THE OPTIMAL MECHANISM IN DIFFERENTIAL PRIVACY

BY

QUAN GENG

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2013

Urbana, Illinois

Doctoral Committee:

Professor Pramod Viswanath, Chair
Professor Bruce Hajek
Professor Rayadurgam Srikant
Professor Nitin Vaidya

ABSTRACT

Differential privacy is a framework to quantify to what extent individual privacy in a statistical database is preserved while releasing useful aggregate information about the database. This dissertation studies the fundamental trade-off between privacy and utility in differential privacy in the most basic problem settings.

We first derive the optimal ϵ-differentially private mechanism for single real-valued query function under a very general utility-maximization (or cost-minimization) framework. The class of noise probability distributions in the optimal mechanism has *staircase-shaped* probability density functions which are symmetric (around the origin), monotonically decreasing and geometrically decaying. The staircase mechanism can be viewed as a *geometric mixture of uniform probability distributions*, providing a simple algorithmic description for the mechanism. Furthermore, the staircase mechanism naturally generalizes to discrete query output settings as well as more abstract settings. We explicitly derive the parameter of the optimal staircase mechanism for ℓ_1 and ℓ_2 cost functions. Comparing the optimal performances with those of the usual Laplacian mechanism, we show that in the high privacy regime (ϵ is small), the Laplacian mechanism is asymptotically optimal as $\epsilon \to 0$; in the low privacy regime (ϵ is large), the minimum magnitude and second moment of noise are $\Theta(\Delta e^{-\frac{\epsilon}{2}})$ and $\Theta(\Delta^2 e^{-\frac{2\epsilon}{3}})$ as $\epsilon \to +\infty$, respectively, while the corresponding figures when using the Laplacian mechanism are $\frac{\Delta}{\epsilon}$ and $\frac{2\Delta^2}{\epsilon^2}$, where Δ is the sensitivity of the query function. We conclude that the gains of the staircase mechanism are more pronounced in the low privacy regime.

We also show the optimality of the staircase mechanism for ϵ-differentially privacy in the multiple dimensional setting where the query output has multiple components, e.g., histogram query function. We prove that when the dimension is two, for the ℓ^1 cost function, the noise probability distribution in the optimal mechanism has a multiple dimensional staircase-shaped probability density function. We explicitly derive the parameter of the optimal two-dimensional staircase mechanism, and study the asymptotical performance of optimal mechanism in the high and low privacy regimes. Comparing the optimal performances with those of the usual Laplacian mechanism, we show that in the high privacy regime (ϵ is small), the Laplacian mechanism is asymptotically optimal as $\epsilon \to 0$; in the low

privacy regime (ϵ is large), the optimal cost is $\Theta(e^{-\frac{\epsilon}{3}})$, while the cost of the Laplacian mechanism is $\frac{2\Delta}{\epsilon}$. We conclude that the gains of the staircase mechanism are more pronounced in the low privacy regime.

Lastly, we study the optimal mechanisms in (ϵ, δ)-differential privacy for integer-valued query functions under a utility-maximization/cost-minimization framework. We show that the (ϵ, δ)-differential privacy is a framework not much more general than the $(\epsilon, 0)$-differential privacy and $(0, \delta)$-differential privacy in the context of ℓ^1 and ℓ^2 cost functions, i.e., minimum expected noise magnitude and noise power. In the same context of ℓ^1 and ℓ^2 cost functions, we show the near-optimality of uniform noise mechanism and discrete Laplacian mechanism in the high privacy regime (as $(\epsilon, \delta) \to (0, 0)$).

To my father, mother and brother.
To my teachers and mentors.
To all the people who have positively changed my perspective on life.

ACKNOWLEDGMENTS

First and foremost, I want to express my sincerest gratitude to my advisor, Professor Pramod Viswanath, from whom I learned everything about research. As a terrific mentor, he is very patient and has meticulously guided me in research and in my career. His sharp thinking, deep intuition, and brilliant insights of research have greatly changed the way I think of doing research. I want to thank him for supporting me, encouraging me, and having faith in me, even though I did not get very good results in my first three years as a PhD student. I will never forget his words "negative result is also positive knowledge," which will continue to motivate me to work actively in every aspect of life even when facing difficulties and setbacks. I will treasure the conversations and interactions I had with him in every meeting in the past four and half years, through each of which I learned not only new knowledge, but also gained his wisdom and sagacious perspective on life.

I would also like to thank Prof. Bruce Hajek, Prof. Rayadurgam Srikant, and Prof. Nitin Vaidya for serving on my dissertation committee and giving me valuable feedback.

I want to thank the faculty of the departments of Electrical and Computer Engineering, Computer Science, and Mathematics for offering great curricula. Taking courses turned out to be a substantial part of my PhD life, and I am fortunate to have had the privilege of interacting with the best instructors in the world, who have made my graduate life an extremely wonderful learning experience. I would like to thank all my course instructors and teachers in the past four and half years (in chronological order): Enlu Zhou, Igor Nikolaev, Richard Blahut, Todd Coleman, Pramod Viswanath, Venugopal Veeravalli, Chandra Chekuri, Derek Robinson, Zoltan Furedi, Zhong-Jin Ruan, M. Burak Erdogan, Sariel Har-Peled, Richard B. Sowers, Eugene M. Lerman, Daniel Liberzon, Timothy E. Weninger, Ryan M. Cunningham, Wade A. Fagen, Dan Roth, Elsa Gunter, Jiawei Han, David M. Nicol, and Rakesh Bobba.

I have been fortunate to work with many talented colleagues and friends in the Coordinated Science Laboratory: Sreekanth Annapureddy, Daniel Cullina, Aly El Gamal, Xun Gong, Xiaoyu Guang, Abhishek Gupta, Dayu Huang, Zhenqi Huang, Tianxiong Ji, Sachin Kadloor, Sreeram Kannan, Farzaneh Khajouei, Yun Li, Jonathan Ligo, Rui Ma, Siva Theja Maguluri, Anshuman Mishra, Jian Ni, Sevinc Figen Oktem, Christopher Quinn, Fardad

Raisali, Adnan Raja, Lili Su, Bo Tan, Shaileshh Venkatakrishnan, Rui Wu, Qiaomin Xie, Jiaming Xu, Yunwen Xu, Tao Yang, Yifei Yang, Ge Yu, Mindi Yuan, Ji Zhu, and many others. The daily conversations and discussions with them have made my graduate life enjoyable. In particular, I would like to thank my academic brother and collaborator, Sreeram Kannan, who has brought so much inspiration to my life. I will always remember the two-hour conversation about religion with him before midnight in the Information Systems Laboratory at Stanford University in 2011, which has completely changed my view on religion. I would also like to thank my roommate, office mate and course mate, Ji Zhu, for sharing with me his endless passion for life, and for helping me mature in many aspects as a man.

During the past four and half years, I have spent three wonderful summers as an intern in three different interesting places. I want to thank my mentors Yi Ma and John Wright (Microsoft Research Asia) for leading me into the exciting area of compressed sensing, my mentors Xinzhou Wu and Sundar Subramanian (Qualcomm Flarion Technologies) for patiently guiding me in doing engineering research in the promising vehicle-to-vehicle wireless communication systems, and my mentors Ronna Cong and David Wang (Tower Research) for introducing me to the wonderland of high-frequency trading and offering me the opportunity to continue to explore this exciting area after graduation.

In my last year in Champaign, I was very fortunate to serve as the president of Tsinghua Alumni Association at UIUC, in which I have met many excellent and brilliant friends. I would like to thank my friends Zhenqi Huang, Peibei Shi, and Xin Zhang for helping me organize so many wonderful social parties and activities for this community. The success of our community would not have been even possible without their support and selfless contributions.

Dancing is another integral part of my graduate life besides research and coursework. I want to thank the Illini Dancesports Team for offering the excellent dance courses, and helping us prepare for various dance competitions. I thank Wei Dong, Allen Gehret, Kyong Noh, and Yan Zhou for patiently helping me improve my dance skills, and thank all my dance partners Danjie Xuan, Jiongjing Guo, Yinghong Zhu, Patricia Carbajales, Conghui Yang, Fanglin Lu and Ye Feng, for sharing with me their love for ballroom dance. In particular, I want to thank my last dance partner, Ye Feng, whose vigorous passion for dance and life continues to inspire me.

I want to thank Wanfei for showing up in my life during the final stage of my PhD life. I will cherish the wonderful time we have spent together in New York, Chicago, and Champaign. I would like to thank her for making me deeply appreciate the beauty of love in every aspect, and making my life as exciting as my research.

Finally, I would like to express my deepest gratitude to my father and mother, for their

unconditional faith, love, and support. Without their support, none of my achievements would have been possible. To them I dedicate this dissertation.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Motivation and Background

Due to the advances of information technology and the prevalence of social networks, vast amounts of personal information can be efficiently collected and processed. The availability of such *big data* enables us to produce a number of useful applications by analyzing the collected information. For example, Netflix, an online DVD-rental service, can use subscribers' preferences for movies to build a movie recommendation system; a user on Amazon.com, the world's largest online retailer, can use other users' reviews and ratings to decide which item might be good and reliable.

While releasing the statistics of collected data has great potential use for analysis, without proper statistical disclosure mechanisms, the privacy of individuals in the dataset can be jeopardized. In 2006, Netflix hosted the Netflix prize contest, an open competition in which contestants designed algorithms to make predictions on user ratings for movies from the released dataset by Netflix. The training dataset provided by Netflix consisted of 100,480,507 ratings that 480,189 users gave to 17,770 movies [1]. To protect the privacy of customers in the released training dataset, Netflix anonymized the dataset by removing users' personal information, e.g., name, from the dataset, and only used integer IDs. However, these anonymization approaches were not sufficient to preserve customers' privacy. By connecting the released anonymized dataset by Netflix and the Internet Movie Database as background knowledge, Narayanan and Shmatikov [2] successfully de-anonymized and identified some Netflix records of known users.

As can be seen from the Netflix prize example, one difficulty for defining a notion of *privacy* which is resilient to attacks is to model the side information of adversaries.

Differential privacy is a recent formal notion of *privacy*, and it separates the issues of modeling adversary side information by requiring the indistinguishability of whether an individual is in the dataset or not based on the released information. The key idea of differential privacy is that the presence or absence of any individual data in the database should not affect the final released statistical information significantly, and thus it can give

strong privacy guarantees against an adversary with arbitrary auxiliary information. For more motivation and background of differential privacy, we refer the readers to the survey by Dwork [3].

The basic problem setting in differential privacy for a statistical database is as follows: suppose a dataset curator is in charge of a statistical database which consists of records of many individuals, and an analyst sends a query request to the curator to get some aggregate information about the whole database. The curator can simply compute the query output by applying the query function to the whole database and send the query output directly to the analyst. However, this approach may not provide privacy guarantees on each individual record in the database. To satisfy the differential privacy constraint, a query-releasing mechanism needs to send a randomized query output to the analyst in a way such that the probability distribution of the query output does not differ too much, whether or not any individual record is in the database.

The standard mechanism to achieve differential privacy is to perturb the query output by adding random noise. If noise is sufficiently large and random, it will help preserve the differential privacy while the utility which the analyst can get from the query output will deteriorate. On the other hand, if the noise is very small, while the analyst can get high utility, it may not satisfy the given privacy constraint. Clearly, there exists a trade-off between privacy and utility.

In many existing works studying the trade-off between accuracy and privacy in differential privacy, the usual metric of accuracy is the variance, or magnitude expectation of the noise added to the query output. For example, Hardt and Talwar [4] study the trade-off between privacy and error for answering a set of linear queries over a histogram in a differentially private way, where the error is defined as the worst expectation of the ℓ^2-norm of the noise among all possible query output. Hardt and Talwar [4] derive lower and upper bounds on the error given the differential privacy constraint. Nikolov, Talwar, and Zhang [5] extend the result on the trade-off between privacy and error to the case of (ϵ, δ)-differential privacy. Li et al. [6] study how to optimize linear counting queries under differential privacy, where the error is measured by the mean squared error of query output estimates, which corresponds to the variance of the noise added to the query output to preserve differential privacy.

More generally, the error can be a general function depending on the additive noise (distortion) to the query output. Ghosh, Roughgarden, and Sundararajan [7] study a very general utility-maximization framework for a single count query with sensitivity one under differential privacy, where the utility (cost) function can be a general function depending on the noise added to the query output. They show that there exists a universally optimal mechanism (adding geometric noise) to preserve differential privacy for a general class of

utility functions under a Bayesian framework. Brenner and Nissim [8] show that for general query functions, no universally optimal differential privacy mechanisms exist. Gupte and Sundararajan [9] generalize the result of [7] to a minimax setting.

The main theme of this dissertation is to delve into fundamental limits of data privacy and derive the optimal mechanisms to preserve differential privacy in the most basic problem settings, as opposed to preserving privacy for each and every application setting, as is done in most works in the literature.

1.2 Our Contribution

In this dissertation, we study the fundamental trade-off between privacy and utility of differential privacy in the most basic problem settings. Our results can be summarized as follows:

- *ϵ-differential privacy in the single dimensional setting*:

  Given the differential privacy constraint, we derive the optimal differentially private mechanism for a single real-valued query function under a general utility-maximization (or cost-minimization) framework. The class of noise probability distributions in the optimal mechanism has *staircase-shaped* probability density functions which are symmetric (around the origin), monotonically decreasing and geometrically decaying. The staircase mechanism can be viewed as a *geometric mixture of uniform probability distributions*, providing a simple algorithmic description for the mechanism. Furthermore, the staircase mechanism naturally generalizes to discrete query output settings as well as more abstract settings. We show that adding query-output independent noise with staircase distribution is optimal among all randomized mechanisms (subject to a mild technical condition) that preserve differential privacy.

  We explicitly derive the optimal noise probability distributions with minimum expectation of noise amplitude and power. Comparing the optimal performances with those of the Laplacian mechanism, we show that in the high privacy regime, the Laplacian mechanism is asymptotically optimal; in the low privacy regime, the staircase mechanism significantly outperforms the Laplacian mechanism. We conclude that the gains are more pronounced in the low privacy regime.

- *ϵ-differential privacy in the multiple dimensional setting*:

  We extend the staircase mechanism from the single dimensional setting to the multiple dimensional setting. We show that for histogram-like query functions, when the

dimension of the query output is two, the multiple dimensional staircase mechanism is optimal for the ℓ^1 cost function. We study the asymptotical performance of optimal mechanisms in the high and low privacy regimes. Comparing the optimal performances with those of the Laplacian mechanism, we conclude that in the multiple dimensional setting, the Laplacian mechanism is asymptotically optimal in the high privacy regime, and the staircase mechanism significantly outperforms the Laplacian mechanism in the low privacy regime.

- (ϵ, δ)-*differential privacy*:

  We study the optimal mechanisms in (ϵ, δ)-differential privacy for integer-valued query functions under a utility-maximization/cost-minimization framework. We show that the (ϵ, δ)-differential privacy is a framework not much more general than the $(\epsilon, 0)$-differential privacy and $(0, \delta)$-differential privacy in the context of ℓ^1 and ℓ^2 cost functions, i.e., minimum expected noise magnitude and noise power. In the same context of ℓ^1 and ℓ^2 cost functions, we show the near-optimality of the uniform noise mechanism and the discrete Laplacian mechanism in the high privacy regime (as $(\epsilon, \delta) \to (0, 0)$).

1.3 Organization

The rest of this dissertation is organized as follows. Chapter 2 studies the optimal mechanism in the standard ϵ-differential privacy setting for a single real-valued query function, and presents our main result on the optimality of the *staircase mechanism* for a general class of cost functions. Chapter 3 shows the optimality of the staircase mechanism in the multi-dimensional setting in which the query output has multiple components for the ℓ^1 cost function. Chapter 4 studies the (approximately) optimal mechanisms in (ϵ, δ)-differential privacy. We concludes this dissertation in Chapter 5.

CHAPTER 2

THE OPTIMAL MECHANISM IN ϵ-DIFFERENTIAL PRIVACY: SINGLE DIMENSIONAL SETTING

In this chapter, we study the optimal mechanism in ϵ-differential privacy under a utility-maximization framework. We first give the background on differential privacy in Section 2.1, then give the precise problem formulation in Section 2.2. Section 2.3 gives an overview of our main results on the optimality mechanism in ϵ-differential privacy. We show the optimality of query-output independent perturbation in Section 2.4, and present the optimal differentially private mechanism, staircase mechanism, in Section 2.5. In Section 2.6, we apply our main result to derive the optimal noise probability distribution with minimum expectation of noise amplitude and power, respectively, and compare the performances with the Laplacian mechanism. Section 2.7 presents the asymptotic properties of γ^* in the staircase mechanism for momentum cost functions, and suggests a heuristic choice of γ that appears to work well for a wide class of cost functions. Section 2.8 generalizes the staircase mechanism for integer-valued query functions in the discrete setting, and Section 2.9 extends the staircase mechanism to the abstract setting. Section 2.10 discusses the connection between our work and the literature.

2.1 Background on Differential Privacy

The basic problem setting in differential privacy for a statistical database is as follows: suppose a dataset curator is in charge of a statistical database which consists of records of many individuals, and an analyst sends a query request to the curator to get some aggregate information about the whole database. Without any privacy concerns, the database curator can simply apply the query function to the dataset, compute the query output, and send the result to the analyst. However, to protect the privacy of individual data in the dataset, the dataset curator should use a randomized query-answering mechanism such that the probability distribution of the query output does not differ too much, whether or not any individual record is in the database.

Formally, consider a real-valued query function

$$q : \mathcal{D}^n \to \mathbb{R},$$

where $\mathcal{D}^n$ is the set of all possible datasets. The real-valued query function q will be applied to a dataset, and the query output is a real number. Two datasets $D_1, D_2 \in \mathcal{D}^n$ are called neighboring datasets if they differ in at most one element, i.e., one is a proper subset of the other and the larger dataset contains just one additional element [3]. A randomized query-answering mechanism $\mathcal{K}$ for the query function q will randomly output a number whose probability distribution depends on query output $q(D)$, where D is the dataset.

Definition 2.1 (ϵ-Differential Privacy [3])**.** *A randomized mechanism $\mathcal{K}$ gives ϵ-differential privacy if for all data sets D_1 and D_2 differing in at most one element, and all $S \subset Range(\mathcal{K})$,*

$$Pr[\mathcal{K}(D_1) \in S] \leq \exp(\epsilon)\ Pr[\mathcal{K}(D_2) \in S], \tag{2.1}$$

where $\mathcal{K}(D)$ is the random output of the mechanism $\mathcal{K}$ when the query function q is applied to the dataset D.

The differential privacy constraint (2.1) essentially requires that for all neighboring datasets, the probability distributions of the output of the randomized mechanism should be approximately the same. Therefore, for any individual record, its presence or absence in the dataset will not significantly affect the output of the mechanism, which makes it hard for adversaries with arbitrary background knowledge to make inferences about any individual from the released query output information. The parameter $\epsilon \in (0, +\infty)$ quantifies how private the mechanism is: the smaller ϵ is , the more private the randomized mechanism is.

2.1.1 Operational Meaning of ϵ-Differential Privacy in the Context of Hypothesis Testing

As shown by [10], one can interpret the differential privacy constraint (2.1) in the context of hypothesis testing in terms of false alarm probability and missing detection probability. Indeed, consider a binary hypothesis-testing problem over two neighboring datasets, $H_0 : D_1$ versus $H_1 : D_2$, where an individual's record is in D_2 only. Given a decision rule, let S be the decision region such that when the released output lies in S, H_1 will be rejected, and when the released output lies in S^C (the complement of S), H_0 will be rejected. The false

alarm probability P_{FA} and the missing detection probability P_{MD} can be written as

$$P_{FA} = P(K(D_1) \in S^C),$$
$$P_{MD} = P(K(D_2) \in S).$$

Therefore, from (2.1) we get

$$1 - P_{FA} \leq e^{\epsilon} P_{MD}.$$

Thus

$$e^{\epsilon} P_{MD} + P_{FA} \geq 1.$$

Switch D_1 and D_2 in (2.1), and we get

$$\Pr[\mathcal{K}(D_2) \in S] \leq \exp(\epsilon) \Pr[\mathcal{K}(D_1) \in S].$$

Therefore,

$$1 - P_{MD} \leq e^{\epsilon} P_{FA},$$

and thus

$$P_{MD} + e^{\epsilon} P_{FA} \geq 1.$$

In conclusion, we have

$$e^{\epsilon} P_{MD} + P_{FA} \geq 1,$$
$$P_{MD} + e^{\epsilon} P_{FA} \geq 1.$$

The ϵ-differential privacy constraint implies that in the context of hypothesis testing, P_{FA} and P_{MD} cannot both be too small. We plot the regions of P_{FA} and P_{MD} under ϵ-differential privacy in Figure 2.1.

2.1.2 Laplacian Mechanism

The standard approach to preserving ϵ-differential privacy is to perturb the query output by adding random noise with Laplacian distribution proportional to the sensitivity Δ of the

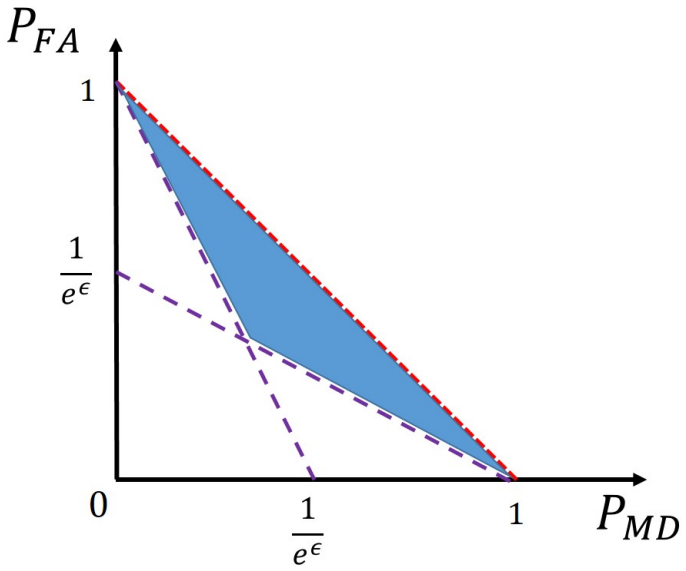


Figure 2.1: The Region of P_{FA} and P_{MD} under ϵ-Differential Privacy

query function q, where the sensitivity of a real-valued query function is defined as

Definition 2.2 (Query Sensitivity [3]). *For a real-valued query function $q : \mathcal{D}^n \to \mathbb{R}$, the sensitivity of q is defined as*

$$\Delta := \max_{D_1, D_2 \in \mathcal{D}^n} |q(D_1) - q(D_2)|, \tag{2.2}$$

for all D_1, D_2 differing in at most one element.

Formally, the Laplacian mechanism is:

Definition 2.3 (Laplacian Mechanism [11]). *For a real-valued query function $q : \mathcal{D}^n \to \mathbb{R}$ with sensitivity Δ, the Laplacian mechanism will output*

$$K(D) := q(D) + Lap(\frac{\Delta}{\epsilon}),$$

where $Lap(\lambda)$ is a random variable with probability density function

$$f(x) = \frac{1}{2\lambda} e^{-\frac{|x|}{\lambda}}, \quad \forall x \in \mathbb{R}.$$

Consider two neighboring datasets D_1 and D_2 where $|q(D_1) - q(D_2)| = \Delta$. It is easy to compute the trade-off between the false alarm probability P_{FA} and the missing detection

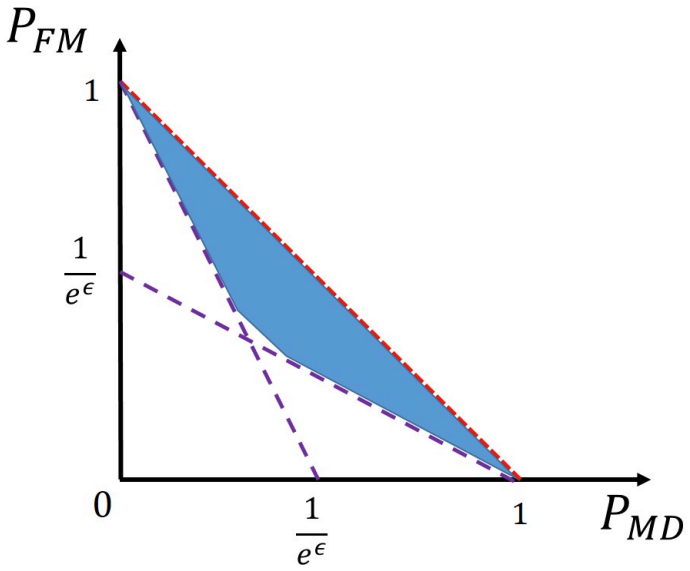


Figure 2.2: The Region of P_{FA} and P_{MD} under the Laplacian Mechanism

probability P_{MD} under the Laplacian mechanism, which is

$$P_{MD} = \begin{cases} 1 - e^{\epsilon} P_{FA} & P_{FA} \in [0, \frac{1}{2}e^{-\epsilon}) \\ \frac{e^{-\epsilon}}{4P_{FA}} & P_{FA} \in [\frac{1}{2}e^{-\epsilon}, \frac{1}{2}) \\ e^{-\epsilon}(1 - P_{FA}) & P_{FA} \in [\frac{1}{2}, 1] \end{cases} \tag{2.3}$$

The region of P_{FA} and F_{MD} under the Laplacian mechanism for two neighboring datasets D_1 and D_2 such that $|q(D_1) - q(D_2)| = \Delta$ is ploted in Figure 2.2.

Since its introduction in [11], the Laplacian mechanism has become the standard tool in differential privacy and has been used as the basic building block in a number of works on differential privacy analysis in other more complex problem settings, e.g., [6, 12–46]. Given this near-routine use of the query-output independent adding of Laplacian noise, the following two questions are natural:

- Is query-output independent perturbation optimal?
- Assuming query-output independent perturbation, is Lapacian noise distribution optimal?

In this dissertation we answer the above two questions. Our main result is that given an ϵ-differential privacy constraint, under a general utility-maximization (equivalently, cost-minimization) model, for a single real-valued query function (assuming local sensitivity is the same as global sensitivity),

- adding query-output independent noise is indeed optimal (under a mild technical condition), and
- the optimal noise distribution is *not* Laplacian distribution; instead, the optimal one has a *staircase-shaped* probability density function.

