) and δ-presence (Nergiz et al. 2007); see also the review of Desfontaines (2020, §2.1). Among those, *differential privacy* (DP), first proposed by Dwork et al. (2006b), has arguably received the most attention among researchers and practitioners.

DP considers databases $D \in \mathcal{D}^n$ with n individuals (or rows), each of which stems from a data universe $\mathcal{D}$ that characterizes the admissible attribute vectors (*e.g.*, the possible values of the predictors and the output for a supervised learner). For any $\varepsilon, \delta \geq 0$, a randomized algorithm $\mathcal{A}$ mapping databases $D \in \mathcal{D}^n$ to random outputs $\omega \in \Omega$ is (ε, δ)-*differentially private* if

$$\mathbb{P}[\mathcal{A}(D) \in A] \leq e^{\varepsilon} \cdot \mathbb{P}[\mathcal{A}(D') \in A] + \delta \qquad \forall (D, D') \in \mathcal{N},\ \forall A \in \mathcal{F},$$

where $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and

$$\mathcal{N} = \{(D, D') \in \mathcal{D}^n \times \mathcal{D}^n \ :\ D' = (D_{-k}, d) \text{ for some } k = 1, \ldots, n \text{ and } d \in \mathcal{D}\}$$

denotes the (symmetric) set of neighboring databases (D, D'), where D' emerges from D by replacing its k-th element with any $d \in \mathcal{D}$ (Dwork et al. 2006a,b). Intuitively, under DP with $\delta = 0$, an adversary cannot confidently estimate any single row of the database D from a single sample $\omega \sim \mathcal{A}(D)$ even if she knows all other rows of D and the implementation of $\mathcal{A}$ (Dwork 2011). In the general case where $\delta > 0$, (ε, δ)-DP is a sufficient (but not necessary) condition for the aforementioned $(\varepsilon, 0)$-DP to hold with a probability of at least $1 - \delta$ (Dinur and Nissim 2003, Meiser 2018, Canonne et al. 2020). Viewed through a Bayesian lens, (ε, δ)-DP allows the adversary to update her prior on D by at most an amount that is bounded by a function of ε and δ upon seeing a realization of $\mathcal{A}(D)$, see Vadhan (2017, §1.6).

Compared to other notions of privacy, DP enjoys several desirable features. The composition theorem (Dwork and Roth 2014, Theorem 3.16), for example, implies that sharing k different statistics, where statistic i has been generated by a $(\varepsilon_i, \delta_i)$-DP mechanism, $i = 1, \ldots, k$, satisfies $(\sum_{i=1}^k \varepsilon_i, \sum_{i=1}^k \delta_i)$-DP. Likewise, the post processing property asserts that any analysis derived from the output of a differentially private mechanism remains differentially private with the same privacy guarantees (Dwork and Roth 2014, Proposition 2.1). Due to these and other features, DP has found manifold recent applications in statistics and machine learning (Chaudhuri and Monteleoni 2009, Friedman and Schuster 2010, Chaudhuri et al. 2011, Abadi et al. 2016, Cai and Kou 2019), optimization (Mangasarian 2011, Hsu et al. 2014, Han et al. 2016, Hsu et al. 2016), mechanism design (McSherry and Talwar 2007) and revenue management (Chen et al. 2022, 2023). DP has also been widely applied in industry, ranging from emoji recommender systems that learn from user behavior (Apple Differential Privacy Team 2017), databases that publish

user interactions on Facebook (Messing et al. 2020) and COVID-19 vaccination search insights at Google (Bavadekar et al. 2021) to insights from LinkedIn's Economic Graph (Rogers et al. 2020), the U.S. Broadband Coverage dataset posted by Microsoft (Pereira et al. 2021) and the earnings distribution published by the U.S. Census Bureau (Foote et al. 2019).

In this paper, we study algorithms that perturb the output of a scalar *query function* $f : \mathcal{D}^n \mapsto \mathbb{R}$ so as to guarantee (ε, δ)-DP. The query function f could be a simple statistical query, such as the average, the median or a quantile of a real-valued attribute across all rows, a count of the rows satisfying a user-specified condition, or it could be part of a machine learning model that is trained on the database. In this setting, the Laplace mechanism (Dwork et al. 2006b) achieves $(\varepsilon, 0)$-DP, also referred to as *pure* differential privacy, by returning $f(D) + \tilde{X}$, where the random variable $\tilde{X}$ follows a zero-mean Laplace distribution with scale parameter $\Delta f/\varepsilon$ and $\Delta f := \sup_{(D,D')\in\mathcal{N}} |f(D) - f(D')|$ denoting the sensitivity of the query f. Pure differential privacy bounds the probability ratio of outputs within any measurable set $A \in \mathcal{F}$, no matter how small the involved probabilities are. This significantly restricts the design of admissible algorithms $\mathcal{A}$, particularly in their tail behavior, and it has spurred research into other DP notions that relax the privacy requirement for unlikely events (Desfontaines and Pejó 2020). The most prominent notion is (ε, δ)-DP, also known as *approximate* differential privacy. The Gaussian mechanism (Dwork and Roth 2014, Appendix A), for example, achieves (ε, δ)-DP for any $\varepsilon, \delta \in (0, 1)$ by returning $f(D) + \tilde{Y}$, where the random variable $\tilde{Y}$ follows a zero-mean Gaussian distribution with variance $2\ln(1.25/\delta)(\Delta f/\varepsilon)^2$.

The Laplace and Gaussian mechanisms are *data independent additive noise mechanisms* as their additive noises $\tilde{X}$ and $\tilde{Y}$ do not depend on the database D. In contrast, the noise of a *data dependent mechanism* may depend on the database D. One of the earliest data dependent mechanisms is the exponential mechanism (McSherry and Talwar 2007), which achieves $(\varepsilon, 0)$-DP for query functions with nominal outputs. Instead of adding data independent noise to the query output, the exponential mechanism ensures that the output is always admissible by randomly selecting one of finitely many outputs according to some score function. For query functions with real-valued outputs, smooth sensitivity mechanisms (Nissim et al. 2007) achieve (ε, δ)-DP at a higher accuracy than their data independent counterparts by adding noise whose variance is smaller for databases in neighborhoods with a low variation of the query outputs.

Algorithms with stronger privacy guarantees tend to offer less utility from data (Alvim et al. 2011). This is clearly seen for the Laplace and Gaussian mechanisms, whose variances increase with smaller values of ε and δ. It is therefore natural to study whether those mechanisms are optimal, that is, whether they minimize a pre-specified loss function among the respective classes of $(\varepsilon, 0)$-DP and (ε, δ)-DP algorithms. The early work on optimal mechanisms has focused on specific query functions (such as count queries), and is reviewed, among others, by Geng and Viswanath (2014, §1) and Sommer (2021, §4.6). Among the first papers that investigate optimal mechanisms for generic query functions is the work of Soria-Comas and Domingo-Ferrer (2013), who show that for large classes of $(\varepsilon, 0)$-DP data independent additive noise mechanisms and loss functions, a necessary optimality condition is that no probability mass in the distribution of the random noise can be moved towards zero without violating the privacy guarantee. The Laplace mechanism violates this condition and is thus not optimal, whereas the condition is

satisfied by piecewise constant 'staircase distributions' that move the probability mass of the Laplace distribution closer to zero. Geng and Viswanath (2014) show that for classes of ℓ_1- and ℓ_2-loss functions, the Laplace mechanism is asymptotically optimal as $\varepsilon \to 0$, but that it can be significantly suboptimal for larger values of ε. They also propose an $(\varepsilon, 0)$-DP data independent additive noise mechanism based on piecewise constant staircase distributions that is optimal across all $(\varepsilon, 0)$-DP algorithms for a large class of symmetric and increasing loss functions. While determining the optimal staircase distribution generally requires the tuning of a single parameter, closed-form characterizations are provided for the special cases of ℓ_1- and ℓ_2-loss functions.

In contrast, the design of optimal mechanisms for $\delta \neq 0$ is much less well understood. For the special case where $\varepsilon = 0$ (also known as additive differential privacy), Geng et al. (2019) show that sampling the additive noise from the product of a uniform and a Bernoulli random variable is optimal for symmetric and increasing loss functions among the class of $(0, \delta)$-DP data independent additive noise mechanisms with decreasing noise distributions. Closed-form characterizations of the optimal distributions are provided for ℓ_p-loss functions, whereas the general case requires the tuning of a single parameter. Balle and Wang (2018) show that the parameter choice of the aforementioned Gaussian mechanism is suboptimal. They propose the analytic Gaussian mechanism, which is optimal among the family of Gaussian mechanisms, by numerically computing the smallest variance that satisfies (ε, δ)-DP. For the general class of (ε, δ)-DP data independent additive noise mechanisms, Geng and Viswanath (2015) show that the suboptimality of uniform and discretized Laplace distributions can be bounded by a multiplicative constant for integer-valued queries under ℓ_1- and ℓ_2-loss functions when $\varepsilon \to 0$ and $\delta \to 0$ simultaneously. Tighter suboptimality bounds have been derived by Geng et al. (2020) for real-valued queries when the data independent additive noise is governed by a truncated Laplace distribution. In their analysis, the authors decompose the support of the distribution into a 'body' that achieves $(\varepsilon, 0)$-DP and a 'tail' that breaches privacy but is limited to a probability mass of δ. The resulting mechanisms are (ε, δ)-DP, and they are asymptotically optimal under ℓ_1- and ℓ_2-loss functions when $\varepsilon \to 0$ and $\delta \to 0$ simultaneously. The authors show that the truncated Laplace mechanism outperforms the aforementioned analytic Gaussian mechanism under ℓ_1- and ℓ_2-loss functions. To our best knowledge, the truncated Laplace mechanism provides the strongest optimality guarantees among the currently known (ε, δ)-DP mechanisms for generic queries.

The definition of optimality is more involved for data dependent mechanisms since their expected losses vary with the database. Minimizing the worst-case expected loss across all potential databases $D \in \mathcal{D}^n$ (which the *minimax optimality* criterion attempts to achieve) is overly conservative since it implies that data independent mechanisms remain optimal under mild conditions (Geng and Viswanath 2014). Instead, instance specific optimality criteria have been proposed that compare the expected loss for each database with a lower bound that is tailored to the database. Local minimax optimality (Asi and Duchi 2020a,b), for example, requires that a mechanism's expected loss for any database $D \in \mathcal{D}^n$ is within a constant factor of the expected loss of any other DP mechanism for at least one database in a neighborhood of D. Local minimax optimality recovers the earlier notion of minimax optimality when the neighborhood contains all databases $D \in \mathcal{D}^n$. Asi and Duchi (2020a,b) show that under the expected ℓ_1-loss, the inverse sensitivity mechanism first proposed by Johnson and Shmatikov

(2013) satisfies local minimax optimality for various query functions, and that it outperforms the Laplace and the smooth Laplace mechanisms under mild conditions.

The aforementioned contributions to the design of optimal (ε, δ)-DP mechanisms have in common that they limit their attention *a priori* to specific classes of mechanisms (such as Gaussian; standard, truncated or discretized Laplace; staircase; uniform-Bernoulli product or uniform distributions) and subsequently prove either optimality among the mechanisms in their respective classes or asymptotic optimality among larger families of mechanisms as $\varepsilon \to 0$ and $\delta \to 0$ simultaneously. In this paper, we propose to formulate and solve the optimal (ε, δ)-DP mechanism design problem as an infinite-dimensional distributionally robust optimization (DRO) problem (Delage and Ye 2010, Wiesemann et al. 2014, Kuhn et al. 2019). To this end, we minimize an expected loss function over all noise distributions, subject to the satisfaction of the DP constraints. In contrast to much of the existing literature, our formulation caters for generic loss functions (including asymmetric ones such as the pinball loss), and it can restrict the noise via support constraints. We show that our formulation affords a strong dual, and we develop hierarchies of finite-dimensional conservative approximations to the primal and dual formulations to derive converging upper and lower bounds on the optimal expected loss. Our upper bounds correspond to implementable perturbations whose optimality gaps can be certified by the lower bounds. Our bounding problems can be solved efficiently via cutting plane techniques that leverage the inherent problem structure. Our numerical experiments show that our optimal mechanisms can outperform the previously best results from the literature on artificial as well as two standard machine learning benchmark problems.

The contributions of this paper may be summarized as follows.

(i) We formulate the data independent additive noise problem as a DRO problem. Our formulation is flexible enough to cater for a large range of loss functions, and it extends to various problem variants such as the data dependent and the instance optimal problem.

(ii) We show that our primal and dual formulations can be bounded from above and below by converging hierarchies of large-scale linear programs that can be solved efficiently via tailored cutting plane techniques.

(iii) In contrast to the existing optimality results, which are either restricted to specific mechanisms or hold asymptotically, our formulation affords optimality guarantees that are non-asymptotic and that apply to any choice of ε and δ. Our numerical results showcase the advantages of our approach on a range of artificial as well as benchmark problems.

In our view, the optimization perspective on DP put forward in this paper opens up several opportunities for future research. On one hand, our proposed hierarchy of primal and dual bounds appears to extend to other existing and new definitions of DP, it may allow for the development of approximation schemes for (ε, δ)-DP with either *a priori* or *a posteriori* optimality guarantees, and it may generalize to multi-dimensional queries. On the other hand, the DP mechanism design problem gives rise to novel classes of DRO problems that have not been studied previously and that may find applications elsewhere. Most importantly, we believe that the design of optimal DP mechanisms should be viewed and addressed as a DRO problem, and the DRO community

has developed a rich arsenal of techniques that can be leveraged beneficially to contribute with novel insights and algorithms.

The remainder of the paper unfolds as follows. Section 2 formulates the data independent additive noise problem as an infinite-dimensional DRO problem, it shows that the problem affords a strong dual, and it develops a hierarchy of converging finite-dimensional upper and lower bounding problems. Section 3 extends our analysis to the data dependent additive noise setting. Section 4 develops a cutting plane algorithm to solve the bounding problems of Sections 2 and 3. We report numerical experiments in Section 5 and offer concluding remarks in Section 6, respectively. For ease of exposition, all proofs are relegated to the appendix.

Notation. Bold lower case letters denote vectors, while scalars are assigned standard lower case letters. The sets $\{1,\ldots,N\}$ and $\{-N,\ldots,N\}$ are abbreviated by $[N]$ and $[\pm N]$, respectively. For a set $\mathcal{S}$ and a scalar $a \in \mathbb{R}$, we let $\mathcal{S}+a = \{s+a \;:\; s \in \mathcal{S}\}$ denote the Minkowski sum of the sets $\mathcal{S}$ and $\{a\}$; similarly, $\mathcal{S}-a$ abbreviates $\mathcal{S}+(-a)$. For a measurable interval $I \subseteq \mathbb{R}^n$, $|I| \in \mathbb{R}\cup\{+\infty\}$ denotes the Lebesgue measure of I. Unless otherwise stated, measures of real sets are defined on the corresponding Borel σ-algebras. The sets of sign-unrestricted and non-negative measures that are defined on the power-set σ-algebra of some set A are denoted by $\mathcal{M}(A)$ and $\mathcal{M}_+(A)$, respectively. Finally, $\mathbb{1}[\mathcal{E}]$ is the indicator function taking value 1 if the condition $\mathcal{E}$ is satisfied and 0 otherwise.

2 Data Independent Noise Optimization

We start with data independent additive noise mechanisms that perturb the query output $f(D)$ of a database $D \in \mathcal{D}^n$ by adding a random noise $\tilde{X}$ independent of D so as to minimize an expected loss while satisfying (ε,δ)-DP. We formalize this problem in Section 2.1, and Section 2.2 derives a converging hierarchy of finite-dimensional upper and lower bounding problems.

2.1 The Data Independent Optimization Problem

We study the problem

$$\begin{array}{ll} \underset{\gamma}{\text{minimize}} & \displaystyle\int_{x\in\mathbb{R}} c(x)\,\mathrm{d}\gamma(x) \\ \text{subject to} & \gamma \in \mathcal{P}_0 \\ & \displaystyle\int_{x\in\mathbb{R}} \mathbb{1}[f(D)+x\in A]\,\mathrm{d}\gamma(x) \le e^{\varepsilon}\cdot\int_{x\in\mathbb{R}} \mathbb{1}[f(D')+x\in A]\,\mathrm{d}\gamma(x)+\delta \\ & \qquad\qquad \forall (D,D')\in\mathcal{N},\ \forall A\in\mathcal{F}, \end{array} \tag{1}$$

where $(\mathbb{R},\mathcal{F})$ is a measurable space with the Borel σ-algebra $\mathcal{F}$ on $\mathbb{R}$ and the set $\mathcal{P}_0$ of probability measures supported on $\mathbb{R}$. Problem (1) selects a probability measure γ governing the random noise $\tilde{X}$ so as to minimize the expected value of the Borel loss function $c:\mathbb{R}\mapsto\mathbb{R}_+$, subject to (ε,δ)-DP for $\varepsilon,\delta>0$. Note in particular that the integrals on both sides of the DP constraint evaluate the probabilities $\mathbb{P}[\mathcal{A}(D)\in A]$ and $\mathbb{P}[\mathcal{A}(D')\in A]$ for the randomized query outputs $\mathcal{A}(D)=f(D)+\tilde{X}$ and $\mathcal{A}(D')=f(D')+\tilde{X}$ with $\tilde{X}\sim\gamma$ as per our definition of (ε,δ)-DP from the previous section. We assume that c satisfies the following regularity conditions.

Assumption 1 (Loss Function). *The loss function $c : \mathbb{R} \mapsto \mathbb{R}_+$ satisfies the following properties:*

(a) Continuity. *c is continuous on $\mathbb{R}$.*

(b) Monotonicity. *$c(x) \leq c(x')$ for all $0 \leq x \leq x'$ and all $x' \leq x \leq 0$.*

(c) Unboundedness. *For any $r \in \mathbb{R}$ we have $c(x) \geq r$ for $|x|$ sufficiently large.*

(d) Polynomial Growth. *There exists a polynomial $g : \mathbb{R} \mapsto \mathbb{R}$ such that $g(x) \geq c(x)$ $\forall x \in \mathbb{R}$.*

Assumption *(a)* enables us to construct discrete approximations to problem (1) and its dual that converge as we refine their granularity. Assumptions *(b)* and *(c)* allow us to restrict these approximations to a bounded support of the involved measures without incurring an unbounded loss. Assumption *(d)*, finally, ensures that problem (1) has a finite optimal value. Loss functions typically used in the literature, such as the noise amplitude (ℓ_1-loss with $c(x) = |x|$) and the noise power (ℓ_2-loss with $c(x) = x^2$), satisfy Assumption 1.

Recall that $\Delta f := \sup\{f(D') - f(D) \ : \ (D, D') \in \mathcal{N}\}$ is the global sensitivity of the query f over the set of neighboring databases $\mathcal{N}$. We assume that f is surjective in the following sense.

Assumption 2 (Query Function). *For each $\varphi \in [-\Delta f, \Delta f]$, we have $f(D') - f(D) = \varphi$ for some $(D, D') \in \mathcal{N}$.*

Assumption 2 is standard (Geng and Viswanath 2014, Geng et al. 2020), and it is satisfied by common descriptive statistics including the mean, median, minimum/maximum and standard deviation, as well as several popular machine learning algorithms (*cf.* Section 5), if the data is numeric. Our theory continues to apply if Assumption 2 is violated, but our reformulation of problem (1) will be conservative as it guarantees DP over the entire range $\varphi \in [-\Delta f, \Delta f]$, as opposed to the subset of query output differences that can actually be observed over $\mathcal{N}$.

Assumption 2 allows us to replace the DP constraint in (1) with

$$\begin{array}{ll} & \displaystyle\int_{x\in\mathbb{R}} \mathbb{1}[x \in A]\,\mathrm{d}\gamma(x) \leq e^{\varepsilon} \cdot \int_{x\in\mathbb{R}} \mathbb{1}[f(D') - f(D) + x \in A]\,\mathrm{d}\gamma(x) + \delta \quad \forall (D, D') \in \mathcal{N},\ \forall A \in \mathcal{F} \\ \iff & \displaystyle\int_{x\in\mathbb{R}} \mathbb{1}[x \in A]\,\mathrm{d}\gamma(x) \leq e^{\varepsilon} \cdot \int_{x\in\mathbb{R}} \mathbb{1}[x + \varphi \in A]\,\mathrm{d}\gamma(x) + \delta \qquad\qquad \forall (\varphi, A) \in \mathcal{E}, \end{array}$$

where $\mathcal{E} := [-\Delta f, \Delta f] \times \mathcal{F}$. Here, the first line holds since $\{A : A \in \mathcal{F}\} = \{A + f(D) : A \in \mathcal{F}\}$ for any $D \in \mathcal{D}^n$, whereas the second line is due to Assumption 2.

In summary, the *data independent noise optimization problem* can be formulated as

$$\begin{array}{ll} \underset{\gamma}{\text{minimize}} & \displaystyle\int_{x\in\mathbb{R}} c(x)\,\mathrm{d}\gamma(x) \\ \text{subject to} & \gamma \in \mathcal{P}_0 \\ & \displaystyle\int_{x\in\mathbb{R}} \mathbb{1}[x \in A]\,\mathrm{d}\gamma(x) \leq e^{\varepsilon} \cdot \int_{x\in\mathbb{R}} \mathbb{1}[x + \varphi \in A]\,\mathrm{d}\gamma(x) + \delta \quad \forall (\varphi, A) \in \mathcal{E}. \end{array} \tag{P}$$

Problem P has uncountably many decision variables and constraints and thus appears to be challenging to solve. The problem is feasible since its constraints are satisfied, for example, by Laplace (Dwork et al. 2006b) and Gaussian measures (Dwork and Roth 2014, Theorem A.1). Convexity of the feasible region implies that mixtures of such measures are also feasible.

Problem P can be interpreted as an uncertainty quantification problem from the distributionally robust optimization literature (Owhadi et al. 2013, Han et al. 2015, Hanasusanto et al. 2015). Under this view, the objective function evaluates the worst-case expected value of the uncertain profit c of a decision maker's action, subject to an uncountable number of moment constraints. Problem P differs from the uncertainty quantification problems typically studied in the literature in both the number and the structure of these moment constraints.

Problem P is also reminiscent of continuous linear programs (Anderson and Nash 1987), which comprise uncountably many decision variables and constraints as well. Owing to their continuous-time control heritage, however, the constraints in continuous linear programs are indexed by a single bounded real scalar $x \in [0, T]$, whereas our constraint indices additionally involve the set of all Borel sets $\mathcal{F}$. Moreover, the decision variable of a continuous linear program has a bounded support $[0, T]$ and is assumed to admit a density, whereas our decision variable γ has an unbounded support $\mathbb{R}$ and may not admit a density. Both of these additional complications imply that we cannot directly use the theory of continuous linear programming and instead have to derive bounding problems and prove their convergence from first principles.

2.2 A Hierarchy of Converging Bounding Problems

To obtain a tractable upper bound on problem P, we first introduce a restriction of P that replaces the generic measure γ with the piecewise constant function

$$\gamma(A) = \sum_{i\in\mathbb{Z}} p(i) \cdot \frac{|A \cap I_i(\beta)|}{\beta} \quad \forall A \in \mathcal{F}, \tag{2}$$

where $p : \mathbb{Z} \mapsto \mathbb{R}_+$ satisfies $\sum_{i\in\mathbb{Z}} p(i) = 1$, and $\{I_i(\beta)\}_{i\in\mathbb{Z}}$ partitions $\mathbb{R}$ into disjoint intervals $I_i(\beta) := [i \cdot \beta, (i+1) \cdot \beta)$, $i \in \mathbb{Z}$, of some pre-selected length $\beta > 0$. To simplify the exposition, we assume that Δf is divisible by β. Under restriction (2), most constraints in P become redundant.

Lemma 1. *Under restriction* (2), P *has the same optimal value as*

$$\begin{array}{ll} \underset{p}{\text{minimize}} & \displaystyle\sum_{i\in\mathbb{Z}} c_i(\beta) \cdot p(i) \\ \text{subject to} & p : \mathbb{Z} \mapsto \mathbb{R}_+, \ \displaystyle\sum_{i\in\mathbb{Z}} p(i) = 1 \\ & \displaystyle\sum_{i\in\mathbb{Z}} \mathbb{1}[I_i(\beta) \subseteq A] \cdot p(i) \le e^{\varepsilon} \cdot \sum_{i\in\mathbb{Z}} \mathbb{1}[I_i(\beta) + \varphi \subseteq A] \cdot p(i) + \delta \quad \forall (\varphi, A) \in \mathcal{E}(\beta), \end{array} \tag{P(β)}$$

where $c_i(\beta) := \beta^{-1} \cdot \int_{x\in I_i(\beta)} c(x)\mathrm{d}x$ *and* $\mathcal{E}(\beta) := \mathscr{B}(\beta) \times \mathcal{F}(\beta)$ *with* $\mathscr{B}(\beta) := \{-\Delta f, -\Delta f + \beta, \ldots, \Delta f\}$ *and* $\mathcal{F}(\beta) := \big\{ \bigcup_{i\in\mathcal{I}} I_i(\beta) \ : \ \mathcal{I} \subseteq \mathbb{Z} \big\}$.

In contrast to P, problem P(β) has countably many decision variables. By Cantor's theorem, however, it still comprises uncountably many constraints since $\mathcal{F}(\beta)$ is indexed by the power set of infinitely many intervals $\{I_i(\beta)\}_{i\in\mathbb{Z}}$. To bound P($\beta$) from above by a finite-dimensional linear optimization problem, we constrain P(β) further by restricting the discrete probability measure

p to a bounded support. Formally, we impose that there is $L \in \mathbb{N}$ such that

$$p(i) = 0 \quad \forall i \in \mathbb{Z} \setminus [\pm L], \tag{3}$$

that is, we bound the overall support to $2L + 1$ intervals $I_i(\beta)$ centered at 0. Restriction (3) allows us to remove from $\mathrm{P}(\beta)$ any privacy constraint that relates to intervals $I_i(\beta)$ and $I_j(\beta)$ whose indices i and j both lie outside the support $[\pm L]$.

Proposition 1. *With the additional constraint* (3), $\mathrm{P}(\beta)$ *has the same optimal value as*

$$\begin{array}{ll} \underset{p}{\text{minimize}} & \displaystyle\sum_{i\in[\pm L]} c_i(\beta)\cdot p(i) \\ \text{subject to} & p : [\pm L] \mapsto \mathbb{R}_+, \ \displaystyle\sum_{i\in[\pm L]} p(i) = 1 \\ & \displaystyle\sum_{i\in[\pm L]} \mathbb{1}[I_i(\beta)\subseteq A]\cdot p(i) \le e^{\varepsilon}\cdot \sum_{i\in[\pm L]} \mathbb{1}[I_i(\beta)+\varphi\subseteq A]\cdot p(i) + \delta \quad \forall (\varphi, A)\in\mathcal{E}(L,\beta), \end{array} \tag{$\mathrm{P}(L,\beta)$}$$

where $\mathcal{E}(L,\beta) := \mathscr{B}(\beta)\times\mathcal{F}(L,\beta)$ *with* $\mathcal{F}(L,\beta) := \big\{\bigcup_{i\in\mathcal{L}} I_i(\beta) \ : \ \mathcal{L}\subseteq[\pm L]\big\}$.

Problem $\mathrm{P}(L,\beta)$ constitutes a large-scale but finite-dimensional linear optimization problem that will serve as a building block to our cutting plane algorithm in Section 4. In terms of optimal values, we have the relationship $\mathrm{P}(L,\beta) \ge \mathrm{P}(\beta) \ge \mathrm{P}$ for all support sizes $L \in \mathbb{N}$ and interval lengths β.

We next derive a lower bound on P. To this end, we employ a strategy widely adopted in distributionally robust optimization and first propose a dual to problem P:

$$\begin{array}{ll} \underset{\theta,\psi}{\text{maximize}} & \theta - \delta\displaystyle\int_{(\varphi,A)\in\mathcal{E}} \mathrm{d}\psi(\varphi,A) \\ \text{subject to} & \theta\in\mathbb{R},\ \psi\in\mathcal{M}_+(\mathcal{E}) \\ & \theta \le c(x) + \displaystyle\int_{(\varphi,A)\in\mathcal{E}} \mathbb{1}[x\in A]\mathrm{d}\psi(\varphi,A) - e^{\varepsilon}\cdot\int_{(\varphi,A)\in\mathcal{E}} \mathbb{1}[x+\varphi\in A]\mathrm{d}\psi(\varphi,A) \\ & \qquad\qquad \forall x\in\mathbb{R}. \end{array} \tag{D}$$

The integrals in this problem are well-defined due to the domain of ψ specified in the first constraint. Problem D affords a natural interpretation: suppose that $\delta = 0$ and replace in the objective function the epigraphical variable θ with

$$\inf_{x\in\mathbb{R}} \ c(x) + \int_{(\varphi,A)\in\mathcal{E}} \mathbb{1}[x\in A]\mathrm{d}\psi(\varphi,A) - e^{\varepsilon}\cdot\int_{(\varphi,A)\in\mathcal{E}} \mathbb{1}[x+\varphi\in A]\mathrm{d}\psi(\varphi,A).$$

Problem D then determines a conic combination of database-event pairs (φ, A) that maximizes the sum of noise-related costs $c(x)$ and cumulative DP shortfall (*i.e.*, the cumulative violation of all DP constraints) under the most benign realization x of the random noise $\tilde{X}$.

We can readily establish weak duality between the problems P and D.

Proposition 2 (Weak Duality)**.** *For any* γ *feasible in* P *and* (θ,ψ) *feasible in* D*, we have*

$$\int_{x\in\mathbb{R}} c(x)\,\mathrm{d}\gamma(x) \ge \theta - \delta\int_{(\varphi,A)\in\mathcal{E}} \mathrm{d}\psi(\varphi,A).$$

Similar to problem P, problem D appears challenging to solve since it comprises uncountably many decision variables and constraints. To construct a tractable lower bound on D, we first remove all variables $\psi(\varphi, A)$ indexed by $(\varphi, A) \in \mathcal{E} \setminus \mathcal{E}(\beta)$, that is, we impose that

$$\int_{(\varphi,A)\in\mathcal{E}\setminus\mathcal{E}(\beta)} \mathrm{d}\psi(\varphi, A) = 0. \tag{4}$$

Restriction (4) can be understood as the dual pendant to our discretization (2); in fact, the dual variables unaffected by (4) correspond precisely to the constraints in problem P(β). Under restriction (4), most constraints of D become redundant.