We also generalize the same result to the discrete setting where the query output is integer-valued and to more abstract settings.

2.2 Problem Formulation

We formulate a utility-maximization (cost-minimization) problem under the differential privacy constraint.

2.2.1 Differential Privacy Constraint

A general randomized releasing mechanism $\mathcal{K}$ is a family of noise probability distributions indexed by the query output (denoted by t), i.e.,

$$\mathcal{K} = \{\mathcal{P}_t : t \in \mathbb{R}\},$$

and given dataset D, the mechanism $\mathcal{K}$ will release the query output $t = q(D)$ corrupted by additive random noise with probability distribution $\mathcal{P}_t$:

$$\mathcal{K}(D) = t + X_t,$$

where X_t is a random variable with probability distribution $\mathcal{P}_t$.

The differential privacy constraint (2.1) on $\mathcal{K}$ is that for any $t_1, t_2 \in \mathbb{R}$ such that $|t_1 - t_2| \leq \Delta$ (corresponding to the query outputs for two neighboring datasets),

$$\mathcal{P}_{t_1}(S) \leq e^{\epsilon}\mathcal{P}_{t_2}(S + t_1 - t_2), \forall \text{ measurable set } S \subset \mathbb{R}, \tag{2.4}$$

where for any $t \in \mathbb{R}$, $S + t := \{s + t \,|\, s \in S\}$.

2.2.2 Utility Model

The utility model we use in this dissertation is a very general one, which is also used in the works by Ghosh, Roughgarden, and Sundararajan [7], Gupte and Sundararajan [9], and Brenner and Nissim [8].

Consider a cost function $\mathcal{L}(\cdot) : \mathbb{R} \to \mathbb{R}$, which is a function of the additive noise. Given additive noise x, the cost is $\mathcal{L}(x)$. Given query output $t \in \mathbb{R}$, the additive noise is a random variable with probability distribution $\mathcal{P}_t$, and thus the expectation of the cost is

$$\int_{x \in \mathbb{R}} \mathcal{L}(x) \mathcal{P}_t(dx).$$

The objective is to minimize the worst-case cost among all possible query outputs $\{t \in \mathbb{R}\}$, i.e.,

$$\text{minimize } \sup_{t \in \mathbb{R}} \int_{x \in \mathbb{R}} \mathcal{L}(x) \mathcal{P}_t(dx). \tag{2.5}$$

2.2.3 Optimization Problem

Combining the differential privacy constraint (2.4) and the objective function (2.5), we formulate a functional optimization problem:

$$\underset{\{\mathcal{P}_t\}_{t \in \mathbb{R}}}{\text{minimize}} \ \sup_{t \in \mathbb{R}} \int_{x \in \mathbb{R}} \mathcal{L}(x) \mathcal{P}_t(dx) \tag{2.6}$$

$$\text{subject to } \mathcal{P}_{t_1}(S) \leq e^{\epsilon} \mathcal{P}_{t_2}(S + t_1 - t_2), \forall \text{ measurable set } S \subseteq \mathbb{R}, \ \forall |t_1 - t_2| \leq \Delta. \tag{2.7}$$

2.3 An Overview of Our Results

2.3.1 Adding Query-Output Independent Noise Is Optimal

Our first result is that under a mild technical condition, adding query-output independent noise is optimal, i.e., we can assume that $\mathcal{P}_t \equiv \mathcal{P}$ for all $t \in \mathbb{R}$ for some probability distribution $\mathcal{P}$.

For any positive integer n, and for any positive real number T, define

$$\mathcal{K}_{T,n} \triangleq \{ \ \{\mathcal{P}_t\}_{t \in \mathbb{R}} \mid \{\mathcal{P}_t\}_{t \in \mathbb{R}} \text{ satisfies (2.7)}, \ \mathcal{P}_t = \mathcal{P}_{k\frac{T}{n}}, \text{for } t \in [k\frac{T}{n}, (k+1)\frac{T}{n}), k \in \mathbb{Z},$$
$$\text{and } \mathcal{P}_{t+T} = \mathcal{P}_t, \forall t \in \mathbb{R} \ \}.$$

Theorem 2.1. *Given any family of probability distribution* $\{\mathcal{P}_t\}_{t\in\mathbb{R}} \in \cup_{T>0}\cup_{n\geq 1}\mathcal{K}_{T,n}$*, there exists a probability distribution* $\mathcal{P}^*$ *such that the family of probability distributions* $\{\mathcal{P}_t^*\}_{t\in\mathbb{R}}$ *with* $\mathcal{P}_t^* \equiv \mathcal{P}^*$ *satisfies the differential privacy constraint* (2.7) *and*

$$\sup_{t\in\mathbb{R}} \int_{x\in\mathbb{R}} \mathcal{L}(x)\mathcal{P}_t^*(dx) \leq \sup_{t\in\mathbb{R}} \int_{x\in\mathbb{R}} \mathcal{L}(x)\mathcal{P}_t(dx).$$

Theorem 2.1 states that if we assume the family of noise probability distributions is piecewise constant (the length of pieces can be arbitrarily small) over t, and periodic (the period can be arbitrary) over t, then in the optimal mechanism we can assume $\mathcal{P}_t$ does not depend on t. We conjecture that the technical condition can be done away with.

2.3.2 Optimal Noise Probability Distribution

Due to Theorem 2.1, adding query-output independent noise is optimal, and thus we only need to study what the optimal noise probability distribution is. Let $\mathcal{P}$ denote the probability distribution of the noise added to the query output. Then the optimization problem (2.6) and (2.7) is reduced to

$$\begin{aligned} &\underset{\mathcal{P}}{\text{minimize}} \int_{x\in\mathbb{R}} \mathcal{L}(x)\mathcal{P}(dx) \\ &\text{subject to } \mathcal{P}(S) \leq e^{\epsilon}\mathcal{P}(S+d), \forall \text{ measurable set } S \subseteq \mathbb{R},\ \forall |d| \leq \Delta. \end{aligned}$$

Consider a staircase-shaped probability distribution with probability density function (p.d.f.) $f_\gamma(\cdot)$ defined as

$$f_\gamma(x) = \begin{cases} a(\gamma) & x \in [0, \gamma\Delta) \\ e^{-\epsilon}a(\gamma) & x \in [\gamma\Delta, \Delta) \\ e^{-k\epsilon}f_\gamma(x-k\Delta) & x \in [k\Delta, (k+1)\Delta) \text{ for } k \in \mathbb{N} \\ f_\gamma(-x) & x < 0, \end{cases}$$

where

$$a(\gamma) \triangleq \frac{1-e^{-\epsilon}}{2\Delta(\gamma + e^{-\epsilon}(1-\gamma))}$$

is a normalizing constant to make $\int_{x\in\mathbb{R}} f_\gamma(x)dx = 1$.

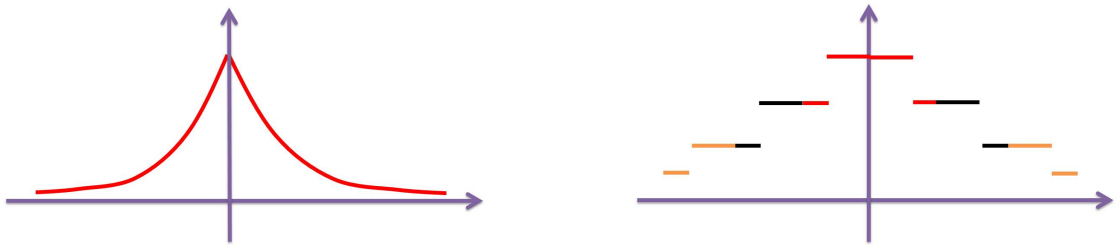
Our main result is

Theorem 2.2. *If the cost function $\mathcal{L}(\cdot)$ is symmetric and increasing, and $\sup_{x\geq T}\frac{\mathcal{L}(x+1)}{\mathcal{L}(x)} < +\infty$ for some $T > 0$, the optimal noise probability distribution has a staircase-shaped probability density function $f_{\gamma^*}(\cdot)$, where*

$$\gamma^* = \arg\min_{\gamma\in[0,1]} \int_{x\in\mathbb{R}} \mathcal{L}(x) f_\gamma(x) dx.$$

We plot the probability density functions of the Laplacian mechanism and the staircase mechanism in Figure 2.3. Figure 2.4 in Section 2.5 gives a precise description of the staircase mechanism.

The staircase mechanism is specified by three parameters: ϵ, Δ, and γ^*, which is determined by ϵ and the cost function $\mathcal{L}(\cdot)$. For certain classes of cost functions, there are closed-form expressions for the optimal γ^*.



(a) Laplacian Mechanism (b) Staircase Mechanism

Figure 2.3: Probability Density Functions of the Laplacian Mechanism and the Staircase Mechanism

2.3.3 Applications: Minimum Noise Magnitude and Noise Power

We apply our main result Theorem 2.4 to two typical cost functions $\mathcal{L}(x) = |x|$ and $\mathcal{L}(x) = x^2$, which measure noise magnitude and noise power, respectively. We derive the closed-form expressions for the optimal parameters γ^* for these two cost functions. Comparing the optimal performances with those of the Laplacian mechanism, we show that in the high privacy regime (ϵ is small), the Laplacian mechanism is asymptotically optimal as $\epsilon \to 0$; in the low privacy regime (ϵ is large), the minimum expectation of noise amplitude and the minimum noise power are $\Theta(\Delta e^{-\frac{\epsilon}{2}})$ and $\Theta(\Delta^2 e^{-\frac{2\epsilon}{3}})$ as $\epsilon \to +\infty$, respectively, while the expectation of noise amplitude and the noise power using the Laplacian mechanism are $\frac{\Delta}{\epsilon}$ and $\frac{2\Delta^2}{\epsilon^2}$, respectively, where Δ is the sensitivity of the query function. We conclude that the gains are more pronounced in the low privacy regime.

2.3.4 Extension to the Discrete Setting

Since for many important practical applications, query functions are integer-valued, we also derive the optimal differentially private mechanisms for answering a single integer-valued query function. We show that adding query-output independent noise is optimal under a mild technical condition, and the optimal noise probability distribution has a staircase-shaped probability mass function, which can be viewed as the discrete variant of the staircase mechanism in the continuous setting.

This result helps us directly compare our work and the existing works [7, 9] on integer-valued query functions. Our result shows that for integer-valued query functions, the optimal noise probability mass function is also staircase-shaped, and in the case the sensitivity $\Delta = 1$, the optimal probability mass function is reduced to the geometric distribution, which was derived in [7, 9]. Therefore, this result can be viewed as a generalization of [7, 9] in the discrete setting for query functions with arbitrary sensitivity.

2.4 Optimality of Query-Qutput Independent Perturbation

Recall that the optimization problem we study in this work is

$$\underset{\{\mathcal{P}_t\}_{t\in\mathbb{R}}}{\text{minimize}} \ \sup_{t\in\mathbb{R}} \int_{x\in\mathbb{R}} \mathcal{L}(x)\mathcal{P}_t(dx) \tag{2.8}$$

$$\text{subject to } \mathcal{P}_{t_1}(S) \le e^{\epsilon}\mathcal{P}_{t_2}(S + t_1 - t_2), \forall \text{ measurable set } S \subseteq \mathbb{R}, \ \forall |t_1 - t_2| \le \Delta, \tag{2.9}$$

where $\mathcal{P}_t$ is the noise probability distribution when the query output is t.

Our claim is that in the optimal family of probability distributions, $\mathcal{P}_t$ can be independent of t, i.e., the probability distribution of noise is independent of the query output. We prove this claim under a technical condition which assumes that $\{\mathcal{P}_t\}_{t\in\mathbb{R}}$ is piecewise constant and periodic (the period can be arbitrary) in terms of t.

For any positive integer n, and for any positive real number T, define

$$\mathcal{K}_{T,n} \triangleq \{ \ \{\mathcal{P}_t\}_{t\in\mathbb{R}} \mid \{\mathcal{P}_t\}_{t\in\mathbb{R}} \text{ satisfies (2.7)}, \ \mathcal{P}_t = \mathcal{P}_{k\frac{T}{n}}, \text{for } t \in [k\frac{T}{n}, (k+1)\frac{T}{n}), k \in \mathbb{Z}, \\ \text{and } \mathcal{P}_{t+T} = \mathcal{P}_t, \forall t \in \mathbb{R} \ \}.$$

Theorem 2.3. *Given any family of probability distribution $\{\mathcal{P}_t\}_{t\in\mathbb{R}} \in \cup_{T>0}\cup_{n\ge 1}\mathcal{K}_{T,n}$, there exists a probability distribution $\mathcal{P}^*$ such that the family of probability distributions $\{\mathcal{P}_t^*\}_{t\in\mathbb{R}}$*

with $\mathcal{P}_t^* \equiv \mathcal{P}^*$ satisfies the differential privacy constraint (2.7) *and*

$$\sup_{t\in\mathbb{R}} \int_{x\in\mathbb{R}} \mathcal{L}(x)\mathcal{P}_t^*(dx) \le \sup_{t\in\mathbb{R}} \int_{x\in\mathbb{R}} \mathcal{L}(x)\mathcal{P}_t(dx).$$

Proof. Here we briefly discuss the main proof technique. For the complete proof, see Appendix A.1. The proof of Theorem 2.3 uses two properties on the family of probability distributions satisfying differential privacy constraint (2.7). First, we show that for any family of probability distributions satisfying (2.7), any translation of the probability distributions will also preserve differential privacy, and the cost is the same. Second, we show that given a collection of families of probability distributions each of which satisfies (2.7), we can take a convex combination of them to construct a new family of probability distributions satisfying (2.7), and the new cost is not worse. Due to these two properties, given any family of probability distributions $\{\mathcal{P}_t\}_{t\in\mathbb{R}} \in \cup_{T>0} \cup_{n\ge 1} \mathcal{K}_{T,n}$, one can take a convex combination of different translations of $\{\mathcal{P}_t\}_{t\in\mathbb{R}}$ to construct $\{\mathcal{P}_t^*\}_{t\in\mathbb{R}}$ with $\mathcal{P}_t^* \equiv \mathcal{P}^*$, and the cost is not worse. □

Theorem 2.3 states that if we assume the family of noise probability distributions is piecewise constant (over intervals with length $\frac{T}{n}$) in terms of t, and periodic over t (with period T), where T, n can be arbitrary, then in the optimal mechanism we can assume $\mathcal{P}_t$ does not depend on t. We conjecture that the technical condition can be done away with.

2.5 Optimal Noise Probability Distribution

Due to Theorem 2.3, to derive the optimal randomized mechanism to preserve differential privacy, we can restrict our attention to noise-adding mechanisms where the noise probability distribution does not depend on the query output. In this section we state our main result Theorem 2.4 on the optimal noise probability distribution.

Let $\mathcal{P}$ denote the probability distribution of the noise added to the query output. Then the optimization problem in (2.6) and (2.7) is reduced to

$$\underset{\mathcal{P}}{\text{minimize}} \int_{x\in\mathbb{R}} \mathcal{L}(x)\mathcal{P}(dx) \tag{2.10}$$

$$\text{subject to } \mathcal{P}(S) \le e^{\epsilon}\mathcal{P}(S+d), \forall \text{ measurable set } S \subseteq \mathbb{R},\ \forall |d| \le \Delta. \tag{2.11}$$

We assume that the cost function $\mathcal{L}(\cdot)$ satisfies two (natural) properties.

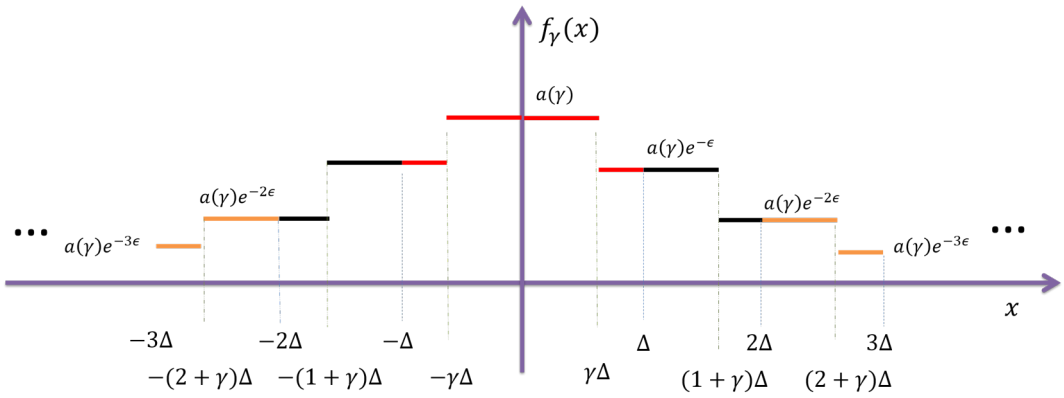


Figure 2.4: The Staircase-Shaped Probability Density Function $f_\gamma(x)$

Property 2.1. *$\mathcal{L}(x)$ is a symmetric function, and monotonically increasing for $x \geq 0$, i.e, $\mathcal{L}(x)$ satisfies*

$$\mathcal{L}(x) = \mathcal{L}(-x), \forall x \in \mathbb{R},$$

and

$$\mathcal{L}(x) \leq \mathcal{L}(y), \forall 0 \leq x \leq y.$$

In addition, we assume $\mathcal{L}(x)$ satisfies a mild technical condition which essentially says that $\mathcal{L}(\cdot)$ does not increase too fast (while still allowing it to be unbounded).

Property 2.2. *There exists a positive integer T such that $\mathcal{L}(T) > 0$ and $\mathcal{L}(x)$ satisfies*

$$\sup_{x \geq T} \frac{\mathcal{L}(x+1)}{\mathcal{L}(x)} < +\infty. \tag{2.12}$$

Consider a staircase-shaped probability distribution with probability density function (p.d.f.) $f_\gamma(\cdot)$ defined as

$$f_\gamma(x) = \begin{cases} a(\gamma) & x \in [0, \gamma\Delta) \\ e^{-\epsilon} a(\gamma) & x \in [\gamma\Delta, \Delta) \\ e^{-k\epsilon} f_\gamma(x - k\Delta) & x \in [k\Delta, (k+1)\Delta) \text{ for } k \in \mathbb{N} \\ f_\gamma(-x) & x < 0, \end{cases} \tag{2.13}$$

where

$$a(\gamma) \triangleq \frac{1-e^{-\epsilon}}{2\Delta(\gamma+e^{-\epsilon}(1-\gamma))}$$

is a normalizing constant to make $\int_{x\in\mathbb{R}} f_\gamma(x)dx = 1$. It is easy to check that for any $\gamma \in [0,1]$, the probability distribution with p.d.f. $f_\gamma(\cdot)$ satisfies the differential privacy constraint (2.11). Indeed, the probability density function $f_\gamma(x)$ satisfies

$$f_\gamma(x) \leq e^\epsilon f_\gamma(x+d), \forall x \in \mathbb{R}, |d| \leq \Delta,$$

which implies (2.11).

Let $\mathcal{SP}$ denote the set of all probability distributions satisfying (2.11). Our main result on the optimal noise probability distribution is:

Theorem 2.4. *If the cost function $\mathcal{L}(x)$ satisfies Property 2.1 and Property 2.2, then*

$$\inf_{\mathcal{P}\in\mathcal{SP}} \int_{x\in\mathbb{R}} \mathcal{L}(x)\mathcal{P}(dx) = \inf_{\gamma\in[0,1]} \int_{x\in\mathbb{R}} \mathcal{L}(x) f_\gamma(x)dx.$$

Proof. Here we briefly discuss the main proof idea and technique. First, by deriving several properties on the probability distributions satisfying the ϵ-differential privacy constraint, we show that without loss of generality, one can "discretize" any valid probability distribution, even for those which do not have probability density functions. Second, we show that to minimize the cost, the probability density function of the discretized probability distribution should be monotonically and geometrically decaying. Lastly, we show that the optimal probability density function should be staircase-shaped. For the complete proof, see Appendix A.2. □

Therefore, the optimal noise probability distribution to preserve ϵ-differential privacy for any real-valued query function has a staircase-shaped probability density function, which is specified by three parameters ϵ, Δ, and $\gamma^* = \arg\min_{\gamma\in[0,1]} \int_{x\in\mathbb{R}} \mathcal{L}(x) f_\gamma(x)dx$.

A natural and simple algorithm to generate random noise with staircase distribution is given in Algorithm 1.

In the formula,

$$X \leftarrow S\left((1-B)\left((G+\gamma U)\Delta\right) + B\left((G+\gamma+(1-\gamma)U)\Delta\right)\right),$$

where

- S determines the sign of the noise,

Algorithm 1 Generation of a Random Variable with Staircase Distribution

Input: ϵ, Δ, and $\gamma \in [0, 1]$.
Output: X, a random variable (r.v.) with staircase distribution specified by ϵ, Δ, and γ.

Generate a r.v. S with $\Pr[S = 1] = \Pr[S = -1] = \frac{1}{2}$.
Generate a geometric r.v. G with $\Pr[G = i] = (1 - b)b^i$ for integer $i \geq 0$, where $b = e^{-\epsilon}$.
Generate a r.v. U uniformly distributed in $[0, 1]$.
Generate a binary r.v. B with $\Pr[B = 0] = \frac{\gamma}{\gamma+(1-\gamma)b}$ and $\Pr[B = 1] = \frac{(1-\gamma)b}{\gamma+(1-\gamma)b}$.
$X \leftarrow S\left((1 - B)\left((G + \gamma U)\Delta\right) + B\left((G + \gamma + (1 - \gamma)U)\Delta\right)\right)$.
Output X.

- G determines which interval $[G\Delta, (G + 1)\Delta)$ the noise lies in,
- B determines which subinterval of $[G\Delta, (G+\gamma)\Delta)$ and $[(G+\gamma)\Delta, (G+1)\Delta)$ the noise lies in,
- U helps to uniformly sample the subinterval.

2.6 Applications

In this section, we apply our main result Theorem 2.4 to derive the parameter γ^* of the staircase mechanism with minimum expectation of noise magnitude and noise second moment, and then compare the performances with the Laplacian mechanism.

2.6.1 Optimal Noise Probability Distribution with Minimum Expectation of Noise Amplitude

To minimize the expectation of amplitude, we have cost function $\mathcal{L}(x) = |x|$, and it is easy to see that it satisfies Property 2.1 and Property 2.2.

To simplify notation, define $b \triangleq e^{-\epsilon}$, and define

$$V(\mathcal{P}) \triangleq \int_{x \in \mathbb{R}} \mathcal{L}(x)\mathcal{P}(dx)$$

for a given probability distribution $\mathcal{P}$.

Theorem 2.5. *To minimize the expectation of the amplitude of noise, the optimal noise probability distribution has probability density function $f_{\gamma^*}(\cdot)$ with*

$$\gamma^* = \frac{1}{1+e^{\frac{\epsilon}{2}}},$$

and the minimum expectation of noise amplitude is

$$V(\mathcal{P}_{\gamma^*}) = \Delta \frac{e^{\frac{\epsilon}{2}}}{e^{\epsilon}-1}.$$

Proof. See Appendix A.3. □

Next, we compare the performances of the optimal noise probability distribution and the Laplacian mechanism. The Laplace distribution has probability density function

$$f(x) = \frac{1}{2\lambda} e^{-\frac{|x|}{\lambda}},$$

where $\lambda = \frac{\Delta}{\epsilon}$. So the expectation of the amplitude of noise with the Laplace distribution is

$$V_{Lap} \triangleq \int_{-\infty}^{+\infty} |x| f(x) dx = \frac{\Delta}{\epsilon}.$$

By comparing $V(\mathcal{P}_{\gamma^*})$ and V_{Lap}, it is easy to see that in the high privacy regime (ϵ is small) the Laplacian mechanism is asymptotically optimal, and the additive gap from optimal value goes to 0 as $\epsilon \to 0$; in the low privacy regime (ϵ is large), $V_{Lap} = \frac{\Delta}{\epsilon}$, while $V(\mathcal{P}_{\gamma^*}) = \Theta(\Delta e^{-\frac{\epsilon}{2}})$. Indeed,

Corollary 2.6. *Consider the cost function $\mathcal{L}(x) = |x|$. In the high privacy regime (ϵ is small),*

$$V_{Lap} - V(\mathcal{P}_{\gamma^*}) = \Delta \left(\frac{\epsilon}{24} - \frac{7\epsilon^3}{5760} + O(\epsilon^5) \right),$$

as $\epsilon \to 0$.

And in the low privacy regime (ϵ is large),

$$V_{Lap} = \frac{\Delta}{\epsilon},$$
$$V(\mathcal{P}_{\gamma^*}) = \Theta(\Delta e^{-\frac{\epsilon}{2}}),$$

as $\epsilon \to +\infty$.

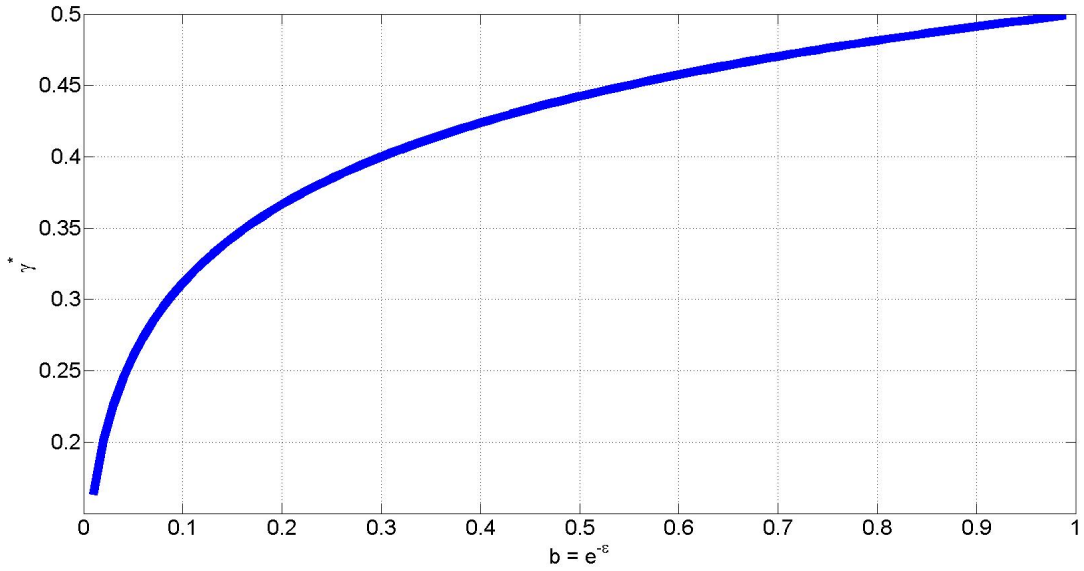


Figure 2.5: Optimal γ^* for the Cost Function $L(x) = x^2$

2.6.2 Optimal Noise Probability Distribution with Minimum Power

Given the probability distribution $\mathcal{P}$ of the noise, the power of noise is defined as $\int_{x\in\mathbb{R}} x^2\mathcal{P}(dx)$. Accordingly, the cost function $\mathcal{L}(x) = x^2$, and it is easy to see it satisfies Property 2.1 and Property 2.2.

Recall $b \triangleq e^{-\epsilon}$.

Theorem 2.7. *To minimize the power of noise (accordingly, $\mathcal{L}(x) = x^2$), the optimal noise probability distribution has probability density function $f_{\gamma^*}(\cdot)$ with*

$$\gamma^* = -\frac{b}{1-b} + \frac{(b - 2b^2 + 2b^4 - b^5)^{1/3}}{2^{1/3}(1-b)^2},$$

and the minimum power of noise is

$$V(\mathcal{P}_{\gamma^*}) = \Delta^2 \frac{2^{-2/3}b^{2/3}(1+b)^{2/3} + b}{(1-b)^2}.$$

Proof. See Appendix A.4. □

We plot γ^* as a function of b for the cost function $\mathcal{L}(x) = x^2$ in Figure 2.5.

Next, we compare the performances of the optimal noise probability distribution and the

Laplacian mechanism. The power of noise with Laplace distribution with $\lambda = \frac{\Delta}{\epsilon}$ is

$$V_{Lap} \triangleq \int_{-\infty}^{+\infty} x^2 \frac{1}{2\lambda} e^{-\frac{|x|}{\lambda}} dx = 2\frac{\Delta^2}{\epsilon^2}.$$

By comparing $V(\mathcal{P}_{\gamma^*})$ and V_{Lap}, it is easy to see that in the high privacy regime (ϵ is small) the Laplacian mechanism is asymptotically optimal, and the additive gap from optimal value is upper bounded by a constant as $\epsilon \to 0$; in the low privacy regime (ϵ is large), $V_{Lap} = \Theta(\frac{2\Delta^2}{\epsilon^2})$, while $V(\mathcal{P}_{\gamma^*}) = \Theta(\Delta^2 e^{-\frac{2\epsilon}{3}})$. Indeed,

Corollary 2.8. *Consider the cost function $\mathcal{L}(x) = x^2$. In the high privacy regime (ϵ is small),*

$$V_{Lap} - V(\mathcal{P}_{\gamma^*}) = \Delta^2 \left(\frac{1}{12} - \frac{\epsilon^2}{720} + O(\epsilon^4) \right),$$

as $\epsilon \to 0$.

And in the low privacy regime (ϵ is large),

$$V_{Lap} = \frac{2\Delta^2}{\epsilon^2},$$
$$V(\mathcal{P}_{\gamma^*}) = \Theta(\Delta^2 e^{-\frac{2\epsilon}{3}}),$$

as $\epsilon \to +\infty$.

2.7 Property of γ^*

In this section, we derive some asymptotic properties of the optimal γ^* for moment cost functions, and give a heuristic choice of γ which only depends on ϵ.

2.7.1 Asymptotic Properties of γ^*

In Section 2.6, we have seen that for the cost functions $\mathcal{L}(x) = |x|$ and $\mathcal{L}(x) = x^2$, the optimal γ^* lies in the interval $[0, \frac{1}{2}]$ for all ϵ and is a monotonically decreasing function of ϵ; and furthermore, $\gamma^* \to \frac{1}{2}$ as ϵ goes to 0, and $\gamma^* \to 0$ as ϵ goes to $+\infty$.

We generalize these asymptotic properties of γ as a function of ϵ to all moment cost functions. More precisely, given $m \in \mathbb{N}$ and $m \geq 1$,

Theorem 2.9. *Consider the cost function $\mathcal{L}(x) = |x|^m$. Let γ^* be the optimal γ in the staircase mechanism for $\mathcal{L}(x)$, i.e.,*

$$\gamma^* = \arg\min_{\gamma\in[0,1]} \int_{x\in\mathbb{R}} |x|^m f_\gamma(x)dx.$$

We have

$$\gamma^* \to \frac{1}{2}, \ as\ \epsilon \to 0,$$
$$\gamma^* \to 0, \ as\ \epsilon \to +\infty.$$

Proof. See Appendix A.5. □

Corollary 2.10. *For all the cost functions $\mathcal{L}(\cdot)$ which can be written as*

$$\mathcal{L}(x) = \sum_{i=1}^{n} \alpha_i |x|^{d_i},$$

where $n \geq 1$, $\alpha_i \in \mathbb{R}, d_i \in \mathbb{N}$ and $\alpha_i, d_i \geq 0$ for all i, the optimal γ^* in the staircase mechanism has the following asymptotic properties:

$$\gamma^* \to \frac{1}{2}, \ as\ \epsilon \to 0,$$
$$\gamma^* \to 0, \ as\ \epsilon \to +\infty.$$

2.7.2 A Heuristic Choice of γ

We have shown that in general the optimal γ^* in the staircase mechanism depends on both ϵ and the cost function $\mathcal{L}(\cdot)$. Here we give a heuristic choice of γ which depends only on ϵ, and show that the performance is reasonably good in the low privacy regime.

Consider a particular choice of γ, which is

$$\tilde{\gamma} := \frac{b}{2} = \frac{e^{-\epsilon}}{2}.$$

It is easy to see that $\tilde{\gamma}$ has the same asymptotic properties as the optimal γ^* for momentum cost functions, i.e.,

$$\tilde{\gamma} \to 0, \text{ as } b \to 0,$$
$$\tilde{\gamma} \to \frac{1}{2}, \text{ as } b \to 1.$$

Furthermore, the probability that the noise magnitude is less than $\frac{e^{-\epsilon}}{2}\Delta$ is approximately $\frac{1}{3}$ in the low privacy regime ($\epsilon \to +\infty$). Indeed,

$$\Pr[|X| \leq \frac{e^{-\epsilon}}{2}\Delta] = \Pr[|X| \leq \tilde{\gamma}\Delta] = 2a(\tilde{\gamma})\tilde{\gamma}\Delta = \frac{1-b}{\tilde{\gamma}+b(1-\tilde{\gamma})}\tilde{\gamma} = \frac{b-b^2}{3b-b^2},$$

which goes to $\frac{1}{3}$ as $\epsilon \to +\infty$ (accordingly, $b \to 0$).

On the other hand, for the Laplacian mechanism,

$$\Pr[|X| \leq \frac{e^{-\epsilon}}{2}\Delta] = \int_{-\frac{e^{-\epsilon}}{2}\Delta}^{\frac{e^{-\epsilon}}{2}\Delta} \frac{1}{2\lambda} e^{-\frac{|x|}{\lambda}} dx = 1 - e^{-\frac{\epsilon e^{-\epsilon}}{2}},$$

which goes to zero as $\epsilon \to +\infty$.

We conclude that in the low privacy regime as $\epsilon \to +\infty$, the staircase mechanism with the heuristic parameter $\tilde{\gamma} = \frac{e^{-\epsilon}}{2}$ can guarantee with probability about $\frac{1}{3}$ the additive noise is very close to zero, while the probability given by the Laplacian mechanism is approximately zero.

2.8 Extension to the Discrete Setting

In this section, we extend our main result Theorem 2.3 and Theorem 2.4 to the discrete settings, and show that the optimal noise-adding mechanism in the discrete setting is a discrete variant of the staircase mechanism in the continuous setting.

2.8.1 Problem Formulation

We first give the problem formulation in the discrete setting.

Consider an integer-valued query function[1]

$$q : \mathcal{D}^n \to \mathbb{Z},$$

where $\mathcal{D}^n$ is the domain of the databases. Let Δ denote the sensitivity of the query function q as defined in (2.2). Clearly, Δ is an integer in this discrete setting.

In the discrete setting, a generic randomized mechanism $\mathcal{K}$ is a family of noise probability

[1]Without loss of generality, we assume that in the discrete setting the query output is integer-valued. Indeed, any uniformly spaced discrete setting can be reduced to the integer-valued setting by scaling the query output.

distributions over the domain $\mathbb{Z}$ indexed by the query output (denoted by i), i.e.,

$$\mathcal{K} = \{\mathcal{P}_i : i \in \mathbb{Z}\},$$

and given dataset D, the mechanism $\mathcal{K}$ will release the query output $i = q(D)$ corrupted by additive random noise with probability distribution $\mathcal{P}_i$:

$$\mathcal{K}(D) = i + X_i,$$

where X_i is a discrete random variable with probability distribution $\mathcal{P}_i$.

Then, the ϵ-differential privacy constraint (2.1) on $\mathcal{K}$ is that for any $i_1, i_2 \in Z$ such that $|i_1 - i_2| \leq \Delta$ (corresponding to the query outputs for two neighboring datasets), and for any subset $S \subset \mathbb{Z}$,

$$\mathcal{P}_{i_1}(j) \leq e^{\epsilon}\mathcal{P}_{i_2}(j + i_1 - i_2), \forall j \in \mathbb{Z}, |i_1 - i_2| \leq \Delta, \tag{2.14}$$

and the goal is to minimize the worst-case cost

$$\sup_{i\in\mathbb{Z}} \sum_{j=-\infty}^{+\infty} \mathcal{L}(j)\mathcal{P}_i(j)$$

subject to the differential privacy constraint (2.14).

2.8.2 Optimality of Query-Qutput Independent Perturbation

In this section, we show that query-output independent perturbation is optimal in the discrete setting.

For any integer $n \geq 1$, define

$$\mathcal{K}_n \triangleq \{\ \{\mathcal{P}_i\}_{i\in\mathbb{Z}} |\ \{\mathcal{P}_i\}_{i\in\mathbb{Z}} \text{ satisfies (2.14), and } \mathcal{P}_{i+n} = \mathcal{P}_i, \forall i \in \mathbb{Z}\}.$$

Theorem 2.11. *Given any family of probability distribution $\{\mathcal{P}_i\}_{i\in\mathbb{Z}} \in \cup_{n\geq 1}\mathcal{K}_n$, there exists a probability distribution $\mathcal{P}^*$ such that the family of probability distributions $\{\mathcal{P}_i^*\}_{i\in\mathbb{Z}}$ with $\mathcal{P}_i^* \equiv \mathcal{P}^*$ satisfies the differential privacy constraint* (2.14) *and*

$$\sup_{i\in\mathbb{Z}} \sum_{j=-\infty}^{+\infty} \mathcal{L}(j)\mathcal{P}_i^*(j) \leq \sup_{i\in\mathbb{Z}} \sum_{j=-\infty}^{+\infty} \mathcal{L}(j)\mathcal{P}_i(j).$$

Proof. The proof is essentially the same as the proof of Theorem 2.3, and thus is omitted. □

Theorem 2.11 states that if we assume the family of noise probability distributions is periodic in terms of i (the period can be arbitrary), then in the optimal mechanism we can assume $\mathcal{P}_i$ does not depend on i. We conjecture that the technical condition can be done away with.

2.8.3 Optimal Noise Probability Distribution

Due to Theorem 2.11, we only need to consider query-output independent perturbation mechanisms.

Let $q(D)$ be the value of the query function evaluated at dataset D. The noise-adding mechanism $\mathcal{K}$ will output

$$\mathcal{K}(D) = q(D) + X,$$

where X is the integer-valued noise added by the mechanism to the output of query function. Let $\mathcal{P}$ be the probability distribution of the noise X. Then the optimization problem we study is

$$\underset{\mathcal{P}}{\text{minimize}} \sum_{i=-\infty}^{+\infty} \mathcal{L}(i)\mathcal{P}(i) \tag{2.15}$$

$$\text{subject to } \mathcal{P}(i) \leq e^{\epsilon}\mathcal{P}(i+d), \forall i \in \mathbb{Z}, d \in \mathbb{Z}, |d| \leq |\Delta|. \tag{2.16}$$

It turns out that when the cost function $\mathcal{L}(\cdot)$ is symmetric and monotonically increasing for $i \geq 0$, the solution to the above optimization problem is a discrete variant of the staircase mechanism in the continuous setting.

As in the continuous setting, we also assume that the cost function $\mathcal{L}(\cdot)$ is symmetric and monotonically increasing for $x \geq 0$, i.e.,

Property 2.3.

$$\mathcal{L}(i) = \mathcal{L}(-i), \forall i \in \mathbb{Z}$$
$$\mathcal{L}(i) \leq \mathcal{L}(i), \forall i, j \in \mathbb{Z}, 0 \leq i \leq j.$$

The easiest case is $\Delta = 1$. In the case that $\Delta = 1$, the solution is the geometric mechanism, which was proposed in [7].

Recall $b \triangleq e^{-\epsilon}$.

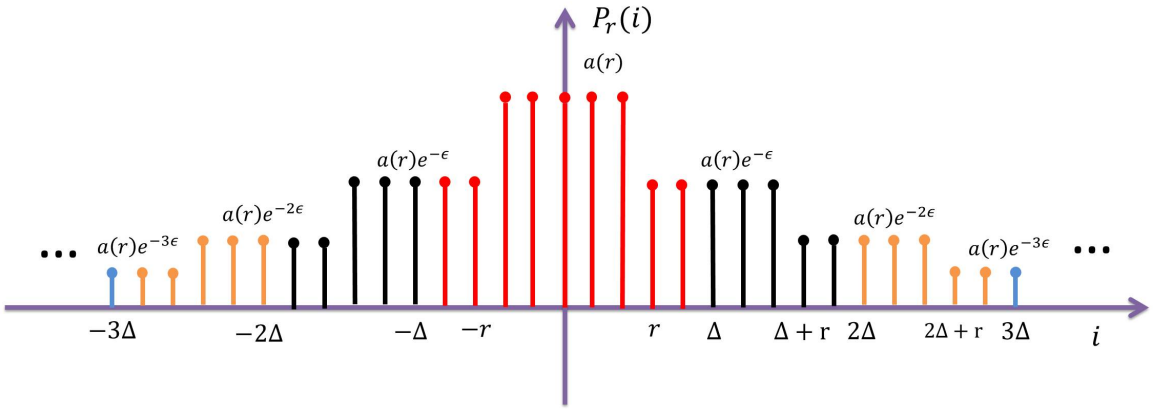


Figure 2.6: The Staircase-Shaped Probability Mass Function $\mathcal{P}_r(i)$

Theorem 2.12. *If the cost function $\mathcal{L}(\cdot)$ satisfies Property 2.3 and $\Delta = 1$, then the geometric mechanism, which has a probability mass function $\mathcal{P}$ with $\mathcal{P}(i) = \frac{1-b}{1+b} b^{|i|}, \forall i \in \mathbb{Z}$, is the optimal solution to (2.15).*

Proof. See Appendix A.6. □

For fixed general $\Delta \geq 2$, consider a class of symmetric and staircase-shaped probability mass functions defined as follows. Given an integer $1 \leq r \leq \Delta$, denote $\mathcal{P}_r$ as the probability mass function defined by

$$\mathcal{P}_r(i) = \begin{cases} a(r) & 0 \leq i < r \\ e^{-\epsilon} a(r) & r \leq i < \Delta \\ e^{-k\epsilon} \mathcal{P}_r(i - k\Delta) & k\Delta \leq i < (k+1)\Delta \text{ for } k \in \mathbb{N} \\ \mathcal{P}_r(-i) & i < 0 \end{cases} \tag{2.17}$$

for $i \in \mathbb{Z}$, where

$$a(r) \triangleq \frac{1-b}{2r + 2b(\Delta - r) - (1-b)}.$$

It is easy to verify that for any $1 \leq r \leq \Delta$, $\mathcal{P}_r$ is a valid probability mass function and it satisfies the ϵ-differential privacy constraint (2.16). We plot the staircase-shaped probability mass function $\mathcal{P}_r(i)$ in Figure 2.6.

Let $\mathcal{SP}$ be the set of all probability mass functions which satisfy the ϵ-differential privacy constraint (2.16).

Theorem 2.13. *For $\Delta \geq 2$, if the cost function $\mathcal{L}(x)$ satisfies Property 2.3, then*

$$\inf_{\mathcal{P}\in\mathcal{SP}} \sum_{i=-\infty}^{+\infty} \mathcal{L}(i)\mathcal{P}(i) = \min_{\{r\in\mathbb{N}|1\leq r\leq\Delta\}} \sum_{i=-\infty}^{+\infty} \mathcal{L}(i)\mathcal{P}_r(i).$$

Proof. See Appendix A.6. □

Therefore, the optimal noise probability distribution to preserve ϵ-differential privacy for integer-valued query function has a staircase-shaped probability mass function, which is specified by three parameters ϵ, Δ, and $r^* = \arg\min_{\{r\in\mathbb{N}|1\leq r\leq\Delta\}} \sum_{i=-\infty}^{+\infty} \mathcal{L}(i)\mathcal{P}_r(i)$. In the case $\Delta = 1$, the staircase-shaped probability mass function is reduced to the geometric mechanism.

2.9 Extension to the Abstract Setting

In this section, we show how to extend the staircase mechanism to the abstract setting. The approach is essentially the same as the exponential mechanism in [47], except that we replace the exponential function by the staircase function.

Consider a privacy mechanism which maps an input from a domain $\mathcal{D}^n$ to some output in a range $\mathcal{R}$. Let μ be the base measure of $\mathcal{R}$. In addition, we have a cost function $\mathcal{C} : \mathcal{D}^n \times \mathcal{R} \to [0, +\infty)$.

Define Δ as

$$\Delta \triangleq \max_{r\in\mathcal{R},\, D_1,D_2\subseteq\mathcal{D}^n:|D_1-D_2|\leq 1} |\mathcal{C}(D_1, r) - \mathcal{C}(D_2, r)|,$$

i.e., the maximum difference of cost function for any two inputs which differ only on one single value over all $r \in \mathcal{R}$ [47].

A randomized mechanism $\mathcal{K}$ achieves ϵ-differential privacy if for any $D_1, D_2 \subseteq \mathcal{D}^n$ such that $|D_1 - D_2| \leq 1$, and for any measurable subset $S \subset \mathcal{R}$,

$$\Pr[\mathcal{K}(D_1) \in S] \leq \exp(\epsilon) \Pr[\mathcal{K}(D_2) \in S].$$

Definition 2.4 (Staircase Mechanism in the Abstract Setting)**.** *For fixed $\gamma \in [0, 1]$, given input $D \in \mathcal{D}^n$, the staircase mechanism in the abstract setting will output an element in $\mathcal{R}$ with the probability distribution defined as*

$$\mathcal{P}_D(S) = \frac{\int_{r\in S} f_\gamma(\mathcal{C}(D, r))\mu(dr)}{\int_{r\in\mathcal{R}} f_\gamma(\mathcal{C}(D, r))\mu(dr)}, \forall \text{ measurable set } S \subset \mathcal{R},$$

where f_γ is the staircase-shaped function defined in (2.13).

Theorem 2.14. *The staircase mechanism in the abstract setting in Definition 2.4 achieves 2ϵ-differential privacy.*

Proof. For any $D_1, D_2 \in \mathcal{D}^n$ such that $|D_1 - D_2| \leq 1$, and for any measurable set $S \subset \mathcal{R}$,

$$\begin{aligned}
\mathcal{P}_{D_1}(S) &= \frac{\int_{r\in S} f_\gamma(\mathcal{C}(D_1, r))\mu(dr)}{\int_{r\in\mathcal{R}} f_\gamma(\mathcal{C}(D_1, r))\mu(dr)} \\
&\leq e^\epsilon \frac{\int_{r\in S} f_\gamma(\mathcal{C}(D_2, r))\mu(dr)}{\int_{r\in\mathcal{R}} f_\gamma(\mathcal{C}(D_1, r))\mu(dr)} \\
&\leq e^\epsilon e^\epsilon \frac{\int_{r\in S} f_\gamma(\mathcal{C}(D_2, r))\mu(dr)}{\int_{r\in\mathcal{R}} f_\gamma(\mathcal{C}(D_2, r))\mu(dr)} \\
&= e^{2\epsilon}\mathcal{P}_{D_2}(S),
\end{aligned}$$

where we have used the property that $f_\gamma(\mathcal{C}(D_1, r)) \leq e^\epsilon f_\gamma(\mathcal{C}(D_2, r))$ and $f_\gamma(\mathcal{C}(D_2, r)) \leq e^\epsilon f_\gamma(\mathcal{C}(D_1, r))$ for all $r \in \mathcal{R}$.

Therefore, the staircase mechanism in the abstract setting achieves 2ϵ-differential privacy for any $\gamma \in [0, 1]$.

□

In the case that the output range $\mathcal{R}$ is the set of real numbers $\mathbb{R}$ and the cost function $\mathcal{C}(d, r) = |r - q(d)|$ for some real-valued query function q, the above mechanism is reduced to the staircase mechanism in the continuous setting.

2.10 Connection to the Literature

In this section, we discuss the relations of our results and some directly related works in the literature, and the implications of our results on other works.

2.10.1 Laplacian Mechanism vs. Staircase Mechanism

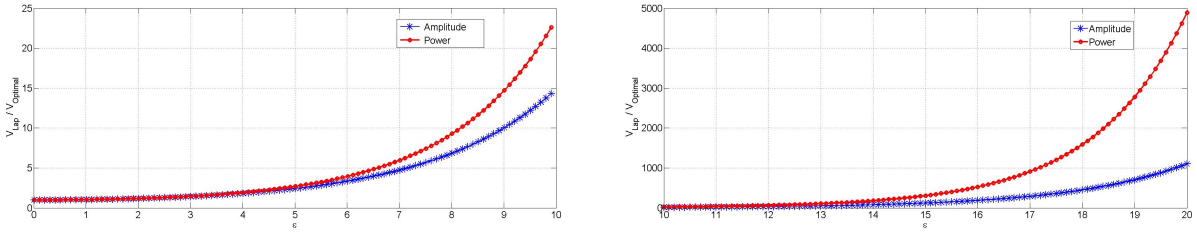
The Laplacian mechanism is specified by two parameters, ϵ and the query function sensitivity Δ. ϵ and Δ completely characterize the differential privacy constraint. On the other hand, the staircase mechanism is specified by three parameters, ϵ, Δ, and γ^*, which is determined by ϵ and the utility function/cost function. For certain classes of utility functions/cost functions, there are closed-form expressions for the optimal γ^*.

From the two examples given in Section 2.6, we can see that although the Laplacian mechanism is not strictly optimal, in the high privacy regime ($\epsilon \to 0$), the Laplacian mechanism is asymptotically optimal:

- For the expectation of noise amplitude, the additive gap from the optimal values goes to 0 as $\epsilon \to 0$,
- For noise power, the additive gap from the optimal values is upper bounded by a constant as $\epsilon \to 0$.

However, in the low privacy regime ($\epsilon \to +\infty$), the multiplicative gap from the optimal values can be arbitrarily large. We conclude that in the high privacy regime, the Laplacian mechanism is nearly optimal, while in the low privacy regime, significant improvement can be achieved by using the staircase mechanism. We plot the multiplicative gain of the staircase mechanism over the Laplacian mechanism for expectation of noise amplitude and noise power in Figure 2.7, where V_{Optimal} is the optimal (minimum) cost, which is achieved by the staircase mechanism, and V_{Lap} is the cost of the Laplacian mechanism. We can see that for $\epsilon \approx 10$, the staircase mechanism has about 15-fold and 23-fold improvement, with noise amplitude and power, respectively.

Since the staircase mechanism is derived under the same problem setting as the Laplacian mechanism, the staircase mechanism can be applied wherever the Laplacian mechanism is used, and it performs strictly better than the Laplacian mechanism (and significantly better in low privacy scenarios).



(a) $0 < \epsilon \leq 10$ (b) $10 \leq \epsilon \leq 20$

Figure 2.7: Multiplicative Gain of the Staircase Mechanism over the Laplacian Mechanism

2.10.2 Relation to Ghosh et al. [7]

Ghosh, Roughgarden, and Sundararajan [7] show that for a single count query with sensitivity $\Delta = 1$, for a general class of utility functions, to minimize the expected cost under a

Bayesian framework the optimal mechanism to preserve differential privacy is the geometric mechanism, which adds noise with geometric distribution.

We discuss the relations and differences between [7] and our work in the following: Both [7] and our work are similar in that, given the query output, the cost function only depends on the additive noise magnitude, and is an increasing function of noise magnitude. On the other hand, there are two main differences:

- Ghosh et al. [7] works under a Bayesian setting, while ours minimizes the worst-case cost.
- Ghosh et al. [7] studies a count query where the query output is integer-valued and bounded, and the sensitivity is unity. In our work, we first study a general real-valued query function where the query output can take any real value, and then generalize the result to the discrete setting where query output is integer valued. In both cases, the sensitivity of query functions can be arbitrary, not restricted to one.

2.10.3 Relation to Gupte and Sundararajan [9]

Gupte and Sundararajan [9] derive the optimal noise probability distributions for a single count query with sensitivity $\Delta = 1$ for minimax (risk-averse) users. Their model is the same as the one in [7] except that their objective function is to minimize the worst-case cost, the same as our objective. Gupte and Sundararajan [9] show that although for a general class of cost functions, there is no universally optimal solution to the minimax optimization problem in [9], each solution (corresponding to different cost functions) can be derived from the same geometric mechanism by randomly remapping.

As in [7], [9] assumes the query-output is bounded. Our results show that when the query sensitivity is one, without any boundedness knowledge about the query-output, the optimal mechanism is to add random noise with geometric distribution to the query output.