Lemma 2. *With the additional constraint* (4)*,* D *has the same optimal value as*

$$\begin{array}{ll} \underset{\theta,\psi}{\text{maximize}} & \theta - \delta \cdot \displaystyle\int_{(\varphi,A)\in\mathcal{E}(\beta)} \mathrm{d}\psi(\varphi, A) \\ \text{subject to} & \theta \in \mathbb{R},\ \psi \in \mathcal{M}_+(\mathcal{E}(\beta)) \\ & \theta \le \underline{c}_i(\beta) + \displaystyle\int_{(\varphi,A)\in\mathcal{E}(\beta)} \mathbb{1}[I_i(\beta) \subseteq A]\mathrm{d}\psi(\varphi, A) - e^{\varepsilon} \cdot \int_{(\varphi,A)\in\mathcal{E}(\beta)} \mathbb{1}[I_i(\beta) + \varphi \subseteq A]\mathrm{d}\psi(\varphi, A) \\ & \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \forall i \in \mathbb{Z}, \\ & \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (\mathrm{D}(\beta)) \end{array}$$

where $\underline{c}_i(\beta) := \inf\{c(x) : x \in I_i(\beta)\},\ i \in \mathbb{Z}$.

In contrast to problem D, which comprises uncountably many constraints, problem D(β) has countably many constraints. However, the problem still contains infinitely many constraints as well as uncountably many variables. To bound D(β) from below by a finite-dimensional linear optimization problem, we set $\psi(\mathcal{E}(\beta) \setminus \mathcal{E}(L, \beta)) = 0$ for some $L \in \mathbb{N}$, that is, we impose that

$$\int_{(\varphi,A)\in\mathcal{E}(\beta)\setminus\mathcal{E}(L,\beta)} \mathrm{d}\psi(\varphi, A) = 0. \tag{5}$$

Restriction (5) is the dual pendant to our support constraint (3). It removes variables associated with events that contain intervals sufficiently far away from 0 since $(\varphi, A) \in \mathcal{E}(\beta)\setminus\mathcal{E}(L, \beta)$ implies that $A \not\subseteq \bigcup_{i\in[\pm L]} I_i(\beta)$.

Proposition 3. *With the additional constraint* (5)*,* D(β) *has the same optimal value as*

$$\begin{array}{ll} \underset{\theta,\psi}{\text{maximize}} & \theta - \delta \cdot \displaystyle\sum_{(\varphi,A)\in\mathcal{E}(L,\beta)} \psi(\varphi, A) \\ \text{subject to} & \theta \in \mathbb{R},\ \psi : \mathcal{E}(L, \beta) \mapsto \mathbb{R}_+ \\ & \theta \le \underline{c}_i(\beta) + \displaystyle\sum_{(\varphi,A)\in\mathcal{E}(L,\beta)} \mathbb{1}[I_i(\beta) \subseteq A] \cdot \psi(\varphi, A) - e^{\varepsilon} \cdot \sum_{(\varphi,A)\in\mathcal{E}(L,\beta)} \mathbb{1}[I_i(\beta) + \varphi \subseteq A] \cdot \psi(\varphi, A) \\ & \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \forall i \in [\pm(L + \Delta f/\beta)]. \\ & \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (\mathrm{D}(L, \beta)) \end{array}$$

Similar to P(L, β), problem D(L, β) constitutes a large-scale but finite-dimensional linear optimization problem that will serve as a building block to our cutting plane algorithm in Section 4. In terms of optimal values, we have the relationship D(L, β) $\le$ D(β) $\le$ D for all support sizes

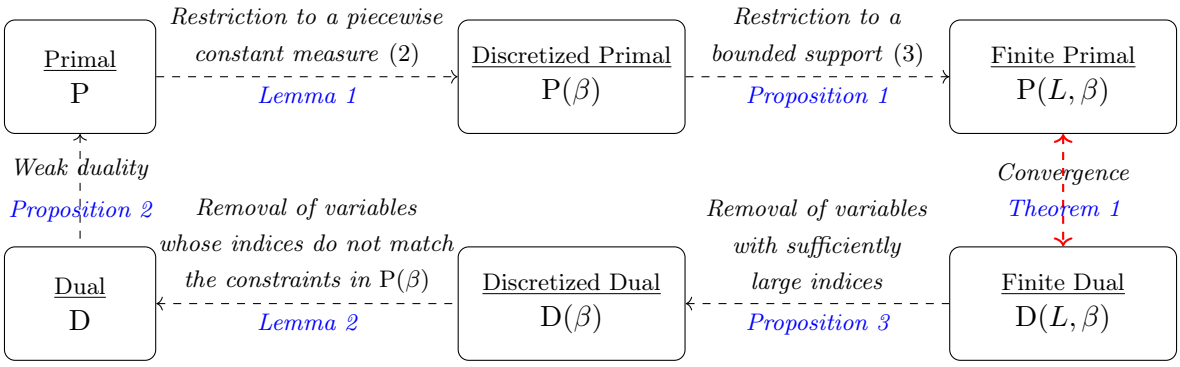

Figure 1: *Summary of the results in Section 2. Directed arrows $x \dashrightarrow y$ indicate upper bound relationships $x \leq y$, whereas the double arrow confirms the convergence of optimal values as L increases and β decreases.*

$L \in \mathbb{N}$ and interval lengths β. In particular, P and D are sandwiched by the finite-dimensional linear optimization problems $\mathrm{P}(L, \beta)$ and $\mathrm{D}(L, \beta)$.

We close this section with an analysis of the convergence of the finite-dimensional linear optimization problems $\mathrm{P}(L, \beta)$ and $\mathrm{D}(L, \beta)$. To this end, recall that by our earlier assumption, β divides Δf, which allows us to equivalently represent β as $\Delta f/k$ for some $k \in \mathbb{N}$.

Theorem 1. *For any $\xi > 0$, there is $\Lambda' \in \mathbb{N}$ and $k' \in \mathbb{N}$ such that*

$$\mathrm{P}(\Lambda \cdot k, \Delta f/k) - \mathrm{D}(\Lambda \cdot k, \Delta f/k) \leq \xi \qquad \forall \Lambda \geq \Lambda',\ \forall k \geq k'.$$

Intuitively, Theorem 1 states that both the discretization granularity $\Delta f/k$ needs to shrink *and* the support $[-\Lambda \cdot \Delta f, \Lambda \cdot (\Lambda + 1/k) \cdot \Delta f)$ of the noise distribution needs to grow for the primal and dual approximations to converge. In particular, keeping the support fixed (which amounts to fixing Λ in Theorem 1) and merely increasing k is *not* sufficient for convergence as the dual approximation provides a lower bound for *all* noise distributions (of potentially unbounded support), as opposed to only the noise distributions that share the same support as the primal approximation. Note that Theorem 1 also implies strong duality of the two infinite-dimensional problems P and D.

Figure 1 summarizes the key results of this section.

3 Data Dependent Noise Optimization

We next study data dependent noise mechanisms whose additive perturbation $\tilde{X}(f(D))$ of the query output $f(D)$ of a database $D \in \mathcal{D}^n$ may depend on $f(D)$. Section 3.1 formalizes this problem, and Section 3.2 develops finite-dimensional upper and lower bounding problems.

3.1 The Data Dependent Optimization Problem

We study the problem

$$\begin{array}{ll} \underset{\gamma}{\text{minimize}} & \displaystyle\int_{\phi\in\Phi} w(\phi)\cdot\left[\int_{x\in\mathbb{R}} c(x)\,\mathrm{d}\gamma(x\mid\phi)\right]\mathrm{d}\phi \\ \text{subject to} & \gamma\in\Gamma \\ & \displaystyle\int_{x\in\mathbb{R}} \mathbb{1}[f(D)+x\in A]\,\mathrm{d}\gamma(x\mid f(D)) \le e^{\varepsilon}\cdot\int_{x\in\mathbb{R}} \mathbb{1}[f(D')+x\in A]\,\mathrm{d}\gamma(x\mid f(D'))+\delta \\ & \qquad\qquad \forall (D,D')\in\mathcal{N},\ \forall A\in\mathcal{F}, \end{array} \tag{6}$$

where $\Phi := \{f(D) : D\in\mathcal{D}^n\}$ is the set of possible query outputs and $(\mathbb{R},\mathcal{F})$ is again a measurable space with the Borel σ-algebra $\mathcal{F}$ on $\mathbb{R}$. Problem (6) selects a family $\{\gamma(\cdot\mid\phi)\}_{\phi\in\Phi}$ of conditional probability measures governing the random noise $\tilde{X}(\cdot)$ so as to minimize an iterated expectation of the Borel loss function $c:\mathbb{R}\mapsto\mathbb{R}_+$ satisfying Assumption 1, subject to satisfaction of (ε,δ)-DP. The inner expectation in the objective function evaluates the expected loss for a specific query output $\phi\in\Phi$, whereas the outer expectation weighs different query outputs according to the continuous probability density function $w:\Phi\mapsto\mathbb{R}_+$. The domain of γ is now defined as

$$\Gamma := \left\{\gamma\ :\ \begin{bmatrix} \gamma(\cdot\mid\phi)\in\mathcal{P}_0, & \phi\in\Phi \\ \phi\mapsto\gamma(A\mid\phi)\text{ measurable}, & A\in\mathcal{F}\end{bmatrix}\right\},$$

where $\mathcal{P}_0$ is again the set of probability measures supported on $\mathbb{R}$. The domain Γ restricts γ to the set of Markov kernels with continuous state $\phi\in\Phi$. The first condition ensures that the inner expectation in the objective function is well-defined, while the second condition ensures that this expectation is measurable in the outer expectation.

To ensure privacy, the weighting w must not reveal anything about the true database (which is ensured, for example, by choosing the uniform distribution over Φ) or it must itself be kept private. In practice, the application domain often guides the choice of w; we discuss this further in Section 5. Similar to Assumption 2, we impose the following surjectivity requirement on f.

Assumption 3 (Query Function)**.** *Φ is a bounded interval, and for each $D\in\mathcal{D}^n$ and $\varphi\in[-\Delta f,\Delta f]\cap(\Phi-f(D))$, we have $f(D')-f(D)=\varphi$ for some $(D,D')\in\mathcal{N}$.*

Assumption 3 reduces to Assumption 2 if we set $\Phi=\mathbb{R}$. To simplify the exposition, however, we assume that Φ is bounded; otherwise, additional steps would have to be taken to restrict the states $\phi\in\Phi$ of γ to bounded intervals without incurring a potentially unbounded loss.

Assumption 3 allows us to replace the DP constraint in (6) with

$$\begin{aligned} &\int_{x\in\mathbb{R}} \mathbb{1}[x\in A]\,\mathrm{d}\gamma(x\mid f(D)) \le e^{\varepsilon}\cdot\int_{x\in\mathbb{R}} \mathbb{1}[f(D')-f(D)+x\in A]\,\mathrm{d}\gamma(x\mid f(D'))+\delta \\ &\qquad\qquad \forall (D,D')\in\mathcal{N},\ \forall A\in\mathcal{F} \\ \iff &\int_{x\in\mathbb{R}} \mathbb{1}[x\in A]\,\mathrm{d}\gamma(x\mid\phi) \le e^{\varepsilon}\cdot\int_{x\in\mathbb{R}} \mathbb{1}[x+\varphi\in A]\,\mathrm{d}\gamma(x\mid\phi+\varphi)+\delta \\ &\qquad\qquad \forall\phi\in\Phi,\ \forall(\varphi,A)\in\mathcal{E}'(\phi), \end{aligned}$$

where $\mathcal{E}'(\phi) := [[-\Delta f, \Delta f] \cap (\Phi - \phi)] \times \mathcal{F}$ for $\phi \in \Phi$. Here, the first line holds since $\{A : A \in \mathcal{F}\} = \{A + f(D) : A \in \mathcal{F}\}$ for any $D \in \mathcal{D}^n$, whereas the second line is due to Assumption 3.

In summary, the *data dependent noise optimization problem* can be formulated as

$$\begin{array}{ll} \underset{\gamma}{\text{minimize}} & \displaystyle\int_{\phi\in\Phi} w(\phi) \cdot \left[\int_{x\in\mathbb{R}} c(x)\,\mathrm{d}\gamma(x \mid \phi)\right] \mathrm{d}\phi \\ \text{subject to} & \gamma \in \Gamma \\ & \displaystyle\int_{x\in\mathbb{R}} \mathbb{1}[x \in A]\,\mathrm{d}\gamma(x \mid \phi) \le e^{\varepsilon} \cdot \int_{x\in\mathbb{R}} \mathbb{1}[x+\varphi \in A]\,\mathrm{d}\gamma(x \mid \phi+\varphi) + \delta \\ & \qquad\qquad \forall \phi \in \Phi,\ \forall (\varphi, A) \in \mathcal{E}'(\phi). \end{array} \tag{P$'$}$$

Problem P$'$ is not a generalization of problem P from Section 2 *per se*, but P would be recovered from P$'$ if we introduced the additional requirement that all $\gamma(\cdot|\phi)$, $\phi \in \Phi$, in P$'$ must coincide. This argument implies that problem P$'$ is guaranteed to be feasible. Note that P$'$ does not decompose into separate problems for $\phi \in \Phi$ since the DP constraint couples the conditional measures of neighbouring databases. Similar to P, problem P$'$ contains uncountably many decision variables and constraints and thus appears challenging to solve.

3.2 A Hierarchy of Converging Bounding Problems

To obtain a tractable upper bound on the data *independent* noise optimization problem, Section 2.2 bounds problem P from above by finite-dimensional approximations that live on a partition $\{I_i(\beta)\}_{i\in\mathbb{Z}}$ of the possible noise realizations into disjoint intervals $I_i(\beta) = [i \cdot \beta, (i+1) \cdot \beta)$ of length $\beta > 0$. In this section, we retain our earlier assumption that Δf is divisible by β, and we additionally stipulate that $\Phi = \bigcup_{k\in[K]} \Phi_k(\beta)$ with $\Phi_k(\beta) := I_{t+k}(\beta)$ for some $t \in \mathbb{Z}$ and $K \in \mathbb{N}$. This will allow us to partition the set of possible query outputs Φ in the same way, using a single granularity parameter β.

To bound problem P$'$ from above, we first restrict the uncountable family $\{\gamma(\cdot|\phi)\}_{\phi\in\Phi}$ of probability measures in P$'$ to a finite subset that is piecewise constant on the intervals $\Phi_k(\beta)$:

$$\gamma(\cdot \mid \phi) = \gamma(\cdot \mid \phi') \quad \forall k \in [K],\ \forall \phi, \phi' \in \Phi_k(\beta). \tag{7a}$$

Under restriction (7a), P$'$ optimizes over finitely many probability measures γ_k, $k \in [K]$, but it still involves uncountably many decision variables and constraints. To further simplify the problem, we restrict each probability measure in P$'$ to a piecewise constant function via

$$\gamma_k(A) = \sum_{i\in\mathbb{Z}} p_k(i) \cdot \frac{|A \cap I_i(\beta)|}{\beta} \quad \forall k \in [K],\ \forall A \in \mathcal{F} \tag{7b}$$

for a family of probability measures $\{p_k : \mathbb{Z} \mapsto \mathbb{R}_+\}_{k\in[K]}$ satisfying $\sum_{i\in\mathbb{Z}} p_k(i) = 1$ for all $k \in [K]$. We also restrict each probability measure γ_k to a bounded support by imposing that

$$p_k(i) = 0 \quad \forall k \in [K],\ \forall i \in \mathbb{Z} \setminus [\pm L] \tag{7c}$$

for some $L \in \mathbb{N}$. The restrictions (7b) and (7c) are akin to (2) and (3) from Section 2, respectively.

Proposition 4. *With the additional constraints* (7), P' *has the same optimal value as*

$$\begin{array}{ll} \underset{p}{\text{minimize}} & \beta \cdot \sum\limits_{k\in[K]} w_k(\beta) \cdot \Big[\sum\limits_{i\in[\pm L]} c_i(\beta)\cdot p_k(i)\Big] \\ \text{subject to} & p_k : [\pm L] \mapsto \mathbb{R}_+, \ \sum\limits_{i\in[\pm L]} p_k(i) = 1, \ k\in[K] \\ & \sum\limits_{i\in[\pm L]} \mathbb{1}[I_i(\beta)\subseteq A]\cdot p_k(i) \le e^\varepsilon \cdot \sum\limits_{i\in[\pm L]} \mathbb{1}[I_i(\beta)+\varphi\subseteq A]\cdot p_m(i) + \delta \\ & \qquad\qquad \forall k,m\in[K],\ \forall(\varphi,A)\in\mathcal{E}'_{km}(L,\beta) \end{array} \tag{$\mathrm{P}'(L,\beta)$}$$

where $w_k(\beta) := \beta^{-1}\cdot\int_{\phi\in\Phi_k(\beta)} w(\phi)\,\mathrm{d}\phi$ *and* $\mathcal{E}'_{km}(L,\beta) := [\mathscr{B}(\beta)\cap\{(m-k-1)\cdot\beta,\ (m-k)\cdot\beta,\ (m-k+1)\cdot\beta\}]\times\mathcal{F}(L,\beta)$ *with* $\mathscr{B}(\beta)$ *and* $\mathcal{F}(L,\beta)$ *defined as in Section 2.*

Problem $\mathrm{P}'(L,\beta)$ constitutes a large-scale but finite-dimensional linear optimization problem that will serve as a building block to our cutting plane algorithm in Section 4. In terms of optimal values, we have the relationship $\mathrm{P}'(L,\beta)\ge\mathrm{P}'$ for all $L\in\mathbb{N}$ and $\beta>0$.

To obtain a lower bound on P', we first propose the dual problem

$$\begin{array}{ll} \underset{\theta,\psi}{\text{maximize}} & \displaystyle\int_{\phi\in\Phi}\left[\theta(\phi)-\delta\cdot\int_{(\varphi,A)\in\mathcal{E}'(\phi)}\mathrm{d}\psi(\varphi,A\mid\phi)\right]\mathrm{d}\phi \\ \text{subject to} & \theta:\Phi\mapsto\mathbb{R}\text{ measurable},\ \psi\in\Psi \\ & \theta(\phi)\le\displaystyle\int_{(\varphi,A)\in\mathcal{E}'(\phi)}\mathbb{1}[x\in A]\,\mathrm{d}\psi(\varphi,A\mid\phi) - e^\varepsilon\cdot\int_{(-\varphi,A)\in\mathcal{E}'(\phi)}\mathbb{1}[x+\varphi\in A]\,\mathrm{d}\psi(\varphi,A\mid\phi-\varphi) \\ & \qquad\qquad +c(x)\cdot w(\phi)\quad\forall\phi\in\Phi,\ \forall x\in\mathbb{R}, \end{array} \tag{D'}$$

where the dual measure is defined over the set

$$\Psi := \left\{\psi\ :\ \left[\begin{array}{c} \psi(\cdot\mid\phi)\in\mathcal{M}_+(\mathcal{E}'(\phi)),\ \phi\in\Phi \\ \exists\psi_0\in\mathcal{M}(\mathcal{E})\ :\ \psi(\cdot\mid\phi)\ll\psi_0,\ \phi\in\Phi\text{ with }(\varphi,A,\phi)\mapsto\dfrac{\mathrm{d}\psi(\varphi,A\mid\phi)}{\mathrm{d}\psi_0(\varphi,A)}\text{ measurable}\end{array}\right]\right\}.$$

Here, the first condition resembles the domain that we imposed on ψ in problem D from Section 2.2. The second condition is new, and it ensures that the integral on the right-hand side of the DP constraint in D', which varies with φ, is well-defined.

We first establish weak duality between the problems P' and D'.

Proposition 5 (Weak Duality)**.** *For any* γ *feasible in* P' *and* (θ,ψ) *feasible in* D'*, we have*

$$\int_{\phi\in\Phi} w(\phi)\cdot\left[\int_{x\in\mathbb{R}} c(x)\,\mathrm{d}\gamma(x\mid\phi)\right]\mathrm{d}\phi \ge \int_{\phi\in\Phi}\left[\theta(\phi)-\delta\cdot\int_{(\varphi,A)\in\mathcal{E}'(\phi)}\mathrm{d}\psi(\varphi,A\mid\phi)\right]\mathrm{d}\phi.$$

To construct a tractable lower bound on D', we impose that

$$\psi(\cdot\mid\phi)=\psi(\cdot\mid\phi')\ \text{ and }\ \theta(\phi)=\theta(\phi')\qquad\forall k\in[K],\ \forall\phi,\phi'\in\Phi_k(\beta), \tag{8a}$$

that is, $\psi(\cdot|\phi)$ and $\theta(\phi)$ have to be piecewise constant over the intervals $\Phi_k(\beta)$, and we remove

all variables $\psi(\varphi, A|\phi)$ indexed by $(\varphi, A) \in \mathcal{E}'(\phi) \setminus \mathcal{E}(L, \beta)$, that is, we impose that

$$\int_{(\varphi,A)\in\mathcal{E}'(\phi)\setminus\mathcal{E}(L,\beta)} \mathrm{d}\psi(\varphi, A \mid \phi) = 0 \qquad \forall \phi \in \Phi, \tag{8b}$$

which implies that $\psi(\cdot|\phi)$ vanishes for all $\phi \in \Phi$ on (φ, A) with φ not divisible by β or A outside $\mathcal{F}(L, \beta)$. Constraint (8b) is reminiscent of the constraints (4) and (5) from Section 2, and it can be interpreted as the dual pendant of the restriction (7c).

Proposition 6. *With the additional constraints* (8), D' *has the same optimal value as*

$$\begin{array}{ll}
\underset{\theta,\psi}{\text{maximize}} & \beta \cdot \left[\displaystyle\sum_{k\in[K]} \theta_k - \delta \cdot \sum_{k\in[K]} \sum_{(\varphi,A)\in\mathcal{E}'_k(L,\beta)} \psi_k(\varphi, A) \right] \\
\text{subject to} & \boldsymbol{\theta} \in \mathbb{R}^K,\ \psi_k : \mathcal{E}'_k(L, \beta) \mapsto \mathbb{R}_+,\ k \in [K] \\
& \theta_k \leq \displaystyle\sum_{(\varphi,A)\in\mathcal{E}'_k(L,\beta)} \mathbb{1}[I_i(\beta) \subseteq A] \cdot \psi_k(\varphi, A) - e^{\varepsilon} \cdot \sum_{(-\varphi,A)\in\mathcal{E}'_k(L,\beta)} \mathbb{1}[I_i(\beta) + \varphi \subseteq A] \cdot \psi_{k-\varphi/\beta}(\varphi, A) \\
& \qquad + \underline{c}_i(\beta) \cdot \underline{w}_k(\beta) \quad \forall k \in [K],\ \forall i \in [\pm(L + \Delta f/\beta)], \\
& \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (\mathrm{D}'(L, \beta))
\end{array}$$

where $\mathcal{E}'_k(L, \beta) := [\mathscr{B}(\beta) \cap (\Phi - \underline{\Phi}_k(\beta))] \times \mathcal{F}(L, \beta)$ *for* $\underline{\Phi}_k(\beta) := \inf\{\phi : \phi \in \Phi_k(\beta)\}$, $\underline{c}_i(\beta) := \inf\{c(x) : x \in I_i(\beta)\}$ *and* $\underline{w}_k(\beta) := \inf\{w(\phi) : \phi \in \Phi_k(\beta)\}$, $k \in [K]$ *and* $i \in \mathbb{Z}$.

The large-scale but finite-dimensional linear optimization problem $\mathrm{D}'(L, \beta)$ will serve as a building block to our cutting plane algorithm in Section 4. In terms of optimal values, we have the relationship $\mathrm{D}'(L, \beta) \leq \mathrm{D}'$ for all $L \in \mathbb{N}$ and $\beta > 0$. Similar to Section 2, P' and D' are sandwiched by the finite-dimensional linear optimization problems $\mathrm{P}'(L, \beta)$ and $\mathrm{D}'(L, \beta)$.

We close this section with an analysis of the convergence of the finite-dimensional linear optimization problems $\mathrm{P}'(L, \beta)$ and $\mathrm{D}'(L, \beta)$. Similarly to Section 2, recall that by our earlier assumption, β divides Δf, which allows us to equivalently represent β as $\Delta f/k$ for some $k \in \mathbb{N}$.

Theorem 2. *For any* $\xi > 0$, *there is* $\Lambda' \in \mathbb{N}$ *and* $k' \in \mathbb{N}$ *such that*

$$\mathrm{P}'(\Lambda \cdot k, \Delta f/k) - \mathrm{D}'(\Lambda \cdot k, \Delta f/k) \leq \xi \qquad \forall \Lambda \geq \Lambda',\ \forall k \geq k'.$$

Intuitively, Theorem 2 has a similar interpretation as Theorem 1, that is, both the discretization granularity $\Delta f/k$ needs to shrink *and* the support $[-\Lambda \cdot \Delta f, (\Lambda + 1/k) \cdot \Delta f)$ of the noise distributions needs to grow for the primal and dual approximations to converge. Here, we additionally observe that shrinking the granularity also results in a larger number of distributions to be optimized over.

4 Iterative Solution of the Bounding Problems

The bounding problems of Sections 2 and 3 employ a uniform partitioning of the noise distribution γ. Motivated by our numerical experiments, which indicate that (near-)optimal noise distributions tend to combine steep peaks around 0 with gradually declining tails, Section 4.1

extends our bounding problems to non-uniform partitions of γ. Non-uniform partitions allow us to compute noise distributions with similar expected losses in shorter computation times.

Unfortunately, even under a non-uniform partitioning the bounding problems cannot be solved monolithically with an off-the-shelf solver due to their exponential scaling in the problem parameters. Instead, Section 4.2 proposes a cutting plane technique that solves those bounding problems iteratively through an increasingly accurate sequence of relaxations. At the heart of our cutting plane technique is the identification of the constraints that our incumbent solutions violate with the largest margins. While a naïve search for these constraints would require an exponential effort, our algorithm scales polynomially in the size of the problem description.

4.1 Upper and Lower Bounding Problems with Non-Uniform Partitions

Recall that our bounding problems $\mathrm{P}(L,\beta)$ and $\mathrm{D}(L,\beta)$ from Section 2 partition the support of the noise distribution γ into $2L+1$ uniform intervals $I_i(\beta) = [i\cdot\beta, (i+1)\cdot\beta)$, $i \in [\pm L]$. Let $\boldsymbol{\pi} \in \{-L,\ldots,L+1\}^{N+1}$ be an index vector satisfying

$$-L = \pi_1 \;<\; \pi_2 \;<\; \ldots \;<\; \pi_N \;<\; \pi_{N+1} = L+1,$$

and let $\Pi_j(\beta) = [\pi_j\cdot\beta, \pi_{j+1}\cdot\beta)$, $j\in[N]$, denote the j-th interval induced by the consecutive elements of $\boldsymbol{\pi}\cdot\beta$. Consider a variant of our upper bounding problem $\mathrm{P}(L,\beta)$ that enforces equality of all decision variables $p(i)$ and $p(i')$, $i,i'\in[\pm L]$, that satisfy $\pi_j \le i,i' < \pi_{j+1}$ for some $j\in[N]$. The revised upper bounding problem is equivalent to

$$\begin{array}{ll}\underset{p}{\text{minimize}} & \displaystyle\sum_{j\in[N]} c_j(\boldsymbol{\pi},\beta)\cdot p(j) \\ \text{subject to} & p:[N]\mapsto\mathbb{R}_+,\ \displaystyle\sum_{j\in[N]} p(j) = 1 \\ & \displaystyle\sum_{j\in[N]} p(j)\cdot\frac{|A\cap\Pi_j(\beta)|}{|\Pi_j(\beta)|} \le e^{\varepsilon}\cdot\sum_{j\in[N]} p(j)\cdot\frac{|(A-\varphi)\cap\Pi_j(\beta)|}{|\Pi_j(\beta)|} + \delta \\ & \qquad\qquad \forall(\varphi,A)\in\mathcal{E}(L,\beta),\end{array} \tag{$\mathrm{P}(\boldsymbol{\pi},\beta)$}$$

where $c_j(\boldsymbol{\pi},\beta) := |\Pi_j(\beta)|^{-1}\cdot\int_{x\in\Pi_j(\beta)} c(x)\,\mathrm{d}x$. The new upper bound $\mathrm{P}(\boldsymbol{\pi},\beta)$ closely resembles the previous bound $\mathrm{P}(L,\beta)$, except that the objective coefficients c_j and the intervals $\Pi_j(\beta)$ in the DP constraints now reflect the new partitioning of γ.

In a similar fashion, we propose the revised lower bound

$$\begin{array}{ll}\underset{p}{\text{minimize}} & \displaystyle\sum_{j\in\mathfrak{N}} \underline{c}_j(\boldsymbol{\pi},\beta)\cdot p(j) \\ \text{subject to} & p:\mathfrak{N}\mapsto\mathbb{R}_+,\ \displaystyle\sum_{j\in\mathfrak{N}} p(j) = 1 \\ & \displaystyle\sum_{j\in\mathfrak{N}} p(j)\cdot\frac{|A\cap\Pi_j(\beta)|}{|\Pi_j(\beta)|} \le e^{\varepsilon}\cdot\sum_{j\in\mathfrak{N}} p(j)\cdot\frac{|(A-\varphi)\cap\Pi_j(\beta)|}{|\Pi_j(\beta)|} + \delta \\ & \qquad\qquad \forall(\varphi,A)\in\mathcal{E}(L,\beta),\end{array} \tag{$\mathrm{D}(\boldsymbol{\pi},\beta)$}$$

where $\underline{c}_j(\boldsymbol{\pi},\beta) := \inf\{c(x) : x\in\Pi_j(\beta)\}$ and the index set $\mathfrak{N} = \{-\frac{\Delta f}{\beta}+1,\ldots,N+\frac{\Delta f}{\beta}\}$ emerges from the previous index set $[N]$ by padding it at both ends with $\Delta f/\beta$ additional elements whose

Algorithm 1: *Cutting plane algorithm for problem* $\mathrm{P}(\boldsymbol{\pi}, \beta)$

input : $\boldsymbol{\pi}$, β, Δf
output: optimal solution $p^\star$ to problem $\mathrm{P}(\boldsymbol{\pi}, \beta)$
Initialize $\mathcal{S} = \emptyset$;
do
  Let $p^\star$ be an optimal solution to the relaxation of $\mathrm{P}(\boldsymbol{\pi}, \beta)$ that only contains the privacy constraints indexed by $(\varphi, A) \in \mathcal{S}$.
  Find a constraint $(\varphi^\star, A^\star)$ with maximum privacy shortfall under the incumbent solution $p^\star$.
  if *constraint* $(\varphi^\star, A^\star)$ *has positive privacy shortfall* **then** update $\mathcal{S} = \mathcal{S} \cup (\varphi^\star, A^\star)$.
while $\mathcal{S}$ *has been updated*;
return $p^\star$.

associated interval indices are set to $\pi_{1-t} = \pi_1 - t$ and $\pi_{N+1+t} = \pi_{N+1} + t$, $t = 1, \ldots, \Delta f/\beta$, with the intervals $\Pi_j(\beta)$ extended to $j \in \mathfrak{N} \setminus [N]$ in the obvious way. Again, the revised lower bound closely resembles the previous bound $\mathrm{D}(L, \beta)$, with minor changes in the objective coefficients $\underline{c}_j$ and the intervals $\Pi_j(\beta)$. The revised bounding problems enjoy convergence properties akin to those from Section 2.