2.10.4 Relation to Brenner and Nissim [8]

While [7] shows that for a single count query with sensitivity $\Delta = 1$, there is a universally optimal mechanism for a general class of utility functions under a Bayesian framework, Brenner and Nissim [8] show that for general query functions, no universally optimal mechanisms exist. Indeed, this follows directly from our results: under our optimization framework, the optimal mechanism is adding noise with staircase-shaped probability distribution which is

specified by three parameters ϵ, Δ, and γ^*, where in general γ^* depends on the cost function. Generally, for different cost functions, the optimal noise probability distributions have staircase-shaped probability density functions specified by different parameters.

2.10.5 Relation to Nissim, Raskhodnikova, and Smith [48]

Nissim, Raskhodnikova, and Smith [48] show that for certain nonlinear query functions, one can improve the accuracy by adding data-dependent noise calibrated to the smooth sensitivity of the query function, which is based on the local sensitivity of the query function. In our model, we use the global sensitivity of the query function only, and assume that the local sensitivity is the same as the global sensitivity, which holds for a general class of query functions, e.g., count, sum.

2.10.6 Relation to Hardt and Talwar [4]

Hardt and Talwar [4] study the trade-off between privacy and error for answering a set of linear queries over a histogram in a differentially private way. The error is defined as the worst expectation of the ℓ^2-norm of the noise. The lower bound given in [4] is $\Omega(\epsilon^{-1}d\sqrt{d})$, where d is the number of linear queries. An immediate consequence of our result is that for fixed d, when $\epsilon \to +\infty$, an upper bound of $\Theta(e^{-\frac{\epsilon}{3d}}d\sqrt{d})$ is achievable by adding independent staircase-shaped noise with parameter $\frac{\epsilon}{d}$ to each component.

2.10.7 Relation to Other Works

Many existing works study how to improve the accuracy for answering more complex queries under differential privacy, in which the basic building block is the standard Laplacian mechanism. For example, Hay et al. [49] show that one can improve the accuracy for a general class of histogram queries, by exploiting the consistency constraints on the query output, and Li et al. [6] study how to optimize linear counting queries under differential privacy by carefully choosing the set of linear queries to be answered. In these works, the error is measured by the mean squared error of query output estimates, which corresponds to the variance of the noise added to the query output to preserve differential privacy. In terms of ϵ, the error bound in these works scales linearly to $\frac{1}{\epsilon^2}$, because of the use of Laplacian noise. If Laplacian distribution is replaced by staircase distribution in these works, one can improve the error bound to $\Theta(e^{-C\epsilon})$ (for some constant C which depends on the number of queries) when $\epsilon \to +\infty$ (corresponding to the low privacy regime).

CHAPTER 3

THE OPTIMAL MECHANISM IN ϵ-DIFFERENTIAL PRIVACY: MULTIPLE DIMENSIONAL SETTING

In this chapter, we extend the staircase mechanism from the single dimensional setting to the multiple dimensional setting. We show that for histogram-like query functions, when the dimension of the query output is two, the multiple dimensional staircase mechanism is optimal for the ℓ^1 cost function. We give the problem formulation in Section 3.1, and present the main result on the opitmality of the multiple dimensional staircase mechanism in Section 3.2. In Section 3.3, we study the asymptotical performance of the optimal mechanism in the high and low privacy regimes. Comparing the optimal performances with those of the Laplacian mechanism, we conclude that in the multiple dimensional setting, the Laplacian mechanism is asymptotically optimal in the high privacy regime, and the staircase mechanism significantly outperforms the Laplacian mechanism in the low privacy regime.

3.1 Problem Formulation

Consider a multiple dimensional real-valued query function

$$q : \mathcal{D}^n \to \mathbb{R}^d,$$

where $\mathcal{D}^n$ is the domain of the databases, and d is the dimension of the query output. Given $D \in \mathcal{D}^n$, the query output can be written as

$$q(D) = (q_1(D), q_2(D), \ldots, q_d(D)),$$

where $q_i(D) \in \mathbb{R}, \forall 1 \leq i \leq d$.

The sensitivity of the query function q is defined as

$$\Delta \triangleq \max_{D_1, D_2 \subseteq \mathcal{D}^n : |D_1 - D_2| \leq 1} \|q(D_1) - q(D_2)\|_1 = \sum_{i=1}^{d} |q_i(D_1) - q_i(D_2)|, \tag{3.1}$$

where the maximum is taken over all possible pairs of neighboring database entries D_1 and

D_2 which differ in at most one element, i.e., one is a proper subset of the other and the larger database contains just one additional element [3].

Definition 3.1 (ϵ-Differential Privacy [3])**.** *A randomized mechanism $\mathcal{K}$ gives ϵ-differential privacy if for all data sets D_1 and D_2 differing in at most one element, and all $S \subset Range(\mathcal{K})$,*

$$Pr[\mathcal{K}(D_1) \in S] \leq e^{\epsilon}\ Pr[\mathcal{K}(D_2) \in S]. \tag{3.2}$$

The standard approach to preserving the differential privacy is to add noise to the output of the query function. Let $q(D)$ be the value of the query function evaluated at $D \subseteq \mathcal{D}^n$. Then the noise-adding mechanism $\mathcal{K}$ will output

$$\mathcal{K}(D) = q(D) + X = (q_1(D) + X_1, \ldots, q_d(D) + X_d),$$

where $X = (X_1, \ldots, X_d) \in \mathbb{R}^d$ is the noise added by the mechanism to the output of the query function.

In the following, we derive the differential privacy constraint on the probability distribution of X from (3.2).

$$\begin{aligned}
&\Pr[\mathcal{K}(D_1) \in S] &&\leq e^{\epsilon} \Pr[\mathcal{K}(D_2) \in S] \\
\Leftrightarrow\ &\Pr[q(D_1) + X \in S] &&\leq e^{\epsilon} \Pr[q(D_2) + X \in S] \\
\Leftrightarrow\ &\Pr[X \in S - q(D_1)] &&\leq e^{\epsilon} \Pr[X \in S - q(D_2)] \\
\Leftrightarrow\ &\Pr[X \in S'] &&\leq e^{\epsilon} \Pr[X \in S' + q(D_1) - q(D_2)],
\end{aligned} \tag{3.3}$$

where $S' \triangleq S - q(D_1) = \{s - q(D_1) | s \in S\}$.

Since (3.2) holds for all measurable sets $S \subseteq \mathbb{R}^d$, and $\|q(D_1) - q(D_2)\|_1 \leq \Delta$, from (3.3) we have

$$\Pr[X \in S'] \leq e^{\epsilon} \Pr[X \in S' + \mathbf{t}], \tag{3.4}$$

for all measurable sets $S' \subseteq \mathbb{R}$ and for all $\mathbf{t} \in \mathbb{R}^d$ such that $\|\mathbf{t}\|_1 \leq \Delta$.

Consider a cost function $\mathcal{L}(\cdot) : \mathbb{R}^d \to \mathbb{R}$ which is a function of the added noise X. Our goal is to minimize the expectation of the cost, subject to the ϵ-differential privacy constraint (3.4).

More precisely, let $\mathcal{P}$ denote the probability distribution of X and let $\mathcal{P}(S)$ denote the probability $\Pr[X \in S]$. The optimization problem we study in this work is

$$\underset{\mathcal{P}}{\text{minimize}} \int\int \ldots \int_{\mathbb{R}^d} \mathcal{L}(x_1, x_2, \ldots, x_d)\mathcal{P}(dx_1 dx_2 \ldots dx_d) \tag{3.5}$$

$$\text{subject to } \mathcal{P}(S) \leq e^{\epsilon}\mathcal{P}(S + \mathbf{t}), \forall \text{ measurable set } S \subseteq \mathbb{R}^d, \ \forall \|\mathbf{t}\|_1 \leq \Delta. \tag{3.6}$$

We solve the above functional optimization problem and derive the optimal noise probability distribution for $\mathcal{L}(x_1, \ldots, x_d) = \sum_{i=1}^{d} |x_i|$, with $d = 2$.

3.2 Main Result

In this section we state our main result, Theorem 3.1. The detailed proof is given in Appendix B.1.

We consider the ℓ^1 cost function:

$$\mathcal{L}(x_1, x_2, \ldots, x_d) = \sum_{i=1}^{d} |x_i|, \forall (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d.$$

Consider a class of multiple dimensional probability distributions with symmetric and staircase-shaped probability density functions defined as follows. Given $\gamma \in [0, 1]$, define $\mathcal{P}_\gamma$ as the probability distribution with probability density function $f_\gamma(\cdot)$ defined as

$$f_\gamma(\mathbf{x}) = \begin{cases} e^{-k\epsilon}a(\gamma) & \|\mathbf{x}\|_1 \in [k\Delta, (k+\gamma)\Delta) \text{ for } k \in \mathbb{N} \\ e^{-(k+1)\epsilon}a(\gamma) & \|\mathbf{x}\|_1 \in [(k+\gamma)\Delta, (k+1)\Delta) \text{ for } k \in \mathbb{N}, \end{cases}$$

where $a(\gamma)$ is the normalization factor to make

$$\int\int \ldots \int_{\mathbb{R}^d} f_\gamma(\mathbf{x}) dx_1 dx_2 \ldots dx_d = 1.$$

Define $b \triangleq e^{-\epsilon}$, and define

$$c_k \triangleq \sum_{i=0}^{+\infty} i^k b^i, \forall k \in \mathbb{N},$$

where by convention 0^0 is defined as 1. Then the closed-form expression for $a(\gamma)$ is

$$a(\gamma) \triangleq \frac{d!}{2^d \Delta^d \sum_{k=1}^{d} \binom{d}{k} c_{d-k}(b + (1-b)\gamma^k)}.$$

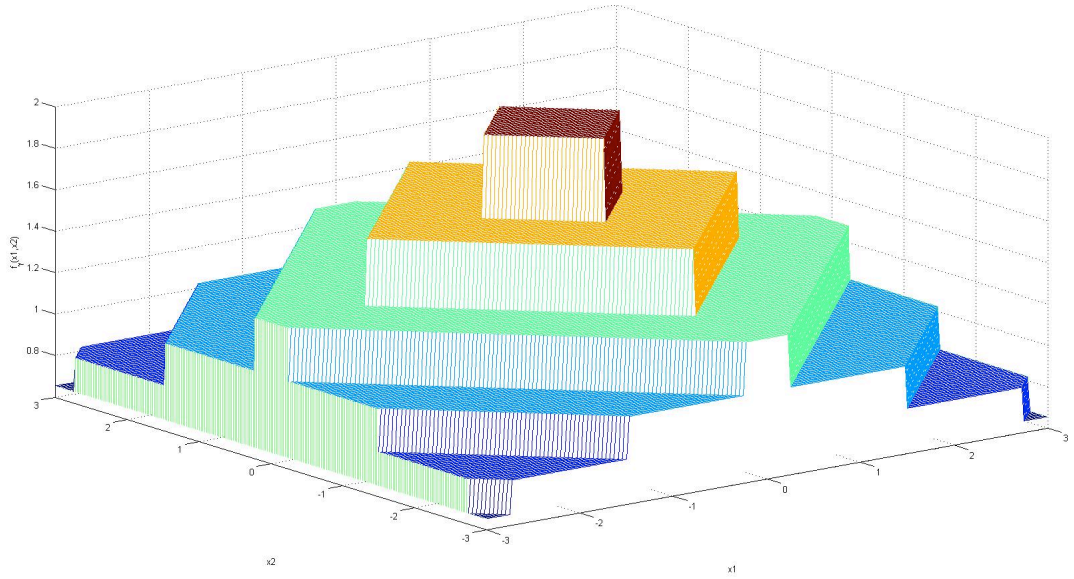


Figure 3.1: Multiple Dimensional Staircase-Shaped Probability Density Function for $d = 2$

It is straightforward to verify that $f_\gamma(\cdot)$ is a valid probability density function and $\mathcal{P}_\gamma$ satisfies the differential privacy constraint (3.6). Indeed, the probability density function $f_\gamma(x)$ satisfies

$$f_\gamma(\mathbf{x}) \leq e^\epsilon f_\gamma(\mathbf{x}+\mathbf{t}), \forall \mathbf{x} \in \mathbb{R}^d, \forall \mathbf{t} \in \mathbb{R}^d \text{ s.t. } \|\mathbf{t}\|_1 \leq \Delta,$$

which implies (3.6).

We plot the probability density function $f_\gamma(\mathbf{x})$ in Figure 3.1 for $d = 2$. It is easy to see that $f_\gamma(\mathbf{x})$ is multiple dimensionally staircase-shaped.

Let $\mathcal{SP}$ be the set of all probability distributions which satisfy the differential privacy constraint (3.6). Our main result is

Theorem 3.1. *For $d = 2$ and the cost function $\mathcal{L}(\mathbf{x}) = \|\mathbf{x}\|_1, \forall \mathbf{x} \in \mathbb{R}^2$, then*

$$\inf_{\mathcal{P} \in \mathcal{SP}} \int\int_{\mathbb{R}^2} \mathcal{L}(\mathbf{x}) \mathcal{P}(dx_1 dx_2) = \inf_{\gamma \in [0,1]} \int\int_{\mathbb{R}^2} \mathcal{L}(\mathbf{x}) f_\gamma(\mathbf{x}) dx_1 dx_2.$$

Proof. See Appendix B.1. □

Therefore, the optimal noise probability distribution to preserve ϵ-differential privacy for multiple dimensional real-valued query functions has a multiple dimensionally staircase-shaped probability density function, which is specified by three parameters ϵ, Δ, and $\gamma^* =$

$\arg\min_{\gamma\in[0,1]} \int\int_{\mathbb{R}^2} \mathcal{L}(x_1,x_2) f_\gamma(\mathbf{x}) dx_1 dx_2$.

3.3 Optimal γ^* and Asymptotic Analysis

Note that the closed-form expressions for c_0, c_1 and c_2 are

$$\begin{aligned} c_0 &= \frac{1}{1-b}, \\ c_1 &= \frac{b}{(1-b)^2}, \\ c_2 &= \frac{b^2+b}{(1-b)^3}. \end{aligned}$$

For $d = 2$, we have

$$\begin{aligned} a(\gamma) &= \frac{1}{2\Delta^2\left(2c_1(b+(1-b)\gamma)+c_0(b+(1-b)\gamma^2)\right)} \\ &= \frac{1}{2\Delta^2\left(\gamma^2+\frac{2b}{1-b}\gamma+\frac{b+b^2}{(1-b)^2}\right)}. \end{aligned}$$

Given the two-dimensional staircase-shaped probability density function $f_\gamma(\mathbf{x})$, the cost is

$$\begin{aligned} V(\mathcal{P}_\gamma) &\triangleq \int\int_{\mathbb{R}^2} (|x_1|+|x_2|) f_\gamma(x_1,x_2) \mathcal{P}(dx_1 dx_2) \\ &= 4\left(\sum_{i=0}^{+\infty}\int_{i\Delta}^{(i+\gamma)\Delta} tta(\gamma)e^{-i\epsilon}dt + \sum_{i=0}^{+\infty}\int_{(i+\gamma)\Delta}^{(i+1)\Delta} tta(\gamma)e^{-(i+1)\epsilon}dt\right) \\ &= \frac{4a(\gamma)\Delta^3}{3}\left(\sum_{i=0}^{+\infty} b^i(3i^2\gamma+3i\gamma^2+\gamma^3) + b\sum_{i=0}^{+\infty} b^i(3i^2+3i+1-3i^2\gamma-3i\gamma^2-\gamma^3)\right) \\ &= \frac{4a(\gamma)\Delta^3}{3}\left(3c_2\gamma+3c_1\gamma^2+c_0\gamma^3+b(3(1-\gamma)c_2+3(1-\gamma^2)c_1+(1-\gamma^3)c_0)\right) \\ &= \frac{2\Delta}{3}\frac{3c_2\gamma+3c_1\gamma^2+c_0\gamma^3+b(3(1-\gamma)c_2+3(1-\gamma^2)c_1+(1-\gamma^3)c_0)}{\gamma^2+\frac{2b}{1-b}\gamma+\frac{b+b^2}{(1-b)^2}} \\ &= \frac{2\Delta}{3}\frac{c_0(1-b)\gamma^3+3c_1(1-b)\gamma^2+3c_2(1-b)\gamma+b(c_0+3c_1+3c_2)}{\gamma^2+\frac{2b}{1-b}\gamma+\frac{b+b^2}{(1-b)^2}} \\ &= \frac{2\Delta}{3}\frac{\gamma^3+\frac{3b}{1-b}\gamma^2+\frac{3(b^2+b)}{(1-b)^2}\gamma+b\frac{1+4b+b^2}{(1-b)^3}}{\gamma^2+\frac{2b}{1-b}\gamma+\frac{b+b^2}{(1-b)^2}}. \end{aligned} \tag{3.7}$$

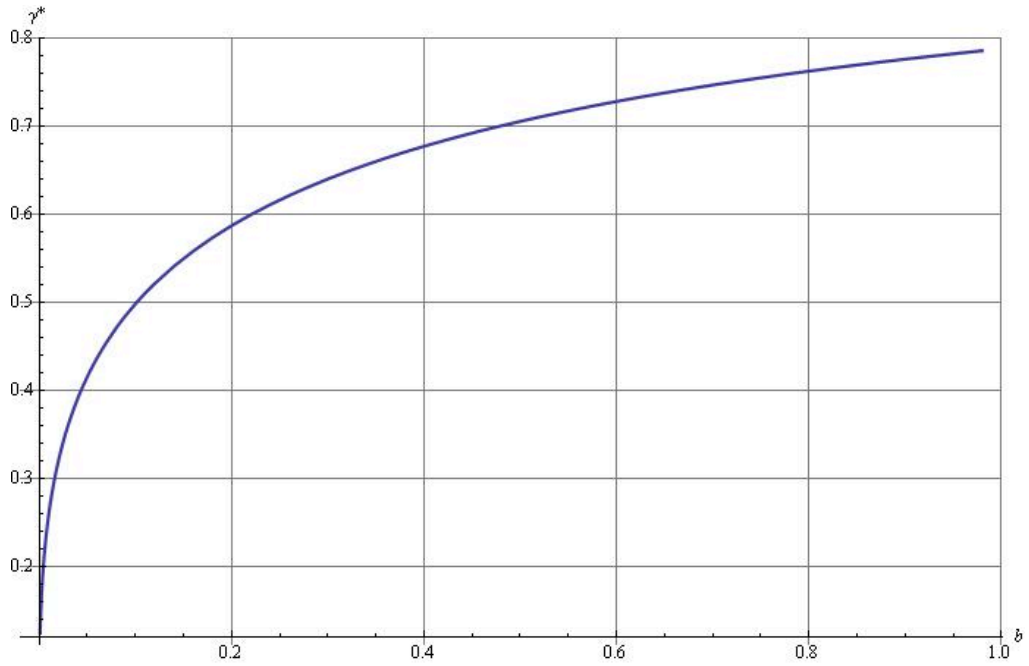


Figure 3.2: The Optimal γ^* as a Function of b

Therefore, in the two-dimensional setting, the optimal γ^* is

$$\gamma^* = \arg\min_{\gamma\in[0,1]} \frac{\gamma^3 + \frac{3b}{1-b}\gamma^2 + \frac{3(b^2+b)}{(1-b)^2}\gamma + b\frac{1+4b+b^2}{(1-b)^3}}{\gamma^2 + \frac{2b}{1-b}\gamma + \frac{b+b^2}{(1-b)^2}}.$$

By setting the derivative of (3.7) to be zero, we use Mathematica to get a closed-form expression for γ^*, which is too complicated to show here. We plot γ^* as a function of b in Figure 3.2.

The optimal cost $V^* = V(\mathcal{P}_{\gamma^*})$. We use Mathematica to analyze the asymptotic behavior of V^* as $\epsilon \to 0$ and $\epsilon \to +\infty$. Indeed, we have

Corollary 3.2. *In the high privacy regime,*

$$V^* = \frac{2\Delta}{\epsilon} - \frac{\Delta\epsilon^2}{36\sqrt{3}} + O(\epsilon^3), \epsilon \to 0,$$

and in the low privacy regime,

$$V^* = \sqrt[3]{2}\Delta e^{-\frac{\epsilon}{3}} + \frac{\Delta e^{-\frac{2\epsilon}{3}}}{\sqrt[3]{2}} + o(e^{-\frac{2\epsilon}{3}}), \epsilon \to +\infty.$$

The Laplacian mechanism adds independent Laplacian noise to each component of the query output, and the cost is $\frac{2\Delta}{\epsilon}$. Therefore, in the high privacy regime, the gap between

optimal cost and the cost achieved by the Laplacian mechanism goes to zero, as $\epsilon \to 0$, and we conclude the Laplacian mechanism is approximately optimal in the high privacy regime. However, in the low privacy regime (as $\epsilon \to +\infty$), the optimal cost is proportional to $e^{-\frac{\epsilon}{3}}$, while the cost of the Laplacian mechanism is proportional to $\frac{1}{\epsilon}$. We conclude the gap is significant in the low privacy regime.

It is natural to compare the performance of the optimal multiple dimensional staircase mechanism and the composite single dimensional staircase mechanism which adds independent staircase noise to each component of the query output. If independent staircase noise is added to each component of the query output, to satisfy the ϵ-differential privacy constraint, the parameter of the staircase noise is $\frac{\epsilon}{2}$ instead of ϵ, and thus the total cost will be proportional to $e^{-\frac{\epsilon}{4}}$, which is worse than the optimal cost $\Theta(e^{-\frac{\epsilon}{3}})$.

CHAPTER 4

THE OPTIMAL MECHANISM IN (ϵ, δ)-DIFFERENTIAL PRIVACY

In this chapter, we study the optimal mechanism in (ϵ, δ)-differential privacy for integer-valued query functions. We show that the (ϵ, δ)-differential privacy is a framework not much more general than the $(\epsilon, 0)$-differential privacy and $(0, \delta)$-differential privacy in the context of ℓ^1 and ℓ^2 cost functions, i.e., minimum expected noise magnitude and noise power. In the same context of ℓ^1 and ℓ^2 cost functions, we show the near-optimality of uniform noise mechanism and the discrete Laplacian mechanism in the high privacy regime (as $(\epsilon, \delta) \to (0, 0)$).

We formulate the utility-maximization/cost-minimization under the (ϵ, δ)-differential privacy constraint as a linear programming problem in Section 4.2. In Section 4.3, we study $(0, \delta)$-differential privacy, and show the near-optimality of the simple uniform noise mechanism. In Section 4.4, we study the optimal mechanisms in (ϵ, δ)-differential privacy, and show the optimality of the uniform noise mechanism and the Laplacian mechanism in the regime $(\epsilon, \delta) \to (0, 0)$ in the context of ℓ^1 and ℓ^2 cost functions. In Section 4.5, we extend the results to the multiple dimensional setting where the query output is a vector of integers.

4.1 Introduction

(ϵ, δ)-differential privacy is a relaxed notion of privacy, compared to the standard ϵ-differential privacy introduced in [11]. (ϵ, δ)-differential privacy includes as special cases:

- $(\epsilon, 0)$-differential privacy. In this standard setting, the optimal mechanism for a general cost minimization framework is the *staircase* mechanism as shown in [50]. In the high privacy regime, the standard discrete Laplacian mechanism also performs well.

- $(0, \delta)$-differential privacy. This setting requires that the total variation of the conditional probability distributions of the query output for neighboring datasets should be bounded by δ. We show that the uniform noise distribution is near-optimal in the $(0, \delta)$-differential privacy setting for a general class of cost functions.

While the (ϵ, δ)-differential privacy setting is more general than the two special cases – $(\epsilon, 0)$ and $(0, \delta)$-differential privacy – our main result in this chapter is to show that it is only more general by *very little*; this is done in the context of ℓ^1 and ℓ^2 cost functions. We show the near-optimality of uniform noise mechanism and discrete Laplacian mechanisms in the high privacy regime (as $(\epsilon, \delta) \to (0, 0)$) for ℓ^1 and ℓ^2 cost functions.

The near-optimality of the two mechanisms (designed for the special cases of $(\epsilon, 0)$ and $(0, \delta)$ differential privacy settings) is proved by demonstrating a uniform bound on the ratio between the costs of these two mechanisms and that of the optimal cost in the (ϵ, δ) differential privacy setting in the high privacy regime, i.e., as $(\epsilon, \delta) \to (0, 0)$ for ℓ^1 and ℓ^2 cost functions.

4.1.1 Summary of Our Results

We summarize our results in the following. Let V_{LB} denote the lower bound we derived for the cost under the differential privacy constraint. Let V_{UB}^{Lap} and V_{UB}^{uniform} denote the upper bounds for the cost achieved by the discrete Laplacian mechanism and the uniform noise mechanism. We show that

- For integer-valued query functions,
  - for $(0, \delta)$-differential privacy with the global sensitivity $\Delta = 1$, the uniform noise mechanism is optimal for all generic cost funtions,
  - for $(0, \delta)$-differential privacy with arbitrary global sensitivity Δ,

    $$\lim_{\delta \to 0} \frac{V_{UB}^{\text{uniform}}}{V_{LB}} = 1$$

    for ℓ^1 and ℓ^2 cost functions,
  - for (ϵ, δ)-differential privacy with ℓ^1 and ℓ^2 cost functions,

    $$\lim_{(\epsilon,\delta) \to (0,0)} \frac{\min(V_{UB}^{\text{Lap}}, V_{UB}^{\text{uniform}})}{V_{LB}} \leq C$$

    for some numerical constant C.
- For multiple dimensional integer-valued query functions,
  - for $(0, \delta)$-differential privacy with the global sensitivity $\Delta = 1$, the multiple dimensional uniform noise mechanism is optimal for ℓ^1 and ℓ^2 cost funtions,

– for $(0, \delta)$-differential privacy with arbitrary global sensitivity Δ, $\lim_{\delta \to 0} \frac{V_{UB}^{\text{uniform}}}{V_{LB}} = 1$ for ℓ^1 and ℓ^2 cost functions,

– for (ϵ, δ)-differential privacy with ℓ^1 and ℓ^2 cost functions,

$$\lim_{(\epsilon,\delta)\to(0,0)} \frac{\min(V_{UB}^{\text{Lap}}, V_{UB}^{\text{uniform}})}{V_{LB}} \leq C$$

for some numerical constant C, which is independent of the dimension of the query function.