Corollary 1. *For any* $\beta > 0$ *and* $\boldsymbol{\pi}$ *satisfying* $-L = \pi_1 < \ldots < \pi_{N+1} = L + 1$ *for some* $L \in \mathbb{N}$, *we have* $\mathrm{P}(\boldsymbol{\pi}, \beta) \geq \mathrm{P} = \mathrm{D} \geq \mathrm{D}(\boldsymbol{\pi}, \beta)$. *Moreover, for any* $\xi > 0$ *there is* $\Lambda' \in \mathbb{N}$ *and* $k' \in \mathbb{N}$ *such that* $\mathrm{P}(\boldsymbol{\pi}, \beta) - \mathrm{D}(\boldsymbol{\pi}, \beta) \leq \xi$ *for any* $\boldsymbol{\pi}$ *whose induced partition* $\{\Pi_j(\beta)\}_{j \in [N]}$ *is a refinement of a uniform partition* $\{I_i(\Delta f/k)\}_{i \in [\pm \Lambda \cdot k]}$ *with* $\Lambda \geq \Lambda'$ *and* $k \geq k'$.

Appendix C.3 presents analogous bounding problems for the data dependent case.

4.2 Cutting Plane Algorithm

Although the revised upper bounding problem $\mathrm{P}(\boldsymbol{\pi}, \beta)$ only contains N decision variables, it remains challenging to solve monolithically since it comprises $\mathcal{O}(2^L \cdot \Delta f/\beta)$ DP constraints. To address this issue, Algorithm 1 solves a sequence of relaxations of $\mathrm{P}(\boldsymbol{\pi}, \beta)$ that only involve those constraints that are active at incumbent solutions. In particular, every iteration of Algorithm 1 determines a constraint $(\varphi^\star, A^\star)$ with maximum *privacy shortfall*, which is the quantity that the DP constraints in $\mathrm{P}(\boldsymbol{\pi}, \beta)$ require to be non-positive:

$$V(\varphi, A) = \sum_{j \in [N]} p(j) \cdot \frac{|A \cap \Pi_j(\beta)|}{|\Pi_j(\beta)|} - e^\varepsilon \cdot \sum_{j \in [N]} p(j) \cdot \frac{|A \cap (\Pi_j(\beta) + \varphi)|}{|\Pi_j(\beta)|} - \delta \quad \text{for } (\varphi, A) \in \mathcal{E}(L, \beta). \tag{9}$$

Since $\mathrm{P}(\boldsymbol{\pi}, \beta)$ contains finitely many constraints, one readily recognizes that Algorithm 1 determines an optimal solution to $\mathrm{P}(\boldsymbol{\pi}, \beta)$ in finitely many iterations.

Observation 1. *Algorithm 1 terminates after a finite number of iterations with an optimal solution* $p^\star$ *to problem* $\mathrm{P}(\boldsymbol{\pi}, \beta)$.

A key step in Algorithm 1 is the identification of a constraint $(\varphi^\star, A^\star) \in \mathcal{E}(L, \beta)$ with maximum privacy shortfall. A naïve implementation of this step would require the evaluation of

Algorithm 2: *Identification of a constraint in* $\mathrm{P}(\boldsymbol{\pi}, \beta)$ *with maximum privacy shortfall*

input : $\boldsymbol{\pi},\ \beta,\ p,\ \Delta f$
output: constraint $(\varphi^\star, A^\star)$ with maximum privacy shortfall $V(\varphi^\star, A^\star)$
Initialize $V^\star = 0$;
for $\varphi \in \big\{(\pi_j - \pi_{j'}) \cdot \beta \ :\ (\pi_j - \pi_{j'}) \cdot \beta \in [-\Delta f, \Delta f] \text{ and } j, j' \in [N]\big\} \cup \{-\Delta f, \Delta f\}$ **do**
  Initialize $A = \emptyset$ and $V = 0$;
  for $j = 1, \ldots, N$ **do**
    Let $A_j = \Pi_j(\beta) \setminus [-L \cdot \beta + \varphi, (L+1) \cdot \beta + \varphi)$ and update

$$A = A \cup A_j, \quad V = V + |A_j| \cdot \frac{p(j)}{|\Pi_j(\beta)|}.$$

    for $j' = 1, \ldots, N$ **do**
      if $p(j)/|\Pi_j(\beta)| > e^\varepsilon \cdot p(j')/|\Pi_{j'}(\beta)|$ **then**
        Let $A_{jj'} = \Pi_j(\beta) \cap (\Pi_{j'}(\beta) + \varphi)$ and update

$$A = A \cup A_{jj'}, \quad V = V + |A_{jj'}| \cdot \left[\frac{p(j)}{|\Pi_j(\beta)|} - e^\varepsilon \cdot \frac{p(j')}{|\Pi_{j'}(\beta)|}\right].$$

      end
    end
  end
  if $V > V^\star$ **then**
    Update $\varphi^\star = \varphi$, $A^\star = A$ and $V^\star = V$.
  end
end
return $(\varphi^\star, A^\star)$ and $V^\star(\varphi, A) = V^\star - \delta$.

$\mathcal{O}(2^L \cdot \Delta f/\beta)$ privacy shortfalls in time $\mathcal{O}(N)$ each. Instead, we employ Algorithm 2 to identify a constraint $(\varphi^\star, A^\star)$ with maximum privacy shortfall in polynomial time.

Proposition 7. *For a fixed solution p, Algorithm 2 can be implemented so as to return a constraint of* $\mathrm{P}(\boldsymbol{\pi}, \beta)$ *with maximum privacy shortfall in time* $\mathcal{O}(N^3)$.

To illustrate the intuition behind Algorithm 2 and Proposition 7, fix any $\varphi \in \mathscr{B}(\beta)$ in problem $\mathrm{P}(\boldsymbol{\pi}, \beta)$. To construct the DP constraint $(\varphi, A) \in \mathcal{E}(L, \beta)$ with maximum privacy shortfall across all $A \in \mathcal{F}(L, \beta)$, we need to decide for each interval $I_i(\beta) = [i \cdot \beta, (i+1) \cdot \beta)$, $i \in [\pm L]$, whether or not to include the interval $I_i(\beta)$ in A. To this end, we first observe that we can include all intervals $I_i(\beta)$ for which $I_i(\beta) \cap (\Pi_{j'}(\beta) + \varphi) = \emptyset$ for all $j' \in [N]$ since the inclusion of those intervals in A cannot decrease $V(\varphi, A)$. For the intervals $I_i(\beta)$ that satisfy both $I_i(\beta) \subseteq \Pi_j(\beta)$ and $I_i(\beta) \subseteq (\Pi_{j'}(\beta) + \varphi)$ for some $j, j' \in [N]$, on the other hand, we compare the magnitude of the positive coefficient $p(j)$ with that of the negative coefficient $-e^\varepsilon \cdot p(j')$ in $V(\varphi, A)$ to decide whether $I_i(\beta)$ should be included in A. Algorithm 2 and Proposition 7 refine this idea by *(i)* iterating only over the intervals $\Pi_j(\beta)$, $j \in [N]$, as opposed to the larger set of intervals $I_i(\beta)$, $i \in [\pm L]$; *(ii)* identifying the relevant interval pairs $(j, j') \in [N]^2$ in linear time $\mathcal{O}(N)$ as opposed to quadratic time $\mathcal{O}(N^2)$; and *(iii)* restricting the search over $\varphi \in \mathscr{B}(\beta)$ with cardinality $\mathcal{O}(\Delta f/\beta)$ to the smaller set in the outer for-loop with cardinality $\mathcal{O}(N^2)$.

This section focused on the cutting plane algorithm for the upper bounding problem $\mathrm{P}(\boldsymbol{\pi}, \beta)$

in the data independent setting. Algorithms 1 and 2 immediately extend to the lower bounding problem $\mathrm{D}(\boldsymbol{\pi}, \beta)$ if we replace the objective coefficients c_j with $\underline{c}_j$ and extend the domain of the decisions p from $[N]$ to $\mathfrak{N}$. Both algorithms also readily extend to the data dependent setting (*cf.* Section 3), where a constraint with maximum privacy shortfall can be identified in time $\mathcal{O}(K^2 \cdot N)$. For the sake of brevity, we relegate the details of that algorithm variant to the GitHub repository accompanying this paper.

5 Numerical Experiments

Our numerical experiments are split into two parts. The first part compares the privacy-accuracy trade-off of our optimization-based approach with popular DP mechanisms from the literature, and it examines the runtime of our cutting plane algorithm as well as the convergence of our bounding problems. Since this part focuses on the quality of our bounds as well as their computation times, we use synthetic instances that give us complete control over all parameters. The second part of our experiments investigates whether the improved accuracy on synthetic instances carries over to a better in-sample and out-of-sample performance in machine learning problems involving standard benchmark instances. To this end, we study differentially private variants of the naïve Bayes classifier and a proximal coordinate descent method for logistic regression.

Our optimization algorithms are implemented in C++ and use the GUROBI 9.5.2 LP solver. The machine learning algorithms from the second part are implemented in Julia, and all data is processed using Python 3. All experiments are conducted on Intel Xeon 2.66GHz cluster processors with 16GB memory in single-core and single-thread mode. All sourcecodes and datasets, together with more detailed descriptions of our numerical experiments, are available open-source on the GitHub repository accompanying this work.[1]

5.1 Synthetic Experiments

In our first experiment, we compare the privacy-accuracy trade-off of our optimization-based DP scheme with that of popular benchmark mechanisms from the literature. To this end, we consider the data independent noise optimization problem and select 100 combinations of $\varepsilon \in [0.005, 5]$ and $\delta \in [0.005, 0.75]$. We intentionally chose conservative combinations of ε and δ (*cf.* Table 1 of Zhao et al. (2019)); larger values of ε yield results that are more favorable to our algorithm. We solve our upper and lower bounding problems $\mathrm{P}(\boldsymbol{\pi}, \beta)$ and $\mathrm{D}(\boldsymbol{\pi}, \beta)$ with $\boldsymbol{\pi}$ and β set appropriately so that their relative optimality gaps, measured as $100\% \cdot (\mathrm{P}(\boldsymbol{\pi}, \beta) - \mathrm{D}(\boldsymbol{\pi}, \beta))/\mathrm{D}(\boldsymbol{\pi}, \beta)$, are strictly less than 1% (with a median gap of 0.28% across our experiments). We then use the midpoint $O := (\mathrm{P}(\boldsymbol{\pi}, \beta) - \mathrm{D}(\boldsymbol{\pi}, \beta))/2$ of both bounds as a substitute to the optimal privacy-accuracy trade-off, and we measure the suboptimality of different benchmark mechanisms from the literature: the analytic Gaussian (Balle and Wang 2018) and the truncated Laplace (Geng et al. 2020) mechanisms as upper bounds and the 'near optimal lower bound' of Geng and Viswanath (2015, Thm 8) and Geng et al. (2020) as lower bound. Table 1 records the optimality gaps of the best performing upper and lower bounds B_{UB} and B_{LB} from the literature, which consistently turn out to be the truncated Laplace mechanism and the lower bound of Geng et al. (2020). The

[1] `https://github.com/selvi-aras/DP-via-DRO`

ε \ δ	**0.005**	**0.010**	**0.020**	**0.050**	**0.100**	**0.200**	**0.250**	**0.300**	**0.500**	**0.750**
0.005	1.87%	1.21%	1.20%	6.10%	18.53%	3.55%	49.59%	23.18%	49.96%	33.34%
0.010	2.89%	1.42%	8.00%	2.32%	17.08%	3.08%	49.18%	23.03%	49.92%	33.35%
0.020	2.84%	2.83%	0.56%	15.09%	14.19%	67.44%	48.39%	22.74%	49.83%	33.37%
0.050	1.85%	2.11%	2.57%	18.36%	6.22%	66.03%	46.15%	21.95%	49.58%	33.42%
0.100	4.78%	4.18%	8.68%	19.37%	33.32%	63.85%	42.87%	20.85%	49.17%	33.50%
0.200	10.43%	9.60%	8.90%	17.29%	19.82%	60.17%	37.70%	19.36%	48.33%	33.63%
0.500	23.10%	21.41%	25.45%	16.36%	38.70%	42.56%	29.52%	18.48%	45.85%	33.79%
1.000	40.11%	39.73%	40.23%	40.41%	28.62%	32.53%	26.63%	21.43%	41.80%	33.36%
2.000	33.96%	34.18%	33.45%	31.63%	32.46%	28.70%	27.12%	25.67%	34.35%	30.51%
5.000	19.32%	19.31%	19.29%	19.23%	19.15%	19.01%	18.94%	18.88%	18.65%	19.06%

Table 1: *Suboptimality of the best performing benchmark mechanisms on synthetic data independent instances with $\Delta f = 1$, ℓ_1-loss and various combinations of ε and δ.*

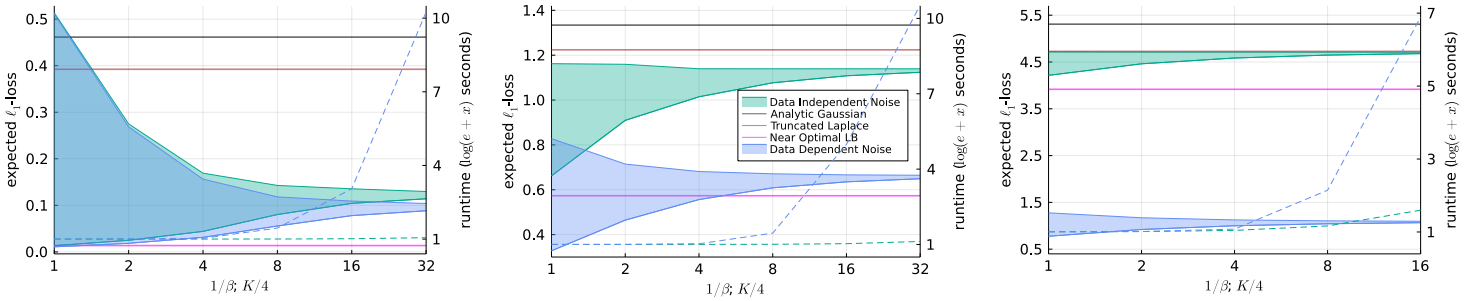

Figure 2: *Comparison of our optimization-based DP schemes with benchmark approaches from the literature on instances with $\Delta f = 2$, ℓ_1-loss and $|\Phi| = 4$ in low-privacy ($\varepsilon = 5$ and $\delta = 0.25$; left), medium-privacy ($\varepsilon = 1$ and $\delta = 0.2$; middle) and high-privacy ($\varepsilon = 0.2$ and $\delta = 0.05$; right) regimes. All computation times are median values over 10 repetitions.*

optimality gaps are reported as $100\% \cdot [(B_{\mathrm{UB}} - B_{\mathrm{LB}})/\max\{O, 1\}]$. A breakdown into separate suboptimalities incurred by the upper and lower bounds is presented in Appendix D. The table shows that the suboptimality of the benchmark approaches increases with ε and δ, and the optimality gaps are significant in most of the considered privacy regimes.[2] We note that if we replace $\max\{O, 1\}$ with O in the denominator of the optimality gap formula, then the gaps in Table 1 increase to more than 700% for $(\varepsilon, \delta) = (5, 0.75)$.

Our second experiment investigates the runtime and the convergence of our optimization-based upper and lower bounds in the data independent and data dependent settings. To this end, we consider three privacy regimes: a low-privacy setting with $(\varepsilon, \delta) = (5, 0.25)$, a medium-privacy setting with $(\varepsilon, \delta) = (1, 0.2)$, and a high-privacy setting with $(\varepsilon, \delta) = (0.2, 0.05)$. As in the

[2]The vigilant reader will observe that the suboptimality is not entirely monotone in ε and δ. This is due to two factors: the computation of the 'near optimal lower bound' involves a non-monotonic parameter rounding, and we approximate the optimal privacy-acurracy trade-off with the midpoint O of our upper and lower bounds.

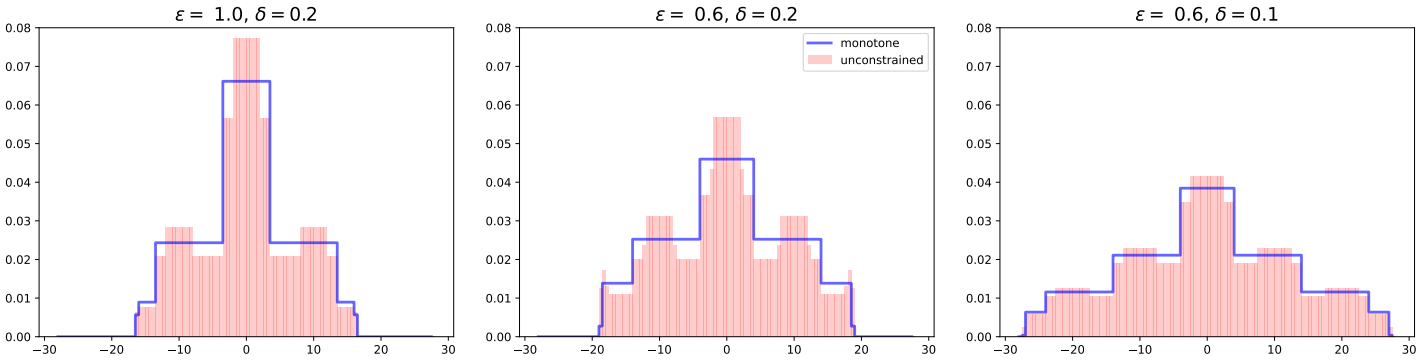

Figure 3: *Optimization-based noise distributions for synthetic data independent instances with* $\Delta f = 10$, $\beta = 0.5$, ℓ_1*-loss and various combinations of* ε *and* δ*. The unconstrained distributions are shown in red shading, whereas the best monotone distributions are shown as blue lines.*

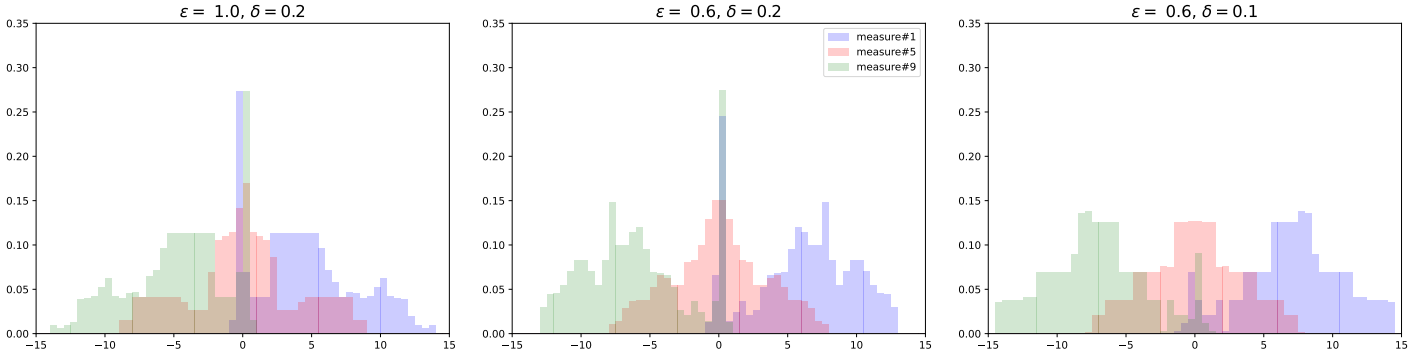

Figure 4: *Optimization-based noise distributions for synthetic data dependent instances with* $\Delta f = 10$, $\beta = 0.5$, ℓ_1*-loss and various combinations of* ε *and* δ*. The set* $\Phi = [0, 18)$ *of query outputs has been partitioned into* 9 *intervals of equal length, resulting in* 9 *noise distributions.*

previous experiment, we compare our optimization-based DP schemes with the analytic Gaussian and the truncated Laplace mechanisms as upper bounds and the 'near optimal lower bound' as lower bound. In our DP schemes, we match the support of the truncated Laplace distribution (with the support bounds rounded to nearest integer values) and compute a hierarchy of refined upper and lower bounds by selecting $\beta \in \{1, 1/2, \ldots, 1/32\}$ and—in the data dependent case—$K \in \{4, 8, \ldots, 128\}$. The results are presented in Figure 2. The figure confirms that the truncated Laplace mechanism is asymptotically optimal among all data independent DP schemes in high-privacy settings. However, the figure also reveals that the truncated Laplace mechanism can be substantially outperformed by our optimization-based data independent noise mechanism in low- and medium-privacy regimes, while it is dominated by our optimization-based data dependent noise mechanism in high-privacy regimes. For low-privacy settings, the difference between our data independent and dependent mechanisms is negligible, but it becomes substantial in medium- and high-privacy regimes, where our data dependent mechanisms significantly outperform the data independent ones. The figure also reveals the computational price to be paid for optimal noise distributions. While the data independent problems were all solved within 2.2 secs, the data dependent problems are more challenging: across all instances, it took up to 5.82 secs (980.44 secs) to reduce the gap between our upper and lower bounds to 10% (5%).

Our final synthetic experiment analyzes the shapes of the noise distributions γ obtained by

our data independent and data dependent noise optimization problems. For clarity of exposition, we present results for uniform partitions $\{I_i(\beta)\}_{i\in[\pm L]}$ of the noise realizations as well as the range Φ of query outputs. We emphasize, however, that better results can normally be obtained by non-uniform partitions that combine finer discretizations around 0 with coarser discretizations of the tails. Figure 3 visualizes optimization-based data independent noise distributions for different privacy regimes. We make multiple observations. Firstly, the optimal noise distribution may not be monotone. Indeed, we verified in separate experiments that the lower bounds of the best monotone noise distributions can strictly exceed the upper bounds of the best non-monotone noise distributions (details are available in the GitHub repository). This is noteworthy as all of the benchmark DP mechanisms employ monotone noise distributions, and monotonicity assumptions are commonly made in the literature without scrutinizing their impact on optimality. Secondly, the shapes of the optimization-based noise distributions are non-trivial, and they appear to depend on the problem parameters in a non-trivial manner. Finally, we note that the shapes of the optimization-based noise distributions differ with the loss function; in particular, asymmetric loss functions (such as the pinball loss) result in asymmetric noise distributions (details are relegated to the GitHub repository). We take our last two observations as an indication of the inherent complexity of optimal noise distributions, which emphasizes the need for optimization-based approaches as opposed to closed-form solutions. Figure 4 reports optimization-based data dependent noise distributions for different privacy regimes. In addition to the previous remarks, which continue to apply to the data dependent setting, we additionally observe that the noise distributions corresponding to different intervals of the query output range Φ differ in non-trivial ways. Again, this confirms our belief that the superior privacy-accuracy trade-offs achieved by optimization-based noise distributions are unlikely to be matched by DP mechanisms relying on closed-form expressions or the tuning of a small number of hyperparameters.

5.2 Differentially Private Naïve Bayes Classification

Given a dataset $(\boldsymbol{x}^i, y^i)_{i=1}^n$ with feature vectors $\boldsymbol{x}^i$ comprising numerical features x_v^i, $v \in \mathcal{V}_{\text{num}}$, and/or categorical features x_v^i, $v \in \mathcal{V}_{\text{cat}}$, as well as categorical outputs $y^i \in \mathcal{C}$, the naïve Bayes classifier employs the class-conditional independence assumption to predict the output $c^\star$ corresponding to the feature values $\boldsymbol{x} = \boldsymbol{\chi}$ of a new sample as the label $c \in \mathcal{C}$ that maximizes

$$\mathbb{P}[y = c \mid \boldsymbol{x}] = \frac{\mathbb{P}[y = c] \cdot \prod_{v\in\mathcal{V}_{\text{num}}\cup\mathcal{V}_{\text{cat}}} \mathbb{P}[x_v = \chi_v \mid y = c]}{\mathbb{P}[\boldsymbol{x}]}.$$

The naïve Bayes classifier replaces the unknown probabilities $\mathbb{P}[y = c]$ and $\mathbb{P}[x_v = \chi_v \,|\, y = c]$, $v \in \mathcal{V}_{\text{cat}}$, with their empirical frequencies in the dataset $(\boldsymbol{x}^i, y^i)_{i=1}^n$, and it makes a normality assumption to replace $\mathbb{P}[x_v = \chi_v \,|\, y = c]$, $v \in \mathcal{V}_{\text{num}}$, with an empirical density in the dataset $(\boldsymbol{x}^i, y^i)_{i=1}^n$ using the empirical means $\mu_{\{x_v|y=c\}}$ and standard deviations $\sigma_{\{x_v|y=c\}}$. We refer to Hastie et al. (2009) for a detailed description of the naïve Bayes classifier.

We follow Vaidya et al. (2013) and Lopuhaä-Zwakenberg et al. (2021) to construct a differentially private naïve Bayes classifier. Assuming that the number of training samples n is public knowledge, the only data-related information used by our classifier is the number $n_{\{y=c\}}$ of training samples with label $c \in \mathcal{C}$, the number $n_{\{x_v=\chi_v\wedge y=c\}}$ of training samples with label

UCI dataset descriptions					In-sample errors					Out-of-sample errors				
Dataset	n	$\lvert\mathcal{V}_{\text{num}}\rvert$	$\lvert\mathcal{V}_{\text{cat}}\rvert$	$\lvert\mathcal{C}\rvert$	GN	TLN	OPT	NB	*Imp*	GN	TLN	OPT	NB	*Imp*
post-operative	86	1	7	2	36.94% (33.41%)	32.42%	**31.43%** (29.61%)	25.65%	*14.62%*	41.28% (39.61%)	39.07%	**38.30%** (37.03%)	35.12%	*19.53%*
adult	45,222	5	8	2	38.78% (21.98%)	21.73%	**20.49%** (18.84%)	17.37%	*28.25%*	38.82% (22.04%)	21.80%	**20.56%** (18.91%)	17.45%	*28.28%*
breast-cancer	683	0	9	2	2.45% (2.23%)	2.20%	**2.17%** (2.14%)	2.12%	*36.80%*	3.82% (3.57%)	3.54%	**3.51%** (3.48%)	3.46%	*36.75%*
contraceptive	1,473	2	7	3	58.84% (52.84%)	52.38%	**51.32%** (50.46%)	49.14%	*32.64%*	59.53% (54.22%)	53.76%	**52.89%** (52.19%)	51.08%	*32.55%*
dermatology	366	2	32	6	1.68% (1.03%)	0.91%	**0.90%** (0.60%)	0.49%	*1.07%*	35.96% (35.61%)	35.53%	**35.52%** (35.28%)	35.25%	*1.33%*
cylinder-bands	539	19	14	2	40.95% (35.63%)	34.69%	**33.98%** (31.46%)	23.43%	*6.35%*	41.83% (37.20%)	36.42%	**35.81%** (33.70%)	26.89%	*6.40%*
annealing	898	6	18	5	13.65% (7.93%)	7.47%	**7.39%** (7.31%)	7.32%	*51.56%*	14.23% (8.38%)	7.89%	**7.80%** (7.69%)	7.70%	*44.41%*
spect	160	0	22	2	26.46% (25.99%)	25.89%	**25.78%** (25.77%)	25.67%	*47.48%*	29.34% (28.77%)	28.66%	**28.59%** (28.57%)	28.50%	*43.74%*
bank	45,211	4	12	2	13.98% (12.40%)	12.35%	**12.32%** (12.24%)	12.13%	*14.20%*	14.05% (12.48%)	12.43%	**12.40%** (12.33%)	12.22%	*14.26%*
abalone	4,177	7	1	2	39.25% (31.85%)	31.36%	**30.64%** (28.75%)	26.46%	*14.77%*	39.42% (32.14%)	31.66%	**30.95%** (29.09%)	26.86%	*14.76%*
spambase	4,601	57	0	2	39.40% (37.09%)	35.50%	**34.99%** (31.56%)	18.09%	*2.91%*	39.26% (36.97%)	35.41%	**34.91%** (31.54%)	18.26%	*2.90%*
ecoli	336	5	2	2	48.09% (31.40%)	27.00%	**26.51%** (18.76%)	3.97%	*2.15%*	48.33% (31.81%)	27.44%	**26.95%** (19.28%)	4.52%	*2.12%*
absent	740	12	8	2	45.95% (30.22%)	28.24%	**28.01%** (26.56%)	24.56%	*2.60%*	47.14% (33.99%)	32.26%	**32.03%** (30.73%)	29.18%	*2.59%*

Table 2: *In-sample and out-of-sample errors of our optimization-based differentially private naïve Bayes classifier as well as several DP mechanisms from the literature on UCI datasets. Bold printing highlights the smallest errors obtained across all data independent DP mechanisms.*

$c \in \mathcal{C}$ whose categorical feature $v \in \mathcal{V}_{\text{cat}}$ attains value χ_v, as well as the conditional empirical means $\mu_{\{x_v|y=c\}}$ and standard deviations $\sigma_{\{x_v|y=c\}}$ of the numerical features $v \in \mathcal{V}_{\text{num}}$. The post processing property (Dwork and Roth 2014, Proposition 2.1) then allows us to design a differentially private naïve Bayes classifier by perturbing these statistics according to their individual sensitivities and using the perturbed statistics to classify new samples. Since $n_{\{y=c\}}$ and $n_{\{x_v=\chi_v \wedge y=c\}}$, $v \in \mathcal{V}_{\text{cat}}$, are simple counting queries, they can change by at most 1 among any two neighboring datasets. For the numerical features $v \in \mathcal{V}_{\text{num}}$, we assume given upper and lower bounds u_v and l_v for the feature values. In this case, the value of x_v can differ by at most $u_v - l_v$ for any two neighboring datasets, which implies that the sensitivity of $\psi_{\{x_v|y=c\}}$ and $\sigma_{\{x_v|y=c\}}$ is bounded from above by $(u_v - l_v)/n_{\{y=c\}}$ and $(u_v - l_v)/\sqrt{n_{\{y=c\}}}$, respectively. We note that $\mu_{\{x_v|y=c\}}$ and $\sigma_{\{x_v|y=c\}}$ satisfy our surjectivity assumption (*cf.* Assumption 2 in Section 2 and Assumption 3 in Section 3), whereas the assumption is violated by the count queries $n_{\{y=c\}}$ and $n_{\{x_v=\chi_v \wedge y=c\}}$. Thus, our optimization-based approaches provide feasible but potentially overly conservative noise distributions.

Table 2 presents the in-sample and out-of-sample errors of our optimization-based differentially private naïve Bayes classifier using the ℓ_1-loss in a data independent ("OPT") as well as data dependent ("(OPT)") setting on the most popular UCI classification datasets (Dua and Graff 2017). The table also compares our results with diffentially private naïve Bayes classifiers employing a Gaussian noise ("GN"), an analytic Gaussian noise ("(GN)") and a truncated Laplace noise ("TLN"), as well as the classical (non-private) naïve Bayes classifier ("NB"). We fix $(\epsilon, \delta) = (1, 0.1)$ in all DP mechanisms. The reported errors are mean errors over 100 random splits of the datasets into training sets (80% of the data) and test sets (20% of the data) as well as, for each split, 1,000 simulations of all differentially private naïve Bayes implementations. The column 'Imp' records the percentage of the gap between NB and the best method from the literature (which turns out to be TLN) that is closed by OPT. The table reveals that our optimization-based data independent and data dependent noise distributions consistently outperform the considered competitors. To see whether this outperformance is statistically significant, we computed the p-values of a t-test with a null hypothesis that the second best approach is as good as OPT. The t-test averages the 1,000 simulated errors for each training set-test set split and considers the differences of the 100 averages corresponding to different training set-test set

splits (Salzberg 1997). In all experiments, the p-values are less than 10^{-7}, except for dermatology where the p-value is 10^{-1}. Further details on the experimental setting as well as the applied the t-tests are relegated to the GitHub repository.

5.3 Differentially Private Proximal Coordinate Descent

Given a dataset $(\boldsymbol{x}^i, y^i)_{i=1}^n$ with feature vectors $\boldsymbol{x}^i \in \mathbb{R}^d$ comprising numerical and/or categorical features as well as binary outputs $y^i \in \{-1, +1\}$, the ℓ_1-regularized logistic regression assumes that $\mathbb{P}[y \,|\, \boldsymbol{x} = \boldsymbol{\chi}] = [1 + \exp(-y \cdot \boldsymbol{h}^{0\top}\boldsymbol{\chi})]^{-1}$ for some unknown hyperplane $\boldsymbol{h}^0 \in \mathbb{R}^d$, and it determines a hyperplane $\boldsymbol{h}^\star \in \mathbb{R}^d$ that minimizes the empirical logistic loss

$$\frac{1}{n} \cdot \sum_{i=1}^{n} \log(1 + \exp(-y^i \cdot \boldsymbol{h}^\top \boldsymbol{x}^i)) + \lambda \cdot ||\boldsymbol{h}||_1,$$

where $\lambda > 0$ is a hyperparameter. Subsequently, the output of a new sample with feature values $\boldsymbol{x} = \boldsymbol{\chi}$ is predicted to be the label $y \in \{-1, +1\}$ that maximizes $[1 + \exp(-y \cdot \boldsymbol{h}^\top \boldsymbol{\chi})]^{-1}$.

To solve the logistic regression problem, proximal coordinate descent (Friedman et al. 2010, Richtárik and Takáč 2014) starts at a randomly selected initial solution $\boldsymbol{h}^0$ and conducts $t = 1, \ldots, T$ iterations, $T \in \mathbb{N}$, each of which applies the proximal operator to a random subset $i_1^t, \ldots, i_K^t \in [d]$ of the components via

$$h_l^{t,k} = \text{prox}_{\lambda|\cdot|}\left(h_l^{t,k-1} - \frac{1}{n} \cdot \left[\sum_{i=1}^{n} \frac{\exp(-y^i \cdot \boldsymbol{h}^{t,k\top}\boldsymbol{x}^i)}{1 + \exp(-y^i \cdot \boldsymbol{h}^{t,k\top}\boldsymbol{x}^i)} \cdot (-y^i \cdot x_l^i)\right]\right) \tag{10}$$

if $l = i_k^t$, and $h_l^{t,k} = h_l^{t,k-1}$ otherwise, for all $k = 1, \ldots, K$. Here, we fix $\boldsymbol{h}^{t,0} = \boldsymbol{h}^{t-1}$, and $\text{prox}_{\lambda|\cdot|}(\cdot)$ denotes the proximal operator (Parikh et al. 2014) associated with the ℓ_1-regularization:

$$\text{prox}_{\lambda|\cdot|}(w_j) := \underset{v \in \mathbb{R}}{\arg\min} \left\{\frac{1}{2} \cdot (w_j - v)^2 + \lambda \cdot |v|\right\} = \begin{cases} w_j - \lambda & \text{if } w_j \geq \lambda \\ w_j + \lambda & \text{if } w_j \leq -\lambda \\ 0 & \text{if } |w_j| \leq \lambda \end{cases}$$

Upon completion of the K applications of the proximal operator in iteration t, we set $\boldsymbol{h}^t = (1/K) \cdot \sum_{k=1}^K \boldsymbol{h}^{t,k}$ and continue with iteration $t + 1$. The algorithm terminates with $\boldsymbol{h}^T$ as an approximately optimal solution to the regularized logistic regression problem.

We follow Mangold et al. (2022) to construct a differentially private proximal coordinate descent method for the logistic regression problem. The only data-related information used by our algorithm is contained in the proximal updates (10). Assuming that the feature vectors $\boldsymbol{x}^i$ are normalized so that $\|\boldsymbol{x}^i\|_\infty \leq 1,\ i \in [n]$, we have

$$\sum_{i=1}^{n} \underbrace{\frac{\exp(-y^i \cdot \boldsymbol{h}^{t,k\top}\boldsymbol{x}^i)}{1 + \exp(-y^i \cdot \boldsymbol{h}^{t,k\top}\boldsymbol{x}^i)}}_{\in(0,1)} \cdot \underbrace{(-y^i \cdot x_l^i)}_{\in[-1,+1]} \in (-n, n),$$

and thus the sensitivity of this summation, which is determined by the maximum change achievable by modifying a single training sample $i \in [n]$, is 2. The post processing property (Dwork and

UCI dataset descriptions			In-sample errors					Out-of-sample errors				
Dataset	n	d	GN	TLN	OPT	PCD	*Imp*	GN	TLN	OPT	PCD	*Imp*
post-operative	86	14	40.23% (34.96%)	33.75%	**33.37%** (30.67%)	27.31%	*5.94%*	45.67% (42.99%)	42.08%	**41.86%** (41.22%)	35.56%	*3.38%*
adult	45,222	57	19.77% (19.77%)	19.75%	**19.73%** (19.67%)	19.75%	*573.31%*	19.79% (19.79%)	19.77%	**19.75%** (19.69%)	19.77%	*640.17%*
breast-cancer	683	26	4.60% (4.36%)	4.36%	**4.34%** (3.93%)	4.31%	*28.05%*	4.75% (4.52%)	4.51%	**4.50%** (4.09%)	4.46%	*22.25%*
contraceptive	1,473	18	38.52% (38.30%)	38.29%	**38.27%** (37.39%)	38.21%	*19.28%*	39.90% (39.71%)	39.70%	**39.69%** (38.93%)	39.64%	*20.17%*
dermatology	366	98	18.09% (14.56%)	14.23%	**14.14%** (7.95%)	13.14%	*8.07%*	21.38% (18.21%)	17.93%	**17.85%** (11.96%)	16.95%	*8.17%*
cylinder-bands	539	63	30.27% (28.74%)	28.65%	**28.59%** (25.30%)	28.20%	*12.96%*	32.65% (31.36%)	31.29%	**31.24%** (28.56%)	30.90%	*14.65%*
annealing	898	42	16.70% (16.27%)	16.30%	**16.29%** (14.74%)	16.20%	*6.16%*	17.52% (17.10%)	17.13%	**17.13%** (15.61%)	17.03%	*0.83%*
spect	160	23	28.89% (23.62%)	22.74%	**22.57%** (18.60%)	19.61%	*5.23%*	31.47% (27.19%)	26.37%	**26.24%** (23.58%)	23.71%	*4.84%*
bank	45,211	44	12.20% (12.20%)	12.20%	**12.20%** (11.69%)	12.22%	*25.43%*	12.21% (12.21%)	12.21%	**12.21%** (11.71%)	12.23%	*20.31%*
abalone	4,177	10	27.45% (27.44%)	27.44%	**27.43%** (27.34%)	27.43%	*87.76%*	27.53% (27.52%)	27.52%	**27.51%** (27.43%)	27.51%	*79.42%*
spambase	4,601	58	39.25% (39.26%)	39.23%	**39.21%** (38.79%)	39.26%	*65.85%*	39.60% (39.60%)	39.57%	**39.55%** (39.13%)	39.61%	*41.07%*
ecoli	336	8	9.38% (7.29%)	7.09%	**7.01%** (6.31%)	6.52%	*14.81%*	9.88% (7.73%)	7.53%	**7.45%** (6.70%)	6.96%	*13.67%*
absent	740	70	33.77% (32.88%)	32.78%	**32.74%** (29.45%)	32.54%	*17.16%*	35.63% (34.83%)	34.74%	**34.68%** (31.95%)	34.52%	*26.34%*
colon-cancer	62	2,000	18.72% (10.85%)	9.20%	**8.62%** (0.03%)	0.00%	*6.34%*	32.09% (31.63%)	31.62%	**31.54%** (30.41%)	30.67%	*7.77%*

Table 3: *In-sample and out-of-sample errors of our optimization-based differentially private logistic classifier as well as several DP mechanisms from the literature on UCI datasets. Bold printing highlights the lowest errors obtained across all data independent DP mechanisms.*

Roth 2014, Proposition 2.1) then allows us to design a differentially private proximal coordinate descent method by perturbing the sum inside the proximal updates (10) accordingly.

Table 3 presents the in-sample and out-of-sample errors of our optimization-based differentially private proximal coordinate descent algorithm in a data independent ("OPT") as well as data dependent ("(OPT)") setting for $T = 100$ iterations and $K = \lceil d/4 \rceil$ proximal updates per iteration, regularization parameter $\lambda = 10^{-8}$ and $(\varepsilon, \delta) = (1, 0.1)$. While our data independent algorithm minimizes the expected ℓ_1-loss, our data dependent algorithm performs much better under a handcrafted loss function that resembles the ℓ_1-loss in the vicinity of the origin but has steep slopes of -1,000 and 1,000 for large negative and positive values away from the origin, respectively. The large slopes penalize switching the sign of gradient, which tends to slow down convergence (recall from Figure 4 that noise distributions associated with negative values tend to have large probability mass on the positive side and vice versa). We use the same datasets as in Section 5.2, but we *(i)* convert non-binary output labels into binary ones via binning (if the output is ordinal) or distinguishing the majority class from all other classes (if the output is nominal) and *(ii)* apply one-hot encoding for the nominal input features. Additionally, as the proximal coordinate descent method is commonly used for datasets where $d \gg n$, we also include the colon-cancer dataset that is available in LIBSVM (Chang and Lin 2011). As in the previous section, we compare our optimization-based algorithms with differentially private proximal coordinate descent methods employing a Gaussian noise ("GN"), an analytic Gaussian noise ("(GN)") and a truncated Laplace noise ("TLN"), as well as the classical (non-private) proximal coordinate descent scheme ("PCD"). As before, the reported errors are mean errors over 100 random splits of the datasets into training sets (80% of the data) and test sets (20% of the data) as well as, for each split, 1,000 simulations of all differentially private algorithm implementations. The column 'Imp' records the percentage of the gap between PCD and the best method from the literature that is closed by OPT. As in the experiment from the previous section, our optimization-based data independent and data dependent noise distributions consistently outperform the considered competitors. The p-values of a t-test similar to the previous section were always smaller than 10^{-7}, except for the annealing, breast-cancer and contraceptive datasets, where the p-values are 10^{-4}, and except for the bank dataset, where there is no significance. Further details on the experimental setting can be found in the GitHub repository. Interestingly, we observe that on

all datasets except for post-operative, our data dependent differentially private classifier (OPT) outperforms the non-private classifier PCD. This result is due to our handcrafted loss function (see above), and it does not hold for ℓ_1- and ℓ_2-losses. Effects of this types have been observed previously in gradient-based learning algorithms (see, *e.g.*, Neelakantan et al. 2015), and they further highlight the benefits of an optimization-based approach towards DP, which readily caters for non-standard loss functions that can be tuned towards the task at hand.

6 Conclusions

With the widespread adoption of analytics, privacy concerns have witnessed a remarkable resurgence in the public discourse. While DP has established itself as a predominant privacy paradigm in both academic research and industrial practice, the existing DP mechanisms almost exclusively focus on privacy to the detriment of accuracy. The few methodological studies on the privacy-accuracy trade-off focus on asymptotic performance and/or restricted classes of mechanisms.

In our view, the privacy-accuracy trade-off is most naturally studied through the lens of optimization theory, which gives rise to infinite-dimensional DP mechanism design problems that are similar to—but at the same time notedly distinct from—continuous linear programs and distributionally robust optimization problems. We developed a hierarchy of converging upper and lower bounds on these problems that result in DP mechanisms with rigorous privacy guarantees and deterministic bounds on their accuracy. Our numerical results demonstrate that our mechanisms can achieve significant improvements on both synthetic and real-world problems.

A key advantage of an optimization-based DP approach is its versatility. Our upper and lower bounds can be readily extended to incorporate monotonicity and symmetry constraints as well as a bounded range for the query output in the case of data dependent mechanisms. We can account for different loss functions, and multiple loss functions can be combined in a multi-objective framework. Our approach also allows us to incorporate tighter bounds on the probability of distinguishing events $A \in \mathcal{F}$, $\mathbb{P}[\mathcal{A}(D) \in A] > 0$ and $\mathbb{P}[\mathcal{A}(D') \in A] = 0$ for some $(D, D') \in \mathcal{N}$, that enable an adversary to exclude certain databases altogether (Dwork and Rothblum 2016, Remark 1.3).

We regard this work as a first step towards optimization-based DP, and it opens up several avenues for future research. Firstly, while our approach generalizes to multi-dimensional query functions f via the composition theorem, better results can be expected if one directly optimizes over multi-dimensional noise distributions. Since a naïve implementation of our algorithms would scale exponentially in that dimension, this may necessitate the development of further approximations. Secondly, our work measures accuracy through a simple loss function. In many interesting practical applications, the query output may be the solution to an optimization problem, and in those cases accuracy may be best measured in terms of the expected performance of the perturbed output (*e.g.*, the expected total discounted reward of the perturbed policy in the context of differentially private Markov decision processes). Finally, DP as currently defined in the literature is tailored to the traditional noise distributions such as the Laplace and the Gaussian mechanisms. The versatility of an optimization-based view on DP allows us to explore other, potentially more general notions of differential privacy as well.

Acknowledgments

Financial support from the EPSRC grants EP/R045518/1 and EP/W003317/1 is gratefully acknowledged. The authors thank Damien Desfontaines, Daniel Kuhn and Juba Ziani for insightful discussions. The first author acknowledges the support of The Alan Turing Institute's Enrichment Scheme.

References

Abadi M, Chu A, Goodfellow I, McMahan HB, Mironov I, Talwar K, Zhang L (2016) Deep learning with differential privacy. *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, 308–318.

Alvim M S, Andres M E, Chatzikokolakis K, Degano P, Palamidessi C (2011) Differential privacy: On the trade-off between utility and information leakage. *Proceedings of the 8th International Workshop on Formal Aspects of Security & Trust*, 39–54.

Anderson EJ, Nash P (1987) *Linear Programming in Infinite-Dimensional Spaces: Theory and Applications* (John Wiley & Sons).

Apple Differential Privacy Team (2017) Learning with privacy at scale. White paper.

Asi H, Duchi JC (2020a) Instance-optimality in differential privacy via approximate inverse sensitivity mechanisms. *Advances in Neural Information Processing Systems*, volume 33, 14106–14117.

Asi H, Duchi JC (2020b) Near instance-optimality in differential privacy. *arXiv preprint 2005.10630* .

Balle B, Wang YX (2018) Improving the Gaussian mechanism for differential privacy: Analytical calibration and optimal denoising. *Proceedings of the 35th International Conference on Machine Learning*, 394–403.

Bavadekar S, Boulanger A, Davis J, Desfontaines D, Gabrilovich E, Gadepalli K, Ghazi B, Griffith T, Gupta JP, Kamath C, Kraft D, Kumar R, Kumok A, Mayer Y, Manurangsi P, Patankar A, Perera IM, Scott C, Shekel T, Miller B, Smith K, Stanton C, Sun M, Young M, Wellenius G (2021) Google COVID-19 vaccination search insights: Anonymization process description. *arXiv preprint 2107.01179* .

Berg M, Kreveld M, Overmars M, Schwarzkopf OC (2000) *Computational Geometry: Algorithms and Applications* (Springer), 2nd edition.

Cai N, Kou S (2019) Econometrics with privacy preservation. *Operations Research* 67(4):905–926.

Canonne CL, Kamath G, Steinke T (2020) The discrete Gaussian for differential privacy. *Advances in Neural Information Processing Systems*, volume 33, 15676–15688.

Chang CC, Lin CJ (2011) LIBSVM: A library for support vector machines. *ACM transactions on Intelligent Systems and Technology* 2(3):1–27.

Chaudhuri K, Monteleoni C (2009) Privacy-preserving logistic regression. *Advances in Neural Information Processing Systems*, volume 21, 289–296.

Chaudhuri K, Monteleoni C, Sarwate AD (2011) Differentially private empirical risk minimization. *Journal of Machine Learning Research* 12(3):1069–1109.

Chen X, Miao S, Wang Y (2023) Differential privacy in personalized pricing with nonparametric demand models. *Operations Research* 71(2):581–602.

Chen X, Simchi-Levi D, Wang Y (2022) Privacy-preserving dynamic personalized pricing with demand learning. *Management Science* 68(7):4878–4898.

Delage E, Ye Y (2010) Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research* 58(3):596–612.

Desfontaines D (2020) *Lowering the Cost of Anonymization*. Ph.D. thesis, ETH Zurich.

Desfontaines D, Pejó B (2020) SoK: Differential privacies. *Proceedings on Privacy Enhancing Technologies*, 288–313.

Dinur I, Nissim K (2003) Revealing information while preserving privacy. *Proceedings of the 22nd ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, 202–210.

Dua D, Graff C (2017) UCI machine learning repository. `http://archive.ics.uci.edu/ml`.

Dwork C (2011) The promise of differential privacy: A tutorial on algorithmic techniques. *Proceedings of the 52nd Annual IEEE Symposium on Foundations of Computer Science*, 1–2.

Dwork C, Kenthapadi K, McSherry F, Mironov I, Naor M (2006a) Our data, ourselves: Privacy via distributed noise generation. *Proceedings of the 25th Annual International Conference on the Theory and Applications of Cryptographic Techniques*, 486–503.

Dwork C, McSherry F, Nissim K, Smith A (2006b) Calibrating noise to sensitivity in private data analysis. *Proceedings of the 3rd Conference on Theory of Cryptography*, 265–284.

Dwork C, Roth A (2014) The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science* 9(3–4):211–407.

Dwork C, Rothblum GN (2016) Concentrated differential privacy. *arXiv preprint 1603.01887* .

Foote AD, Machanavajjhala A, McKinney K (2019) Releasing earnings distributions using differential privacy. *Journal of Privacy and Confidentiality* 9(2):1–19.

Friedman A, Schuster A (2010) Data mining with differential privacy. *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 493–502.

Friedman J, Hastie T, Tibshirani R (2010) Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* 33(1):1–22.

Geng Q, Ding W, Guo R, Kumar S (2019) Optimal noise-adding mechanism in additive differential privacy. *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics*, 11–20.

Geng Q, Ding W, Guo R, Kumar S (2020) Tight analysis of privacy and utility tradeoff in approximate differential privacy. *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics*, 89–99.

Geng Q, Viswanath P (2014) The optimal mechanism in differential privacy. *Proceedings of the IEEE International Symposium on Information Theory*, 2371–2375.

Geng Q, Viswanath P (2015) Optimal noise adding mechanisms for approximate differential privacy. *IEEE Transactions on Information Theory* 62(2):952–969.

Han S, Tao M, Topcu U, Owhadi H, Murray RM (2015) Convex optimal uncertainty quantification. *SIAM Journal on Optimization* 25(3):1368–1387.

Han S, Topcu U, Pappas GJ (2016) Differentially private distributed constrained optimization. *IEEE Transactions on Automatic Control* 62(1):50–64.

Hanasusanto GA, Roitch V, Kuhn D, Wiesemann W (2015) A distributionally robust perspective on uncertainty quantification and chance constrained programming. *Mathematical Programming* 151(1):35–62.

Hastie T, Tibshirani R, Friedman J (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer).

Heffetz O, Ligett K (2014) Privacy and data-based research. *Journal of Economic Perspectives* 28(2):75–98.

Hsu J, Huang Z, Roth A, Wu ZS (2016) Jointly private convex programming. *Proceedings of the 27th Annual ACM-SIAM Symposium on Discrete Algorithms*, 580–599.

Hsu J, Roth A, Roughgarden T, Ullman J (2014) Privately solving linear programs. *Proceedings of the 41st International Colloquium on Automata, Languages, and Programming*, 612–624.

Johnson A, Shmatikov V (2013) Privacy-preserving data exploration in genome-wide association studies. *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1079–1087.

Kuhn D, Mohajerin Esfahani P, Nguyen VA, Shafieezadeh-Abadeh S (2019) Wasserstein distributionally robust optimization: Theory and applications in machine learning. *INFORMS TutORials in Operations Research*, 130–166.

Lopuhaä-Zwakenberg M, Alishahi M, Kivits J, Klarenbeek J, van der Velde GJ, Zannone N (2021) Comparing classifiers' performance under differential privacy. *Proceedings of the International Conference on Security and Cryptography*, 50–61.

Machanavajjhala A, Kifer D, Gehrke J, Venkitasubramaniam M (2007) ℓ-diversity: Privacy beyond k-anonymity. *ACM Transactions on Knowledge Discovery from Data* 1(1):1556–4681.

Mangasarian OL (2011) Privacy-preserving linear programming. *Optimization Letters* 5(1):165–172.

Mangold P, Bellet A, Salmon J, Tommasi M (2022) Differentially private coordinate descent for composite empirical risk minimization. *Proceedings of the 39th International Conference on Machine Learning*, 14948–14978.

McSherry F, Talwar K (2007) Mechanism design via differential privacy. *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science*, 94–103.

Meiser S (2018) Approximate and probabilistic differential privacy definitions. Cryptology ePrint Archive Preprint 2018/277.

Messing S, DeGregorio C, Hillenbrand B, King G, Mahanti S, Mukerjee Z, Nayak C, Persily N, State B, Wilkins A (2020) Facebook Privacy-Protected Full URLs Data Set. `https://doi.org/10.7910/DVN/TDOAPG`.

Neelakantan A, Vilnis L, Le QV, Sutskever I, Kaiser L, Kurach K, Martens J (2015) Adding gradient noise improves learning for very deep networks. *arXiv preprint 1511.06807* .

Nergiz ME, Atzori M, Clifton C (2007) Hiding the presence of individuals from shared databases. *Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data*, 665–676.

Nissim K, Raskhodnikova S, Smith A (2007) Smooth sensitivity and sampling in private data analysis. *Proceedings of the 39th Annual ACM Symposium on Theory of Computing*, 75–84.

Owhadi H, Scovel C, Sullivan TJ, McKerns M, Ortiz M (2013) Optimal uncertainty quantification. *SIAM Review* 55(2):271–345.

Parikh N, Boyd S, et al. (2014) Proximal algorithms. *Foundations and Trends® in Optimization* 1(3):127–239.

Pereira M, Kim A, Allen J, White K, Ferres JL, Dodhia R (2021) U.S. broadband coverage data set: A differentially private data release. *arXiv preprint 2103.14035* .

Richtárik P, Takáč M (2014) Iteration complexity of randomized block-coordinate descent methods for minimizing a composite function. *Mathematical Programming* 144(1):1–38.

Rogers R, Cardoso AR, Mancuhan K, Kaura A, Gahlawat N, Jain N, Ko P, Ahammad P (2020) A members first approach to enabling LinkedIn's labor market insights at scale. *arXiv preprint 2010.13981* .

Salzberg SL (1997) On comparing classifiers: Pitfalls to avoid and a recommended approach. *Data Mining and Knowledge Discovery* 1(3):317–328.

Sommer DM (2021) *Fighting Uphill Battles: Improvements in Personal Data Privacy*. Ph.D. thesis, ETH Zurich.

Soria-Comas J, Domingo-Ferrer J (2013) Optimal data-independent noise for differential privacy. *Information Sciences* 250:200–214.

Sweeney L (1997) Weaving technology and policy together to maintain confidentiality. *The Journal of Law, Medicine & Ethics* 25(2-3):98–110.

Sweeney L (2001) *Computational Disclosure Control: A Primer on Data Privacy Protection*. Ph.D. thesis, Massachusetts Institute of Technology.

Sweeney L (2002) k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 10(5):557–570.

Vadhan S (2017) The complexity of differential privacy. *Tutorials on the Foundations of Cryptography*, 347–450 (Springer).

Vaidya J, Shafiq B, Basu A, Hong Y (2013) Differentially private naïve Bayes classification. *Proceedings of the 12th IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technologies*, 571–576.

Wiesemann W, Kuhn D, Sim M (2014) Distributionally robust convex optimization. *Operations Research* 62(6):1358–1376.

Zhao J, Wang T, Bai T, Lam KY, Xu Z, Shi S, Ren X, Yang X, Liu Y, Yu H (2019) Reviewing and improving the Gaussian mechanism for differential privacy. *arXiv preprint 1911.12060* .

Appendices

A Proofs of Section 2

A.1 Proof of Lemma 1

The proof of Lemma 1 relies on three auxiliary results, which we state and prove first. Lemma A.1 shows that under restriction (2), problem P can be expressed entirely in terms of the decision variables p. The resulting problem coincides with P(β) in terms of the decision variables, but it still comprises a larger set of constraints. Lemma A.2 identifies for each query output difference $\varphi \in [-\Delta f, \Delta f]$ a constraint $(\varphi, A^\star(\varphi))$ that weakly dominates all constraints (φ, A), $A \in \mathcal{F}$, and Lemma A.3 identifies a set of values for φ such that the associated constraints $(\varphi, A^\star(\varphi))$ weakly dominate all constraints of the problem. Lemma 1 then combines these results to prove the equivalence between the problems P and P(β) under restriction (2).

Lemma A.1. *Under restriction* (2)*,* P *has the same optimal value as*

$$\begin{array}{ll} \underset{p}{\text{minimize}} & \displaystyle\sum_{i\in\mathbb{Z}} c_i(\beta)\cdot p(i) \\ \text{subject to} & p:\mathbb{Z}\mapsto\mathbb{R}_+,\ \displaystyle\sum_{i\in\mathbb{Z}} p(i)=1 \\ & \displaystyle\sum_{i\in\mathbb{Z}} p(i)\cdot\frac{|A\cap I_i(\beta)|}{\beta} \le e^\varepsilon\cdot\sum_{i\in\mathbb{Z}} p(i)\cdot\frac{|(A-\varphi)\cap I_i(\beta)|}{\beta}+\delta \quad \forall(\varphi,A)\in\mathcal{E}. \end{array} \tag{11}$$

Proof. We use restriction (2) to replace γ in problem P with the new decision variables p. To this end, observe first that under restriction (2), γ affords a density function via

$$\begin{aligned} \gamma(A) = \sum_{i\in\mathbb{Z}} p(i)\cdot\frac{|A\cap I_i(\beta)|}{\beta} &= \sum_{i\in\mathbb{Z}} p(i)\cdot\int_{x\in\mathbb{R}}\frac{\mathbb{1}[x\in A]\cdot\mathbb{1}[x\in I_i(\beta)]}{\beta}\,\mathrm{d}x \\ &= \int_{x\in\mathbb{R}}\mathbb{1}[x\in A]\cdot\left(\sum_{i\in\mathbb{Z}} p(i)\cdot\frac{\mathbb{1}[x\in I_i(\beta)]}{\beta}\right)\mathrm{d}x, \end{aligned}$$

$A\in\mathcal{F}$, where the last step holds by Fubini's theorem since $\gamma(A)\in[0,1]$. This derivation shows that $\sum_{i\in\mathbb{Z}} p(i)\cdot\mathbb{1}[x\in I_i(\beta)]/\beta$ is the density function of γ. Using this density function, the objective function of P can be rewritten as

$$\begin{aligned} \int_{x\in\mathbb{R}} c(x)\,\mathrm{d}\gamma(x) = \int_{x\in\mathbb{R}} c(x)\cdot\left(\sum_{i\in\mathbb{Z}} p(i)\cdot\frac{\mathbb{1}[x\in I_i(\beta)]}{\beta}\right)\mathrm{d}x &= \sum_{i\in\mathbb{Z}}\frac{p(i)}{\beta}\cdot\int_{x\in\mathbb{R}} c(x)\cdot\mathbb{1}[x\in I_i(\beta)]\,\mathrm{d}x \\ &= \sum_{i\in\mathbb{Z}} c_i(\beta)\cdot p(i), \end{aligned}$$

where the second equality holds by Fubini's theorem, and the last step substitutes $c_i(\beta) = \beta^{-1}\cdot\int_{x\in I_i(\beta)} c(x)\mathrm{d}x$ for all $i\in\mathbb{Z}$. The final expression coincides with the objective of problem (11).

In view of the DP constraints in problem P, we note that

$$\int_{x\in\mathbb{R}}\mathbb{1}[x\in A]\,\mathrm{d}\gamma(x) = \gamma(A) = \sum_{i\in\mathbb{Z}} p(i)\cdot\frac{|A\cap I_i(\beta)|}{\beta}$$

as well as

$$\int_{x\in\mathbb{R}} \mathbb{1}[x+\varphi\in A]\,\mathrm{d}\gamma(x) = \gamma(A-\varphi) = \sum_{i\in\mathbb{Z}} p(i)\cdot\frac{|(A-\varphi)\cap I_i(\beta)|}{\beta},$$

and the right-hand side expressions coincide with the expressions on either side of the DP constraints of problem (11). This concludes the proof. □

To reduce the number of DP constraints in problem (11), we first characterize the tightest privacy constraint $(\varphi, A)\in\mathcal{E}$ in (11) for a fixed decision p and a fixed query output difference $\varphi\in[-\Delta f,\Delta f]$. To this end, we define the privacy shortfall as

$$V(\varphi,A) = \sum_{i\in\mathbb{Z}} p(i)\cdot|A\cap I_i(\beta)| - e^{\varepsilon}\cdot\sum_{i\in\mathbb{Z}} p(i)\cdot|(A-\varphi)\cap I_i(\beta)|.$$

Note that $V(\varphi,A)$ coincides with the slack of the DP constraint (φ,A) in problem (11), shifted by $-\delta$ and scaled by β.[3] In particular, maximizers $(\varphi,A)\in\mathcal{E}$ of the privacy shortfall V correspond to the tightest constraints in problem (11).

Observation A.1. *The privacy shortfall is linear over partitions of A, that is, we have $V(\varphi,A) = \sum_\ell V(\varphi,A_\ell)$ for any partition $\{A_\ell\}_\ell$ of A.*

We next show that for any fixed $\varphi\in[-\Delta f,\Delta f]$, the largest privacy shortfall $\sup\{V(\varphi,A) : A\in\mathcal{F}\}$ is attained by a worst-case event $A^\star(\varphi)\in\mathcal{F}$ of a simple structure.

Definition 1. *For any $i\in\mathbb{Z}$, let $j := \lceil i-\varphi/\beta\rceil$ be the unique integer satisfying $\varphi + j\cdot\beta\in I_i(\beta)$, and define $I_i^1(\varphi,\beta)$ and $I_i^2(\varphi,\beta)$ as the following partition of $I_i(\beta)$:*

(i) $I_i^1(\varphi,\beta) := I_i(\beta)\cap(I_{j-1}(\beta)+\varphi) = [i\cdot\beta, \varphi + j\cdot\beta)$

(ii) $I_i^2(\varphi,\beta) := I_i(\beta)\setminus I_i^1(\varphi,\beta) = I_i(\beta)\cap(I_j(\beta)+\varphi) = [\varphi+j\cdot\beta, (i+1)\cdot\beta)$.