4.2 Problem Formulation

Consider an integer-valued query function

$$q : \mathcal{D}^n \to \mathbb{Z},$$

where $\mathcal{D}^n$ is the domain of the databases.

The sensitivity of the query function q is defined as

$$\Delta \triangleq \max_{D_1, D_2 \subseteq \mathcal{D}^n : |D_1 - D_2| \leq 1} |q(D_1) - q(D_2)|, \tag{4.1}$$

where the maximum is taken over all possible pairs of neighboring database entries D_1 and D_2 which differ in at most one element, i.e., one is a proper subset of the other and the larger database contains just one additional element [3]. Clearly, Δ is an integer in this discrete setting.

Definition 4.1 ((ϵ, δ)-Differential Privacy [18])**.** *A randomized mechanism $\mathcal{K}$ gives ϵ-differential privacy if for all data sets D_1 and D_2 differing in at most one element, and all $S \subset Range(\mathcal{K})$,*

$$Pr[\mathcal{K}(D_1) \in S] \leq \exp(\epsilon)\ Pr[\mathcal{K}(D_2) \in S] + \delta. \tag{4.2}$$

4.2.1 Operational Meaning of (ϵ, δ)-Differential Privacy in the Context of Hypothesis Testing

As shown by [10], one can interpret the differential privacy constraint (4.2) in the context of hypothesis testing in terms of false alarm probability and missing detection probability. Indeed, consider a binary hypothesis testing problem over two neighboring datasets, $H_0 : D_1$

versus $H_1 : D_2$, where an individual's record is in D_2 only. Given a decision rule, let S be the decision region such that when the released output lies in S, H_1 will be rejected, and when the released output lies in S^C (the complement of S), H_0 will be rejected. The false-alarm probability P_{FA} and the missing-detection probability P_{MD} can be written as

$$P_{FA} = P(K(D_1) \in S^C),$$
$$P_{MD} = P(K(D_2) \in S).$$

Therefore, from (4.2) we get

$$1 - P_{FA} \leq e^{\epsilon} P_{MD} + \delta.$$

Thus

$$e^{\epsilon} P_{MD} + P_{FA} \geq 1 - \delta.$$

Switch D_1 and D_2 in (4.2), and we get

$$\Pr[\mathcal{K}(D_2) \in S] \leq \exp(\epsilon) \Pr[\mathcal{K}(D_1) \in S] + \delta.$$

Therefore,

$$1 - P_{MD} \leq e^{\epsilon} P_{FA} + \delta,$$

and thus

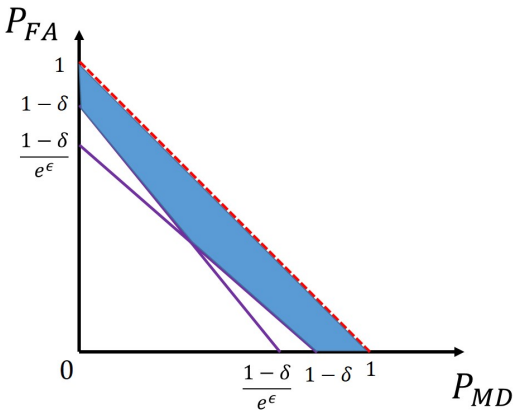
$$P_{MD} + e^{\epsilon} P_{FA} \geq 1 - \delta.$$

In conclusion, we have

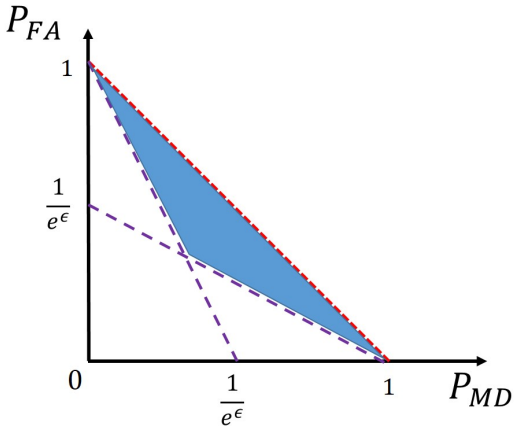
$$e^{\epsilon} P_{MD} + P_{FA} \geq 1 - \delta,$$
$$P_{MD} + e^{\epsilon} P_{FA} \geq 1 - \delta.$$

The (ϵ, δ)-differential privacy constraint implies that in the context of hypothesis testing, P_{FA} and P_{MD} cannot both be too small.

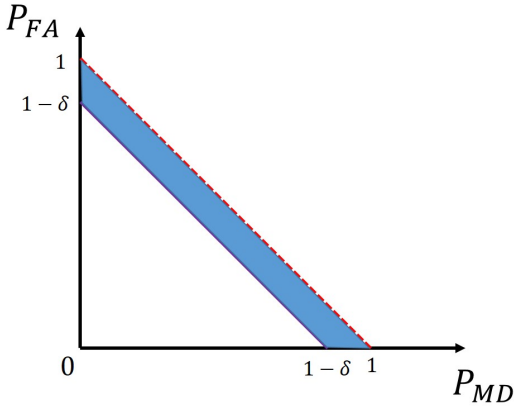
We plot the regions of P_{FA} and P_{MD} under (ϵ, δ)-differential privacy, and under two special cases: $(\epsilon, 0)$ and $(0, \delta)$-differential privacy, in Figure 4.1.



(a) (ϵ, δ)-Differential Privacy



(b) $(\epsilon, 0)$-Differential Privacy



(c) $(0, \delta)$-Differential Privacy

Figure 4.1: Regions of P_{MD} and P_{FA} in (ϵ, δ), $(\epsilon, 0)$ and $(0, \delta)$-Differential Privacy

4.2.2 Cost-Minimization/Utility-Maximization Formulation

The standard approach to preserving the differential privacy is to add noise to the output of query function. Let $q(D)$ be the value of the query function evaluated at $D \subseteq \mathcal{D}^n$, the noise-adding mechanism $\mathcal{K}$ will output

$$\mathcal{K}(D) = q(D) + X,$$

where X is the noise added by the mechanism to the output of query function. To make the output of the mechanism be valid, i.e., $q(D) + X \in \mathbb{Z}$, X can only take integer values.

Let $\mathcal{P}$ be the probability mass function of the noise X, and let $\mathcal{P}_i$ denote $\Pr[X = i]$. For a set $S \subset \mathbb{Z}$, denote $\Pr[X \in S]$ by $\mathcal{P}_S$.

In the following we derive the differential privacy constraint on the probability distribu-

tion of X from (4.2).

$$\begin{aligned} & \Pr[\mathcal{K}(D_1) \in S] && \leq \exp(\epsilon) \Pr[\mathcal{K}(D_2) \in S] + \delta \\ \Leftrightarrow\ & \Pr[q(D_1) + X \in S] && \leq \exp(\epsilon) \Pr[q(D_2) + X \in S] + +\delta \\ \Leftrightarrow\ & \mathcal{P}_{S-q(D_1)} && \leq \exp(\epsilon)\, \mathcal{P}_{S-q(D_2)} + \delta \\ \Leftrightarrow\ & \mathcal{P}_{S'} && \leq \exp(\epsilon)\, \mathcal{P}_{S'+q(D_1)-q(D_2)} + \delta, \end{aligned} \tag{4.3}$$

where $S' \triangleq S - q(D_1) = \{s - q(D_1) | s \in S\}$.

Since (4.2) holds for any set $S \subseteq \mathbb{Z}$, and $|q(D_1) - q(D_2)| \leq \Delta$, from (4.3), we have

$$\mathcal{P}_S \leq \exp(\epsilon)\, \mathcal{P}_{S+d} + \delta, \tag{4.4}$$

for any set $S \subseteq \mathbb{Z}$ and for all $|d| \leq \Delta$.

Consider a cost function $\mathcal{L}(\cdot) : \mathbb{Z} \to \mathbb{R}$, which is a function of the added noise X. Our goal is to minimize the expectation of the cost subject to the (ϵ, δ)-differential privacy constraint (4.4):

$$\begin{aligned} V^* := & \min_{\mathcal{P}} \sum_{i=-\infty}^{+\infty} \mathcal{L}(i)\mathcal{P}(i) \\ & \text{subject to } \mathcal{P}_S \leq \exp(\epsilon)\, \mathcal{P}_{S+d} + \delta, \forall S \subset \mathbb{Z}, d \in \mathbb{Z}, |d| \leq |\Delta|. \end{aligned}$$

We restrict our attention to the scenario when the cost function $\mathcal{L}(k)$ is symmetric (around $k = 0$) and monotonically increasing for $k \geq 0$. Furthermore, without loss of generality, we assume $\mathcal{L}(0) = 0$. Using the same argument in Lemma 28 in [50], we only need to consider symmetric noise probability distributions.

4.3 $(0, \delta)$-Differential Privacy

We first consider the simple case when $\epsilon = 0$, i.e., $(0, \delta)$-differential privacy. The $(0, \delta)$-differential privacy constraint requires that the total variation of the conditional probability distributions of the query output for neighboring datasets should be bounded by δ.

In the differential privacy constraint (4.4), by choosing the subset $S = S_k := \{\ell : \ell \geq k\}$ for $k \in \mathbb{N}$ and $d = \Delta$, we see that the noise probability distribution $\mathcal{P}$ must satisfy the

constraints

$$\sum_{\ell=0}^{\Delta-1} \mathcal{P}_{k+\ell} \leq \delta, \quad \forall k \in \mathbb{N}. \tag{4.5}$$

4.3.1 $\Delta = 1$

In the special case $\Delta = 1$, the constraints in (4.5) are particularly simple:

$$p_k \leq \delta; \quad \forall k \geq 0.$$

For symmetric cost functions $\mathcal{L}(k)$ that are monotonically increasing in $k \geq 0$, we can now readily argue that the *uniform* probability distribution is optimal.

To avoid integer rounding issues, assume $\frac{1}{2\delta}$ is an integer.

Theorem 4.1. *If $\Delta = 1$, then*

$$V^* = \sum_{k=-\frac{1}{2\delta}}^{\frac{1}{2\delta}-1} \delta\mathcal{L}(k),$$

and the optimal noise probability distribution is

$$\mathcal{P}_k = \begin{cases} \delta & -\frac{1}{2\delta} \leq k \leq \frac{1}{2\delta} - 1 \\ 0 & \textit{otherwise} \end{cases}. \tag{4.6}$$

4.3.2 General Lower Bound for $\Delta \geq 2$

We now turn to understanding near-optimal $(0, \delta)$ privacy mechanisms in terms of minimizing the expected loss when the sensitivity $\Delta \geq 2$.

Recall that in $(0, \delta)$-differential privacy, the minimum cost V^* is the result of the following optimization problem, which is a linear program:

$$\begin{aligned} V^* := \min \quad & \sum_{k=-\infty}^{+\infty} \mathcal{L}(k)\mathcal{P}_k \\ \text{such that} \quad & p_k \geq 0 \quad \forall k \in N \end{aligned}$$

$$\sum_{k=-\infty}^{+\infty} \mathcal{P}_k = 1$$

$$\mathcal{P}_S \leq \ \mathcal{P}_{S+d} + \delta, \forall S \subset \mathbb{Z}, d \in \mathbb{Z}, |d| \leq |\Delta|. \tag{4.7}$$

Since $\mathcal{L}(\cdot)$ is a symmetric function, we can assume $\mathcal{P}$ is a symmetric probability distribution. In addition, we relax the constraint (4.7) by choosing $d = \Delta$ and $S = S_k$ for $k \in \mathbb{N}$. Then we get a relaxed linear program, the solution of which is a lower bound for V^*. More precisely,

$$V_{LB} := \min \quad 2\sum_{k=1}^{\infty} \mathcal{L}(k)\mathcal{P}_k \tag{4.8}$$

$$\text{such that} \quad \mathcal{P}_k \geq 0 \quad \forall k \in N$$

$$\frac{\mathcal{P}_0}{2} + \sum_{k=1}^{\infty} \mathcal{P}_k \geq \frac{1}{2} \tag{4.9}$$

$$-\sum_{\ell=0}^{\Delta-1} \mathcal{P}_{k+\ell} \geq -\delta, \quad \forall k \in \mathbb{N}. \tag{4.10}$$

To avoid integer rounding issues, assume $\frac{1}{2\delta}$ is a positive integer.

Theorem 4.2. *If*

$$\mathcal{L}(1 + \frac{\Delta}{2\delta}) \geq 2\left(\mathcal{L}(1) + \sum_{i=1}^{\frac{1}{2\delta}} (\mathcal{L}(1 + i\Delta) - \mathcal{L}(i\Delta))\right), \tag{4.11}$$

then

$$V^* \geq V_{LB} = 2\delta \sum_{i=0}^{\frac{1}{2\delta}-1} \mathcal{L}(1 + i\Delta). \tag{4.12}$$

Proof. See Appendix C.1. □

4.3.3 Uniform Noise Mechanism

Consider the noise with the *uniform* probability distribution:

$$\mathcal{P}_k = \begin{cases} \frac{\delta}{\Delta} & \forall -\frac{\Delta}{2\delta} \leq k \leq \frac{\Delta}{2\delta} - 1 \\ 0 & \text{otherwise} \end{cases} \tag{4.13}$$

It is readily verified that this noise probability distribution satisfies the $(0, \delta)$ differential privacy constraint. Therefore, an upper bound for V^* is

Theorem 4.3.

$$V^* \leq V_{UB} \triangleq 2 \sum_{i=1}^{\frac{\Delta}{2\delta}-1} \frac{\delta}{\Delta} \mathcal{L}(i) + \frac{\delta}{\Delta} \mathcal{L}(\frac{\Delta}{2\delta}). \tag{4.14}$$

4.3.4 Comparison of V_{LB} and V_{UB}

We first apply the lower bound (4.12) and upper bound (4.14) to the ℓ^1 and ℓ^2 cost functions, i.e., $\mathcal{L}(i) = |i|$ and $\mathcal{L}(i) = i^2$, in which V^* corresponds to the minimum expected noise amplitude and minimum noise power, respectively.

Note that in the case $\mathcal{L}(i) = |i|$, the condition (4.11) in Theorem 4.2 is

$$\frac{\Delta}{2\delta} \geq \frac{1}{\delta} + 1. \tag{4.15}$$

When $\Delta \geq 3$, (4.11) holds.

Corollary 4.4. *For the cost function $\mathcal{L}(i) = |i|$,*

$$V_{LB} = \frac{\Delta}{4\delta} + 1 - \frac{\Delta}{2},$$
$$V_{UB} = \frac{\Delta}{4\delta},$$

and thus the additive gap

$$V_{UB} - V_{LB} = \frac{\Delta}{2} - 1$$

is a constant independent of δ.

In the case $\mathcal{L}(i) = i^2$, the condition (4.11) in Theorem 4.2 is

$$\frac{\Delta}{2\delta^2}(\frac{\Delta}{2} - 1) \geq \frac{1}{\delta} + 1. \tag{4.16}$$

When $\Delta \geq 3$, (4.16) holds.

Corollary 4.5. *For the cost function $\mathcal{L}(i) = i^2$,*

$$V_{LB} = \frac{\Delta^2}{12\delta^2} - \frac{\Delta^2}{4\delta} + \Delta(\frac{1}{2\delta} - 1) + \frac{\Delta^2}{6} + 1,$$
$$V_{UB} = \frac{\Delta^2}{12\delta^2} + \frac{1}{6},$$

and thus the multiplicative gap

$$\lim_{\delta \to 0} \frac{V_{UB}}{V_{LB}} = 1.$$

Proof. See Appendix C.2. □

Corollary 4.6. *Given a positive integer m, consider the cost function $\mathcal{L}(i) = |i|^m$. Then*

$$\lim_{\delta \to 0} \frac{V_{UB}}{V_{LB}} = 1.$$

Proof. By induction, it is easy to show that $\sum_{i=1}^{n} i^m = \Theta(\frac{n^{m+1}}{m+1})$, and

$$\lim_{n \to +\infty} \frac{\sum_{i=1}^{n} i^m}{\frac{n^{m+1}}{m+1}} = 1.$$

Therefore,

$$\begin{aligned}
\lim_{\delta \to 0} \frac{V_{UB}}{V_{LB}} &= \lim_{\delta \to 0} \frac{2\frac{\delta}{\Delta} \sum_{i=1}^{\frac{\Delta}{2\delta}-1} i^m + \frac{\delta}{\Delta} \frac{\Delta^m}{(2\delta)^m}}{2\delta \sum_{i=0}^{\frac{1}{2\delta}-1} (1 + i\Delta)^m} \\
&= \lim_{\delta \to 0} \frac{2\frac{\delta}{\Delta} \frac{\frac{\Delta^{m+1}}{(2\delta)^{m+1}}}{m+1}}{2\delta\Delta^m \frac{(\frac{1}{2\delta})^{m+1}}{m+1}} \\
&= 1.
\end{aligned}$$

□

For general cost functions, we have the following bound on the multiplicative gap between the lower bound and upper bound.

Corollary 4.7. *Given a cost function $\mathcal{L}(\cdot)$ satisfying*

$$\sup_{k\geq T}\frac{\mathcal{L}(k)}{\mathcal{L}(k-\Delta+1)}\leq C,$$

for some integer $T\in\mathbb{N}$, and some positive number $C\in\mathbb{R}$, then

$$\lim_{\delta\to 0}\frac{V_{UB}}{V_{LB}}\leq 1+(1+\frac{1}{2\Delta})C.$$

Proof. See Appendix C.3. □

4.4 (ϵ,δ)-Differential Privacy

Recall that since $\mathcal{L}(\cdot)$ is a symmetric function, without loss of generality, we can restrict ourselves to symmetric noise probability distributions, i.e.,

$$\mathcal{P}_k=\mathcal{P}_{-k},\forall k\in\mathbb{Z}. \tag{4.17}$$

The differential privacy constraint in (4.4) can be understood in some detail by choosing the subset $S=S_k:=\{\ell:\ell\geq k\}$ for $k\in\mathbb{N}$. In this case we see that the noise probability distribution must satisfy the following constraints. For $k=0$ and $d=\Delta$,

$$\mathcal{P}_{S_0}\leq e^{\epsilon}\mathcal{P}_{S_\Delta}+\delta. \tag{4.18}$$

By using the symmetry condition in (4.17) and the fact that $\sum_{\ell=-\infty}^{+\infty}\mathcal{P}_\ell=1$, from (4.18) we get

$$\mathcal{P}_0\frac{1+e^{\epsilon}}{2}+e^{\epsilon}\sum_{\ell=1}^{\Delta-1}\mathcal{P}_\ell\leq\delta+\frac{e^{\epsilon}-1}{2}.$$

For $k=1$ and $d=\Delta$, we have

$$\mathcal{P}_{S_1}\leq e^{\epsilon}\mathcal{P}_{S_{\Delta+1}}+\delta,$$

and thus

$$\mathcal{P}_0\frac{e^{\epsilon}-1}{2}+e^{\epsilon}\sum_{\ell=1}^{\Delta}\mathcal{P}_\ell\leq\delta+\frac{e^{\epsilon}-1}{2}.$$

For general $k \geq 2$ and $d = \Delta$, we have

$$\mathcal{P}_{S_k} \leq e^{\epsilon} \mathcal{P}_{S_{\Delta+k}} + \delta,$$

and thus

$$\mathcal{P}_0 \frac{e^{\epsilon} - 1}{2} + (e^{\epsilon} - 1) \sum_{\ell=1}^{k-1} \mathcal{P}_\ell + e^{\epsilon} \sum_{\ell=k}^{k+\Delta-1} \mathcal{P}_\ell \leq \delta + \frac{e^{\epsilon} - 1}{2}.$$

4.4.1 Lower Bound

By restricting the set S in (4.4) to be $S_k := \{\ell : \ell \geq k\}$ for $k \in \mathbb{Z}$ and restricting d to be Δ, we get the following relaxed linear program, the solution of which is a lower bound for V^*:

$$\begin{aligned}
V_{LB} := \min \quad & 2 \sum_{k=1}^{\infty} \mathcal{L}(k) \mathcal{P}_k \\
\text{such that} \quad & \mathcal{P}_k \geq 0 \quad \forall k \in N \\
& \frac{\mathcal{P}_0}{2} + \sum_{k=1}^{\infty} \mathcal{P}_k \geq \frac{1}{2} && (4.19) \\
& \mathcal{P}_0 \frac{1 + \epsilon^{\epsilon}}{2} + e^{\epsilon} \sum_{k=1}^{\Delta-1} \mathcal{P}_k \leq \delta + \frac{e^{\epsilon} - 1}{2} && (4.20) \\
& \mathcal{P}_0 \frac{e^{\epsilon} - 1}{2} + e^{\epsilon} \sum_{k=1}^{\Delta} \mathcal{P}_k \leq \delta + \frac{e^{\epsilon} - 1}{2} && (4.21) \\
& \mathcal{P}_0 \frac{e^{\epsilon} - 1}{2} + (e^{\epsilon} - 1) \sum_{k=1}^{i-1} \mathcal{P}_k + e^{\epsilon} \sum_{k=i}^{i+\Delta-1} \mathcal{P}_k \leq \delta + \frac{e^{\epsilon} - 1}{2}, \forall i \geq 2. && (4.22)
\end{aligned}$$

Define

$$\begin{aligned}
a &\triangleq \frac{\delta + \frac{e^{\epsilon}-1}{2}}{e^{\epsilon}}, \\
b &\triangleq e^{-\epsilon}.
\end{aligned}$$

To avoid integer rounding issues, assume that there exists an integer n such that

$$\sum_{k=0}^{n-1} a b^k = \frac{1}{2}.$$

Theorem 4.8. *If*

$$\sum_{i=1}^{n-1} e^{-i\epsilon}(2\mathcal{L}(i\Delta) - \mathcal{L}(1+(i-1)\Delta) - \mathcal{L}(1+i\Delta)) \geq \mathcal{L}(1),$$

then we have

$$V^* \geq V_{LB} = 2\sum_{k=0}^{n-1} \frac{\delta + \frac{e^\epsilon - 1}{2}}{e^\epsilon} e^{-k\epsilon}\mathcal{L}(1+k\Delta). \tag{4.23}$$

Proof. See Appendix C.4. □

4.4.2 Upper Bound: the Uniform Noise Mechanism and the Discrete Laplacian Mechanism

Since $(0,\delta)$-differential privacy implies (ϵ,δ)-differential privacy, we can use the uniform noise mechanism with noise probability distribution defined in (4.13) to preserve (ϵ,δ)-differential privacy, and the corresponding upper bound is

Theorem 4.9. *For (ϵ,δ)-differential privacy, we have*

$$V^* \leq V_{UB}^{uniform} = 2\sum_{i=1}^{\frac{\Delta}{2\delta}-1} \frac{\delta}{\Delta}\mathcal{L}(i) + \frac{\delta}{\Delta}\mathcal{L}(\frac{\Delta}{2\delta}). \tag{4.24}$$

On the other hand, if we simply ignore the parameter δ (i.e., set $\delta = 0$), we can use a discrete variant of Laplacian distribution to satisfy the $(\epsilon,0)$-differential privacy, which implies (ϵ,δ)-differential privacy.

More precisely, define $\lambda \triangleq e^{-\frac{\epsilon}{\Delta}}$.

Theorem 4.10. *The probability distribution $\mathcal{P}$ with*

$$p_k \triangleq \frac{1-\lambda}{1+\lambda}\lambda^{|k|}, \forall k \in \mathbb{Z},$$

satisfies the (ϵ,δ)-differential privacy constraint, and the corresonding cost is

$$\sum_{k=-\infty}^{+\infty} p_k\mathcal{L}(k) = 2\sum_{k=1}^{+\infty} \frac{1-\lambda}{1+\lambda}\lambda^k\mathcal{L}(k).$$

Corollary 4.11.

$$V^* \leq V_{UB}^{Lap} \triangleq 2\sum_{k=1}^{+\infty} \frac{1-\lambda}{1+\lambda}\lambda^k \mathcal{L}(k). \tag{4.25}$$

4.4.3 Comparison of Lower Bound and Upper Bound

In this section, we compare the lower bound (4.23) and the upper bounds V_{UB}^{uniform} and V_{UB}^{Lap} for (ϵ, δ)-differential privacy for the ℓ^1 and ℓ^2 cost functions, i.e., $\mathcal{L}(i) = |i|$ and $\mathcal{L}(i) = i^2$, in which V^* corresponds to the minimum expected noise amplitude and minimum noise power, respectively. We show that the multiplicative gap between the lower bound and upper bound is bounded by a constant as $(\epsilon, \delta) \to (0, 0)$.

$\epsilon \leq \delta$ Regime

We first compare the gap between the lower bound V_{LB} and the upper bound V_{UB}^{uniform} in the regime $\epsilon \leq \delta$ as $\delta \to 0$.

Corollary 4.12. *For the cost function $\mathcal{L}(k) = |k|$, in the regime $\epsilon \leq \delta$, we have*

$$\lim_{\delta \to 0} \frac{V_{UB}^{uniform}}{V_{LB}} \leq \frac{1}{4(1 - 2\log\frac{3}{2})} \approx 1.32$$

Proof. See Appendix C.5. □

Corollary 4.13. *For the cost function $\mathcal{L}(k) = k^2$, in the regime $\epsilon \leq \delta$, we have*

$$\lim_{\delta \to 0} \frac{V_{UB}^{uniform}}{V_{LB}} \leq \frac{1}{12(2 - 4\log(\frac{3}{2}) - 2(\log(\frac{3}{2}))^2)} \approx \frac{5}{3}.$$

Proof. See Appendix C.6. □

$\delta \leq \epsilon$ Regime

We then compare the gap between the lower bound V_{LB} and the upper bound V_{UB}^{Lap} in the regime $\delta \leq \epsilon$ as $\epsilon \to 0$.

Corollary 4.14. *For the cost function $\mathcal{L}(k) = |k|$, in the regime $\delta \leq \epsilon$, we have*

$$\lim_{\epsilon \to 0} \frac{V_{UB}^{Lap}}{V_{LB}} \leq \frac{1}{1 - 2\log\frac{3}{2}} \approx 5.29.$$

Proof. See Appendix C.7. □

Corollary 4.15. *For the cost function $\mathcal{L}(k) = k^2$, in the regime $\epsilon \leq \delta$, we have*

$$\lim_{\delta \to 0} \frac{V_{UB}^{Lap}}{V_{LB}} \leq \frac{2}{(2 - 4\log(\frac{3}{2}) - 2(\log(\frac{3}{2}))^2)} \approx 40.$$

Proof. See Appendix C.8. □

4.5 (ϵ, δ)-Differential Privacy in the Multiple Dimensional Setting

In this section we consider the (ϵ, δ)-differential privacy in the multiple dimensional setting, where the query output has multiple components and the global sensitivity Δ is defined as the maximum ℓ^1 norm of the difference of the query outputs over two neighboring datasets.