For any $i\in\mathbb{Z}$, Definition 1 implies that $I_i^1(\varphi,\beta)\cap I_{i'}(\beta) = \emptyset$ for all $i'\neq i$. Moreover, since $I_i^1(\varphi,\beta)\subseteq I_{j-1}(\beta)+\varphi$ it also follows that $(I_i^1(\varphi,\beta)-\varphi)\subseteq I_{j-1}(\beta)$ and therefore $(I_i^1(\varphi,\beta)-\varphi)\cap I_{j'}(\beta) = \emptyset$ for all $j'\neq j-1$. Applying a similar reasoning also to $I_i^2(\varphi,\beta)$ allows us to simplify the expressions for the privacy shortfall over subsets of $I_i^1(\varphi,\beta)$ and $I_i^2(\varphi,\beta)$.

Observation A.1. *For any $i\in\mathbb{Z}$ and $\varphi\in[-\Delta f,\Delta f]$ we have*

$$V(\varphi,A) = \begin{cases} |A|\cdot(p(i) - e^{\varepsilon}\cdot p(j-1)) & \text{for all } A\subseteq I_i^1(\varphi,\beta), \\ |A|\cdot(p(i) - e^{\varepsilon}\cdot p(j)) & \text{for all } A\subseteq I_i^2(\varphi,\beta), \end{cases}$$

where $j = \lceil i-\varphi/\beta\rceil$.

Figure 5 illustrates the intuition underlying Observation A.1. We now show that there is always a worst-case event $A^\star(\varphi)$ that constitutes a union of intervals $I_i^1(\varphi,\beta)$ and $I_i^2(\varphi,\beta)$, $i\in\mathbb{Z}$.

[3] Our definition here differs slightly from a later definition of privacy shortfall in Section 4.2. The context will always make it clear which definition is being used.

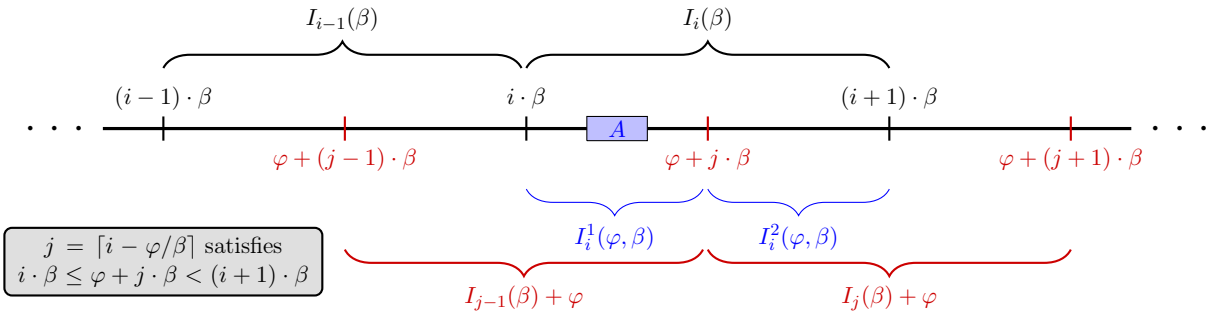

Figure 5: *The interval $I_i^1(\varphi,\beta)$ satisfies $I_i^1(\varphi,\beta)\subseteq I_i(\beta)$ as well as $I_i^1(\varphi,\beta)\subseteq I_{j-1}(\beta)+\varphi$. Therefore, for any $A\subseteq I_i^1(\varphi,\beta)$ we have $|A\cap I_i(\beta)|=|A|$ and $|A\cap I_{i'}(\beta)|=0$ for all $i'\neq i$; similarly, $|(A-\varphi)\cap I_{j-1}(\beta)|=|A|$ and $|(A-\varphi)\cap I_{j'-1}(\beta)|=0$ for all $j'\neq j$. This shows the first case in Observation A.1; the second case can be verified analogously.*

Lemma A.2. *For any $\varphi\in[-\Delta f,\Delta f]$, there is an event*

$$A^\star(\varphi)=\bigcup_{i\in\mathcal{I}_1} I_i^1(\varphi,\beta)\cup\bigcup_{i\in\mathcal{I}_2} I_i^2(\varphi,\beta)\in\mathcal{F}\qquad \text{for some } \mathcal{I}_1,\mathcal{I}_2\subseteq\mathbb{Z} \tag{12}$$

that attains the largest privacy shortfall $\sup\{V(\varphi,A):A\in\mathcal{F}\}$.

Proof. By Definition 1, $I_i^1(\varphi,\beta)$ and $I_i^2(\varphi,\beta)$ partition $I_i(\beta)$ for any $i\in\mathbb{Z}$, and thus $\{I_i^1(\varphi,\beta)\cup I_i^2(\varphi,\beta)\}_{i\in\mathbb{Z}}$ partitions $\mathbb{R}$. Observation A.1 and the sub-additivity of the supremum operator then imply that

$$\begin{aligned}\sup_{A\in\mathcal{F}}\{V(\varphi,A)\}&=\sup_{A\in\mathcal{F}}\left\{\sum_{i\in\mathbb{Z}}V(\varphi,A\cap I_i^1(\varphi,\beta))+\sum_{i\in\mathbb{Z}}V(\varphi,A\cap I_i^2(\varphi,\beta))\right\}\\&\le\sum_{i\in\mathbb{Z}}\sup_{A\subseteq I_i^1(\varphi,\beta)}\{V(\varphi,A)\}+\sum_{i\in\mathbb{Z}}\sup_{A\subseteq I_i^2(\varphi,\beta)}\{V(\varphi,A)\}.\end{aligned}$$

We will show that each supremum on the right-hand side of the inequality is attained and then construct $A^\star(\varphi)=\bigcup_{i\in\mathbb{Z}}A_i^1(\varphi)\cup\bigcup_{i\in\mathbb{Z}}A_i^2(\varphi)$, where $A_i^1(\varphi),A_i^2(\varphi)\in\mathcal{F}$ are defined as

$$A_i^1(\varphi)\in\operatorname*{arg\,max}_{A\subseteq I_i^1(\varphi,\beta)}\{V(\varphi,A)\}\quad\text{and}\quad A_i^2(\varphi)\in\operatorname*{arg\,max}_{A\subseteq I_i^2(\varphi,\beta)}\{V(\varphi,A)\}.$$

The statement then follows from the fact that $A^\star(\varphi)\in\mathcal{F}$ by construction.

In view of $A_i^1(\varphi)$, Observation A.1 implies that $V(\varphi,A)$ is maximized by

$$A_i^1(\varphi)=\begin{cases}I_i^1(\varphi,\beta) & \text{if } p(i)-e^\varepsilon\cdot p(j-1)>0,\\ \emptyset & \text{otherwise.}\end{cases}$$

Applying a similar reasoning to $A_i^2(\varphi)$, we observe that

$$A_i^2(\varphi)=\begin{cases}I_i^2(\varphi,\beta) & \text{if } p(i)-e^\varepsilon\cdot p(j)>0,\\ \emptyset & \text{otherwise.}\end{cases}$$

The statement of the lemma thus follows. □

The next result shows that $V(\varphi, A^\star(\varphi))$ is maximized by $\varphi^\star = k \cdot \beta$ for some $k \in \mathbb{Z}$.

Lemma A.3. *The function $\varphi \mapsto V(\varphi, A^\star(\varphi))$ is affine over each interval $[k\cdot\beta, (k+1)\cdot\beta]$, $k \in \mathbb{Z}$.*

Proof. For any $k \in \mathbb{Z}$, the construction of $A^\star(\varphi)$ in the proof of Lemma A.2 implies that the sets $\mathcal{I}_1$ and $\mathcal{I}_2$ in the statement of the lemma coincide for all $\varphi \in [k \cdot \beta, (k+1) \cdot \beta) = I_k(\beta)$. Therefore, for all $\varphi \in I_k(\beta)$ we have

$$\begin{aligned} V(\varphi, A^\star(\varphi)) &= \sum_{i\in\mathcal{I}_1} V(\varphi, I_i^1(\varphi,\beta)) + \sum_{i\in\mathcal{I}_2} V(\varphi, I_i^2(\varphi,\beta)) \\ &= \sum_{i\in\mathcal{I}_1} |I_i^1(\varphi,\beta)| \cdot (p(i) - e^\varepsilon \cdot p(j-1)) + \sum_{i\in\mathcal{I}_2} |I_i^2(\varphi,\beta)| \cdot (p(i) - e^\varepsilon \cdot p(j)) \\ &= (\varphi \bmod \beta) \cdot \sum_{i\in\mathcal{I}_1} (p(i) - e^\varepsilon \cdot p(j-1)) + (\beta - (\varphi \bmod \beta)) \cdot \sum_{i\in\mathcal{I}_2} (p(i) - e^\varepsilon \cdot p(j)) \\ &= (\varphi \bmod \beta) \cdot \left[\sum_{i\in\mathcal{I}_1} (p(i) - e^\varepsilon \cdot p(j-1)) - \sum_{i\in\mathcal{I}_2} (p(i) - e^\varepsilon \cdot p(j)) \right] + \beta \cdot \sum_{i\in\mathcal{I}_2} (p(i) - e^\varepsilon \cdot p(j)), \end{aligned} \tag{13}$$

where $j = \lceil i - \varphi/\beta \rceil$ as specified by Definition 1. Here, the first equality follows from Lemma A.2 and Observation A.1, the second equality is due to Observation A.1, the third equality holds since

$$|I_i^1(\varphi,\beta)| = \varphi + \lceil i - \varphi/\beta \rceil \cdot \beta - i \cdot \beta = \varphi - \beta \cdot \lfloor \varphi/\beta \rfloor = (\varphi \bmod \beta)$$

and $|I_i^2(\varphi,\beta)| = \beta - |I_i^1(\varphi,\beta)| = \beta - (\varphi \bmod \beta)$. In the final expression, all terms except for $(\varphi \bmod \beta)$ are independent of φ, and $\varphi \mapsto (\varphi \bmod \beta)$ is affine over $\varphi \in I_k(\beta)$. We thus conclude that $\varphi \mapsto V(\varphi, A^\star(\varphi))$ is affine over $\varphi \in I_k(\beta)$.

To conclude the proof, we show that the result holds for the closure of $I_k(\beta)$, that is, $\varphi \mapsto V(\varphi, A^\star(\varphi))$ is not discontinuous at $\overline{\varphi} = (k+1) \cdot \beta$. In other words, we show that

$$\lim_{\varphi \to \overline{\varphi}} V(\varphi, A^\star(\varphi)) = V(\overline{\varphi}, A^\star(\overline{\varphi})).$$

To this end, we first note that $(\overline{\varphi} \bmod \beta) = 0$ as well as $j = \lceil i - \overline{\varphi}/\beta \rceil = i - k - 1$, and hence (13) implies that

$$\begin{aligned} V(\overline{\varphi}, A^\star(\overline{\varphi})) = \beta \cdot \sum_{i\in\mathcal{I}_2} (p(i) - e^\varepsilon \cdot p(j)) &= \beta \cdot \sum_{i\in\mathcal{I}_2} (p(i) - e^\varepsilon \cdot p(i-k-1)) \\ &= \beta \cdot \sum_{i\in\mathbb{Z}} \max\{p(i) - e^\varepsilon \cdot p(i-k-1), 0\}. \end{aligned}$$

Since we also have $j = \lceil i - \varphi/\beta \rceil = i - k$ for all $k \cdot \beta \le \varphi < (k+1) \cdot \beta$, it follows that

$$\begin{aligned} \lim_{\varphi\to\overline{\varphi}} V(\varphi, A^\star(\varphi)) &= \lim_{\varphi\to\overline{\varphi}} \Big((\varphi \bmod \beta) \cdot \Big[\sum_{i\in\mathcal{I}_1} (p(i) - e^\varepsilon \cdot p(j-1)) - \sum_{i\in\mathcal{I}_2} (p(i) - e^\varepsilon \cdot p(j)) \Big] + \beta \cdot \sum_{i\in\mathcal{I}_2} (p(i) - e^\varepsilon \cdot p(j)) \Big) \\ &= \Big[\sum_{i\in\mathcal{I}_1} (p(i) - e^\varepsilon \cdot p(j-1)) - \sum_{i\in\mathcal{I}_2} (p(i) - e^\varepsilon \cdot p(j)) \Big] \cdot \lim_{\varphi\to\overline{\varphi}} \{(\varphi \bmod \beta)\} + \beta \cdot \sum_{i\in\mathcal{I}_2} (p(i) - e^\varepsilon \cdot p(j)) \\ &= \beta \cdot \sum_{i\in\mathcal{I}_1} (p(i) - e^\varepsilon \cdot p(j-1)) = \beta \cdot \sum_{i\in\mathcal{I}_1} (p(i) - e^\varepsilon \cdot p(i-k-1)) \end{aligned}$$

$$=\beta \cdot \sum_{i\in\mathbb{Z}} \max\{p(i) - e^{\varepsilon} \cdot p(i-k-1), 0\} = V(\overline{\varphi}, A^\star(\overline{\varphi})),$$

where the first equality follows from (13), the second equality exploits the linearity of limits, the third equality holds since $\lim_{\varphi\to\overline{\varphi}} (\varphi \bmod \beta) = \beta$, and the final equalities follow from substituting $j = i - k$ and using the construction of $\mathcal{I}_1$, which includes all indices $i \in \mathbb{Z}$ for which the incremental privacy shortfall $p(i) - e^{\varepsilon} \cdot p(i-k-1)$ is positive. This shows that $\varphi \mapsto V(\varphi, A^\star(\varphi))$ is not discontinuous at $\overline{\varphi} = (k+1) \cdot \beta$ and therefore concludes the proof. □

We can now prove Lemma 1 by showing that problem (11) in the statement of Lemma A.1 has the same optimal value as problem P(β).

Proof of Lemma 1. First notice that the DP constraints of P(β) can be written as

$$\sum_{i\in\mathbb{Z}} p(i) \cdot \frac{|A \cap I_i(\beta)|}{\beta} \leq e^{\varepsilon} \cdot \sum_{i\in\mathbb{Z}} p(i) \cdot \frac{|(A-\varphi) \cap I_i(\beta)|}{\beta} + \delta \quad \forall (\varphi, A) \in \mathcal{E}(\beta)$$

since for any $(\varphi, A) \in \mathcal{E}(\beta)$ we have $|A \cap I_i(\beta)|/\beta = \mathbb{1}[I_i(\beta) \subseteq A]$ as well as $|(A-\varphi) \cap I_i(\beta)|/\beta = \mathbb{1}[I_i(\beta) + \varphi \subseteq A]$ by definition. This shows that P(β) is a relaxation of problem (11) since $\mathcal{E}(\beta) \subset \mathcal{E}$. Hence, if P($\beta$) is infeasible, then so is problem (11), and the result follows. To complete the proof, we show that any p feasible in P(β) is also feasible in problem (11).

Fix any feasible solution p to P(β), and assume to the contrary that p violates a DP constraint $(\varphi, A) \in \mathcal{E}$ in problem (11). In that case, Lemmas A.2 and A.3 imply that there is a constraint $(\varphi^\star, A^\star(\varphi^\star)) \in \mathcal{E}(\beta)$ with a weakly higher privacy shortfall than (φ, A). Indeed, Lemma A.3 and our earlier assumption that Δf is divisible by β imply that $\varphi^\star$ can without loss of generality be chosen such that $\varphi^\star \in [-\Delta f, \Delta f] \cap \{k \cdot \beta\}_{k\in\mathbb{Z}} = \mathscr{B}(\beta)$. Since such $\varphi^\star$ is a multiple of β, we have $I_i^1(\varphi, \beta) = \emptyset$ and $I_i^2(\varphi, \beta) = I_i(\beta)$ for all $i \in \mathbb{Z}$. Hence, Lemma A.2 implies that $A^\star(\varphi^\star)$ can be chosen such that $A^\star(\varphi^\star) = \bigcup_{i\in\mathcal{I}_2} I_i(\beta)$ for some $\mathcal{I}_2 \subseteq \mathbb{Z}$, which in turn implies that $A^\star(\varphi) \in \mathcal{F}(\beta)$. Thus, there must be a violated DP constraint $(\varphi^\star, A^\star(\varphi^\star))$ such that $\varphi^\star \in \mathscr{B}(\beta)$ and $A^\star(\varphi) \in \mathcal{F}(\beta)$, that is, $(\varphi^\star, A^\star(\varphi^\star)) \in \mathcal{E}(\beta)$. This contradicts our earlier assumption that p is feasible in P(β), and thus the result follows. □

A.2 Proof of Proposition 1

Appending the additional constraint (3) to problem P(β) yields

$$\begin{array}{ll} \underset{p}{\text{minimize}} & \displaystyle\sum_{i\in[\pm L]} c_i(\beta) \cdot p(i) \\ \text{subject to} & p : [\pm L] \mapsto \mathbb{R}_+, \ \displaystyle\sum_{i\in[\pm L]} p(i) = 1 \\ & \displaystyle\sum_{i\in[\pm L]} \mathbb{1}[I_i(\beta) \subseteq A] \cdot p(i) \leq e^{\varepsilon} \cdot \sum_{i\in[\pm L]} \mathbb{1}[I_i(\beta) + \varphi \subseteq A] \cdot p(i) + \delta \quad \forall (\varphi, A) \in \mathcal{E}(\beta). \end{array}$$

The result follows if we show that any DP constraint $(\varphi, A) \in \mathcal{E}(\beta) \setminus \mathcal{E}(L, \beta)$ is redundant in the above problem. To this end, fix any p that satisfies the first constraint. For any $(\varphi, A) \in \mathcal{E}(\beta)$, define $A_L := A \cap [-L \cdot \beta, (L+1) \cdot \beta)$ so that $(\varphi, A_L) \in \mathcal{E}(L, \beta)$. We show that if p satisfies the

DP constraint (φ, A_L), then it also satisfies the DP constraint (φ, A). Indeed, we observe that

$$\begin{aligned}
&\sum_{i\in[\pm L]} \mathbb{1}[I_i(\beta) \subseteq A] \cdot p(i) - e^{\varepsilon} \cdot \sum_{i\in[\pm L]} \mathbb{1}[(I_i(\beta)+\varphi) \subseteq A] \cdot p(i) \\
=&\sum_{i\in[\pm L]} \mathbb{1}[I_i(\beta) \subseteq A_L] \cdot p(i) - e^{\varepsilon} \cdot \sum_{i\in[\pm L]} \mathbb{1}[(I_i(\beta)+\varphi) \subseteq A] \cdot p(i) \\
\leq&\sum_{i\in[\pm L]} \mathbb{1}[I_i(\beta) \subseteq A_L] \cdot p(i) - e^{\varepsilon} \cdot \sum_{i\in[\pm L]} \mathbb{1}[(I_i(\beta)+\varphi) \subseteq A_L] \cdot p(i),
\end{aligned}$$

where the equality follows from

$$\sum_{i\in[\pm L]} \mathbb{1}[I_i(\beta) \subseteq A] \cdot p(i) = \sum_{i\in[\pm L]} \mathbb{1}[I_i(\beta) \subseteq A_L] \cdot p(i) + \sum_{i\in[\pm L]} \underbrace{\mathbb{1}[I_i(\beta) \subseteq A \setminus A_L] \cdot p(i)}_{=0}$$

which holds since no $i \in [\pm L]$ can satisfy $I_i(\beta) \subseteq A \setminus A_L$. The inequality in the third row, on the other hand, follows from $A_L \subseteq A$. Thus, the DP constraints $\mathcal{E}(\beta) \setminus \mathcal{E}(L, \beta)$ are redundant since they are weakly dominated by the DP constraints $(\varphi, A_L) \in \mathcal{E}(L, \beta)$. □

A.3 Proof of Proposition 2

As $(\theta, \psi) \in \mathbb{R} \times \mathcal{M}_+(\mathcal{E})$ is feasible in D and $\delta > 0$, we have that $\int_{(\varphi,A)\in\mathcal{E}} \mathrm{d}\psi(\varphi, A) < \infty$ from the objective function of D, which shows that $\mathcal{E}$ is σ-finite with measure ψ. Moreover, γ is a probability measure on $\mathbb{R}$, and hence it is also σ-finite. We now observe that

$$\begin{aligned}
&\int_{x\in\mathbb{R}} c(x)\,\mathrm{d}\gamma(x) \\
\geq&\int_{x\in\mathbb{R}} \left[\theta - \int_{(\varphi,A)\in\mathcal{E}} \mathbb{1}[x \in A]\,\mathrm{d}\psi(\varphi, A) + e^{\varepsilon} \cdot \int_{(\varphi,A)\in\mathcal{E}} \mathbb{1}[x+\varphi \in A]\,\mathrm{d}\psi(\varphi, A)\right] \mathrm{d}\gamma(x) \\
=&\theta - \int_{(\varphi,A)\in\mathcal{E}} \left[\int_{x\in\mathbb{R}} \mathbb{1}[x \in A]\,\mathrm{d}\gamma(x) - e^{\varepsilon} \cdot \int_{x\in\mathbb{R}} \mathbb{1}[x+\varphi \in A]\,\mathrm{d}\gamma(x)\right] \mathrm{d}\psi(\varphi, A) \\
\geq&\theta - \int_{(\varphi,A)\in\mathcal{E}} \delta\,\mathrm{d}\psi(\varphi, A).
\end{aligned}$$

Here, the first inequality follows from the constraints of problem D. The equality follows from Fubini's theorem, which is applicable since the indicator functions are integrable on $\mathbb{R} \times \mathcal{E}$ with the associated product measure and the fact that $\int_{x\in\mathbb{R}} \mathrm{d}\gamma(x) = 1$. The second inequality follows from the constraints of problem P as well as the fact that ψ is a non-negative measure. □

A.4 Proof of Lemma 2

We first show that under the additional constraint (4), the DP constraints in D reduce to

$$\theta \leq c(x) + \int_{(\varphi,A)\in\mathcal{E}(\beta)} \mathbb{1}[I_i(\beta) \subseteq A]\mathrm{d}\psi(\varphi, A) - e^{\varepsilon} \cdot \int_{(\varphi,A)\in\mathcal{E}(\beta)} \mathbb{1}[I_i(\beta)+\varphi \subseteq A]\mathrm{d}\psi(\varphi, A) \qquad \forall i \in \mathbb{Z},\ \forall x \in I_i(\beta). \quad (14)$$

We then argue that for every $i \in \mathbb{Z}$, all constraints (14) indexed by (i, x), $x \in I_i(\beta)$, are simultaneously satisfied if and only if the DP constraint indexed by i is satisfied in D(β). The

result then follows since both the decision variables and the objective function in D coincide with their counterparts in D(β), restricted to the elements $(\varphi, A) \in \mathcal{E}(\beta)$ as stipulated by (4).

In view of the first step, fix any $x \in \mathbb{R}$, and select $i \in \mathbb{Z}$ such that $x \in I_i(\beta)$. Under the additional constraint (4), the first integral in the DP constraint of D indexed by x reduces to

$$\int_{(\varphi,A)\in\mathcal{E}} \mathbb{1}[x \in A] \mathrm{d}\psi(\varphi, A) = \int_{(\varphi,A)\in\mathcal{E}(\beta)} \mathbb{1}[x \in A] \mathrm{d}\psi(\varphi, A) + \underbrace{\int_{(\varphi,A)\in\mathcal{E}\setminus\mathcal{E}(\beta)} \mathbb{1}[x \in A] \mathrm{d}\psi(\varphi, A)}_{=0}$$
$$= \int_{(\varphi,A)\in\mathcal{E}(\beta)} \mathbb{1}[I_i(\beta) \subseteq A] \mathrm{d}\psi(\varphi, A).$$

Here, the last integral in the first row vanishes due to (4), whereas the second equality holds since for all $(\varphi, A) \in \mathcal{E}(\beta)$, the requirement that $A \in \mathcal{F}(\beta)$ implies that $x \in A$ only if $I_i(\beta) \subseteq A$. Note that the integral in the second row above coincides with the first integral in (14) indexed by (i, x). A similar argument shows that under the additional constraint (4), the second integral in the DP constraint of D indexed by x reduces to the second integral in (14) indexed by (i, x). In summary, under the additional constraint (4) the DP constraints in D indeed reduce to (14).

As for the second step, note that for any fixed $i \in \mathbb{Z}$, the constraints (14) indexed by (i, x), $x \in I_i(\beta)$, only differ in their additive terms $c(x)$. Thus, for any fixed $i \in \mathbb{Z}$, all constraints (14) indexed by (i, x), $x \in I_i(\beta)$, are satisfied if and only if they are satisfied for the smallest value $c(x)$, $x \in I_i(\beta)$, which is precisely what the DP constraint in D(β) indexed by i stipulates. □

A.5 Proof of Proposition 3

First observe that the additional constraint (5) reduces the uncountable set $\mathcal{E}(\beta)$ in the definition of the decision variables as well as the objective function and the constraints of D(β) to the finite subset $\mathcal{E}(L, \beta)$, which allows us to replace the measure $\psi \in \mathcal{M}_+(\mathcal{E}(\beta))$ in D(β) with the discrete map $\psi : \mathcal{E}(L, \beta) \mapsto \mathbb{R}_+$ in D(L, β) as well as replace all integrals in D(β) with sums in D(L, β).

The result now follows if we show that under the additional constraint (5), all DP constraints in D(β) indexed by $i \in \mathbb{Z} \setminus [\pm(L + \Delta f/\beta)]$ are weakly dominated by DP constraints indexed by $i \in [\pm(L + \Delta f/\beta)]$. Indeed, observe that the DP constraints indexed by $i \in \mathbb{Z} \setminus [\pm(L + \Delta f/\beta)]$ simplify to

$$\theta \leq \underline{c}_i(\beta) \quad \forall i \in \mathbb{Z} \setminus [\pm(L + \Delta f/\beta)] \tag{15a}$$

since $\mathbb{1}[I_i(\beta) \subseteq A] = \mathbb{1}[I_i(\beta) + \varphi \subseteq A] = 0$ for all $(\varphi, A) \in \mathcal{E}(L, \beta)$ whenever $i \in \mathbb{Z} \setminus [\pm(L + \Delta f/\beta)]$. In contrast, the constraints indexed by $i \in [\pm(L + \Delta f/\beta)] \setminus [\pm L]$ simplify to

$$\theta \leq -e^{\varepsilon} \cdot \sum_{(\varphi,A)\in\mathcal{E}(L,\beta)} \mathbb{1}[I_i(\beta) + \varphi \subseteq A] \cdot \psi(\varphi, A) + \underline{c}_i(\beta) \tag{15b}$$

since $\mathbb{1}[I_i(\beta) \subseteq A] = 0$ for all $(\varphi, A) \in \mathcal{E}(L, \beta)$ whenever $i \in [\pm(L + \Delta f/\beta)] \setminus [\pm L]$. Note that $\underline{c}_i(\beta)$ inherits monotonicity from $c_i(\beta)$, that is, we have $\underline{c}_i(\beta) \leq \underline{c}_{i+1}(\beta)$ for all $i \geq 0$ as well as $\underline{c}_i(\beta) \leq \underline{c}_{i-1}(\beta)$ for all $i \leq 0$. This property, along with the non-negativity of ψ, shows that the constraints (15a) are implied by constraints (15b), and the result thus follows. □

A.6 Proof of Theorem 1

The proof of Theorem 1 relies on the feasibility and monotonicity of the upper bounding problems $\mathrm{P}(L,\beta)$, which they inherit from the upper bounding problems $\mathrm{P}(\beta)$. We prove these results first in Sections A.6.1 and A.6.2, and we subsequently prove Theorem 1 in Section A.6.3.

A.6.1 Monotonicity and Feasibility of $\mathrm{P}(\beta)$

The upper bound $\mathrm{P}(\beta)$ employs a discretization that is parametrized by β. We first show that the optimal value of this problem is monotonically non-decreasing in β in the following sense.

Lemma A.4. *For any $\varepsilon > 0$, $\delta > 0$ and $\beta > 0$, the optimal value of problem $\mathrm{P}(\beta)$ satisfies $\mathrm{P}(\beta) \geq \mathrm{P}(\beta/k)$ for all $k \in \mathbb{N}$.*

Proof. The result trivially holds if $\mathrm{P}(\beta)$ is infeasible. Assume therefore that $\mathrm{P}(\beta)$ is feasible and fix an arbitrary feasible solution p' in $\mathrm{P}(\beta)$. For any $k \in \mathbb{N}$, problem $\mathrm{P}(\beta/k)$ can be written as

$$\begin{array}{ll} \underset{p}{\text{minimize}} & \displaystyle\sum_{i\in\mathbb{Z}}\sum_{l\in[k]} c_{il}(\beta)\cdot p(i,l) \\ \text{subject to} & p:\mathbb{Z}\times[k]\mapsto\mathbb{R}_+,\ \displaystyle\sum_{i\in\mathbb{Z}}\sum_{l\in[k]} p(i,l)=1 \\ & \displaystyle\sum_{i\in\mathbb{Z}}\sum_{l\in[k]} \mathbb{1}[I_{il}(\beta)\subseteq A]\cdot p(i,l) \leq e^{\varepsilon}\cdot\sum_{i\in\mathbb{Z}}\sum_{l\in[k]} \mathbb{1}[I_{il}(\beta)+\varphi\subseteq A]\cdot p(i,l)+\delta \\ & \qquad\qquad \forall(\varphi,A)\in\mathcal{E}(\beta/k), \\ & \qquad\qquad (\mathrm{P}(\beta/k)) \end{array}$$

where $I_{il}(\beta) := [(i+(l-1)/k)\cdot\beta, (i+l/k)\cdot\beta)$ and $c_{il}(\beta) := (\beta/k)^{-1}\cdot\int_{x\in I_{il}(\beta)} c(x)\,\mathrm{d}x$. One readily verifies that $p''(i,l) = p'(i)/k$, $i\in\mathbb{Z}$ and $l\in[k]$, is feasible in problem $\mathrm{P}(\beta/k)$ and attains the same objective value as p' in $\mathrm{P}(\beta)$. The statement thus follows. □

We next show that problem $\mathrm{P}(\beta)$ is feasible. Our result makes use of the following technical result about non-zero polynomials (*i.e.*, polynomials with at least one non-zero coefficient), which we state and prove first.

Lemma A.5. *For any non-zero polynomial $g:\mathbb{R}\mapsto\mathbb{R}$ and any constant $c\in\mathbb{R}$, we have*

$$\lim_{t\to\infty}\frac{g(t+c)}{g(t)} = 1.$$

Proof. Let d be the degree of g such that $g(t) = a_d t^d + a_{d-1}t^{d-1} + \ldots + a_0$ with $a_d \neq 0$. We have

$$\lim_{t\to\infty}\frac{g(t+c)}{g(t)} = \lim_{t\to\infty}\frac{a_d(t+c)^d + a_{d-1}(t+c)^{d-1} + \ldots + a_0}{a_d t^d + a_{d-1}t^{d-1} + \ldots + a_0} = \lim_{t\to\infty}\frac{a_d t^d + o(t^d)}{a_d t^d + o(t^d)} = \frac{a_d}{a_d} = 1,$$

where $o(\cdot)$ is the standard small-oh notation. Here, the first identity uses the definition of g, the second identity holds since the highest-order term of t in $a_d(t+c)^d$ is $a_d t^d$, and the third identity follows from dividing both the numerator and denominator by t^d and taking the limit. □

Lemma A.6. *For any $\varepsilon > 0$, there is $M\in\mathbb{R}$ such that $\mathrm{P}(\Delta f/k)\leq M$ for all $\delta > 0$ and $k\in\mathbb{N}$.*

Proof. Fix $\varepsilon > 0$ and denote by $\mathrm{P}_0(\Delta f)$ the variant of $\mathrm{P}(\Delta f)$ that replaces δ with 0. We show that there exists $M \in \mathbb{R}$ such that $\mathrm{P}_0(\Delta f) \leq M$. The statement then follows since for any $\delta > 0$ and $\beta = \Delta f/k,\ k \in \mathbb{N}$, we have $\mathrm{P}_0(\Delta f) \geq \mathrm{P}(\Delta f) \geq \mathrm{P}(\beta)$, where the first inequality is direct and the second inequality is due to Lemma A.4.

Theorem 9 of Soria-Comas and Domingo-Ferrer (2013) implies that there exists some $m > 0$ such that the following is a feasible solution to problem $\mathrm{P}_0(\Delta f)$:

$$p(i) = \begin{cases} \Delta f \cdot m & \text{if } i \in \{-1, 0\} \\ \Delta f \cdot m \cdot e^{-(i-1)\cdot\varepsilon} & \text{if } i \geq 1 \\ \Delta f \cdot m \cdot e^{(i+2)\cdot\varepsilon} & \text{if } i \leq -2 \end{cases} \qquad \forall i \in \mathbb{Z}.$$

Let $M = \sum_{i\in\mathbb{Z}} c_i(\Delta f) \cdot p(i)$ be the objective value that is attained by p in problem $\mathrm{P}_0(\Delta f)$. To see that $M < \infty$, we note that

$$M = \Big[\underbrace{\sum_{i\in\{-1,0\}} c_i(\Delta f) \cdot \Delta f \cdot m}_{(i)} \Big] + \Big[\underbrace{\sum_{i\geq 1} c_i(\Delta f) \cdot \Delta f \cdot m \cdot e^{-(i-1)\varepsilon}}_{(ii)} \Big] + \Big[\underbrace{\sum_{i\leq -2} c_i(\Delta f) \cdot \Delta f \cdot m \cdot e^{(i+2)\varepsilon}}_{(iii)} \Big].$$

Term (i) is a finite sum of finite terms. We show that term (ii) is a convergent series. Indeed, we can bound it from above as follows:

$$\begin{aligned} \sum_{i\geq 1} c_i(\Delta f) \cdot \Delta f \cdot m \cdot e^{-(i-1)\varepsilon} &\leq \sum_{i\geq 1} c((i+1)\cdot\Delta f) \cdot \Delta f \cdot m \cdot e^{-(i-1)\varepsilon} \\ &\leq \sum_{i\geq 1} g((i+1)\cdot\Delta f) \cdot \Delta f \cdot m \cdot e^{-(i-1)\varepsilon}. \end{aligned} \tag{16}$$

Here, the first inequality holds since $c_i(\Delta f) = \Delta f^{-1} \cdot \int_{x\in I_i(\Delta f)} c(x)\mathrm{d}x \leq \Delta f^{-1} \cdot \int_{x\in I_i(\Delta f)} c((i+1)\cdot\Delta f)\mathrm{d}x = c((i+1)\cdot\Delta f)$ by Assumption 1 *(b)* and since $\Delta f \cdot m \cdot e^{-(i-1)\cdot\varepsilon} > 0$ for all i. The second inequality follows from Assumption 1 *(c)*. Note that (16) can be bounded from above by the ratio test,

$$\lim_{i\to\infty} \frac{g((i+2)\cdot\Delta f) \cdot \Delta f \cdot m \cdot e^{-i\cdot\varepsilon}}{g((i+1)\cdot\Delta f) \cdot \Delta f \cdot m \cdot e^{-(i-1)\cdot\varepsilon}} = \lim_{i\to\infty} \frac{g((i+2)\cdot\Delta f)}{g((i+1)\cdot\Delta f)} \cdot e^{-\varepsilon} = e^{-\varepsilon} \cdot \lim_{t\to\infty} \frac{g(t+\Delta f)}{g(t)} = e^{-\varepsilon} < 1,$$

where the equalities follow from simplifying the division, taking the constant $e^{-\varepsilon}$ out of the limit and substituting $t := (i+1)\cdot\Delta f$, and taking the limit (*cf.* Lemma A.5), respectively.

A similar argument shows that term (iii) can be bounded from above as well. The solution p therefore attains an objective value that is bounded from above by a finite scalar M. As problem $\mathrm{P}_0(\Delta f)$ is a minimization problem, $\mathrm{P}_0(\Delta f) \leq M$ follows. □

Lemma A.6 implies that $\mathrm{P}(\Delta f/k)$ is feasible for any fixed $\varepsilon, \delta > 0$ and $k \in \mathbb{N}$.

A.6.2 Monotonicity and Feasibility of $\mathrm{P}(L, \beta)$

We first show that problem $\mathrm{P}(L, \beta)$ is monotonically non-increasing in L and monotonically non-decreasing in β in the following sense.

Lemma A.7. *For any $\varepsilon > 0$, $\delta > 0$, $L' \in \mathbb{N}$ and $\beta > 0$, the optimal value of problem $\mathrm{P}(L', \beta)$ satisfies $\mathrm{P}(L', \beta) \geq \mathrm{P}(L, \beta/k)$ for all $k \in \mathbb{N}$ and $L \geq L' \cdot k + k - 1$.*

Proof. The result trivially holds if $\mathrm{P}(L', \beta)$ is infeasible. We thus assume that $\mathrm{P}(L', \beta)$ is feasible, and we fix any $k \in \mathbb{N}$ as well as $L = L' \cdot k + k - 1$. We proceed in two steps. We first derive an upper bound $\mathrm{P}'(L, \beta/k)$ to $\mathrm{P}(L, \beta/k)$ in which the noise distribution has the same support as in $\mathrm{P}(L', \beta)$. We then show that $\mathrm{P}(L', \beta)$ bounds $\mathrm{P}'(L, \beta/k)$ from above. The result then follows from the fact that $\mathrm{P}(L, \beta/k)$ is monotonically non-increasing in L for any fixed β and k.

In view of the first step, we construct the upper bound $\mathrm{P}'(L, \beta/k)$ to problem $\mathrm{P}(L, \beta/k)$ by adding to $\mathrm{P}(L, \beta/k)$ the constraint that $p(i) = 0$ for $i = -(L' \cdot k + k - 1), \ldots, -(L' \cdot k + 1)$, that is, we remove the first $k - 1$ elements from the domain of p. This ensures that despite its finer interval granularity of β/k, the support of the noise distribution in problem $\mathrm{P}'(L, \beta/k)$ is the same as in the more coarsely discretized problem $\mathrm{P}(L', \beta)$, namely $[-L' \cdot \beta, (L' + 1) \cdot \beta)$.

As for the second step, note that the upper bound $\mathrm{P}'(L, \beta/k)$ can be formulated as

$$\begin{array}{ll} \underset{p}{\text{minimize}} & \displaystyle\sum_{i\in\mathbb{Z}} \sum_{l\in[k]} c_{il}(\beta) \cdot p(i,l) \\ \text{subject to} & p : [\pm L'] \times [k] \mapsto \mathbb{R}_+, \ \displaystyle\sum_{i\in\mathbb{Z}} \sum_{l\in[k]} p(i,l) = 1 \\ & \displaystyle\sum_{i\in[\pm L']} \sum_{l\in[k]} \mathbb{1}[I_{il}(\beta) \subseteq A] \cdot p(i,l) \leq e^{\varepsilon} \cdot \sum_{i\in[\pm L']} \sum_{l\in[k]} \mathbb{1}[I_{il}(\beta) + \varphi \subseteq A] \cdot p(i,l) + \delta \\ & \qquad\qquad \forall (\varphi, A) \in \mathcal{E}(\beta/k), \end{array}$$

where $I_{il}(\beta) = [(i + (l-1)/k) \cdot \beta, (i + l/k) \cdot \beta)$ and $c_{il}(\beta) = (\beta/k)^{-1} \cdot \int_{x \in I_{il}(\beta)} c(x)\, \mathrm{d}x$. Fix any feasible solution p' in problem $\mathrm{P}(L', \beta)$. One readily observes that the solution $p''(i,l) = p'(i)/k$, $i \in [\pm L']$ and $l \in [k]$, is feasible in $\mathrm{P}'(L, \beta/k)$ and attains the same objective value as p' in $\mathrm{P}(L', \beta)$. We thus conclude that $\mathrm{P}(L', \beta)$ bounds $\mathrm{P}(L, \beta/k)$ from above, as desired. □

We next bound the maximum constraint violation of a solution p' in $\mathrm{P}(\beta)$ that is obtained by truncating any feasible solution p in $\mathrm{P}(\beta)$ to a bounded domain. This will later enable us to determine values of L that ensure the feasibility of $\mathrm{P}(L, \beta)$ for any fixed β.

Lemma A.8. *Let p be an arbitrary feasible solution to problem $\mathrm{P}(\beta)$ and fix $L \in \mathbb{N}$ such that $\sum_{i\in[\pm L]} p(i) \geq 1 - \tau$ for some $\tau > 0$. Construct another candidate solution p' to $\mathrm{P}(\beta)$ where $p'(i) = 0$ for all $i \in \mathbb{Z} \setminus [\pm L]$, $p'(L) = \sum_{i \geq L} p(i)$ as well as $p'(-L) = \sum_{i \leq -L} p(i)$, and $p'(i) = p(i)$ otherwise. Then p' violates the DP constraints of problem $\mathrm{P}(\beta)$ by at most $(1 + e^{\varepsilon}) \cdot \tau$, that is,*

$$\sup_{(\varphi,A)\in\mathcal{E}(\beta)} \left\{ \sum_{i\in\mathbb{Z}} \mathbb{1}[I_i(\beta) \subseteq A] \cdot p'(i) - e^{\varepsilon} \cdot \sum_{i\in\mathbb{Z}} \mathbb{1}[(I_i(\beta) + \varphi) \subseteq A] \cdot p'(i) - \delta \right\} \leq (1 + e^{\varepsilon}) \cdot \tau.$$

Note that the constant L in the statement of Lemma A.8 exists since for any probability measure $\gamma \in \mathcal{P}_0$ and any $\tau > 0$, there is $L' \in \mathbb{N}$ such that $\gamma([-L, L]) \geq 1 - \tau$ for all $L \geq L'$.

Proof of Lemma A.8. Since p is feasible in problem $\mathrm{P}(\beta)$, it satisfies

$$\sum_{i\in\mathbb{Z}} \mathbb{1}[I_i(\beta) \subseteq A] \cdot p(i) \leq e^{\varepsilon} \cdot \sum_{i\in\mathbb{Z}} \mathbb{1}[(I_i(\beta) + \varphi) \subseteq A] \cdot p(i) + \delta \quad \forall (\varphi, A) \in \mathcal{E}(\beta).$$

On the other hand, for any $(\varphi, A) \in \mathcal{E}(\beta)$, the constructed solution p' satisfies

$$\begin{aligned}
&\sum_{i\in\mathbb{Z}} \mathbb{1}[I_i(\beta) \subseteq A] \cdot p'(i) - e^\varepsilon \cdot \sum_{i\in\mathbb{Z}} \mathbb{1}[(I_i(\beta)+\varphi) \subseteq A] \cdot p'(i) - \delta \\
=&\sum_{i\in\mathbb{Z}} \mathbb{1}[I_i(\beta) \subseteq A] \cdot [p(i) + (p'(i) - p(i))] - e^\varepsilon \cdot \sum_{i\in\mathbb{Z}} \mathbb{1}[(I_i(\beta)+\varphi) \subseteq A] \cdot [p(i) + (p'(i) - p(i))] - \delta \\
=&\underbrace{\sum_{i\in\mathbb{Z}} \mathbb{1}[I_i(\beta) \subseteq A] \cdot p(i) - e^\varepsilon \cdot \sum_{i\in\mathbb{Z}} \mathbb{1}[(I_i(\beta)+\varphi) \subseteq A] \cdot p(i) - \delta}_{\leq 0 \text{ as } p \text{ is feasible in } (\mathrm{P}(\beta))} + \\
&\underbrace{\sum_{i\in\mathbb{Z}} \mathbb{1}[I_i(\beta) \subseteq A] \cdot (p'(i) - p(i))}_{\leq \tau} - e^\varepsilon \cdot \underbrace{\sum_{i\in\mathbb{Z}} \mathbb{1}[(I_i(\beta)+\varphi) \subseteq A] \cdot (p'(i) - p(i))}_{\geq -\tau} \leq (1+e^\varepsilon)\cdot\tau,
\end{aligned}$$

which implies the statement. □

We can now prove the feasibility of $\mathrm{P}(L, \beta)$.

Lemma A.9. *For any $\varepsilon > 0$, $\delta > 0$ and $\beta > 0$, there exists $L' \in \mathbb{N}$ such that problem $\mathrm{P}(L, \beta)$ is feasible for all $L \geq L'$.*

Proof. Denote by $\mathrm{P}_0(\beta)$ the variant of $\mathrm{P}(\beta)$ that replaces δ with 0. Fix any feasible solution p in problem $\mathrm{P}_0(\beta)$, whose existence is guaranteed by the proof of Lemma A.6, and choose $L \in \mathbb{N}$ large enough so that $\sum_{i\in[\pm L]} p(i) \geq 1 - \delta/(1+e^\varepsilon)$. Lemma A.8 then allows us to construct a solution p' from p that violates the DP constraints of $\mathrm{P}_0(L, \beta)$ by at most δ. By construction, p' is thus feasible in $\mathrm{P}(L, \beta)$, and the statement follows. □

A.6.3 Proof of Theorem 1

To show our convergence result, we define the following auxiliary problem:

$$\begin{aligned}
\underset{p}{\text{minimize}} \quad & \sum_{i\in[\pm(L+\Delta f/\beta)]} c_i(\beta)\cdot p(i) \\
\text{subject to} \quad & p : [\pm(L+\Delta f/\beta)] \mapsto \mathbb{R}_+, \quad \sum_{i\in[\pm(L+\Delta f/\beta)]} p(i) = 1 \\
& \sum_{i\in[\pm(L+\Delta f/\beta)]} \mathbb{1}[I_i(\beta) \subseteq A] \cdot p(i) \leq e^\varepsilon \cdot \sum_{i\in[\pm(L+\Delta f/\beta)]} \mathbb{1}[I_i(\beta)+\varphi \subseteq A] \cdot p(i) + \delta \\
& \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \forall (\varphi, A) \in \mathcal{E}(L, \beta).
\end{aligned} \tag{$\mathrm{M}(L,\beta)$}$$

In the following, we will show that *(i)* problem $\mathrm{M}(L, \beta)$ differs from $\mathrm{P}(L, \beta)$ only in the domain of the decision variable p; *(ii)* problem $\mathrm{M}(L, \beta)$ differs from the strong dual of $\mathrm{D}(L, \beta)$ only in the objective coefficients; and *(iii)* the relationship $\mathrm{P}(L, \beta) \geq \mathrm{M}(L, \beta) \geq \mathrm{D}(L, \beta)$ holds for all $L \in \mathbb{N}$ and $\beta > 0$. Thus, instead of analyzing the convergence of $\mathrm{P}(L, \beta)$ and $\mathrm{D}(L, \beta)$ directly, we can analyze separately the convergence of $\mathrm{P}(L, \beta)$ and $\mathrm{M}(L, \beta)$ (*cf.* Lemma A.11) as well as of $\mathrm{M}(L, \beta)$ and the strong dual of $\mathrm{D}(L, \beta)$ (*cf.* Lemma A.12).

Since $\mathrm{M}(L, \beta)$ coincides with $\mathrm{P}(L, \beta)$ except for the additional decision variables $p(i)$, $i \in [\pm(L + \Delta f/\beta)] \setminus [\pm L]$, we have $\mathrm{P}(L, \beta) \geq \mathrm{M}(L, \beta)$. To show convergence of both problems

(*cf.* Lemma A.11), we need to ensure that these additional decision variables take sufficiently small values in optimal solutions to $\mathrm{M}(L,\beta)$. This is guaranteed by the next result.

Lemma A.10. *For any $\varepsilon > 0$, $\delta > 0$ and $\tau > 0$, there exists $L' \in \mathbb{N}$ such that for all $k \in \mathbb{N}$ and $L \geq L' \cdot k + k - 1$, any optimal solution $p^\star$ to $\mathrm{M}(L, \Delta f/k)$ satisfies $\sum_{i \in [\pm(L+\Delta f/\beta)] \setminus [\pm L]} p^\star(i) < \tau$.*

Proof. Fix $\varepsilon > 0$, $\delta > 0$ and $\tau > 0$. By Lemma A.9, there is $L_1 \in \mathbb{N}$ such that problem $\mathrm{P}(L, \Delta f)$ is feasible for all $L \geq L_1$. Select $L_2 \in \mathbb{N}$ large enough such that

$$\min\{c(L_2 \cdot \Delta f),\ c(-L_2 \cdot \Delta f)\} > \frac{\mathrm{P}(L_1, \Delta f) - (1-\tau) \cdot c(0)}{\tau}; \tag{17}$$

such values exist due to Assumption 1 *(c)*. We claim that the statement of the lemma holds for $L' = \max\{L_1, L_2\}$. To see this, fix any $k \in \mathbb{N}$ and $L \geq L' \cdot k + k - 1$.

Take any optimal solution $p^\star : [\pm(L + \Delta f/\beta)] \mapsto \mathbb{R}_+$ to problem $M(L, \Delta f/k)$ and assume to the contrary that $\sum_{i \in [\pm(L+\Delta f/\beta)] \setminus [\pm L]} p^\star(i) \geq \tau$. We then observe that

$$\begin{aligned}
\sum_{i \in [\pm(L+\Delta f/\beta)]} c_i(\Delta f/k) \cdot p^\star(i) &= \sum_{i \in [\pm L]} c_i(\Delta f/k) \cdot p^\star(i) + \sum_{i \in [\pm(L+\Delta f/\beta)] \setminus [\pm L]} c_i(\Delta f/k) \cdot p^\star(i) \\
&\geq \sum_{i \in [\pm L]} c(0) \cdot p^\star(i) + \sum_{i \in [\pm(L+\Delta f/\beta)] \setminus [\pm L]} \min\{c(L_2 \cdot \Delta f), c(-L_2 \cdot \Delta f)\} \cdot p^\star(i) \\
&> (1-\tau) \cdot c(0) + \tau \cdot \frac{\mathrm{P}(L_1, \Delta f) - (1-\tau) \cdot c(0)}{\tau} \\
&= \mathrm{P}(L_1, \Delta f) \geq \mathrm{P}(L, \Delta f/k) \geq \mathrm{M}(L, \Delta f/k).
\end{aligned} \tag{18}$$

Here, the first inequality is due to Assumption 1 *(b)*. Indeed, we have $c_i(\Delta f/k) \geq c(0)$ for all $i \in \mathbb{Z}$, as well as

$$\begin{aligned}
c_i(\Delta f/k) \ &= \ \left(\frac{\Delta f}{k}\right)^{-1} \cdot \int_{x \in I_i(\Delta f/k)} c(x) \mathrm{d}x \ \geq \ \left(\frac{\Delta f}{k}\right)^{-1} \cdot \int_{x \in I_i(\Delta f/k)} c(L \cdot \Delta f/k) \mathrm{d}x \\
&= \ c(L \cdot \Delta f/k) \ \geq \ c([L_2 \cdot k + k - 1] \cdot \Delta f/k) \ \geq \ c(L_2 \cdot \Delta f) \qquad \forall i > L
\end{aligned}$$

and, similarly, $c_i(\Delta f/k) \geq c(-L_2 \cdot \Delta f)$ for all $i < -L$. The second inequality in (18) follows from (17) and the fact that $\sum_{i \in [\pm(L+\Delta f/\beta)] \setminus [\pm L]} p^\star(i) \geq \tau$. The third inequality in (18) is due to Lemma A.7 and the fact that $L \geq L_1 \cdot k + k - 1$, and the last inequality in (18) holds by construction of problem $\mathrm{M}(L, \Delta f/k)$. We thus conclude that $p^\star$ cannot be optimal in problem $M(L, \Delta f/k)$, which yields the desired contradiction. □

Lemma A.10 allows us to prove the convergence between problems $\mathrm{M}(L,\beta)$ and $\mathrm{P}(L,\beta)$.

Lemma A.11. *For any $\varepsilon > 0$, $\delta > 0$ and $\xi > 0$, there exists $L' \in \mathbb{N}$ such that $\mathrm{P}(L, \Delta f/k) - \mathrm{M}(L, \Delta f/k) \leq \xi$ for all $k \in \mathbb{N}$ and all $L \geq L' \cdot k + k - 1$.*

Intuitively, Lemma A.11 shows that $\mathrm{P}(L,\beta) - \mathrm{M}(L,\beta) \to 0$ for any $\beta > 0$ as long as L grows sufficiently quickly relative to $1/\beta$. Recall that the size of the support of the noise distribution γ is $L \cdot \beta$. Thus, $\mathrm{P}(L,\beta) - \mathrm{M}(L,\beta) \to 0$ for any $\beta > 0$ as long as the support of γ grows large.

Proof of Lemma A.11. Fix $\varepsilon > 0$, $\delta > 0$ and $\xi > 0$, select any $\alpha \in (0, \delta)$, set $\hat{\delta} = \delta - \alpha$ and denote by $\mathrm{P}_{\hat{\delta}}(L, \Delta f/k)$ the variant of $\mathrm{P}(L, \Delta f/k)$ that replaces δ with $\hat{\delta}$. We invoke Lemma A.9 to select $L_0 \in \mathbb{N}$ so that $\mathrm{P}_{\hat{\delta}}(L, \Delta f)$ is feasible for all $L \geq L_0$, and we denote by M the optimal value of $\mathrm{P}_{\hat{\delta}}(L_0, \Delta f)$. We next invoke Lemma A.7 to conclude that $\mathrm{P}_{\hat{\delta}}(L, \Delta f/k)$ remains feasible with an optimal value bounded from above by M for all $k \in \mathbb{N}$ and all $L \geq L_0 \cdot k + k - 1$.

If $\xi \geq M - c(0)$, then the statement vacuously holds since

$$\mathrm{P}(L, \Delta f/k) - \mathrm{M}(L, \Delta f/k) \leq \mathrm{P}_{\hat{\delta}}(L, \Delta f/k) - \mathrm{M}(L, \Delta f/k) \leq M - c(0) \leq \xi$$

for all $k \in \mathbb{N}$ and $L \geq L_0 \cdot k + k - 1$. We thus assume that $M - c(0) > \xi$ and set $\tau = \dfrac{\xi \cdot \alpha}{M - c(0) - \xi}$.

The remainder of the proof shows the statement in four steps. Step 1 constructs a solution p_τ to problem $\mathrm{P}(L, \Delta f/k)$ whose expected loss is bounded from above by the optimal value of $\mathrm{M}(L, \Delta f/k)$, but that may violate the DP constraints in $\mathrm{P}(L, \Delta f/k)$ by up to τ. Step 2 then constructs a convex combination $p^\star$ of p_τ and $p_{\hat{\delta}}$, an optimal solution to problem $\mathrm{P}_{\hat{\delta}}(L, \Delta f/k)$. Step 3 shows that the convex combination $p^\star$ is feasible in $\mathrm{P}(L, \Delta f/k)$, and Step 4 shows that the expected loss of $p^\star$ in $\mathrm{P}(L, \Delta f/k)$ is bounded from above by $\mathrm{M}(L, \Delta f/k) + \xi$, as desired.

In view of Step 1, note that problem $\mathrm{M}(L, \Delta f/k)$ is feasible by construction, and it is bounded since the objective coefficients are non-negative. Lemma A.10 then ensures the existence of $L_1 \in \mathbb{N}$ such that any optimal solution to $\mathrm{M}(L, \Delta f/k)$, $L \geq L_1 \cdot k + k - 1$, places a probability of strictly less than $\tau/(1 + e^\varepsilon)$ outside the index range $[\pm L]$. We claim that $L' = \max\{L_0, L_1\}$ satisfies the statement of this lemma. Take any $k \in \mathbb{N}$ and any $L \geq L' \cdot k + k - 1$, and denote by p_{M} an optimal solution to problem $\mathrm{M}(L, \Delta f/k)$, which satisfies $\sum_{i \in [\pm(L + \Delta f/\beta)] \setminus [\pm L]} p_{\mathrm{M}}(i) < \tau/(1 + e^\varepsilon)$ by the selection of L'. Construct a new truncated solution p_τ via

$$p_\tau(i) = \begin{cases} 0 & \text{if } i \in [\pm(L + \Delta f/\beta)] \setminus [\pm L] \\ \sum_{i' \geq L} p_{\mathrm{M}}(i') & \text{if } i = L \\ \sum_{i' \leq -L} p_{\mathrm{M}}(i') & \text{if } i = -L \\ p_{\mathrm{M}}(i) & \text{otherwise} \end{cases} \qquad \forall i \in [\pm(L + \Delta f/\beta)].$$

By Assumption 1 *(b)*, we have $\sum_{i \in [\pm L]} c_i \cdot p_\tau(i) \leq \sum_{i \in [\pm L]} c_i \cdot p_{\mathrm{M}}(i)$, that is, the truncated solution p_τ achieves a weakly smaller objective value in problem $\mathrm{P}(L, \Delta f/k)$ than the optimal value of $\mathrm{M}(L, \Delta f/k)$. However, a similar reasoning as in the proof of Lemma A.8 shows that p_τ can violate the DP constraints in $\mathrm{P}(L, \Delta f/k)$ by up to τ.

As for Step 2, we define $p_{\hat{\delta}}$ as an optimal solution to problem $\mathrm{P}_{\hat{\delta}}(L, \Delta f/k)$, which exists as $\mathrm{P}_{\hat{\delta}}(L, \Delta f/k)$ is feasible by the selection of L. We then construct the solution $p^\star$ via

$$p^\star = \lambda \cdot p_\tau + (1 - \lambda) \cdot p_{\hat{\delta}} \quad \text{for } \lambda = \frac{\alpha}{\tau + \alpha}.$$

The next two steps will show that $p^\star$ is feasible in problem $\mathrm{P}(L, \Delta f/k)$ and that its expected loss is bounded from above by $\mathrm{M}(L, \Delta f/k) + \xi$.

In view of Step 3, first notice that $p^\star(i) = 0$ for all $i \in [\pm(L + \Delta f/\beta)] \setminus [\pm L]$ since $p^\star$ is a convex combination of two solutions, neither of which places positive probability on indices $i \in [\pm(L + \Delta f/\beta)] \setminus [\pm L]$. Consider now any DP constraint (φ, A) in problem $\mathrm{P}(L, \Delta f/k)$. We

observe that

$$\begin{aligned}
&\sum_{i\in[\pm L]} \mathbb{1}[I_i(\Delta f/k) \subseteq A] \cdot p^\star(i) - e^\varepsilon \cdot \sum_{i\in[\pm L]} \mathbb{1}[I_i(\Delta f/k) + \varphi \subseteq A] \cdot p^\star(i) \\
=&\lambda \cdot \left(\sum_{i\in[\pm L]} \mathbb{1}[I_i(\Delta f/k) \subseteq A] \cdot p_\tau(i) - e^\varepsilon \cdot \sum_{i\in[\pm L]} \mathbb{1}[I_i(\Delta f/k) + \varphi \subseteq A] \cdot p_\tau(i) \right) + \\
&\quad (1-\lambda) \cdot \left(\sum_{i\in[\pm L]} \mathbb{1}[I_i(\Delta f/k) \subseteq A] \cdot p_{\hat\delta}(i) - e^\varepsilon \cdot \sum_{i\in[\pm L]} \mathbb{1}[I_i(\Delta f/k) + \varphi \subseteq A] \cdot p_{\hat\delta}(i) \right) \\
\leq&\lambda \cdot (\delta + \tau) + (1-\lambda) \cdot \hat\delta \\
=&\lambda \cdot (\delta + \tau) + (1-\lambda) \cdot (\delta - \alpha) = \lambda \cdot (\tau + \alpha) + \delta - \alpha = \delta,
\end{aligned}$$

where the first equality uses the definition of $p^\star$, the inequality holds since p_τ violates the DP constraints in problem $\mathrm{P}(L,\beta)$ by up to τ and $p_{\hat\delta}$ is feasible in $\mathrm{P}_{\hat\delta}(L, \Delta f/k)$, and the final equalities follow from the definition of $\hat\delta$, rearranging terms and from the definition of λ, respectively. We thus conclude that $p^\star$ satisfies the DP constraint (φ, A) in problem $\mathrm{P}(L, \Delta f/k)$, and since the choice of (φ, A) was arbitrary, $p^\star$ must indeed be feasible in $\mathrm{P}(L, \Delta f/k)$.

As for Step 4, first note that the solution $p^\star$ achieves an objective value of $\sum_{i\in[\pm L]} c_i(\Delta f/k) \cdot p^\star(i)$ in $\mathrm{P}(L, \Delta f/k)$, which itself satisfies

$$\begin{aligned}
\sum_{i\in[\pm L]} c_i(\Delta f/k) \cdot p^\star(i) &= \lambda \sum_{i\in[\pm L]} c_i(\Delta f/k) \cdot p_\tau(i) + (1-\lambda) \cdot \sum_{i\in[\pm L]} c_i(\Delta f/k) \cdot p_{\hat\delta}(i) \\
&\leq \lambda \sum_{i\in[\pm L]} c_i(\Delta f/k) \cdot p_{\mathrm{M}}(i) + (1-\lambda) \cdot \sum_{i\in[\pm L]} c_i(\Delta f/k) \cdot p_{\hat\delta}(i). \qquad (19)
\end{aligned}$$

We can use (19) to bound the difference $\mathrm{P}(L, \Delta f/k) - \mathrm{M}(L, \Delta f/k)$ as follows:

$$\begin{aligned}
\mathrm{P}(L, \Delta f/k) - \mathrm{M}(L, \Delta f/k) &\leq \sum_{i\in[\pm L]} c_i(\Delta f/k) \cdot p^\star(i) - \sum_{i\in[\pm L]} c_i(\Delta f/k) \cdot p_{\mathrm{M}}(i) \\
&\leq (1-\lambda) \left(\sum_{i\in[\pm L]} c_i(\Delta f/k) \cdot p_{\hat\delta}(i) - \sum_{i\in[\pm L]} c_i(\Delta f/k) \cdot p_{\mathrm{M}}(i) \right) \\
&\leq (1-\lambda) \cdot (M - c(0)) = \frac{\tau}{\tau + \alpha} \cdot (M - c(0)) = \xi.
\end{aligned}$$

Here, the first inequality holds since $p^\star$ is feasible in $\mathrm{P}(L, \Delta f/k)$ and p_{M} is optimal in $\mathrm{M}(L, \Delta f/k)$. The second inequality employs (19). The third inequality bounds the objective value of $p_{\hat\delta}$ from above by M and the objective value of p_{M} from below by $c(0)$, respectively. The two identities, finally, follow from substituting back the definitions of λ and τ, respectively. □

We next prove the convergence between problems $\mathrm{M}(L,\beta)$ and $\mathrm{D}(L,\beta)$.