Let d be the dimension of the query output. Hence, the query output $q(D) \in \mathbb{Z}^d$. Let $\mathcal{P}$ be the probability mass function of the additive noise over the domain $\mathbb{Z}^d$. Then the (ϵ, δ)-differential privacy constraint on $\mathcal{P}$ in the multiple dimensional setting is that

$$\mathcal{P}_S \leq \mathcal{P}_{S+\mathbf{v}} + \delta, \forall S \subset \mathbb{Z}^d, \mathbf{v} \in \mathbb{Z}^d, \|\mathbf{v}\|_1 \leq \Delta. \tag{4.26}$$

Consider a cost function $\mathcal{L}(\cdot) : \mathbb{Z}^d \to \mathbb{R}$, which is a function of the added noise X. Our goal is to minimize the expectation of the cost subject to the (ϵ, δ)-differential privacy constraint (4.26):

$$\begin{aligned} V^* := \min_{\mathcal{P}} \ & \sum_{\mathbf{v} \in \mathbb{Z}^d} \mathcal{L}(\mathbf{v})\mathcal{P}(\mathbf{v}) \\ \text{subject to } & \mathcal{P}_S \leq \mathcal{P}_{S+\mathbf{v}} + \delta, \forall S \subset \mathbb{Z}^d, \mathbf{v} \in \mathbb{Z}^d, \|\mathbf{v}\|_1 \leq \Delta. \end{aligned}$$

4.5.1 $(0, \delta)$-Differential Privacy

We first consider the simple case when $\epsilon = 0$, i.e., $(0, \delta)$-differential privacy. The $(0, \delta)$-differential privacy constraint requires that the total variation of the conditional probability distributions of the query output for neighboring datasets should be bounded by δ.

In the differential privacy constraint (4.26), by choosing the subset

$$S = S_k^m := \{(i_1, i_2, \ldots, i_d) \in \mathbb{Z}^d \mid i_m \geq k\}$$

for $k \in \mathbb{N}$, $m \in \{1, 2, \ldots, d\}$, and choosing $\mathbf{v}$ such that only one compoment is Δ and all other components are zero, we see that the noise probability distribution $\mathcal{P}$ must satisfy the constraints

$$\sum_{(i_1,i_2,\ldots,i_d)\in\mathbb{Z}^d : k \leq i_m \leq k+\Delta-1} \mathcal{P}(i_1, i_2, \ldots, i_d) \leq \delta, \quad \forall k \in \mathbb{N}, \forall m \in \{1, 2, \ldots, d\}.$$

To avoid integer-rounding issues, we assume that $\frac{1}{2\delta}$ is an integer.

Lower Bound on V^*

We relax the constraint (4.26) by choosing S to be S_k^m and choosing $\mathbf{v}$ such that only one compoment is Δ and all other components are zero. Then we get a relaxed linear program, the solution of which is a lower bound for V^*. More precisely,

$$\begin{aligned} V^* \geq V_{LB} := \min \quad & \sum_{\mathbf{i}\in\mathbb{Z}^d} \mathcal{P}(\mathbf{i})\mathcal{L}(\mathbf{i}) \\ \text{such that} \quad & \mathcal{P}(\mathbf{i}) \geq 0 \quad \forall \mathbf{i} \in \mathbb{Z}^d \\ & \sum_{\mathbf{i}\in\mathbb{Z}^d} \mathcal{P}(\mathbf{i}) \geq 1 \\ & \sum_{(i_1,i_2,\ldots,i_d)\in\mathbb{Z}^d : k \leq i_m \leq k+\Delta-1} \mathcal{P}(i_1, i_2, \ldots, i_d) \leq \delta, \quad \forall k \in \mathbb{N}, \forall m \in \{1, 2, \ldots, d\}. \end{aligned} \tag{4.27}$$

Theorem 4.16. *In the case $\mathcal{L}(\mathbf{i}) = \|\mathbf{i}\|_1, \forall \mathbf{i} \in \mathbb{Z}^d$, we have*

$$V_{LB} \geq \frac{d\Delta}{4\delta} - \frac{\Delta - 1}{2} d.$$

Proof. See Appendix C.9. □

Theorem 4.17. *In the case $\mathcal{L}(\mathbf{i}) = \|\mathbf{i}\|_2^2 = \sum_{m=1}^d i_m^2, \forall \mathbf{i} = (i_1, \ldots, i_d) \in \mathbb{Z}^d$, we have*

$$V_{LB} \geq \frac{d\Delta^2}{12\delta^2} + (\frac{1}{\Delta} - 1)\frac{d\Delta^2}{4\delta} + \frac{1 - \Delta}{2} d + \frac{d\Delta^2}{6}.$$

Proof. See Appendix C.10. □

Uniform Noise Mechanism in the Multiple Dimensional Setting

Consider the noise with the *uniform* probability distribution:

$$\mathcal{P}(i_1, i_2, \ldots, i_d) = \begin{cases} \frac{\delta^d}{\Delta^d} & -\frac{\Delta}{2\delta} \leq i_m \leq \frac{\Delta}{2\delta} - 1, \forall m \in \{1, 2, \ldots, d\} \\ 0 & \text{otherwise} \end{cases}. \tag{4.28}$$

It is readily verified that this noise probability distribution satisfies the $(0, \delta)$ differential privacy constraint (4.26). Therefore, an upper bound for V^* is

Theorem 4.18.

$$V^* \leq V_{UB} \triangleq \sum_{(i_1, i_2, \ldots, i_d) \in \mathbb{Z}^d \mid -\frac{\Delta}{2\delta} \leq i_m \leq \frac{\Delta}{2\delta} - 1, \forall m \in \{1, 2, \ldots, d\}} \frac{\delta^d}{\Delta^d} \mathcal{L}(i_1, i_2, \ldots, i_d). \tag{4.29}$$

Corollary 4.19. *In the case* $\mathcal{L}(\mathbf{i}) = \|\mathbf{i}\|_1, \forall \mathbf{i} \in \mathbb{Z}^d$*, we have*

$$V_{UB} = \frac{d\Delta}{4\delta}.$$

Proof.

$$\begin{aligned} V_{UB} &= \sum_{(i_1, i_2, \ldots, i_d) \in \mathbb{Z}^d \mid -\frac{\Delta}{2\delta} \leq i_m \leq \frac{\Delta}{2\delta} - 1, \forall m \in \{1, 2, \ldots, d\}} \frac{\delta^d}{\Delta^d} \mathcal{L}(i_1, i_2, \ldots, i_d) \\ &= \sum_{i_1 = -\frac{\Delta}{2\delta}}^{\frac{\Delta}{2\delta} - 1} \cdots \sum_{i_d = -\frac{\Delta}{2\delta}}^{\frac{\Delta}{2\delta} - 1} \frac{\delta^d}{\Delta^d} (|i_1| + \cdots + |i_d|) \\ &= d \sum_{i_1 = -\frac{\Delta}{2\delta}}^{\frac{\Delta}{2\delta} - 1} \cdots \sum_{i_d = -\frac{\Delta}{2\delta}}^{\frac{\Delta}{2\delta} - 1} \frac{\delta^d}{\Delta^d} |i_1| \\ &= d \left(\frac{\Delta}{\delta} \right)^{d-1} \sum_{i_1 = -\frac{\Delta}{2\delta}}^{\frac{\Delta}{2\delta} - 1} \frac{\delta^d}{\Delta^d} |i_1| \\ &= d \left(\frac{\Delta}{\delta} \right)^{d-1} \frac{\delta^d}{\Delta^d} \left(\frac{(1 + \frac{\Delta}{2\delta}) \frac{\Delta}{2\delta}}{2} + \frac{\frac{\Delta}{2\delta} (\frac{\Delta}{2\delta} - 1)}{2} \right) \\ &= \frac{d\Delta}{4\delta}. \end{aligned}$$

□

Corollary 4.20. *In the case* $\mathcal{L}(\mathbf{i}) = \|\mathbf{i}\|_2^2 \triangleq \sum_{m=1}^{d} i_m^2, \forall \mathbf{i} = (i_1, \ldots, i_d) \in \mathbb{Z}^d$*, we have*

$$V_{UB} = \frac{d\Delta^2}{12\delta^2} + \frac{d}{6}.$$

Proof.

$$\begin{aligned}
V_{UB} &= \sum_{(i_1,i_2,\ldots,i_d)\in\mathbb{Z}^d \mid -\frac{\Delta}{2\delta} \le i_m \le \frac{\Delta}{2\delta}-1, \forall m\in\{1,2,\ldots,d\}} \frac{\delta^d}{\Delta^d}\mathcal{L}(i_1, i_2, \ldots, i_d) \\
&= \sum_{i_1=-\frac{\Delta}{2\delta}}^{\frac{\Delta}{2\delta}-1} \cdots \sum_{i_d=-\frac{\Delta}{2\delta}}^{\frac{\Delta}{2\delta}-1} \frac{\delta^d}{\Delta^d}(|i_1|^2 + \cdots + |i_d|^2) \\
&= d \sum_{i_1=-\frac{\Delta}{2\delta}}^{\frac{\Delta}{2\delta}-1} \cdots \sum_{i_d=-\frac{\Delta}{2\delta}}^{\frac{\Delta}{2\delta}-1} \frac{\delta^d}{\Delta^d}|i_1|^2 \\
&= d\left(\frac{\Delta}{\delta}\right)^{d-1} \sum_{i_1=-\frac{\Delta}{2\delta}}^{\frac{\Delta}{2\delta}-1} \frac{\delta^d}{\Delta^d}|i_1|^2 \\
&= d\left(\frac{\Delta}{\delta}\right)^{d-1} \frac{\delta^d}{\Delta^d}\left(\frac{\frac{\Delta}{2\delta}(1+\frac{\Delta}{2\delta})(\frac{\Delta}{\delta}+1)}{6} + \frac{(\frac{\Delta}{2\delta}-1)\frac{\Delta}{2\delta}(\frac{\Delta}{\delta}-1)}{6}\right) \\
&= \frac{d\Delta^2}{12\delta^2} + \frac{d}{6}.
\end{aligned}$$

□

Comparison of Lower Bound and Upper Bound for the ℓ^1 Cost Function

Corollary 4.21. *For the cost function* $\mathcal{L}(\mathbf{i}) = \|\mathbf{i}\|_1$*,*

$$\begin{aligned}
V_{LB} &\ge \frac{d\Delta}{4\delta} - \frac{\Delta-1}{2}d, \\
V_{UB} &= \frac{d\Delta}{4\delta},
\end{aligned}$$

and thus the additive gap

$$V_{UB} - V_{LB} \le \frac{\Delta-1}{2}d,$$

which is a constant independent of δ*.*

In the case that $\Delta = 1$, the additive gap $\frac{\Delta-1}{2}d$ is zero, and thus $V_{LB} = V_{UB}$.

Corollary 4.22. *For the cost function* $\mathcal{L}(\mathbf{i}) = \|\mathbf{i}\|_1$*, if* $\Delta = 1$*, then*

$$V^* = V_{UB} = V_{LB} = \frac{d\Delta}{4\delta},$$

and thus the uniform noise mechanism is optimal in this setting.

Corollary 4.23. *For the cost function* $\mathcal{L}(\mathbf{i}) = \|\mathbf{i}\|_2^2$*,*

$$V_{LB} \geq \frac{d\Delta^2}{12\delta^2} + (\frac{1}{\Delta} - 1)\frac{d\Delta^2}{4\delta} + \frac{1-\Delta}{2}d + \frac{d\Delta^2}{6},$$
$$V_{UB} = \frac{d\Delta^2}{12\delta^2} + \frac{d}{6},$$

and thus

$$\lim_{\delta \to 0} \frac{V_{UB}}{V_{LB}} = 1.$$

In the case that $\Delta = 1$,

$$V_{LB} \geq \frac{d}{12\delta^2} + \frac{d}{6} = V_{UB},$$

and thus $V_{LB} = V_{UB}$.

Corollary 4.24. *For the cost function* $\mathcal{L}(\mathbf{i}) = \|\mathbf{i}\|_2^2$*, if* $\Delta = 1$*, then*

$$V^* = V_{UB} = V_{LB} = \frac{d}{12\delta^2} + \frac{d}{6},$$

and thus the uniform noise mechanism is optimal in this setting.

4.5.2 (ϵ, δ)-Differential Privacy

The (ϵ, δ)-differential privacy constraint on the probability mass function $\mathcal{P}$ in the multiple dimensional setting is that

$$\mathcal{P}_S \leq e^{\epsilon}\mathcal{P}_{S+\mathbf{v}} + \delta, \forall S \subset \mathbb{Z}^d, \mathbf{v} \in \mathbb{Z}^d, \|\mathbf{v}\|_1 \leq \Delta.$$

We relax this constraint by choosing S to be S_k^m and choosing $\mathbf{v}$ such that only one compoment is Δ and all other components are zero. Then we get a relaxed linear program, the solution of which is a lower bound for V^*. More precisely,

$$
\begin{aligned}
V^* \geq V_{LB} := \min \quad & \sum_{\mathbf{i}\in\mathbb{Z}^d} \mathcal{P}(\mathbf{i})\mathcal{L}(\mathbf{i}) \\
\text{such that} \quad & \mathcal{P}(\mathbf{i}) \geq 0 \quad \forall \mathbf{i} \in \mathbb{Z}^d \\
& \sum_{\mathbf{i}\in\mathbb{Z}^d} \mathcal{P}(\mathbf{i}) \geq 1 \\
& \forall k \in \mathbb{N}, \forall m \in \{1, 2, \ldots, d\}, \\
& \sum_{(i_1,i_2,\ldots,i_d)\in\mathbb{Z}^d : k \leq i_m \leq k+\Delta-1} \mathcal{P}(i_1, i_2, \ldots, i_d) - (e^\epsilon - 1) \sum_{(i_1,i_2,\ldots,i_d)\in\mathbb{Z}^d : i_m \geq k+\Delta} \mathcal{P}(i_1, i_2, \ldots, i_d) \leq \delta.
\end{aligned} \tag{4.30}
$$

We are interested in characterizing V^* for the ℓ^1 and ℓ^2 cost functions in the high privacy regime when $(\epsilon, \delta) \to (0, 0)$.

Lower Bound for the ℓ^1 Cost Function

The dual linear program of (4.30) for the ℓ^1 cost function $\mathcal{L}(\mathbf{i}) = \|\mathbf{i}\|_1$ is that

$$
V_{LB} := \max \quad \mu - \delta \left(\sum_{i_1\in\mathbb{Z}} y_{i_1}^{(1)} + \sum_{i_2\in\mathbb{Z}} y_{i_2}^{(2)} + \cdots + \sum_{i_d\in\mathbb{Z}} y_{i_d}^{(d)} \right)
$$

$$
\begin{aligned}
\text{such that} \quad & y_{i_1}^{(1)}, y_{i_2}^{(2)}, \ldots, y_{i_d}^{(d)} \geq 0, \forall i_1 \in \mathbb{Z}, i_2 \in \mathbb{Z}, \ldots, i_d \in \mathbb{Z} \\
& \mu - \sum_{i_1\in[k_1-\Delta+1,k_1]} y_{i_1}^{(1)} + (e^\epsilon - 1) \sum_{i_1\leq k_1-\Delta} y_{i_1}^{(1)} \\
& - \cdots - \sum_{i_d\in[k_d-\Delta+1,k_d]} y_{i_d}^{(d)} + (e^\epsilon - 1) \sum_{i_d\leq k_d-\Delta} y_{i_d}^{(d)} \\
& \leq |k_1| + |k_2| + \cdots + |k_d|, \forall (k_1, \ldots, k_d) \in \mathbb{Z}^d.
\end{aligned}
$$

Given the parameters (ϵ, δ), let $\beta = \max(\epsilon, \delta)$. Since (β, β)-differential privacy is a relaxed version of (ϵ, δ)-differential privacy, in the above dual program we can replace both ϵ and δ by β, and the optimal value of the objective function will still be a lower bound of V^*. More precisely,

$$
V^* \geq V'_{LB} := \max \quad \mu - \beta \left(\sum_{i_1\in\mathbb{Z}} y_{i_1}^{(1)} + \sum_{i_2\in\mathbb{Z}} y_{i_2}^{(2)} + \cdots + \sum_{i_d\in\mathbb{Z}} y_{i_d}^{(d)} \right)
$$

$$\begin{aligned}\text{such that}\quad & y_{i_1}^{(1)}, y_{i_2}^{(2)}, \ldots, y_{i_d}^{(d)} \geq 0, \forall i_1 \in \mathbb{Z}, i_2 \in \mathbb{Z}, \ldots, i_d \in \mathbb{Z} \\ & \mu - \sum_{i_1 \in [k_1 - \Delta + 1, k_1]} y_{i_1}^{(1)} + (e^\beta - 1) \sum_{i_1 \leq k_1 - \Delta} y_{i_1}^{(1)} \\ & - \cdots - \sum_{i_d \in [k_d - \Delta + 1, k_d]} y_{i_d}^{(d)} + (e^\beta - 1) \sum_{i_d \leq k_d - \Delta} y_{i_d}^{(d)} \\ & \leq |k_1| + |k_2| + \cdots + |k_d|, \forall (k_1, \ldots, k_d) \in \mathbb{Z}^d.\end{aligned}$$

Theorem 4.25. *For the ℓ^1 cost function,*

$$\lim_{\max(\epsilon,\delta) \to 0} \frac{V'_{LB}}{\frac{d\Delta}{\max(\epsilon,\delta)}} \geq \log \frac{9}{8} \approx 0.1178.$$

Proof. See Appendix C.11. □

Similarly, for the ℓ^2 cost function, we have the lower bound

$$V^* \geq V'_{LB} := \max \quad \mu - \beta \left(\sum_{i_1 \in \mathbb{Z}} y_{i_1}^{(1)} + \sum_{i_2 \in \mathbb{Z}} y_{i_2}^{(2)} + \cdots + \sum_{i_d \in \mathbb{Z}} y_{i_d}^{(d)} \right)$$

$$\begin{aligned}\text{such that}\quad & y_{i_1}^{(1)}, y_{i_2}^{(2)}, \ldots, y_{i_d}^{(d)} \geq 0, \forall i_1 \in \mathbb{Z}, i_2 \in \mathbb{Z}, \ldots, i_d \in \mathbb{Z} \\ & \mu - \sum_{i_1 \in [k_1 - \Delta + 1, k_1]} y_{i_1}^{(1)} + (e^\beta - 1) \sum_{i_1 \leq k_1 - \Delta} y_{i_1}^{(1)} \\ & - \cdots - \sum_{i_d \in [k_d - \Delta + 1, k_d]} y_{i_d}^{(d)} + (e^\beta - 1) \sum_{i_d \leq k_d - \Delta} y_{i_d}^{(d)} \\ & \leq |k_1|^2 + |k_2|^2 + \cdots + |k_d|^2, \forall (k_1, \ldots, k_d) \in \mathbb{Z}^d.\end{aligned}$$

Theorem 4.26. *For the ℓ^2 cost function,*

$$\lim_{\max(\epsilon,\delta) \to 0} \frac{V'_{LB}}{\frac{d\Delta^2}{\beta^2}} \geq 0.0177.$$

Proof. See Appendix C.12. □

Upper Bounds: the Uniform Noise Mechanism and the Discrete Laplacian Mechanism

Since $(0, \delta)$-differential privacy implies (ϵ, δ)-differential privacy and we have shown that the uniform noise mechanism defined in (4.28) satisfies $(0, \delta)$-differential privacy, an upper bound for V^* for the ℓ^1 cost function is

$$V^* \leq V_{UB}^{\text{uniform}} = \frac{d\Delta}{4\delta} \tag{4.31}$$

by Corollary 4.19.

In addition, $(\epsilon, 0)$-differential privacy also implies (ϵ, δ)-differential privacy, and the discrete Laplacian mechanism satisfies $(\epsilon, 0)$-differential privacy. Consider the discrete Laplacian mechanism in the multiple dimensional setting with probability mass function $\mathcal{P}$ defined as

$$\mathcal{P}(i_1, i_2, \ldots, i_d) = \left(\frac{1-\lambda}{1+\lambda}\right)^d \lambda^{|i_1|+|i_2|+\cdots+|i_d|}, \forall (i_1, \ldots, i_d) \in \mathbb{Z}^d,$$

where $\lambda \triangleq e^{-\frac{\epsilon}{\Delta}}$.

The corresponding cost achieved by the Laplacian mechanism for the ℓ^1 cost function is

$$\begin{aligned} V_{UB}^{\text{Lap}} &= \sum_{(i_1, i_2, \ldots, i_d) \in \mathbb{Z}^d} \left(\frac{1-\lambda}{1+\lambda}\right)^d \lambda^{|i_1|+|i_2|+\cdots+|i_d|} (|i_1| + |i_2| + \cdots + |i_d|) \\ &= \frac{2d\lambda}{1-\lambda^2} \\ &= \frac{2de^{-\frac{\epsilon}{\Delta}}}{1-e^{-2\frac{\epsilon}{\Delta}}} \\ &= \Theta(\frac{d\Delta}{\epsilon}), \end{aligned} \tag{4.32}$$

as $\epsilon \to 0$.

Similarly, for the ℓ^2 cost function, we have

$$V_{UB}^{\text{uniform}} = \frac{d\Delta^2}{12\delta^2} + \frac{d}{6},$$

and

$$\begin{aligned} V_{UB}^{\text{Lap}} &= \sum_{(i_1, i_2, \ldots, i_d) \in \mathbb{Z}^d} \left(\frac{1-\lambda}{1+\lambda}\right)^d \lambda^{|i_1|+|i_2|+\cdots+|i_d|} (|i_1|^2 + |i_2|^2 + \cdots + |i_d|^2) \\ &= \frac{2d\lambda}{(1-\lambda)^2} \end{aligned}$$

$$= \Theta(\frac{2d\Delta^2}{\epsilon^2}).$$

Comparison of Lower Bound and Upper Bounds

Compare the lower bound in Theorem 4.25 and the upper bounds (4.31) and (4.32), and we conclude that for the ℓ^1 cost function, the multiplicative gap between the upper bound and lower bound is upper bounded by a constant as $(\epsilon, \delta) \to (0, 0)$. More precisely,

Corollary 4.27. *For the ℓ^1 cost function, we have*

$$V'_{LB} \le V^* \le \min(V_{UB}^{uniform}, V_{UB}^{Lap}),$$

and as $(\epsilon, \delta) \to (0, 0)$,

$$\lim_{(\epsilon,\delta)\to(0,0)} \frac{\min(V_{UB}^{uniform}, V_{UB}^{Lap})}{V'_{LB}} \le \frac{1}{\log \frac{9}{8}} \approx 8.49.$$

Similarly, for the ℓ^2 cost function, we have

Corollary 4.28. *For the ℓ^2 cost function, we have*

$$V'_{LB} \le V^* \le \min(V_{UB}^{uniform}, V_{UB}^{Lap}),$$

and as $(\epsilon, \delta) \to (0, 0)$,

$$\lim_{(\epsilon,\delta)\to(0,0)} \frac{\min(V_{UB}^{uniform}, V_{UB}^{Lap})}{V'_{LB}} \le \frac{2}{0.0177} \approx 113.$$

CHAPTER 5

CONCLUSION

Differential privacy is a framework to quantify to what extent individual privacy in a statistical database is preserved while releasing useful aggregate information about the database. The purpose of this dissertation is to delve into fundamental limits of data privacy and derive the optimal mechanisms to preserve differential privacy in the most basic problem settings, as opposed to doing privacy for each and every application setting as in most works in the literature. The main contributions of this dissertation can be summarized as follows.

- *ϵ-differential privacy in the single dimensional setting*:

  Given the differential privacy constraint, we derive the optimal differentially private mechanism for a single real-valued query function under a general utility-maximization (or cost-minimization) framework. The class of noise probability distributions in the optimal mechanism has *staircase-shaped* probability density functions which are symmetric (around the origin), monotonically decreasing, and geometrically decaying. The staircase mechanism can be viewed as a *geometric mixture of uniform probability distributions*, providing a simple algorithmic description for the mechanism. Furthermore, the staircase mechanism naturally generalizes to discrete query output settings as well as more abstract settings. We show that adding query-output independent noise with the staircase distribution is optimal among all randomized mechanisms (subject to a mild technical condition) that preserve differential privacy.

  We explicitly derive the optimal noise probability distributions with minimum expectation of noise amplitude and power. Comparing the optimal performances with those of the Laplacian mechanism, we show that in the high privacy regime, the Laplacian mechanism is asymptotically optimal; in the low privacy regime, the staircase mechanism significantly outperforms the Laplacian mechanism. We conclude that the gains are more pronounced in the low privacy regime.

- *ϵ-differential privacy in the multiple dimensional setting*:

  We extend the staircase mechanism from the single dimensional setting to the multiple dimensional setting. We show that for histogram-like query functions, when the

dimension of query output is two, the multiple dimensional staircase mechanism is optimal for the ℓ^1 cost function. We explicitly derive the parameter of the optimal two-dimensional staircase mechanism, and study the asymptotical performance of the optimal mechanism in the high and low privacy regimes. Comparing the optimal performances with those of the usual Laplacian mechanisms, we show that in the high privacy regime (ϵ is small), the Laplacian mechanism is asymptotically optimal as $\epsilon \to 0$; in the low privacy regime (ϵ is large), the optimal cost is $\Theta(e^{-\frac{\epsilon}{3}})$, while the cost of the Laplacian mechanism is $\frac{2\Delta}{\epsilon}$. We conclude that the gains of the staircase mechanism are more pronounced in the low privacy regime.