Lemma A.12. *For any $\varepsilon > 0$, $\delta > 0$, $\xi > 0$ and $\Lambda \in \mathbb{N}$, there exists $k' \in \mathbb{N}$ such that $\mathrm{M}(\Lambda \cdot k, \Delta f/k) - \mathrm{D}(\Lambda \cdot k, \Delta f/k) \leq \xi$ for all $k \geq k'$.*

Intuitively, Lemma A.12 shows that $\mathrm{M}(L,\beta) - \mathrm{D}(L,\beta) \to 0$ for any fixed size of the support of the noise distribution γ as long as the discretization granularity β vanishes to zero.

Proof of Lemma A.12. Fix $\varepsilon > 0$, $\delta > 0$, $\xi > 0$ and $\Lambda \in \mathbb{N}$. We will show that there is $k' \in \mathbb{N}$ such that $\mathrm{M}(\Lambda \cdot k, \Delta f/k) - \overline{D}(\Lambda \cdot k, \Delta f/k) \leq \xi$ for all $k \geq k'$, where $\overline{D}(\Lambda \cdot k, \Delta f/k)$ denotes the dual to $\mathrm{D}(\Lambda \cdot k, \Delta f/k)$. Indeed, strong duality holds between $\mathrm{D}(\Lambda \cdot k, \Delta f/k)$ and $\overline{D}(\Lambda \cdot k, \Delta f/k)$ since (θ, ψ) with $\theta = \min\{\underline{c}_i(\beta) : i \in [\pm(\Lambda \cdot k + \Delta f/\beta)]\}$ and $\psi(\varphi, A) = 0$ for all $\mathcal{E}(\Lambda \cdot k, \beta)$ is feasible in $\mathrm{D}(\Lambda \cdot k, \Delta f/k)$. The dual problem $\overline{D}(\Lambda \cdot k, \Delta f/k)$ can be formulated as

$$
\begin{array}{lll}
\underset{p}{\text{minimize}} & \displaystyle\sum_{i \in [\pm(\Lambda \cdot k + k)]} \underline{c}_i(\Delta f/k) \cdot p(i) & \\
\text{subject to} & p : [\pm(\Lambda \cdot k + k)] \mapsto \mathbb{R}_+, \quad \displaystyle\sum_{i \in [\pm(\Lambda \cdot k + k)]} p(i) = 1 & \\
& \displaystyle\sum_{i \in [\pm(\Lambda \cdot k + k)]} \mathbb{1}[I_i(\Delta f/k) \subseteq A] \cdot p(i) \leq e^{\varepsilon} \cdot \sum_{i \in [\pm(\Lambda \cdot k + k)]} \mathbb{1}[I_i(\Delta f/k) + \varphi \subseteq A] \cdot p(i) + \delta & \\
& \qquad\qquad \forall (\varphi, A) \in \mathcal{E}(\Lambda \cdot k, \Delta f/k). & \\
& \qquad\qquad (\overline{\mathrm{D}}(\Lambda \cdot k, \Delta f/k)) &
\end{array}
$$

Note that $\overline{\mathrm{D}}(\Lambda \cdot k, \Delta f/k)$ and $\mathrm{M}(\Lambda \cdot k, \Delta f/k)$ only differ in their objective coefficients $\underline{c}_i(\beta)$ and $c_i(\beta)$, respectively. We now show that the difference between those two coefficient sets can be made arbitrarily small, uniformly across all $i \in [\pm(\Lambda \cdot k + k)]$, by increasing k. Indeed, the loss function c is continuous by Assumption 1 *(a)*, and it is therefore uniformly continuous over the (closure of the) finite interval $\bigcup\{I_i(\Delta f/k) : i \in [\pm(\Lambda \cdot k + k)]\}$ by the Heine-Cantor theorem. (Note in particular that for any k, this interval is contained in the set $[-(\Lambda+1) \cdot \Delta f, (\Lambda+2) \cdot \Delta f)$ that is independent of k, which justifies our use of the Heine-Cantor theorem.) For the selected $\xi > 0$, we can thus find $k' \in \mathbb{N}$ such that for all $k \geq k'$, we have

$$
\begin{aligned}
c_i(\Delta f/k) - \underline{c}_i(\Delta f/k) \;&=\; \left(\frac{\Delta f}{k}\right)^{-1} \cdot \int_{x \in I_i(\Delta f/k)} c(x)\,\mathrm{d}x - \inf_{x \in I_i(\Delta f/k)} c(x) \\
&\leq \sup_{x \in I_i(\Delta f/k)} c(x) - \inf_{x \in I_i(\Delta f/k)} c(x) \;\leq\; \xi,
\end{aligned}
$$

uniformly across all $i \in [\pm(\Lambda \cdot k + k)]$. Here, the identity replaces $c_i(\Delta f/k)$ and $\underline{c}_i(\Delta f/k)$ with their respective definitions, the first inequality exploits that $c(x) \leq \sup_{x \in I_i(\Delta f/k)} c(x)$ for all $x \in I_i(\Delta f/k)$, and the last inequality makes use of the uniform continuity of c.

To bound $\mathrm{M}(\Lambda \cdot k, \Delta f/k) - \overline{\mathrm{D}}(\Lambda \cdot k, \Delta f/k)$, take any optimal solution $p^\star$ to problem $\overline{\mathrm{D}}(\Lambda \cdot k, \Delta f/k)$ and notice that $p^\star$ is feasible in $\mathrm{M}(\Lambda \cdot k, \Delta f/k)$ since both problems only differ in their objective coefficients. We thus have

$$
\mathrm{M}(\Lambda \cdot k, \Delta f/k) - \overline{\mathrm{D}}(\Lambda \cdot k, \Delta f/k) \;\leq \sum_{i \in [\pm(\Lambda \cdot k + k)]} [c_i(\Delta f/k) - \underline{c}_i(\Delta f/k)] \cdot p^\star(i) \;\leq\; \xi,
$$

where the first inequality holds since $p^\star$ is optimal in $\overline{\mathrm{D}}(\Lambda \cdot k, \Delta f/k)$ but feasible (and possibly not optimal) in $\mathrm{M}(\Lambda \cdot k, \Delta f/k)$, and the second inequality is due to our earlier uniform bound on $c_i(\Delta f/k) - \underline{c}_i(\Delta f/k)$ and the fact that $p^\star$ is a probability distribution. □

Proof of Theorem 1. Fix $\xi > 0$ as well as any $\xi_1, \xi_2 > 0$ satisfying $\xi_1 + \xi_2 = \xi$. Since

$$
\mathrm{P}(\Lambda \cdot k, \Delta f/k) - \mathrm{D}(\Lambda \cdot k, \Delta f/k) = [\mathrm{P}(\Lambda \cdot k, \Delta f/k) - \mathrm{M}(\Lambda \cdot k, \Delta f/k)] + [\mathrm{M}(\Lambda \cdot k, \Delta f/k) - \mathrm{D}(\Lambda \cdot k, \Delta f/k)]\,,
$$

for any $\Lambda \in \mathbb{N}$ and $k \in \mathbb{N}$, it suffices to show that there is $\Lambda' \in \mathbb{N}$ and $k' \in \mathbb{N}$ such that

$$\mathrm{P}(\Lambda \cdot k, \Delta f/k) - \mathrm{M}(\Lambda \cdot k, \Delta f/k) \leq \xi_1 \quad \text{and} \quad \mathrm{M}(\Lambda \cdot k, \Delta f/k) - \mathrm{D}(\Lambda \cdot k, \Delta f/k) \leq \xi_2 \tag{20}$$

simultaneously hold for all $\Lambda \geq \Lambda'$ and $k \geq k'$.

In view of the first inequality in (20), we invoke Lemma A.11 to select $L' \in \mathbb{N}$ such that $\mathrm{P}(L, \Delta f/k) - \mathrm{M}(L, \Delta f/k) \leq \xi_1$ for all $k \in \mathbb{N}$ and all $L \geq L' \cdot k + k - 1$. Fix $\Lambda' = L' + 1$ and notice that $\Lambda' \cdot k = (L' + 1) \cdot k \geq L' \cdot k + k - 1$. For our choice of Λ', we thus have $\mathrm{P}(\Lambda' \cdot k, \Delta f/k) - \mathrm{M}(\Lambda' \cdot k, \Delta f/k) \leq \xi_1$ for all $k \in \mathbb{N}$. As for the second inequality in (20), we invoke Lemma A.12 to select $k' \in \mathbb{N}$ such that $\mathrm{M}(\Lambda' \cdot k, \Delta f/k) - \mathrm{D}(\Lambda' \cdot k, \Delta f/k) \leq \xi_2$ for all $k \geq k'$. So far, we have shown that there exists $\Lambda' \in \mathbb{N}$ and $k' \in \mathbb{N}$ such that $\mathrm{P}(\Lambda' \cdot k, \Delta f/k) - \mathrm{D}(\Lambda' \cdot k, \Delta f/k) \leq \xi$ holds for all $k \geq k'$. Since we have $\mathrm{P}(\Lambda' \cdot k, \Delta f/k) \geq \mathrm{P}(\Lambda \cdot k, \Delta f/k)$ and $\mathrm{D}(\Lambda' \cdot k, \Delta f/k) \leq \mathrm{D}(\Lambda \cdot k, \Delta f/k)$ for all $\Lambda \geq \Lambda'$, we can conclude the proof. □

B Proofs of Section 3

B.1 Proof of Proposition 4

The proof of Proposition 4 relies on three auxiliary lemmas, each of which is devoted to the inclusion of one of the restrictions (7a), (7b) and (7c) into problem P′. We state and prove these auxiliary lemmas first.

Lemma B.13. *With the additional constraint* (7a)*,* P′ *has the same optimal value as*

$$\begin{array}{ll} \underset{\gamma}{\text{minimize}} & \beta \cdot \sum\limits_{k \in [K]} w_k(\beta) \cdot \int_{x \in \mathbb{R}} c(x) \, \mathrm{d}\gamma_k(x) \\ \text{subject to} & \gamma_k \in \mathcal{P}_0, \ k \in [K] \\ & \int_{x \in \mathbb{R}} \mathbb{1}[x \in A] \, \mathrm{d}\gamma_k(x) \leq e^{\varepsilon} \cdot \int_{x \in \mathbb{R}} \mathbb{1}[x + \varphi \in A] \, \mathrm{d}\gamma_m(x) + \delta \\ & \qquad \forall k, m \in [K], \ \forall (\varphi, A) \in \mathcal{E}''_{km}(\beta), \end{array} \tag{21}$$

where $\mathcal{E}''_{km}(\beta) := [[-\Delta f, \Delta f] \cap (\Phi_m(\beta) - \Phi_k(\beta))] \times \mathcal{F}$ *and* $w_k(\beta) := \beta^{-1} \cdot \int_{\phi \in \Phi_k(\beta)} w(\phi) \, \mathrm{d}\phi$.

Proof. We use restriction (7a) to replace the uncountable family of measures $\{\gamma(\cdot|\phi)\}_{\phi \in \Phi}$ in problem P′ with the finite set of measures $\{\gamma_k\}_{k \in [K]}$. Note that in this case, the requirement $\gamma \in \Gamma$ simplifies to $\gamma_k \in \mathcal{P}_0$ for all $k \in [K]$.

We can now equivalently reformulate the objective function of problem P′ as

$$\begin{aligned} \int_{\phi \in \Phi} w(\phi) \cdot \left[\int_{x \in \mathbb{R}} c(x) \, \mathrm{d}\gamma(x \mid \phi) \right] \mathrm{d}\phi &= \sum_{k \in [K]} \int_{\phi \in \Phi_k(\beta)} w(\phi) \cdot \Big[\int_{x \in \mathbb{R}} c(x) \, \mathrm{d} \underbrace{\gamma(x \mid \phi)}_{=\gamma_k(x)} \Big] \mathrm{d}\phi \\ &= \sum_{k \in [K]} \Big[\underbrace{\int_{\phi \in \Phi_k(\beta)} w(\phi) \, \mathrm{d}\phi}_{=\beta \cdot w_k(\beta)} \Big] \cdot \Big[\int_{x \in \mathbb{R}} c(x) \, \mathrm{d}\gamma_k(x) \Big], \end{aligned}$$

which coincides with the objective function of (21).

Next, fix any DP constraint (ϕ, φ, A) in P′, and note that $(\phi, \phi+\varphi) \in \Phi_k(\beta) \times \Phi_m(\beta)$ for a unique pair $(k, m) \in [K]$. In terms of our new measures $\{\gamma_k\}_{k\in[K]}$, the constraint becomes

$$\int_{x\in\mathbb{R}} \mathbb{1}[x \in A]\, \mathrm{d}\gamma_k(x) \leq e^{\varepsilon} \cdot \int_{x\in\mathbb{R}} \mathbb{1}[x+\varphi \in A]\, \mathrm{d}\gamma_m(x) + \delta,$$

and this constraint is indeed included in (21) since $\varphi \in [-\Delta f, \Delta f] \cap (\Phi_m(\beta) - \Phi_k(\beta))$. Similarly, one readily verifies that all DP constraints in (21) have corresponding constraints in problem P′. We thus conclude that P′ and (21) are indeed equivalent under restriction (7a), as desired. □

In contrast to P′, problem (21) comprises finitely many probability measures $\{\gamma_k\}_{k\in[K]}$. The next result shows that the restriction (7b) allows us to equivalently represent each measure γ_k by countably many decision variables.

Lemma B.14. *With the additional constraint* (7b)*, problem* (21) *has the same optimal value as*

$$\begin{array}{ll} \underset{p}{\text{minimize}} & \beta \cdot \sum\limits_{k\in[K]} w_k(\beta) \cdot \sum\limits_{i\in\mathbb{Z}} c(i) \cdot p_k(i) \\ \text{subject to} & p_k : \mathbb{Z} \mapsto \mathbb{R}_+,\ \sum\limits_{i\in\mathbb{Z}} p_k(i) = 1,\ k \in [K] \\ & \sum\limits_{i\in\mathbb{Z}} \mathbb{1}[I_i(\beta) \subseteq A] \cdot p_k(i) \leq e^{\varepsilon} \cdot \sum\limits_{i\in\mathbb{Z}} \mathbb{1}[I_i(\beta)+\varphi \subseteq A] \cdot p_m(i) + \delta \\ & \qquad\qquad \forall k, m \in [K],\ \forall (\varphi, A) \in \mathcal{E}'_{km}(\beta), \end{array} \tag{P′(β)}$$

where $\mathcal{E}'_{km}(\beta) := [\mathscr{B}(\beta) \cap \{(m-k-1)\cdot\beta,\ (m-k)\cdot\beta,\ (m-k+1)\cdot\beta\}] \times \mathcal{F}(\beta)$.

Proof. Similar arguments as in the proof of Lemma A.1 allow us to replace each measure γ_k with a countable set of decision variables $p_k : \mathbb{Z} \mapsto \mathbb{R}_+$, $k \in [K]$. The resulting reformulation of problem (21) under restriction (7b) reads as follows.

$$\begin{array}{ll} \underset{p}{\text{minimize}} & \beta \cdot \sum\limits_{k\in[K]} w_k(\beta) \cdot \sum\limits_{i\in\mathbb{Z}} c(i) \cdot p_k(i) \\ \text{subject to} & \sum\limits_{i\in\mathbb{Z}} p_k(i) = 1,\ p_k : \mathbb{Z} \mapsto \mathbb{R}_+,\ k \in [K] \\ & \sum\limits_{i\in\mathbb{Z}} p_k(i) \cdot \dfrac{|A \cap I_i(\beta)|}{\beta} \leq e^{\varepsilon} \cdot \sum\limits_{i\in\mathbb{Z}} p_m(i) \cdot \dfrac{|(A-\varphi) \cap I_i(\beta)|}{\beta} + \delta \\ & \qquad\qquad \forall k, m \in [K],\ \forall (\varphi, A) \in \mathcal{E}''_{km}(\beta) \end{array}$$

In contrast to P′(β), this problem still employs the larger constraint set $(\varphi, A) \in \mathcal{E}''_{km}(\beta)$. Applying similar arguments as in the proof of Lemma A.2 shows that the constraints (φ, A) satisfying $A \in \mathcal{F} \setminus \mathcal{F}(\beta)$ are weakly dominated by the constraints (φ, A') satisfying $A' \in \mathcal{F}(\beta)$, and arguments similar to those in the proof of Lemma A.3 show that the constraints (φ, A) satisfying $A \in \mathcal{F}(\beta)$ and $\varphi \in [-\Delta f, \Delta f] \cap (\Phi_m(\beta) - \Phi_k(\beta))$ are weakly dominated by the constraints (φ', A) satisfying $\varphi' \in \mathscr{B}(\beta) \cap \mathrm{cl}(\Phi_m(\beta) - \Phi_k(\beta))$ whenever $[-\Delta f, \Delta f] \cap (\Phi_m(\beta) - \Phi_k(\beta))$ is nonempty. We can thus replace in the above optimization problem the constraints $(\varphi, A) \in \mathcal{E}''_{km}(\beta)$ with the smaller set of constraints $(\varphi, A) \in \mathcal{E}'_{km}(\beta)$. In that case, however, the identities

$$\frac{|A \cap I_i(\beta)|}{\beta} = \mathbb{1}[I_i(\beta) \subseteq A] \quad \text{and} \quad \frac{|(A-\varphi) \cap I_i(\beta)|}{\beta} = \mathbb{1}[I_i(\beta)+\varphi \subseteq A]$$

hold for all $i \in \mathbb{Z}$ (*cf.* the proof of Lemma 1), which concludes the proof. □

Problem $\mathrm{P}'(\beta)$ still comprises a countably infinite number of decision variables and uncountably many constraints. The next result shows that the restriction (7c) allows us to reformulate $\mathrm{P}'(\beta)$ as a finite-dimensional linear program.

Lemma B.15. *With the additional constraint* (7c)*,* $\mathrm{P}'(\beta)$ *has the same optimal value as the finite-dimensional linear program* $\mathrm{P}'(L, \beta)$*.*

Proof. Under restriction (7c), we can reduce each countable set of decisions $p_k : \mathbb{Z} \mapsto \mathbb{R}_+$ to a finite set $p_k : [\pm L] \mapsto \mathbb{R}_+$, $k \in [K]$. Moreover, similar arguments as in the proof of Proposition 1 allow us to show that in the resulting problem, each constraint $(\varphi, A) \in \mathcal{E}'_{km}(\beta) \setminus \mathcal{E}'_{km}(L, \beta)$ is weakly dominated by the a constraint $(\varphi, A_L) \in \mathcal{E}'_{km}(L, \beta)$. This concludes the proof. □

Proof of Proposition 4. The proof directly follows from Lemmas B.13, B.14 and B.15. □

B.2 Proof of Proposition 5

Fix any γ feasible in P' and any (θ, ψ) feasible in D'. We then have

$$
\begin{aligned}
&\int_{\phi\in\Phi} w(\phi) \cdot \left[\int_{x\in\mathbb{R}} c(x)\,\mathrm{d}\gamma(x \mid \phi)\right] \mathrm{d}\phi \\
&= \int_{\phi\in\Phi}\int_{x\in\mathbb{R}} \Big[w(\phi)\cdot c(x)\Big]\,\mathrm{d}\gamma(x \mid \phi)\,\mathrm{d}\phi \\
&\geq \int_{\phi\in\Phi}\int_{x\in\mathbb{R}} \left[\theta(\phi) - \int_{(\varphi,A)\in\mathcal{E}'(\phi)} \mathbb{1}[x\in A]\,\mathrm{d}\psi(\varphi, A \mid \phi) + \right. \\
&\qquad\qquad \left. e^{\varepsilon}\cdot \int_{(-\varphi,A)\in\mathcal{E}'(\phi)} \mathbb{1}[x+\varphi\in A]\,\mathrm{d}\psi(\varphi, A \mid \phi-\varphi)\right] \mathrm{d}\gamma(x \mid \phi)\,\mathrm{d}\phi \\
&= \underbrace{\int_{\phi\in\Phi}\int_{x\in\mathbb{R}} \theta(\phi)\,\mathrm{d}\gamma(x \mid \phi)\,\mathrm{d}\phi}_{(i)} - \underbrace{\int_{\phi\in\Phi}\int_{x\in\mathbb{R}}\int_{(\varphi,A)\in\mathcal{E}'(\phi)} \mathbb{1}[x\in A]\,\mathrm{d}\psi(\varphi, A \mid \phi)\,\mathrm{d}\gamma(x \mid \phi)\,\mathrm{d}\phi}_{(ii)} \\
&\qquad + e^{\varepsilon}\cdot \underbrace{\int_{\phi\in\Phi}\int_{x\in\mathbb{R}}\int_{(-\varphi,A)\in\mathcal{E}'(\phi)} \mathbb{1}[x+\varphi\in A]\,\mathrm{d}\psi(\varphi, A \mid \phi-\varphi)\,\mathrm{d}\gamma(x \mid \phi)\,\mathrm{d}\phi}_{(iii)},
\end{aligned}
$$

where the inequality follows from the constraints in D' and the fact that $\gamma(\cdot|\phi)$ is a non-negative measure, and the final equality is due to the linearity of the integration operator.

The above term *(i)* simplifies to

$$
\int_{\phi\in\Phi}\int_{x\in\mathbb{R}} \theta(\phi)\,\mathrm{d}\gamma(x \mid \phi)\,\mathrm{d}\phi = \int_{\phi\in\Phi} \theta(\phi)\left[\int_{x\in\mathbb{R}} \mathrm{d}\gamma(x \mid \phi)\right]\mathrm{d}\phi = \int_{\phi\in\Phi} \theta(\phi)\,\mathrm{d}\phi,
$$

where we used the fact that $\gamma(\cdot|\phi)$ is a probability measure for every $\phi \in \Phi$.

The above term *(ii)* can be reformulated as

$$
\int_{\phi\in\Phi}\int_{x\in\mathbb{R}}\int_{(\varphi,A)\in\mathcal{E}'(\phi)} \mathbb{1}[x\in A]\,\mathrm{d}\psi(\varphi, A \mid \phi)\,\mathrm{d}\gamma(x \mid \phi)\,\mathrm{d}\phi
$$

$$= \int_{\phi\in\Phi} \int_{(\varphi,A)\in\mathcal{E}'(\phi)} \left[\int_{x\in\mathbb{R}} \mathbb{1}[x \in A]\, \mathrm{d}\gamma(x \mid \phi) \right] \mathrm{d}\psi(\varphi, A \mid \phi)\, \mathrm{d}\phi,$$

where we used Fubini's theorem (whose applicability follows from similar arguments as in the proof of Proposition 2) to change the order of integration.

Finally, the above term *(iii)* can be rewritten as

$$\begin{aligned}
&\int_{\phi\in\Phi} \int_{x\in\mathbb{R}} \int_{(-\varphi,A)\in\mathcal{E}'(\phi)} \mathbb{1}[x+\varphi \in A]\, \mathrm{d}\psi(\varphi, A \mid \phi-\varphi)\, \mathrm{d}\gamma(x \mid \phi)\, \mathrm{d}\phi \\
&= \int_{\phi\in\Phi} \int_{(-\varphi,A)\in\mathcal{E}'(\phi)} \left[\int_{x\in\mathbb{R}} \mathbb{1}[x+\varphi \in A]\, \mathrm{d}\gamma(x \mid \phi) \right] \mathrm{d}\psi(\varphi, A \mid \phi-\varphi)\, \mathrm{d}\phi \\
&= \int_{\phi\in\Phi} \int_{(\varphi,A)\in\mathcal{E}'(\phi)} \left[\int_{x\in\mathbb{R}} \mathbb{1}[x+\varphi \in A]\, \mathrm{d}\gamma(x \mid \phi+\varphi) \right] \mathrm{d}\psi(\varphi, A \mid \phi)\, \mathrm{d}\phi.
\end{aligned}$$

where the first equality follows from Fubini's theorem. The second equality is due to a change of variables. Specifically, we use the definition of $\mathcal{E}'(\phi)$ to rewrite the region of integration as

$$\left\{(\phi,\varphi,A) \;:\; \phi\in\Phi, (-\varphi,A)\in\mathcal{E}'(\phi)\right\} = \left\{(\phi,\varphi,A)\in\Phi\times[-\Delta f,\Delta f]\times\mathcal{F} \;:\phi-\varphi\in\Phi\right\}.$$

Introducing a new variable $\phi' = \phi - \varphi$, we observe that $\varphi + \phi' = \phi \in \Phi$. Hence the region of integration region can be expressed as

$$\left\{(\phi',\varphi,A)\in\Phi\times[-\Delta f,\Delta f]\times\mathcal{F} \;:\varphi\in\Phi-\phi'\right\} = \left\{(\phi',\varphi,A) \;:\; \phi'\in\Phi, (\varphi,A)\in\mathcal{E}'(\phi')\right\}$$

if we replace ϕ with $\phi'+\varphi$ in the integrals. The second equality now holds if we relabel ϕ' as ϕ.

Replacing the terms *(i)–(iii)* with their equivalent expressions derived above, we obtain

$$\begin{aligned}
&\int_{\phi\in\Phi} w(\phi) \cdot \left[\int_{x\in\mathbb{R}} c(x)\, \mathrm{d}\gamma(x \mid \phi) \right] \mathrm{d}\phi \\
&\geq \int_{\phi\in\Phi} \theta(\phi)\, \mathrm{d}\phi \;- \int_{\phi\in\Phi} \int_{(\varphi,A)\in\mathcal{E}'(\phi)} \left[\int_{x\in\mathbb{R}} \mathbb{1}[x\in A]\, \mathrm{d}\gamma(x \mid \phi) - \right. \\
&\qquad\qquad \left. e^{\varepsilon} \cdot \int_{x\in\mathbb{R}} \mathbb{1}[x+\varphi\in A]\, \mathrm{d}\gamma(x \mid \phi+\varphi) \right] \mathrm{d}\psi(\varphi, A \mid \phi)\, \mathrm{d}\phi \\
&\geq \int_{\phi\in\Phi} \left[\theta(\phi) - \delta \cdot \int_{(\varphi,A)\in\mathcal{E}'(\phi)} \mathrm{d}\psi(\varphi, A \mid \phi) \right] \mathrm{d}\phi,
\end{aligned}$$

where the final inequality is due to the constraints in P′ and the fact that $\psi(\cdot|\phi)$ is a non-negative measure. The last expression coincides with the objective function of D′, as desired. □

B.3 Proof of Proposition 6

Our proof proceeds in two steps. We first use restriction (8a) to reduce the number of decision variables in D′, and we subsequently use restriction (8b) to remove the integrals as well as reduce the number of constraints in D′. This will yield the formulation in the statement of Proposition 6.

In view of the first step, we use restriction (8a) to replace the measure $\theta : \Phi \mapsto \mathbb{R}$ with a vector $\boldsymbol{\theta} \in \mathbb{R}^K$ and ψ with a finite family of unconditional measures $\psi_k \in \mathcal{M}_+(\mathcal{E}'(\phi))$, $k \in [K]$.

Under those substitutions, problem D′ simplifies to the following formulation.

$$\begin{array}{ll}\underset{\theta,\psi}{\text{maximize}} & \beta \cdot \sum_{k\in[K]} \theta_k - \delta \cdot \sum_{k\in[K]} \int_{\phi\in\Phi_k(\beta)} \int_{(\varphi,A)\in\mathcal{E}'(\phi)} \mathrm{d}\psi_k(\varphi,A)\,\mathrm{d}\phi \\ \text{subject to} & \boldsymbol{\theta}\in\mathbb{R}^K,\ \psi_k\in\mathcal{M}_+(\mathcal{E}'(\phi)),\ k\in[K] \\ & \theta_k \le \int_{(\varphi,A)\in\mathcal{E}'(\phi)} \mathbb{1}[x\in A]\,\mathrm{d}\psi_k(\varphi,A) - \\ & \qquad e^{\varepsilon}\cdot\sum_{m\in[K]}\int_{(-\varphi,A)\in\mathcal{E}'(\phi)} \mathbb{1}[x+\varphi\in A]\cdot\mathbb{1}[\phi-\varphi\in\Phi_m(\beta)]\,\mathrm{d}\psi_m(\varphi,A) \\ & \qquad\qquad +c(x)\cdot w(\phi) \quad \forall k\in[K],\ \forall\phi\in\Phi_k(\beta),\ \forall x\in\mathbb{R}\end{array}$$

Here, our reformulation uses the fact that $\{\Phi_k(\beta)\}_{k\in[K]}$ partitions Φ, and the reformulated first term of the objective function additionally exploits that $|\Phi_k(\beta)|=\beta$ for all $k\in[K]$.

As for the second step, we note that under restriction (8b), the second expression in the objective function simplifies to

$$\begin{aligned}\sum_{k\in[K]}\int_{\phi\in\Phi_k(\beta)}\int_{(\varphi,A)\in\mathcal{E}'(\phi)}\mathrm{d}\psi_k(\varphi,A)\,\mathrm{d}\phi &= \sum_{k\in[K]}\int_{\phi\in\Phi_k(\beta)}\int_{(\varphi,A)\in\mathcal{E}'_k(L,\beta)}\mathrm{d}\psi_k(\varphi,A)\,\mathrm{d}\phi \\ &= \beta\cdot\sum_{k\in[K]}\int_{(\varphi,A)\in\mathcal{E}'_k(L,\beta)}\mathrm{d}\psi_k(\varphi,A),\end{aligned}$$

where the first equality is due to restriction (8b) and the fact that $\mathcal{E}'(\phi)\cap\mathcal{E}(L,\beta)=\mathcal{E}'_k(L,\beta)$, and the second equality follows from taking the inner integral outside (as it is not parameterized by ϕ) and from $|\Phi_k(\beta)|=\beta$. The resulting objective function coincides with that of problem D′(L,β). For any fixed $k\in[K]$, $\phi\in\Phi_k(\beta)$ and $x\in\mathbb{R}$, the first integral in the constraints simplifies to

$$\int_{(\varphi,A)\in\mathcal{E}'(\phi)}\mathbb{1}[x\in A]\,\mathrm{d}\psi_k(\varphi,A) = \int_{(\varphi,A)\in\mathcal{E}'_k(L,\beta)}\mathbb{1}[x\in A]\,\mathrm{d}\psi_k(\varphi,A).$$

Likewise, the second integral simplifies to

$$\begin{aligned}&\sum_{m\in[K]}\int_{(-\varphi,A)\in\mathcal{E}'(\phi)}\mathbb{1}[x+\varphi\in A]\cdot\mathbb{1}[\phi-\varphi\in\Phi_m(\beta)]\,\mathrm{d}\psi_m(\varphi,A)\\ =&\sum_{m\in[K]}\int_{(-\varphi,A)\in\mathcal{E}'_k(L,\beta)}\mathbb{1}[x+\varphi\in A]\cdot\mathbb{1}[\phi-\varphi\in\Phi_m(\beta)]\,\mathrm{d}\psi_m(\varphi,A)\\ =&\int_{(-\varphi,A)\in\mathcal{E}'_k(L,\beta)}\mathbb{1}[x+\varphi\in A]\,\mathrm{d}\psi_{k-\varphi/\beta}(\varphi,A),\end{aligned}$$

where the first equality is due to restriction (8b) and the second equality exploits the fact that for $\phi\in\Phi_k(\beta)$ and $-\varphi\in\mathscr{B}(\beta)$, we have $\phi-\varphi\in\Phi_m(\beta)$ if and only if $m=k-\varphi/\beta$. In summary, the constraints simplify to

$$\begin{aligned}\theta_k \le \int_{(\varphi,A)\in\mathcal{E}'_k(L,\beta)}\mathbb{1}[x\in A]\,\mathrm{d}\psi_k(\varphi,A) - e^{\varepsilon}\cdot\int_{(-\varphi,A)\in\mathcal{E}'_k(L,\beta)}\mathbb{1}[x+\varphi\in A]\,\mathrm{d}\psi_{k-\varphi/\beta}(\varphi,A)\\ +c(x)\cdot w(\phi)\quad\forall k\in[K],\ \forall\phi\in\Phi_k(\beta),\ \forall x\in\mathbb{R}.\end{aligned}$$

Note that the index $\phi\in\Phi_k(\beta)$ only affects the last term in this constraint, and the constraint is

thus equivalent to

$$\theta_k \leq \int_{(\varphi,A)\in\mathcal{E}'_k(L,\beta)} \mathbb{1}[x \in A]\,\mathrm{d}\psi_k(\varphi,A) - e^{\varepsilon}\cdot \int_{(-\varphi,A)\in\mathcal{E}'_k(L,\beta)} \mathbb{1}[x+\varphi \in A]\,\mathrm{d}\psi_{k-\varphi/\beta}(\varphi,A) + c(x)\cdot \underline{w}_k(\beta) \quad \forall k\in[K],\ \forall x\in\mathbb{R}.$$

We can equivalently express the index $x \in \mathbb{R}$ in this constraint with the double index $(i,x) \in \mathbb{Z}\times I_i(\beta)$, and arguments similar to those in the proof of Lemma 2 show that the constraint subsequently simplifies to

$$\theta_k \leq \int_{(\varphi,A)\in\mathcal{E}'_k(L,\beta)} \mathbb{1}[I_i(\beta) \subseteq A]\,\mathrm{d}\psi_k(\varphi,A) - e^{\varepsilon}\cdot \int_{(-\varphi,A)\in\mathcal{E}'_k(L,\beta)} \mathbb{1}[I_i(\beta)+\varphi \subseteq A]\,\mathrm{d}\psi_{k-\varphi/\beta}(\varphi,A) + \underline{c}_i(\beta)\cdot \underline{w}_k(\beta) \quad \forall k\in[K],\ \forall i\in\mathbb{Z}.$$

Similar arguments as in the proof of Proposition 3 allow us to further restrict the index $i \in \mathbb{Z}$ in the above constraint to $i \in [\pm(L+\Delta f/\beta)]$. The restriction (8b) also allows us to replace the measures $\psi_k \in \mathcal{M}_+(\mathcal{E}'(\phi))$ with discrete measures $\psi_k : \mathcal{E}'_k(L,\beta) \mapsto \mathbb{R}_+$, $k\in[K]$, and replace the integrals in the objective function and the constraints with sums. This results in the formulation $\mathrm{D}'(L,\beta)$ and thus concludes the proof. □

B.4 Proof of Theorem 2

We employ the same strategy as in the proof of Theorem 1. We define the auxiliary problem

$$\begin{array}{ll}
\underset{p}{\text{minimize}} & (\Delta f/\ell)\cdot \sum_{k\in[K]} w_k(\Delta f/\ell)\cdot \Big[\sum_{i\in[\pm(\Lambda\cdot\ell+\ell)]} c_i(\Delta f/\ell)\cdot p_k(i)\Big] \\
\text{subject to} & p_k : [\pm(\Lambda\cdot\ell+\ell)] \mapsto \mathbb{R}_+,\ \sum_{i\in[\pm(\Lambda\cdot\ell+\ell)]} p_k(i) = 1,\ k\in[K] \\
& \sum_{i\in[\pm(\Lambda\cdot\ell+\ell)]} \mathbb{1}[I_i(\Delta f/\ell)\subseteq A]\cdot p_k(i) \leq e^{\varepsilon}\cdot \sum_{i\in[\pm(\Lambda\cdot\ell+\ell)]} \mathbb{1}[I_i(\Delta f/\ell)+\varphi\subseteq A]\cdot p_m(i) + \delta \\
& \qquad \forall k,m\in[K],\ \forall(\varphi,A)\in\mathcal{E}'_{km}(\Lambda\cdot\ell,\Delta f/\ell), \qquad (\mathrm{M}'(\Lambda\cdot\ell,\Delta f/\ell))
\end{array}$$

and we show that the optimal values of $\mathrm{P}'(\Lambda\cdot\ell,\Delta f/\ell)$ and $\mathrm{D}'(\Lambda\cdot\ell,\Delta f/\ell)$ converge to that of $\mathrm{M}'(\Lambda\cdot\ell,\Delta f/\ell)$ when ℓ increases (which, in return, refines the granularity $\Delta f/\ell$) and Λ increases (which, in return, increases the support $[-\Lambda\cdot\Delta f,\ (\Lambda+1/\ell)\cdot\Delta f)$). Note that the number K of intervals in Φ depends on ℓ due to the partitioning $\Phi = \bigcup_{k\in[K]}\Phi_k(\Delta f/\ell)$.

To see that the optimal value of $\mathrm{P}'(\Lambda\cdot\ell,\Delta f/\ell)$ converges to that of $\mathrm{M}'(\Lambda\cdot\ell,\Delta f/\ell)$, we first note that $\mathrm{M}'(\Lambda\cdot\ell,\Delta f/\ell)$ differs from $\mathrm{P}'(\Lambda\cdot\ell,\Delta f/\ell)$ only in the existence of the additional decision variables $p_k(i)$, $i\in[\pm(\Lambda\cdot\ell+\ell)]\setminus[\pm\Lambda\cdot\ell]$, which also implies that $\mathrm{P}'(\Lambda\cdot\ell,\Delta f/\ell) \geq \mathrm{M}'(\Lambda\cdot\ell,\Delta f/\ell)$. Using similar arguments as in the proof of Lemma A.10, we can show that for any $\varepsilon>0$, $\delta>0$ and $\tau>0$, there is $\Lambda'\in\mathbb{N}$ such that any optimal solution $p^\star$ to $\mathrm{M}'(\Lambda\cdot\ell,\Delta f/\ell)$ satisfies $\sum_{i\in[\pm(\Lambda\cdot\ell+\ell)]\setminus[\pm\Lambda\cdot\ell]} p^\star_k(i) < \tau$ for all $k\in[K]$, $\ell\in\mathbb{N}$ and $\Lambda\geq\Lambda'$. Similar arguments as in the proof of Lemma A.11 then allow us to show that there is $\Lambda'\in\mathbb{N}$ such that $\mathrm{P}'(\Lambda'\cdot\ell,\Delta f/\ell) - \mathrm{M}'(\Lambda'\cdot\ell,\Delta f/\ell) \leq \xi$ for all $\ell\in\mathbb{N}$. For the remainder of the proof, we fix such a value of Λ'.

To see that the optimal value of $\mathrm{D}'(\Lambda' \cdot \ell, \Delta f/\ell)$ converges to that of $\mathrm{M}'(\Lambda' \cdot \ell, \Delta f/\ell)$, on the other hand, we note that $\mathrm{M}'(\Lambda' \cdot \ell, \Delta f/\ell)$ differs from the strong dual of $\mathrm{D}'(\Lambda' \cdot \ell, \Delta f/\ell)$ essentially only in the objective coefficients, which change from $\underline{w}_k(\Delta f/\ell)$ and $\underline{c}_i(\Delta f/\ell)$ in the strong dual of $\mathrm{D}'(\Lambda' \cdot \ell, \Delta f/\ell)$ to $w_k(\Delta f/\ell)$ and $c_i(\Delta f/\ell)$ in $\mathrm{M}'(\Lambda' \cdot \ell, \Delta f/\ell)$, respectively. Similar arguments as in the proof of Lemma A.12 show that for any $\varepsilon > 0$, $\delta > 0$ and $\xi > 0$, there exists $\ell' \in \mathbb{N}$ such that $\mathrm{M}'(\Lambda' \cdot \ell, \Delta f/\ell) - \mathrm{D}'(\Lambda' \cdot \ell, \Delta f/\ell) \leq \xi$ for all $\ell \geq \ell'$. Here, the uniform continuity of c ensures the convergence of the terms $c_i(\Delta f/\ell)$ and $\underline{c}_i(\Delta f/\ell)$, while our earlier assumption that w is a continuous probability density function ensures the convergence of the terms $w_k(\Delta f/\ell)$ and $\underline{w}_k(\Delta f/\ell)$. Moreover, since $\{w_k(\Delta f/\ell)\}_\ell$ and $\{\underline{w}_k(\Delta f/\ell)\}_\ell$ are non-negative and sum up to at most 1, the overall objective functions of both problem converge despite the growing number K of subsets of Φ.

So far, we have shown that there exists $\Lambda' \in \mathbb{N}$ and $\ell' \in \mathbb{N}$ such that $\mathrm{P}'(\Lambda' \cdot \ell, \Delta f/\ell) - \mathrm{D}'(\Lambda' \cdot \ell, \Delta f/\ell) \leq \xi$ holds for all $\ell \geq \ell'$. Since we have $\mathrm{P}'(\Lambda' \cdot \ell, \Delta f/\ell) \geq \mathrm{P}'(\Lambda \cdot \ell, \Delta f/\ell)$ and $\mathrm{D}'(\Lambda' \cdot \ell, \Delta f/\ell) \leq \mathrm{D}'(\Lambda \cdot \ell, \Delta f/\ell)$ for all $\Lambda \geq \Lambda'$, we can conclude the proof. Further details of this proof are relegated to the GitHub supplement.

C Proofs of Section 4

C.1 Proof of Corollary 1

$\mathrm{P}(\boldsymbol{\pi}, \beta)$ and $\mathrm{D}(\boldsymbol{\pi}, \beta)$ sandwich P and D from above and below since $\mathrm{P}(\boldsymbol{\pi}, \beta)$ and $\mathrm{D}(\boldsymbol{\pi}, \beta)$ constitute restrictions of the earlier problems $\mathrm{P}(L, \beta)$ and $\mathrm{D}(L, \beta)$ that satisfy the same inequality (*cf.* Lemmas 1 and 2 as well as Propositions 1 and 3).

In view of the second part of the statement, recall from Theorem 1 that there is $\Lambda' \in \mathbb{N}$ and $k' \in \mathbb{N}$ such that $\mathrm{P}(\Lambda \cdot k, \Delta f/k) - \mathrm{D}(\Lambda \cdot k, \Delta f/k) \leq \xi$ holds for all $\Lambda \geq \Lambda'$ and $k \geq k'$. Moreover, we have $\mathrm{P}(\boldsymbol{\pi}, \beta) \leq \mathrm{P}(\Lambda \cdot k, \Delta f/k)$ if $\{\Pi_j(\beta)\}_{j \in [N]}$ is a refinement of $\{I_i(\Delta f/k)\}_{i \in [\pm \Lambda \cdot k]}$. Indeed, $\mathrm{P}(\Lambda \cdot k, \Delta f/k)$ is equivalent to a variant of $\mathrm{P}(\boldsymbol{\pi}, \beta)$ that includes the additional constraints $p(j) = p(j')$ for all $j, j' \in [N]$ satisfying $\Pi_j(\beta), \Pi_{j'}(\beta) \subseteq I_i(\Delta f/k)$ for some $i \in [\pm \Lambda \cdot k]$. A similar argument shows that $\mathrm{D}(\boldsymbol{\pi}, \beta) \geq \mathrm{D}(\Lambda \cdot k, \Delta f/k)$, which concludes the proof.

C.2 Proof of Proposition 7

We prove Proposition 7 via three auxiliary results. Lemma C.16 proves that each inner loop over j in Algorithm 2 determines a worst-case event $A \in \arg\max\{V(\varphi, A) : A \in \mathcal{F}(L, \beta)\}$ for the query output difference $\varphi \in \mathscr{B}(\beta)$ fixed by the outer loop. Subsequently, Lemma C.17 proves that each outer loop over φ determines a maximally violated constraint (φ, A), which concludes the correctness of Algorithm 2. Finally, Lemma C.18 shows that Algorithm 2 can be implemented such that it determines a maximally violated constraint (φ, A) in time $\mathcal{O}(N^3)$.

Lemma C.16. *For any $\varphi \in \mathscr{B}(\beta)$ fixed by the outer loop of Algorithm 2, the event A constructed in the inner loop satisfies $A \in \arg\max\{V(\varphi, A) : A \in \mathcal{F}(L, \beta)\}$.*

Proof. Fix an arbitrary $\varphi \in \mathscr{B}(\beta)$ and recall that for any $A \in \mathcal{F}(L, \beta)$, the privacy shortfall

$V(\varphi, A)$ can be expressed as

$$\sum_{j\in[N]} p(j)\cdot\frac{|A\cap\Pi_j(\beta)|}{|\Pi_j(\beta)|} - e^{\varepsilon}\cdot\sum_{j\in[N]} p(j)\cdot\frac{|A\cap(\Pi_j(\beta)+\varphi)|}{|\Pi_j(\beta)|}$$
$$= \sum_{i\in[\pm L]} \mathbb{1}[I_i(\beta)\subseteq A]\cdot\left[\sum_{j\in[N]} p(j)\cdot\frac{|I_i(\beta)\cap\Pi_j(\beta)|}{|\Pi_j(\beta)|} - e^{\varepsilon}\cdot\sum_{j\in[N]} p(j)\cdot\frac{|(I_i(\beta)-\varphi)\cap\Pi_j(\beta)|}{|\Pi_j(\beta)|}\right]$$
$$= \sum_{i\in[\pm L]} \mathbb{1}[I_i(\beta)\subseteq A]\cdot\beta\cdot\left[\sum_{j\in[N]} p(j)\cdot\frac{\mathbb{1}[I_i(\beta)\subseteq\Pi_j(\beta)]}{|\Pi_j(\beta)|} - e^{\varepsilon}\cdot\sum_{j\in[N]} p(j)\cdot\frac{\mathbb{1}[I_i(\beta)\subseteq(\Pi_j(\beta)+\varphi)]}{|\Pi_j(\beta)|}\right],$$

where we disregard the constant $-\delta$ since it does not affect the relative order of privacy shortfalls across the constraints (φ, A). Here, the first identity exploits that $A = \bigcup_{i\in\mathcal{L}} I_i(\beta)$ for some $\mathcal{L}\subseteq[\pm L]$. The second identity holds since $|I_i(\beta)| = \beta$ and $I_i(\beta)$ is either entirely contained in or intersection free with $\Pi_j(\beta)$ and $\Pi_j(\beta)+\varphi$, $i\in[\pm L]$ and $j\in[N]$. Thus, $I_i(\beta)$ must be contained in the worst-case event A whenever

$$\sum_{j\in[N]} p(j)\cdot\frac{\mathbb{1}[I_i(\beta)\subseteq\Pi_j(\beta)]}{|\Pi_j(\beta)|} - e^{\varepsilon}\cdot\sum_{j'\in[N]} p(j')\cdot\frac{\mathbb{1}[I_i(\beta)\subseteq\Pi_{j'}(\beta)+\varphi]}{|\Pi_{j'}(\beta)|} \tag{22}$$

is strictly positive; $I_i(\beta)$ can be (but does not have to be) included in A if the above quantity is zero; and it must not be contained in A if the above quantity is negative. In the remainder, we fix $i\in[\pm L]$ and distinguish between two cases: *(i)* there is no $j'\in[N]$ satisfying $I_i(\beta)\subseteq\Pi_{j'}(\beta)+\varphi$; and *(ii)* there is $j'\in[N]$ satisfying $I_i(\beta)\subseteq\Pi_{j'}(\beta)+\varphi$.

In case *(i)*, we can include $I_i(\beta)$ in the worst-case event A since the second term in (22) vanishes, whereas the first term is always non-zero by construction. Note that the events A_j, $j\in[N]$, constructed in the first part of the inner loop of Algorithm 2 comprise precisely all intervals $I_i(\beta)$ falling under case *(i)*.

In case *(ii)*, the existence of $j'\in[N]$ satisfying $I_i(\beta)\subseteq\Pi_{j'}(\beta)+\varphi$ implies that (22) equals to $p(j)/|\Pi_j(\beta)| - e^{\varepsilon}\cdot p(j')/|\Pi_{j'}(\beta)|$ for some $j\in[N]$, and $I_i(\beta)$ should be included in the worst-case event if this quantity is positive. Note that the events $A_{jj'}$, $j,j'\in[N]$, constructed in the second part of the inner loop of Algorithm 2 comprise precisely all intervals $I_i(\beta)$ falling under case *(ii)* that satisfy $p(j)/|\Pi_j(\beta)| - e^{\varepsilon}\cdot p(j')/|\Pi_{j'}(\beta)| > 0$. □

Lemma C.17. *Algorithm 2 returns a constraint (φ, A) with maximum privacy shortfall.*

Proof. Recall that the DP constraints of $\mathrm{P}(\boldsymbol{\pi},\beta)$ are indexed by $(\varphi, A)\in\mathcal{E}(L,\beta) = \mathscr{B}(\beta)\times\mathcal{F}(L,\beta)$ and that Lemma C.16 proved that for any fixed $\varphi\in\mathscr{B}(\beta)$, the inner loop of Algorithm 2 constructs a worst-case event A associated with φ. We show in this proof that it is sufficient to consider the values $\varphi\in\mathscr{B}(\beta,\boldsymbol{\pi})\subseteq\mathscr{B}(\beta)$, where

$$\mathscr{B}(\beta,\boldsymbol{\pi}) := \{(\pi_j-\pi_{j'})\cdot\beta \;:\; (\pi_j-\pi_{j'})\cdot\beta\in[-\Delta f,\Delta f] \text{ and } j,j'\in[N]\}\cup\{-\Delta f,\Delta f\},$$

which is what the outer loop of Algorithm 2 does.

Our earlier arguments of this section have shown that for any $\varphi\in\mathscr{B}(\beta)$, the maximum

privacy shortfall $\max\{V(\varphi, A) : A \in \mathcal{F}(L, \beta)\}$ satisfies

$$\sum_{j\in[N]} |A_j(\varphi)| \cdot \frac{p(j)}{|\Pi_j(\beta)|} + \sum_{j,j'\in[N]} |A_{jj'}(\varphi)| \cdot \left[\frac{p(j)}{|\Pi_j(\beta)|} - e^{\epsilon} \cdot \frac{p(j')}{|\Pi_{j'}(\beta)|}\right]^+ - \delta,$$

where $[x]^+ = \max\{0, x\}$ and the only quantities varying with φ are

$$A_j(\varphi) = \Pi_j(\beta) \setminus [-L \cdot \beta + \varphi, (L+1) \cdot \beta + \varphi) \quad \text{and} \quad A_{jj'}(\varphi) = \Pi_j(\beta) \cap (\Pi_{j'}(\beta) + \varphi), \;\; j, j' \in [N].$$

One readily verifies that both $|A_j(\varphi)|$ and $|A_{jj'}(\varphi)|$, $j, j' \in [N]$, are affine between any two consecutive points in $\mathscr{B}(\beta, \boldsymbol{\pi})$. In other words, the maximum privacy shortfall is piecewise affine with breakpoints $\mathscr{B}(\beta, \boldsymbol{\pi})$ or a subset thereof, which implies that its maximum must be attained at one of the points $\varphi \in \mathscr{B}(\beta, \boldsymbol{\pi})$. This concludes the proof. □

Lemma C.18. *Algorithm 2 can be implemented such that it terminates in time $\mathcal{O}(N^3)$.*

Proof. Since all individual steps in Algorithm 2 take constant time, the runtime of the algorithm is determined by the numbers of iterations in the outer and inner loops. In the naïve implementation of the main text, both loops comprise $\mathcal{O}(N^2)$ iterations, and thus the overall complexity of that implementation is $\mathcal{O}(N^4)$. We show in this proof that the inner loop can be implemented such that it comprises $\mathcal{O}(N)$ iterations only, which implies the statement of the lemma.

Note that the inner loop in Algorithm 2 constructs all events A_j, $j \in [N]$, in time $\mathcal{O}(N)$, and thus we only need to consider the construction of the events $A_{jj'}$, $j, j' \in [N]$. Instead of the naïve implementation from the main text, which probes all pairs of subsets $(j, j') \in [N]^2$, we consider the following variant of the Bentley-Ottmann algorithm used to identify crossings in a set of line segments (Berg et al. 2000, §2): We merge the two lists of tuples $\{(\pi_j \cdot \beta, 1) : j \in [N+1]\}$ and $\{(\pi_{j'} \cdot \beta + \varphi, 2) : j' \in [N+1]\}$ in order of non-decreasing first component; since each list is already sorted, this can be achieved in time $\mathcal{O}(N)$. We initialize the two index counters $j_1 = j_2 = 0$ and loop through the entire merged list of tuples once in sorted order. Whenever we encounter an element $(\pi_j, 1)$, we update $j_1 \leftarrow \pi_j$; whenever we encounter an element $(\pi_{j'}, 2)$, we update $j_2 \leftarrow \pi_{j'}$. After each update, we consider the intersection $A_{jj'} = \Pi_{j_1}(\beta) \cap (\Pi_{j_2}(\beta) + \varphi)$ for possible inclusion in the worst-case event A whenever $(j_1, j_2) \neq (0, 0)$. Since the merged list of tuples has length $2N + 2$, the overall algorithm evidently runs in time $\mathcal{O}(N)$. □

Proof of Proposition 7 The proof immediately follows from Lemmas C.16, C.17 and C.18. □

C.3 Bounding P′ and D′ with Non-Uniform Partitions

We next extend the non-uniform upper and lower bounding problems $\mathrm{P}(\boldsymbol{\pi}, \beta)$ and $\mathrm{D}(\boldsymbol{\pi}, \beta)$ of Section 4.1 to the data dependent case. We obtain the following upper bound on problem P′,

$$\begin{array}{ll} \underset{p}{\text{minimize}} & \beta \cdot \sum\limits_{k\in[K]} \omega_k(\beta) \cdot \Big[\sum\limits_{j\in[N]} c_j(\boldsymbol{\pi}, \beta) \cdot p_k(j) \Big] \\ \text{subject to} & p_k : [N] \mapsto \mathbb{R}_+, \ \sum\limits_{j\in[N]} p_k(j) = 1, \ k \in [K] \\ & \sum\limits_{j\in[N]} p_k(j) \cdot \dfrac{|A \cap \Pi_j(\beta)|}{|\Pi_j(\beta)|} \leq e^{\varepsilon} \cdot \sum\limits_{j\in[N]} p_m(j) \cdot \dfrac{|(A-\varphi) \cap \Pi_j(\beta)|}{|\Pi_j(\beta)|} + \delta \\ & \qquad\qquad \forall k, m \in [K], \ \forall (\varphi, A) \in \mathcal{E}'_{km}(L, \beta), \end{array} \tag{$\mathrm{P}'(\boldsymbol{\pi}, \beta)$}$$

as well as the following lower bound on problem D′,

$$\begin{array}{ll} \underset{p}{\text{minimize}} & \beta \cdot \sum\limits_{k\in[K]} \omega_k(\beta) \cdot \Big[\sum\limits_{j\in\mathfrak{N}} \underline{c}_j(\boldsymbol{\pi}, \beta) \cdot p_k(j) \Big] \\ \text{subject to} & p_k : \mathfrak{N} \mapsto \mathbb{R}_+, \ \sum\limits_{j\in\mathfrak{N}} p_k(j) = 1, \ k \in [K] \\ & \sum\limits_{j\in\mathfrak{N}} p_k(j) \cdot \dfrac{|A \cap \Pi_j(\beta)|}{|\Pi_j(\beta)|} \leq e^{\varepsilon} \cdot \sum\limits_{j\in\mathfrak{N}} p_m(j) \cdot \dfrac{|(A-\varphi) \cap \Pi_j(\beta)|}{|\Pi_j(\beta)|} + \delta \\ & \qquad\qquad \forall k, m \in [K], \ \forall (\varphi, A) \in \mathcal{E}'_{km}(L, \beta). \end{array} \tag{$\mathrm{D}'(\boldsymbol{\pi}, \beta)$}$$

Algorithm 3 extends the ideas of Algorithm 2 to the data dependent setting; as before, extending the domain of the decisions p from $[N]$ to $\mathfrak{N}$ allows us to employ the same algorithm to solve the lower bounding problem $\mathrm{D}'(\boldsymbol{\pi}, \beta)$ as well. Similar arguments as in the previous section show that Algorithm 3 terminates in time $\mathcal{O}(K^2 \cdot N)$. We explain in the GitHub supplement of this paper how the bounding problems $\mathrm{P}'(\boldsymbol{\pi}, \beta)$ and $\mathrm{D}'(\boldsymbol{\pi}, \beta)$, as well as Algorithm 3, can be generalized to account for non-uniform partitions of the set of possible query outputs Φ as well; this reduces computation times when the granularity parameter β is small.

Algorithm 3: *Identification of a constraint in* $\mathrm{P}'(\boldsymbol{\pi}, \beta)$ *with maximum privacy shortfall*

input : $\boldsymbol{\pi},\ \beta,\ p,\ \Delta f$
output: constraint $(\varphi^\star, A^\star)$ with maximum privacy shortfall $V(\varphi^\star, A^\star)$
Initialize $V^\star = 0$;
for $k, m \in [K]$ **do**
  for $\varphi \in \{(m-k-1)\cdot\beta,\ (m-k)\cdot\beta,\ (m-k+1)\cdot\beta\} \cap [-\Delta f, \Delta f]$ **do**
    Initialize $A = \emptyset$ and $V = 0$;
    for $j = 1, \ldots, N$ **do**
      Let $A_j = \Pi_j(\beta) \setminus [-L\cdot\beta + \varphi, (L+1)\cdot\beta + \varphi)$ and update

$$A = A \cup A_j, \quad V = V + |A_j| \cdot \frac{p_k(j)}{|\Pi_j(\beta)|}.$$

      for $j' = 1, \ldots, N$ **do**
        if $p_k(j)/|\Pi_j(\beta)| > e^\varepsilon \cdot p_m(j')/|\Pi_{j'}(\beta)|$ **then**
          Let $A_{jj'} = \Pi_j(\beta) \cap (\Pi_{j'}(\beta) + \varphi)$ and update

$$A = A \cup A_{jj'}, \quad V = V + |A_{jj'}| \cdot \left[\frac{p_k(j)}{|\Pi_j(\beta)|} - e^\varepsilon \cdot \frac{p_m(j')}{|\Pi_{j'}(\beta)|}\right].$$

        end
      end
    end
    if $V > V^\star$ **then**
      Update $\varphi^\star = \varphi$, $A^\star = A$ and $V^\star = V$.
    end
  end
end
return $(\varphi^\star, A^\star)$ and $V^\star(\varphi, A) = V^\star - \delta$.

D Additional Numerical Experiments

Table 1 summarized the suboptimality of the best performing benchmark mechanisms by summing the upper and lower bound gaps. Tables 4 and 5 report these gaps separately.

ε \ δ	**0.005**	**0.010**	**0.020**	**0.050**	**0.100**	**0.200**	**0.250**	**0.300**	**0.500**	**0.750**
0.005	0.17%	0.12%	0.03%	0.26%	0.26%	0.65%	0.64%	0.63%	0.49%	8.37%
0.010	0.41%	0.33%	0.58%	0.54%	0.54%	0.68%	0.55%	0.55%	0.55%	5.26%
0.020	0.28%	0.41%	0.04%	0.58%	0.58%	0.73%	0.61%	1.24%	0.61%	8.53%
0.050	0.50%	0.27%	0.38%	0.45%	0.73%	0.89%	0.80%	1.48%	0.87%	8.73%
0.100	0.38%	0.45%	0.58%	0.54%	0.90%	1.14%	1.15%	1.88%	1.25%	9.12%
0.200	0.78%	0.87%	1.02%	0.92%	0.96%	1.66%	1.92%	2.61%	1.80%	10.24%
0.500	1.81%	1.91%	2.09%	2.58%	2.71%	3.35%	3.92%	4.08%	4.33%	12.91%
1.000	4.70%	4.72%	4.66%	4.84%	5.10%	5.82%	6.03%	6.35%	8.09%	16.38%
2.000	7.99%	8.02%	8.09%	8.39%	8.70%	9.60%	10.18%	10.86%	13.05%	19.74%
5.000	12.25%	12.28%	12.33%	12.50%	12.78%	13.37%	13.67%	13.97%	14.86%	16.74%

Table 4: *Upper bound suboptimality of the best performing benchmark mechanisms on synthetic data independent instances with $\Delta f = 1$, ℓ_1-loss and various combinations of ε and δ.*

ε \ δ	**0.005**	**0.010**	**0.020**	**0.050**	**0.100**	**0.200**	**0.250**	**0.300**	**0.500**	**0.750**
0.005	1.71%	1.09%	1.18%	5.84%	18.27%	2.90%	48.94%	22.55%	49.47%	24.97%
0.010	2.48%	1.09%	7.42%	1.77%	16.55%	2.40%	48.63%	22.48%	49.36%	28.09%
0.020	2.56%	2.42%	0.52%	14.50%	13.61%	66.71%	47.78%	21.50%	49.23%	24.84%
0.050	1.35%	1.84%	2.18%	17.92%	5.50%	65.14%	45.35%	20.47%	48.72%	24.69%
0.100	4.40%	3.73%	8.10%	18.83%	32.42%	62.71%	41.72%	18.97%	47.92%	24.38%
0.200	9.65%	8.73%	7.88%	16.37%	18.86%	58.51%	35.79%	16.74%	46.53%	23.39%
0.500	21.29%	19.50%	23.36%	13.78%	35.99%	39.20%	25.60%	14.41%	41.52%	20.88%
1.000	35.41%	35.01%	35.57%	35.56%	23.52%	26.71%	20.59%	15.08%	33.71%	16.98%
2.000	25.97%	26.16%	25.36%	23.24%	23.76%	19.10%	16.94%	14.81%	21.29%	10.77%
5.000	7.07%	7.03%	6.96%	6.74%	6.37%	5.64%	5.27%	4.91%	3.79%	2.33%

Table 5: *Lower bound suboptimality of the best performing benchmark mechanisms on synthetic data independent instances with $\Delta f = 1$, ℓ_1-loss and various combinations of ε and δ.*