- (ϵ, δ)-*differential privacy*:

  We study the optimal mechanisms in (ϵ, δ)-differential privacy for integer-valued query functions under a utility-maximization/cost-minimization framework. We show that the (ϵ, δ)-differential privacy is a framework not much more general than the $(\epsilon, 0)$-differential privacy and $(0, \delta)$-differential privacy in the context of ℓ^1 and ℓ^2 cost functions, i.e., minimum expected noise magnitude and noise power. In the same context of ℓ^1 and ℓ^2 cost functions, we show the near-optimality of the uniform noise mechanism and the discrete Laplacian mechanism in the high privacy regime (as $(\epsilon, \delta) \to (0, 0)$).

REFERENCES

[1] Wikipedia, "Netflix prize." [Online]. Available: http://en.wikipedia.org/wiki/Netflix_Prize

[2] A. Narayanan and V. Shmatikov, "Robust de-anonymization of large sparse datasets," in *Proceedings of the 2008 IEEE Symposium on Security and Privacy*, ser. SP '08. Washington, DC, USA: IEEE Computer Society, 2008, pp. 111–125.

[3] C. Dwork, "Differential privacy: A survey of results," in *Proceedings of the 5th International Conference on Theory and Applications of Models of Computation*, ser. TAMC'08. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 1–19.

[4] M. Hardt and K. Talwar, "On the geometry of differential privacy," in *Proceedings of the 42nd ACM Symposium on Theory of Computing*, ser. STOC '10. New York, NY, USA: ACM, 2010, pp. 705–714.

[5] A. Nikolov, K. Talwar, and L. Zhang, "The geometry of differential privacy: The sparse and approximate cases," in *Proceedings of the 45th Annual ACM Symposium on Symposium on Theory of Computing*, ser. STOC '13. New York, NY, USA: ACM, 2013, pp. 351–360.

[6] C. Li, M. Hay, V. Rastogi, G. Miklau, and A. McGregor, "Optimizing linear counting queries under differential privacy," in *Proceedings of the Twenty-Ninth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, ser. PODS '10. New York, NY, USA: ACM, 2010, pp. 123–134.

[7] A. Ghosh, T. Roughgarden, and M. Sundararajan, "Universally utility-maximizing privacy mechanisms," in *Proceedings of the 41st Annual ACM Symposium on Theory of Computing*, ser. STOC '09. New York, NY, USA: ACM, 2009, pp. 351–360.

[8] H. Brenner and K. Nissim, "Impossibility of differentially private universally optimal mechanisms," in *Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science*, ser. FOCS '10, Oct. 2010, pp. 71 –80.

[9] M. Gupte and M. Sundararajan, "Universally optimal privacy mechanisms for minimax agents," in *Symposium on Principles of Database Systems*, 2010, pp. 135–146.

[10] L. Wasserman and S. Zhou, "A statistical framework for differential privacy," *Journal of the American Statistical Association*, vol. 105, no. 489, pp. 375–389, 2010.

[11] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Theory of Cryptography*, ser. Lecture Notes in Computer Science, S. Halevi and T. Rabin, Eds. Springer Berlin / Heidelberg, 2006, vol. 3876, pp. 265–284.

[12] M. Hardt, K. Ligett, and F. McSherry, "A simple and practical algorithm for differentially private data release," in *Advances in Neural Information Processing Systems*, P. Bartlett, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, Eds., 2012, pp. 2348–2356.

[13] F. McSherry and I. Mironov, "Differentially private recommender systems: Building privacy into the net," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '09. New York, NY, USA: ACM, 2009, pp. 627–636.

[14] X. Xiao, G. Wang, and J. Gehrke, "Differential privacy via wavelet transforms," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 8, pp. 1200–1214, 2011.

[15] Z. Huang, S. Mitra, and G. Dullerud, "Differentially private iterative synchronous consensus," in *Proceedings of the 2012 ACM Workshop on Privacy in the Electronic Society*, ser. WPES '12. New York, NY, USA: ACM, 2012, pp. 81–90.

[16] F. McSherry, "Privacy integrated queries: An extensible platform for privacy-preserving data analysis," *Commun. ACM*, vol. 53, no. 9, pp. 89–97, Sep. 2010.

[17] B. Barak, K. Chaudhuri, C. Dwork, S. Kale, F. McSherry, and K. Talwar, "Privacy, accuracy, and consistency too: A holistic solution to contingency table release," in *Proceedings of the Twenty-Sixth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, ser. PODS '07. New York, NY, USA: ACM, 2007, pp. 273–282.

[18] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor, "Our data, ourselves: Privacy via distributed noise generation," in *Proceedings of the 24th Annual International Conference on the Theory and Applications of Cryptographic Techniques*, ser. EUROCRYPT '06. Berlin, Heidelberg: Springer-Verlag, 2006, pp. 486–503.

[19] C. Dwork and J. Lei, "Differential privacy and robust statistics," in *Proceedings of the 41st Annual ACM symposium on Theory of Computing*, ser. STOC '09. New York, NY, USA: ACM, 2009, pp. 371–380.

[20] A. Roth and T. Roughgarden, "Interactive privacy via the median mechanism," in *Proceedings of the 42nd ACM Symposium on Theory of Computing*, ser. STOC '10. New York, NY, USA: ACM, 2010, pp. 765–774.

[21] A. Smith, "Privacy-preserving statistical estimation with optimal convergence rates," in *Proceedings of the 43rd Annual ACM Symposium on Theory of Computing*, ser. STOC '11. New York, NY, USA: ACM, 2011, pp. 813–822.

[22] K. Chaudhuri and C. Monteleoni, "Privacy-preserving logistic regression," in *Advances in Neural Information Processing Systems*, Vancouver, B.C., Canada, 2008, pp. 289–296.

[23] C. Dwork, M. Naor, T. Pitassi, and G. N. Rothblum, "Differential privacy under continual observation," in *Proceedings of the 42nd ACM Symposium on Theory of Computing*, ser. STOC '10. New York, NY, USA: ACM, 2010, pp. 715–724.

[24] B. Ding, M. Winslett, J. Han, and Z. Li, "Differentially private data cubes: optimizing noise sources and consistency," in *Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data*, ser. SIGMOD '11. New York, NY, USA: ACM, 2011, pp. 217–228.

[25] M. Hardt and G. N. Rothblum, "A multiplicative weights mechanism for privacy-preserving data analysis," in *Proceedings of the 2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, ser. FOCS '10. Washington, DC, USA: IEEE Computer Society, 2010, pp. 61–70.

[26] M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi, "Geo-indistinguishability: Differential privacy for location-based systems," *ArXiv e-prints*, December 2012.

[27] S. P. Kasiviswanathan, H. K. Lee, K. Nissim, S. Raskhodnikova, and A. Smith, "What can we learn privately?" *SIAM J. Comput.*, vol. 40, no. 3, pp. 793–826, June 2011.

[28] I. Mironov, "On significance of the least significant bits for differential privacy," in *Proceedings of the 2012 ACM conference on Computer and Communications Security*, ser. CCS '12. New York, NY, USA: ACM, 2012, pp. 650–661.

[29] R. Sarathy and K. Muralidhar, "Evaluating Laplace noise addition to satisfy differential privacy for numeric data," *Trans. Data Privacy*, vol. 4, no. 1, pp. 1–17, Apr. 2011.

[30] X. Xiao, G. Bender, M. Hay, and J. Gehrke, "iReduct: Differential privacy with reduced relative errors," in *Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data*, ser. SIGMOD '11. New York, NY, USA: ACM, 2011, pp. 229–240.

[31] F. K. Dankar and K. El Emam, "The application of differential privacy to health data," in *Proceedings of the 2012 Joint EDBT/ICDT Workshops*, ser. EDBT-ICDT '12. New York, NY, USA: ACM, 2012, pp. 158–166.

[32] A. Friedman and A. Schuster, "Data mining with differential privacy," in *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '10. New York, NY, USA: ACM, 2010, pp. 493–502.

[33] J. Zhang, Z. Zhang, X. Xiao, Y. Yang, and M. Winslett, "Functional mechanism: Regression analysis under differential privacy," *Proceedings of the VLDB Endowment*, vol. 5, no. 11, pp. 1364–1375, 2012.

[34] J. Lei, "Differentially private m-estimators," in *Proceedings of the 23rd Annual Conference on Neural Information Processing Systems*, Granada, Spain, 2011, pp. 361–369.

[35] C. Dwork, M. Naor, T. Pitassi, G. N. Rothblum, and S. Yekhanin, "Pan-private streaming algorithms," in *Proceedings of the First Symposium on Innovations in Computer Science*, ser. ICS '10, Beijing, China, 2010.

[36] A. Gupta, K. Ligett, F. McSherry, A. Roth, and K. Talwar, "Differentially private combinatorial optimization," in *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms*, ser. SODA '10. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2010, pp. 1106–1125.

[37] A. Blum and A. Roth, "Fast private data release algorithms for sparse queries," *arXiv preprint arXiv:1111.6842*, 2011.

[38] J. Hsu, S. Khanna, and A. Roth, "Distributed private heavy hitters," in *Proceedings of the 39th International Colloquium Conference on Automata, Languages, and Programming - Volume Part I*, ser. ICALP '12. Berlin, Heidelberg: Springer-Verlag, 2012, pp. 461–472.

[39] J. Hsu, A. Roth, and J. Ullman, "Differential privacy for the analyst via private equilibrium computation," *arXiv preprint arXiv:1211.0877*, 2012.

[40] J. Blocki, A. Blum, A. Datta, and O. Sheffet, "The Johnson-Lindenstrauss transform itself preserves differential privacy," in *Proceedings of the 2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*, ser. FOCS '12. Washington, DC, USA: IEEE Computer Society, 2012, pp. 410–419.

[41] M. Hardt and A. Roth, "Beyond worst-case analysis in private singular vector computation," in *Proceedings of the 45th Annual ACM Symposium on Symposium on Theory of Computing*, ser. STOC '13. New York, NY, USA: ACM, 2013, pp. 331–340.

[42] M. Hardt, G. N. Rothblum, and R. A. Servedio, "Private data release via learning thresholds," in *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM, 2012, pp. 168–187.

[43] A. Gupta, A. Roth, and J. Ullman, "Iterative constructions and private data release," *Theory of Cryptography*, pp. 339–356, 2012.

[44] S. P. Kasiviswanathan, K. Nissim, S. Raskhodnikova, and A. Smith, "Analyzing graphs with node differential privacy," in *Theory of Cryptography*. Springer, 2013, pp. 457–476.

[45] V. Karwa, S. Raskhodnikova, A. Smith, and G. Yaroslavtsev, "Private analysis of graph structure," *Proceedings of the VLDB Endowment*, vol. 4, no. 11, pp. 1146–1157, 2011.

[46] G. Cormode, C. Procopiuc, D. Srivastava, E. Shen, and T. Yu, "Differentially private spatial decompositions," in *Proceedings of the 2012 IEEE 28th International Conference on Data Engineering*. IEEE, 2012, pp. 20–31.

[47] F. McSherry and K. Talwar, "Mechanism design via differential privacy," in *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science*, ser. FOCS '07. Washington, DC, USA: IEEE Computer Society, 2007, pp. 94–103.

[48] K. Nissim, S. Raskhodnikova, and A. Smith, "Smooth sensitivity and sampling in private data analysis," in *Proceedings of the Thirty-Ninth annual ACM Symposium on Theory of Computing*, ser. STOC '07. New York, NY, USA: ACM, 2007, pp. 75–84.

[49] M. Hay, V. Rastogi, G. Miklau, and D. Suciu, "Boosting the accuracy of differentially private histograms through consistency," *Proceedings of the VLDB Endowment*, vol. 3, no. 1-2, pp. 1021–1032, Sep. 2010.

[50] Q. Geng and P. Viswanath, "The optimal mechanism in differential privacy," *ArXiv e-prints*, Dec. 2012.

APPENDIX A

PROOFS FOR CHAPTER 2

A.1 Proof of Theorem 2.3

We first give two lemmas on the properties of $\{\mathcal{P}_t\}_{t\in\mathbb{R}}$ which satisfies (2.7).

Lemma A.1. *Given $\{\mathcal{P}_t\}_{t\in\mathbb{R}}$ satisfying (2.7), and given any scalar $\alpha \in \mathbb{R}$, consider the family of noise probability measures $\{\mathcal{P}_t^{(\alpha)}\}_{t\in\mathbb{R}}$ defined by*

$$\mathcal{P}_t^{(\alpha)} \triangleq \mathcal{P}_{t+\alpha}, \forall t \in \mathbb{R}. \tag{A.1}$$

Then $\{\mathcal{P}_t^{(\alpha)}\}_{t\in\mathbb{R}}$ also satisfies the differential privacy constraint, i.e., $\forall |t_1 - t_2| \leq \Delta$,

$$\mathcal{P}_{t_1}^{(\alpha)}(S) \leq e^{\epsilon}\mathcal{P}_{t_2}^{(\alpha)}(S + t_1 - t_2). \tag{A.2}$$

Furthermore, $\{\mathcal{P}_t\}_{t\in\mathbb{R}}$ and $\{\mathcal{P}_t^{(\alpha)}\}_{t\in\mathbb{R}}$ have the same cost, i.e.,

$$\sup_{t\in\mathbb{R}} \int_{x\in\mathbb{R}} \mathcal{L}(x)\mathcal{P}_t(dx) = \sup_{t\in\mathbb{R}} \int_{x\in\mathbb{R}} \mathcal{L}(x)\mathcal{P}_t^{(\alpha)}(dx). \tag{A.3}$$

Proof. Since by definition the family of probability measures $\{\mathcal{P}_t^{(\alpha)}\}_{t\in\mathbb{R}}$ is a shifted version of $\{\mathcal{P}_t\}_{t\in\mathbb{R}}$, (A.3) holds.

Next we show that $\{\mathcal{P}_t^{(\alpha)}\}_{t\in\mathbb{R}}$ satisfies (A.2). Given any t_1, t_2 such that $|t_1 - t_2| \leq \Delta$, then for any measurable set $S \subset \mathbb{R}$, we have

$$\begin{aligned}
\mathcal{P}_{t_1}^{(\alpha)} &= \mathcal{P}_{t_1+\alpha}(S) \\
&\leq e^{\epsilon}\mathcal{P}_{t_2+\alpha}(S + (t_1 + \alpha) - (t_2 + \alpha)) \\
&= e^{\epsilon}\mathcal{P}_{t_2+\alpha}(S + t_1 - t_2) \\
&= e^{\epsilon}\mathcal{P}_{t_2}^{(\alpha)}(S + t_1 - t_2).
\end{aligned}$$

This completes the proof. □

Next we show that given a collection of families of probability measures each of which satisfies the differential privacy constraint (2.7), we can take a convex combination of them to construct a new family of probability measures satisfying (2.7) and the new cost is not worse. More precisely,

Lemma A.2. *Given a collection of finite number of families of probability measures* $\{\mathcal{P}_t^{[i]}\}_{t\in\mathbb{R}}$ *(*$i \in \{1, 2, 3, \ldots, n\}$*), such that for each* i, $\{\mathcal{P}_t^{[i]}\}_{t\in\mathbb{R}}$ *satisfies* (2.7) *and*

$$\sup_{t\in\mathbb{R}} \int_{x\in\mathbb{R}} \mathcal{L}(x)\mathcal{P}_t^{[i]}(dx) = Q, \forall i,$$

for some real number Q*, consider the family of probability measures* $\{\tilde{\nu}_t\}_{t\in\mathbb{R}}$ *defined by*

$$\tilde{\nu}_t \triangleq \sum_{i=1}^{n} c_i \mathcal{P}_t^{[i]}, \forall t \in \mathbb{R},$$

i.e., for any measurable set $S \subset \mathbb{R}$*,*

$$\tilde{\nu}_t(S) = \sum_{i=1}^{n} c_i \mathcal{P}_t^{[i]}(S),$$

where $c_i \geq 0$*, and* $\sum_{i=1}^{n} c_i = 1$*.*

Then $\{\tilde{\nu}_t\}_{t\in\mathbb{R}}$ *also satisfies the differential privacy constraint* (2.7)*, and*

$$\sup_{t\in\mathbb{R}} \int_{x\in\mathbb{R}} \mathcal{L}(x)\tilde{\nu}_t(dx) \leq Q.$$

Proof. First we show that $\{\tilde{\nu}_t\}_{t\in\mathbb{R}}$ also satisfies the differential privacy constraint (2.7). Indeed, $\forall\, |t_1 - t_2| \leq \Delta$, $\forall$ measurable set $S \subset \mathbb{R}$,

$$\begin{aligned}
\tilde{\nu}_{t_1}(S) &= \sum_{i=1}^{n} c_i \mathcal{P}_{t_1}^{[i]}(S) \\
&\leq \sum_{i=1}^{n} c_i e^{\epsilon} \mathcal{P}_{t_2}^{[i]}(S + t_1 - t_2) \\
&= e^{\epsilon} \tilde{\nu}_{t_2}(S + t_1 - t_2).
\end{aligned}$$

Next we show that the cost of $\{\tilde{\nu}_t\}_{t\in\mathbb{R}}$ is no bigger than Q. Indeed, for any $t \in \mathbb{R}$,

$$\int_{x\in\mathbb{R}} \mathcal{L}(x)\tilde{\nu}_t(dx) = \sum_{i=1}^{n} c_i \int_{x\in\mathbb{R}} \mathcal{L}(x)\tilde{\nu}_t^{[i]}(dx)$$

$$\leq \sum_{i=1}^{n} c_i Q$$
$$= Q.$$

Therefore,

$$\sup_{t\in\mathbb{R}} \int_{x\in\mathbb{R}} \mathcal{L}(x)\tilde{\nu}_t(dx) \leq Q.$$

□

Applying Lemma A.1 and Lemma A.2, we can prove the conjecture under the assumption that the family of probability measures $\{\mathcal{P}_t\}_{t\in\mathbb{R}}$ is piecewise constant and periodic over t.

Proof of Theorem 2.3. We first prove that for any family of probability measures $\{\mathcal{P}_t\}_{t\in\mathbb{R}} \in \mathcal{K}_{T,n}$, there exists a new family of probability measures $\{\tilde{\mathcal{P}}_t\}_{t\in\mathbb{R}} \in \mathcal{K}_{T,n}$ such that $\tilde{\mathcal{P}}_t = \tilde{\mathcal{P}}$ for all $t \in \mathbb{R}$, i.e., the added noise is independent of query output t, and

$$\sup_{t\in\mathbb{R}} \int_{x\in\mathbb{R}} \mathcal{L}(x)\tilde{\mathcal{P}}_t(dx) \leq \sup_{t\in\mathbb{R}} \int_{x\in\mathbb{R}} \mathcal{L}(x)\mathcal{P}_t(dx).$$

Indeed, consider the collection of probability measures $\{\mathcal{P}_t^{(i\frac{T}{n})}\}_{t\in\mathbb{R}}$ for $i \in \{0, 1, 2, \ldots, n-1\}$, where $\{\mathcal{P}_t^{(\alpha)}\}$ is defined in (A.1). Due to Lemma A.1, for all i, $\{\mathcal{P}_t^{(i\frac{T}{n})}\}_{t\in\mathbb{R}}$ satisfies the differential privacy constraint (2.7), and the cost is the same as the cost of $\{\mathcal{P}_t\}_{t\in\mathbb{R}}$.

Define

$$\tilde{\mathcal{P}}_t = \sum_{i=0}^{n-1} \frac{1}{n} \mathcal{P}_t^{(i\frac{T}{n})}.$$

Then due to Lemma A.2, $\{\tilde{\mathcal{P}}_t\}_{t\in\mathbb{R}}$ satisfies (2.7), and the cost of is not worse, i.e.,

$$\sup_{t\in\mathbb{R}} \int_{x\in\mathbb{R}} \mathcal{L}(x)\tilde{\mathcal{P}}_t(dx) \leq \sup_{t\in\mathbb{R}} \int_{x\in\mathbb{R}} \mathcal{L}(x)\mathcal{P}_t(dx).$$

Furthermore, since $\{\mathcal{P}_t\}_{t\in\mathbb{R}} \in \mathcal{K}_{T,n}$, for any $t \in \mathbb{R}$,

$$\tilde{\mathcal{P}}_t = \sum_{i=0}^{n-1} \frac{1}{n} \mathcal{P}_t^{(i\frac{T}{n})} = \sum_{i=0}^{n-1} \frac{1}{n} \mathcal{P}_{i\frac{T}{n}}.$$

Hence, $\tilde{\mathcal{P}}_t$ is independent of t.

Therefore, among the collection of probability measures in $\cup_{T>0} \cup_{n\geq 1} \mathcal{K}_{T,n}$, to minimize the cost we only need to consider the families of noise probability measures which are independent of the query output t. Then due to Theorem 2.4, the staircase mechanism is optimal among all query-output independent noise-adding mechanisms. This completes the proof of Theorem 2.3. □

A.2 Proof of Theorem 2.4

In this section, we give detailed and rigorous proof of Theorem 2.4.

A.2.1 Outline of Proof

The key idea of the proof is to use a sequence of probability distributions with piecewise constant probability density functions to approximate any probability distribution satisfying the differential privacy constraint (2.11). The proof consists of 8 steps in total, and in each step we narrow down the set of probability distributions where the optimal probability distribution should lie:

- Step 1 proves that we only need to consider symmetric probability distributions.
- Step 2 and Step 3 prove that we only need to consider probability distributions which have symmetric and piecewise constant probability density functions.
- Step 4 proves that we only need to consider those symmetric and piecewise constant probability density functions which are monotonically decreasing for $x \geq 0$.
- Step 5 proves that the optimal probability density function should periodically decay.
- Step 6, Step 7, and Step 8 prove that the optimal probability density function over the interval $[0, \Delta)$ is a step function, and they conclude the proof of Theorem 2.4.

A.2.2 Step 1

Define

$$V^* \triangleq \inf_{\mathcal{P}\in\mathcal{SP}} \int_{x\in\mathbb{R}} \mathcal{L}(x)\mathcal{P}(dx).$$

Our goal is to prove that $V^* = \inf_{\gamma\in[0,1]} \int_{x\in\mathbb{R}} \mathcal{L}(x)\mathcal{P}_\gamma(dx)$.

If $V^* = +\infty$, then due to the definition of V^*, we have

$$\inf_{\gamma\in[0,1]} \int_{x\in\mathbb{R}} \mathcal{L}(x)\mathcal{P}_\gamma(dx) \geq V^* = +\infty,$$

and thus $\inf_{\gamma\in[0,1]} \int_{x\in\mathbb{R}} \mathcal{L}(x) = V^* = +\infty$. So we only need to consider the case $V^* < +\infty$, i.e., V^* is finite. Therefore, in the rest of the proof, we assume V^* is finite.

First, we prove that we only need to consider symmetric probability measures.

Lemma A.3. *Given $\mathcal{P} \in \mathcal{SP}$, define a symmetric probability distribution $\mathcal{P}_{sym}$ as*

$$\mathcal{P}_{sym}(S) \triangleq \frac{\mathcal{P}(S) + \mathcal{P}(-S)}{2}, \forall \text{ measurable set } S \subseteq \mathbb{R}, \tag{A.4}$$

where the set $-S \triangleq \{-x \mid x \in S\}$. Then $\mathcal{P}_{sym} \in \mathcal{SP}$, i.e., $\mathcal{P}_{sym}$ satisfies the differential privacy constraint (2.11)*, and*

$$\int_{x\in\mathbb{R}} \mathcal{L}(x)\mathcal{P}_{sym}(dx) = \int_{x\in\mathbb{R}} \mathcal{L}(x)\mathcal{P}(dx).$$

Proof. It is easy to verify that $\mathcal{P}_{\text{sym}}$ is a valid probability distribution. Due to the definition of $\mathcal{P}_{\text{sym}}$ in (A.4), we have

$$\mathcal{P}_{\text{sym}}(S) = \frac{\mathcal{P}(S) + \mathcal{P}(-S)}{2} = \mathcal{P}_{\text{sym}}(-S),$$

for any measurable set $S \subseteq \mathbb{R}$. Thus, $\mathcal{P}_{\text{sym}}$ is a symmetric probability distribution.

Next, we show that $\mathcal{P}_{\text{sym}}$ satisfies (2.11). Indeed, $\forall$ measurable set $S \subseteq \mathbb{R}$ and $\forall |d| \leq \Delta$,

$$\begin{aligned}
\mathcal{P}_{\text{sym}}(S) &= \frac{\mathcal{P}(S) + \mathcal{P}(-S)}{2} \\
&\leq \frac{e^\epsilon \mathcal{P}(S+d) + e^\epsilon \mathcal{P}(-S-d)}{2} \qquad \text{(A.5)} \\
&= \frac{e^\epsilon \mathcal{P}(S+d) + e^\epsilon \mathcal{P}(-(S+d))}{2} \\
&= e^\epsilon \mathcal{P}_{\text{sym}}(S+d),
\end{aligned}$$

where in (A.5) we use the facts $\mathcal{P}(S) \leq e^\epsilon \mathcal{P}(S+d)$ and $\mathcal{P}(-S) \leq e^\epsilon \mathcal{P}(-S-d)$.

Lastly, since $\mathcal{L}(x)$ is symmetric,

$$\int_{x\in\mathbb{R}} \mathcal{L}(x)\mathcal{P}(dx) = \int_{x\in\mathbb{R}} \frac{\mathcal{L}(x) + \mathcal{L}(-x)}{2}\mathcal{P}(dx)$$

$$= \int_{x \in \mathbb{R}} \mathcal{L}(x) \mathcal{P}_{\text{sym}}(dx).$$

□

Therefore, if we define

$$\mathcal{SP}_{\text{sym}} \triangleq \{\mathcal{P}_{\text{sym}} | \mathcal{P} \in \mathcal{SP}\},$$

due to Lemma A.3,

Lemma A.4.

$$V^* = \inf_{\mathcal{P} \in \mathcal{SP}_{sym}} \int_{x \in \mathbb{R}} \mathcal{L}(x) \mathcal{P}(dx).$$

A.2.3 Step 2

Next we prove that for any probability distribution $\mathcal{P}$ satisfying differential privacy constraint (2.11), the probability $\Pr(X = x) = 0, \forall x \in \mathbb{R}$, and $\mathcal{P}([y, z]) \neq 0$ for any $y < z \in \mathbb{R}$.

Lemma A.5. $\forall \mathcal{P} \in \mathcal{SP}, \forall x \in \mathbb{R}$, $\mathcal{P}(\{x\}) = 0$. *And, for any* $y < z \in \mathbb{R}$, $\mathcal{P}([y, z]) \neq 0$.

Proof. Given $\mathcal{P} \in \mathcal{SP}$, suppose $\mathcal{P}(\{x_0\}) = p_0 > 0$, for some $x_0 \in \mathbb{R}$. Then for any $x \in [x_0, x_0 + \Delta]$,

$$\mathcal{P}(\{x\}) \geq e^{-\epsilon},$$

due to (2.11).

So $\mathcal{P}(\{x\})$ is strictly lower bounded by a positive constant for an uncountable number of x, and thus $\mathcal{P}([x_0, x_0 + \Delta]) = +\infty$, which contradicts with the fact $\mathcal{P}$ is a probability distribution.

Therefore, $\forall \mathcal{P} \in \mathcal{SP}, \forall x \in \mathbb{R}$, $\mathcal{P}(\{x\}) = 0$.

Suppose $\mathcal{P}([y, z]) = 0$ for some $y < z \in \mathbb{R}$. Then from (2.11) we have for any $|d| \leq \Delta$,

$$\mathcal{P}([y + d, z + d]) \leq e^{\epsilon} \mathcal{P}([y, z]) = 0,$$

and thus $\mathcal{P}([y + d, z + d]) = 0$. By induction, for any $k \in \mathbb{Z}$, $\mathcal{P}([y + kd, z + kd]) = 0$, which implies that $\mathcal{P}((-\infty, +\infty)) = 0$. Contradiction. So for any $y < z \in \mathbb{R}$, $\mathcal{P}([y, z]) \neq 0$. □

A.2.4 Step 3

In this subsection, we show that for any $\mathcal{P} \in \mathcal{SP}_{\text{sym}}$ with

$$V(\mathcal{P}) \triangleq \int_{x\in\mathbb{R}} \mathcal{L}(x)\mathcal{P}(dx) < +\infty,$$

we can use a sequence of probability measures $\{\mathcal{P}_i \in \mathcal{SP}_{\text{sym}}\}_{i\geq 1}$ with symmetric and piecewise constant probability density functions to approximate $\mathcal{P}$ with $\lim_{i\to+\infty} V(\mathcal{P}_i) = V(\mathcal{P})$.

Lemma A.6. *Given $\mathcal{P} \in \mathcal{SP}_{sym}$ with $V(\mathcal{P}) < +\infty$, any positive integer $i \in N$, define $\mathcal{P}_i$ as the probability distribution with a symmetric and piecewise constant probability density function $f_i(x)$ defined as*

$$f_i(x) = \begin{cases} a_k \triangleq \frac{\mathcal{P}([k\frac{D}{i},(k+1)\frac{D}{i})}{\frac{D}{i}} & x \in [k\frac{D}{i},(k+1)\frac{D}{i}) \text{ for } k \in \mathbb{N} \\ f_i(-x) & x < 0 \end{cases}.$$

Then $\mathcal{P}_i \in \mathcal{SP}_{sym}$ and

$$\lim_{i\to+\infty} V(\mathcal{P}_i) = V(\mathcal{P}).$$

Proof. First we prove that $\mathcal{P}_i \in \mathcal{SP}_{\text{sym}}$, i.e., $\mathcal{P}_i$ is symmetric and satisfies the differential privacy constraint (2.11).

By definition $f_i(x)$ is a symmetric and nonnegative function, and

$$\begin{aligned} \int_{-\infty}^{+\infty} f_i(x)dx &= 2\int_{0}^{+\infty} f_i(x)dx \\ &= 2\int_{x\in[0,+\infty)} \mathcal{P}(dx) \\ &= 2\int_{x\in(0,+\infty)} \mathcal{P}(dx) \qquad \text{(A.6)} \\ &= 1, \qquad \text{(A.7)} \end{aligned}$$

where in (A.6) we used the fact $\mathcal{P}(\{0\}) = 0$ due to Lemma A.5. In addition, due to Lemma A.5, $a_k > 0, \forall k \in \mathbb{N}$.

So $f_i(x)$ is a valid symmetric probability density function, and thus $\mathcal{P}_i$ is a valid symmetric probability distribution.

Define the density sequence of $\mathcal{P}_i$ as the sequence $\{a_0, a_1, a_2, \ldots, a_n, \ldots\}$. Since $\mathcal{P}$ satisfies

(2.11), it is easy to see that

$$a_j \leq e^{\epsilon} a_{j+k} \text{ and } a_{j+k} \leq e^{\epsilon} a_j, \forall j \geq 0, 0 \leq k \leq i.$$

Therefore, for any x, y such that $|x - y| \leq \Delta$, we have

$$f_i(x) \leq e^{\epsilon} f_i(y) \text{ and } f_i(y) \leq e^{\epsilon} f_i(x),$$

which implies that $\mathcal{P}_i$ satisfies (2.11). Hence, $\mathcal{P}_i \in \mathcal{SP}_{\text{sym}}$.

Next we show that

$$\lim_{i \to +\infty} V(\mathcal{P}_i) = V(\mathcal{P}).$$

Since $\mathcal{L}(x)$ satisfies Property 2.2, we can assume there exists a constant $B > 0$ such that

$$\mathcal{L}(x+1) \leq B\mathcal{L}(x), \forall x \geq T.$$

Given $\delta > 0$, since $V(\mathcal{P})$ is finite, there exists integer $T^* > T$ such that

$$\int_{x \geq T^*} \mathcal{L}(x)\mathcal{P}(dx) < \frac{\delta}{B}.$$

For any integers $i \geq 1$, $N \geq T^*$,

$$\begin{aligned}
\int_{x \in [N, N+1)} \mathcal{L}(x)\mathcal{P}_i(dx) &\leq \mathcal{P}_i([N, N+1))\mathcal{L}(N+1) \\
&= \mathcal{P}([N, N+1))\mathcal{L}(N+1) \\
&\leq \int_{x \in [N, N+1)} B\mathcal{L}(x)\mathcal{P}(dx).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\int_{x \in [T^*, +\infty)} \mathcal{L}(x)\mathcal{P}_i(dx) &\leq \int_{x \in [T^*, +\infty)} B\mathcal{L}(x)\mathcal{P}(dx) \\
&\leq B\frac{\delta}{B} \\
&= \delta.
\end{aligned}$$

For $x \in [0, T^*)$, $\mathcal{L}(x)$ is a bounded function, and thus by the definition of Riemann-Stieltjes

integral, we have

$$\lim_{i\to\infty}\int_{x\in[0,T^*)}\mathcal{L}(x)\mathcal{P}_i(dx)=\int_{x\in[0,T^*)}\mathcal{L}(x)\mathcal{P}(dx).$$

So there exists a sufficiently large integer i^* such that for all $i\geq i^*$

$$\left|\int_{x\in[0,T^*)}\mathcal{L}(x)\mathcal{P}_i(dx)-\int_{x\in[0,T^*)}\mathcal{L}(x)\mathcal{P}(dx)\right|\leq\delta.$$

Hence, for all $i\geq i^*$

$$\begin{aligned}
&|V(\mathcal{P}_i)-V(\mathcal{P})|\\
=&\left|\int_{x\in\mathbb{R}}\mathcal{L}(x)\mathcal{P}_i(dx)-\int_{x\in\mathbb{R}}\mathcal{L}(x)\mathcal{P}(dx)\right|\\
=&2\left|\int_{x\in[0,T^*)}\mathcal{L}(x)\mathcal{P}_i(dx)-\int_{x\in[0,T^*)}\mathcal{L}(x)\mathcal{P}(dx)\right.\\
&\left.+\int_{x\in[T^*,+\infty)}\mathcal{L}(x)\mathcal{P}_i(dx)-\int_{x\in[T^*,+\infty)}\mathcal{L}(x)\mathcal{P}(dx)\right|\\
\leq&2\left|\int_{x\in[0,T^*)}\mathcal{L}(x)\mathcal{P}_i(dx)-\int_{x\in[0,T^*)}\mathcal{L}(x)\mathcal{P}(dx)\right|\\
&+2\int_{x\in[T^*,+\infty)}\mathcal{L}(x)\mathcal{P}_i(dx)+2\int_{x\in[T^*,+\infty)}\mathcal{L}(x)\mathcal{P}(dx)\\
\leq&2(\delta+\delta+\frac{\delta}{B})\\
\leq&(4+\frac{2}{B})\delta.
\end{aligned}$$

Therefore,

$$\lim_{i\to+\infty}\int_{x\in\mathbb{R}}\mathcal{L}(x)\mathcal{P}_i(dx)=\int_{x\in\mathbb{R}}\mathcal{L}(x)\mathcal{P}(dx).$$

□

Define $\mathcal{SP}_{i,\mathrm{sym}}\triangleq\{\mathcal{P}_i|\mathcal{P}\in\mathcal{SP}_{\mathrm{sym}}\}$ for $i\geq 1$, i.e., $\mathcal{SP}_{i,\mathrm{sym}}$ is the set of probability distributions satisfying differential privacy constraint (2.11) and having symmetric and piecewise constant (over intervals $[k\frac{\Delta}{i},(k+1)\frac{\Delta}{i})\ \forall k\in\mathbb{N}$) probability density functions.

Due to Lemma A.6,

Lemma A.7.

$$V^* = \inf_{\mathcal{P} \in \cup_{i=1}^{\infty} \mathcal{SP}_{i,sym}} \int_{x \in \mathbb{R}} \mathcal{L}(x) \mathcal{P}(dx).$$

Therefore, to characterize V^*, we only need to study probability distributions with symmetric and piecewise constant probability density functions.

A.2.5 Step 4

Next we show that indeed we only need to consider those probability distributions with symmetric and piecewise constant probability density functions which are *monotonically decreasing* when $x \geq 0$.

Lemma A.8. *Given $\mathcal{P}_a \in \mathcal{SP}_{i,sym}$ with a symmetric and piecewise constant probability density function $f(\cdot)$, let $\{a_0, a_1, \dots, a_n, \dots\}$ be the density sequence of $f(\cdot)$, i.e,*

$$f(x) = a_k, x \in [k\frac{\Delta}{i}, (k+1)\frac{\Delta}{i}) \; \forall k \in \mathbb{N}.$$

Then we can construct a new probability distribution $\mathcal{P}_b \in \mathcal{SP}_{i,sym}$ the probability density function of which is monotonically decreasing when $x \geq 0$, and

$$\int_{x \in \mathbb{R}} \mathcal{L}(x) \mathcal{P}_b(dx) \leq \int_{x \in \mathbb{R}} \mathcal{L}(x) \mathcal{P}_a(dx).$$

Proof. Since $a_k > 0, \forall k \in \mathbb{N}$, and

$$\sum_{k=0}^{+\infty} a_k \frac{\Delta}{i} = \frac{1}{2},$$

we have $\lim_{k \to +\infty} a_k = 0$.

Given the density sequence $\{a_0, a_1, \dots, a_n, \dots\}$, construct a new monotonically decreasing density sequence $\{b_0, b_1, \dots, b_n, \dots\}$ and a bijective mapping $\pi : \mathbb{N} \to \mathbb{N}$ as follows

$$I_0 = \arg\max_{k \in \mathbb{N}} a_k, \tag{A.8}$$

$$\begin{aligned} \pi(0) &= \min_{n \in I_0} n, \text{i.e., the smallest element in } I_0, \\ b_0 &= a_{\pi(0)}, \end{aligned} \tag{A.9}$$

$$\forall m \in \mathbb{N} \text{ and } m \geq 1,$$

$$
\begin{aligned}
I_m &= \underset{k\in\mathbb{N}\setminus\{\pi(j)|j<m\}}{\arg\max} a_k, \qquad (A.10)\\
\pi(m) &= \min_{n\in I_m} n, \text{i.e., the smallest element in } I_m,\\
b_m &= a_{\pi(m)}.
\end{aligned}
$$

Since the sequence $\{a_k\}$ converges to 0, the maximum of $\{a_k\}$ always exists in (A.8) and (A.10). Therefore, I_m is well defined for all $m \in \mathbb{N}$.

Note that since $\sum_{k=0}^{\infty} a_k \frac{\Delta}{i} = \frac{1}{2}$ and the sequence $\{b_k\}_{k\in\mathbb{N}}$ is simply a permutation of $\{a_k\}_{k\in\mathbb{N}}$, $\sum_{k=1}^{\infty} b_k \frac{\Delta}{i} = \frac{1}{2}$.

Therefore, if we define a function $g(\cdot)$ as

$$
g(x) = \begin{cases} b_k & x \in [k\frac{D}{i}, (k+1)\frac{D}{i}) \text{ for } k \in \mathbb{N} \\ g(-x) & x < 0 \end{cases}
$$

then $g(\cdot)$ is a valid symmetric probability density function, and

$$
\int_{x\in\mathbb{R}} \mathcal{L}(x)g(x)dx \le \int_{x\in\mathbb{R}} \mathcal{L}(x)f(x)dx.
$$

Next, we prove that the probability distribution $\mathcal{P}_b$ with probability density function $g(\cdot)$ satisfies the differential privacy constraint (2.11). Since $\{b_k\}_{k\in\mathbb{N}}$ is a monotonically decreasing sequence, it is sufficient and necessary to prove that for all $k \in \mathbb{N}$,

$$
\frac{b_k}{b_{k+i}} \le e^{\epsilon}.
$$

To simplify notation, given k, we define

$$
a^*(k) = \min_{k\le j\le k+i} a_k,
$$

i.e., $a^*(k)$ denotes the smallest number of $\{a_k, a_{k+1}, \ldots, a_{k+i}\}$.

First, when $k = 0$, it is easy to prove that $\frac{b_0}{b_i} \le e^{\epsilon}$. Indeed, recall that $b_0 = a_{\pi(0)}$ and consider the $i+1$ consecutive numbers $\{a_{\pi(0)}, a_{\pi(0)+1}, \ldots, a_{\pi(0)+i}\}$ in the original sequence $\{a_k\}_{k\in\mathbb{N}}$. Then $a^*(0) \le b_i$, since b_i is the $(i+1)$th largest number in the sequence $\{a_k\}_{k\in\mathbb{N}}$. Therefore,

$$
\frac{b_0}{b_i} = \frac{a_{\pi(0)}}{b_i} \le \frac{a_{\pi(0)}}{a^*(0)} \le e^{\epsilon}.
$$

For $k = 1$, $b_1 = a_{\pi(1)}$ and consider the $i+1$ consecutive numbers $\{a_{\pi(1)}, a_{\pi(1)+1}, \ldots, a_{\pi(1)+i}\}$.

If $\pi(0) \notin [\pi(1), \pi(1)+i]$, then $a^*(\pi(1)) \le b_{i+1}$, and thus

$$\frac{b_1}{b_{i+1}} = \frac{a_{\pi(1)}}{b_{1+i}} \le \frac{a_{\pi(1)}}{a^*(\pi(1))} \le e^\epsilon.$$

If $\pi(0) \in [\pi(1), \pi(1)+i]$, then $a^*(\pi(0)) \le b_{i+1}$ and $\frac{a_{\pi(0)}}{a^*(\pi(0))} \le e^\epsilon$. Therefore,

$$\frac{b_1}{b_{i+1}} \le \frac{b_0}{b_{1+i}} \le \frac{b_0}{a^*(\pi(0))} \le e^\epsilon.$$

Hence, $\frac{b_k}{b_{k+i}} \le e^\epsilon$ holds for $k = 1$.

In general, given k, we prove $\frac{b_k}{b_{k+i}} \le e^\epsilon$ as follows. First, if $\pi_j \notin [\pi(k), \pi(k)+i], \forall j < k$, then $a^*\pi(k) \le b_{k+i}$, and hence

$$\frac{b_k}{b_{i+k}} = \frac{a_{\pi(k)}}{b_{i+k}} \le \frac{a_{\pi(k)}}{a^*(\pi(k))} \le e^\epsilon.$$

If there exists $j < k$ and $\pi_j \in [\pi(k)+1, \pi(k)+i]$, we use Algorithm 2 to compute a number j^* such that $j^* < k$ and $\pi_j \notin [\pi(j^*)+1, \pi(j^*)+i], \forall j < k$.

Algorithm 2

$j^* \leftarrow k$
while there exists some $j < k$ and $\pi_j \in [\pi(j^*)+1, \pi(j^*)+i]$ **do**
  $j^* \leftarrow j$
end while
Output j^*

It is easy to show that the loop in Algorithm 2 will terminate after at most k steps.

After finding j^*, we have $j^* < k$, and $a^*(\pi(j^*)) \le b_{k+i}$. Therefore

$$\frac{b_k}{b_{i+k}} \le \frac{a_{\pi(j^*)}}{b_{i+k}} \le \frac{a_{\pi(j^*)}}{a^*(\pi(j^*))} \le e^\epsilon.$$

So $\frac{b_k}{b_{k+i}} \le e^\epsilon$ holds for all $k \in \mathbb{N}$. Therefore, $\mathcal{P}_b \in \mathcal{SP}_{i,\text{sym}}$.

This completes the proof of Lemma A.8.

□

Therefore, if we define

$$\mathcal{SP}_{i,\text{md}} \triangleq \{\mathcal{P} | \mathcal{P} \in \mathcal{SP}_{i,\text{sym}}, \text{ and the density sequence of } \mathcal{P} \text{ is monotonically decreasing}\},$$

then due to Lemma A.8,

Lemma A.9.

$$V^* = \inf_{\mathcal{P} \in \cup_{i=1}^{\infty} \mathcal{SP}_{i,md}} \int_{x \in \mathbb{R}} \mathcal{L}(x) \mathcal{P}(dx).$$

A.2.6 Step 5

Next we show that among all symmetric and piecewise constant probability density functions, we only need to consider those which are periodically decaying.

More precisely, given positive integer i,

$$\mathcal{SP}_{i,\text{pd}} \triangleq \{ \mathcal{P} \mid \mathcal{P} \in \mathcal{SP}_{i,\text{md}}, \text{ and } \mathcal{P} \text{ has density sequence } \{a_0, a_1, \ldots, a_n, \ldots, \} \text{ satisfying} \frac{a_k}{a_{k+i}} = e^{\epsilon}, \forall k \in \mathbb{N}\},$$

then

Lemma A.10.

$$V^* = \inf_{\mathcal{P} \in \cup_{i=1}^{\infty} \mathcal{SP}_{i,pd}} \int_{x \in \mathbb{R}} \mathcal{L}(x) \mathcal{P}(dx).$$

Proof. Due to Lemma A.9, we only need to consider probability distributions with symmetric and piecewise constant probability density functions which are monotonically decreasing for $x \geq 0$.

We first show that given $\mathcal{P}_a \in \mathcal{SP}_{i,\text{md}}$ with density sequence $\{a_0, a_1, \ldots, a_n, \ldots, \}$, if $\frac{a_0}{a_i} < e^{\epsilon}$, then we can construct a probability distributions $\mathcal{P}_b \in \mathcal{SP}_{i,\text{md}}$ with density sequence $\{b_0, b_1, \ldots, b_n, \ldots, \}$ such that $\frac{b_0}{b_i} = e^{\epsilon}$ and

$$V(\mathcal{P}_a) \geq V(\mathcal{P}_b).$$

Define a new sequence $\{b_0, b_1, \ldots, b_n, \ldots\}$ by scaling up a_0 and scaling down $\{a_1, a_2, \ldots\}$. More precisely, let $\delta = \frac{i}{2D((\frac{i}{2D} - a_0)e^{-\epsilon}\frac{a_0}{a_i} + a_0)} - 1 > 0$, and set

$$\begin{aligned} b_0 &= a_0(1+\delta), \\ b_k &= a_k(1-\delta'), \forall\ k \geq 1, \end{aligned}$$

where $\delta' \triangleq \frac{a_0\delta}{\frac{i}{2D} - a_0} > 0$, and we have chosen δ such that $\frac{b_0}{b_i} = \frac{a_0}{a_k}\frac{\frac{i}{2D} - a_0}{\frac{i}{2D(1+\delta)} - a_0} = e^{\epsilon}$.

It is easy to see the sequence $\{b_0, b_1, \ldots, b_n, \ldots, \}$ corresponds to a valid probability

density function and it also satisfies the differential privacy constraint (2.11), i.e.,

$$\frac{b_k}{b_{k+i}} \leq e^{\epsilon}, \forall k \geq 0.$$

Let $\mathcal{P}_b$ be the probability distribution with $\{b_0, b_1, \ldots, b_n, \ldots,\}$ as the density sequence of its probability density function. Next we show $V(\mathcal{P}_b) \leq V(\mathcal{P}_a)$.

It is easy to compute $V(\mathcal{P}_a)$, which is

$$V(\mathcal{P}_a) = 2\frac{\Delta}{i}\left(a_0 \int_0^{\frac{\Delta}{i}} \mathcal{L}(x)dx + \sum_{k=1}^{\infty} a_k \int_{k\frac{\Delta}{i}}^{(k+1)\frac{\Delta}{i}} \mathcal{L}(x)dx\right).$$

Similarly, we can compute $V(\mathcal{P}_b)$ by

$$\begin{aligned}
V(\mathcal{P}_b) &= 2\frac{\Delta}{i}\left(b_0 \int_0^{\frac{\Delta}{i}} \mathcal{L}(x)dx + \sum_{k=1}^{\infty} b_k \int_{k\frac{\Delta}{i}}^{(k+1)\frac{\Delta}{i}} \mathcal{L}(x)dx\right) \\
&= V(\mathcal{P}_a) + 2\frac{\Delta}{i}\left(a_0\delta \int_0^{\frac{D}{i}} \mathcal{L}(x)dx - \delta' \sum_{k=1}^{\infty} a_k \int_{k\frac{D}{i}}^{(k+1)\frac{D}{i}} \mathcal{L}(x)dx\right) \\
&= V(\mathcal{P}_a) + 2\frac{\Delta}{i}\frac{a_0\delta}{\frac{i}{2\Delta} - a_0}\left(\sum_{k=1}^{\infty} a_k \int_0^{\frac{\Delta}{i}} \mathcal{L}(x)dx - \sum_{k=1}^{\infty} a_k \int_{k\frac{\Delta}{i}}^{(k+1)\frac{\Delta}{i}} \mathcal{L}(x)dx\right) \\
&= V(\mathcal{P}_a) + 2\frac{\Delta}{i}\frac{a_0\delta}{\frac{i}{2\Delta} - a_0}\sum_{k=1}^{\infty} a_k \left(\int_0^{\frac{\Delta}{i}} \mathcal{L}(x)dx - \int_{k\frac{\Delta}{i}}^{(k+1)\frac{\Delta}{i}} \mathcal{L}(x)dx\right) \\
&\leq V(\mathcal{P}_a),
\end{aligned}$$

where in the last step we used the fact that $\left(\int_0^{\frac{\Delta}{i}} \mathcal{L}(x)dx - \int_{k\frac{\Delta}{i}}^{(k+1)\frac{\Delta}{i}} \mathcal{L}(x)dx\right) \leq 0$, since $\mathcal{L}(\cdot)$ is a monotonically increasing function for $x \geq 0$.

Therefore, for given $i \in \mathbb{N}$, we only need to consider $\mathcal{P} \in \mathcal{SP}_{i,\mathrm{md}}$ with density sequence $\{a_0, a_1, \ldots, a_n, \ldots\}$ satisfying $\frac{a_0}{a_i} = e^{\epsilon}$.

Next, we argue that among all probability distributions $\mathcal{P} \in \mathcal{SP}_{i,\mathrm{md}}$ with density sequence $\{a_0, a_1, \ldots, a_n, \ldots,\}$ satisfying $\frac{a_0}{a_i} = e^{\epsilon}$, we only need to consider those probability distributions with density sequence also satisfying $\frac{a_1}{a_{i+1}} = e^{\epsilon}$.

Given $\mathcal{P}_a \in \mathcal{SP}_{i,\mathrm{md}}$ with density sequence $\{a_0, a_1, \ldots, a_n, \ldots\}$ satisfying $\frac{a_0}{a_i} = e^{\epsilon}$ and $\frac{a_1}{a_{i+1}} < e^{\epsilon}$, we can construct a new probability distribution $\mathcal{P}_b \in \mathcal{SP}_{i,\mathrm{md}}$ with density sequence $\{b_0, b_1, \ldots, b_n, \ldots\}$ satisfying

$$\frac{b_0}{b_i} = e^{\epsilon},$$

$$\frac{b_1}{b_{i+1}} = e^\epsilon,$$

and $V(\mathcal{P}_a) \geq V(\mathcal{P}_b)$.

First, it is easy to see a_1 is strictly less than a_0, since if $a_0 = a_1$, then $\frac{a_1}{a_{i+1}} = \frac{a_0}{a_{i+1}} \geq \frac{a_0}{a_i} = e^\epsilon$. Then we construct a new density sequence by increasing a_1 and decreasing a_{i+1}. More precisely, we define a new sequence $\{b_0, b_1, \dots, b_n, \dots\}$ as

$$\begin{aligned} b_k &= a_k, \forall k \neq 1, k \neq i+1, \\ b_1 &= a_1 + \delta, \\ b_{i+1} &= a_{i+1} - \delta, \end{aligned}$$

where $\delta = \frac{e^\epsilon a_{i+1} - a_1}{1+e^\epsilon}$ and thus $\frac{b_1}{b_{i+1}} = e^\epsilon$.

It is easy to verify that $\{b_0, b_1, \dots, b_n, \dots\}$ is a valid probability density sequence and the corresponding probability distribution $\mathcal{P}_b$ satisfies the differential privacy constraint (2.11). Moreover, $V(\mathcal{P}_a) \geq V(\mathcal{P}_b)$. Therefore, we only need to consider $\mathcal{P} \in \mathcal{SP}_{i,\mathrm{md}}$ with density sequences $\{a_0, a_1, \dots, a_n, \dots\}$ satisfying $\frac{a_0}{a_i} = e^\epsilon$ and $\frac{a_1}{a_{i+1}} = e^\epsilon$.

Using the same argument, we can show that we only need to consider $\mathcal{P} \in \mathcal{SP}_{i,\mathrm{md}}$ with density sequences $\{a_0, a_1, \dots, a_n, \dots\}$ satisfying

$$\frac{a_k}{a_{i+k}} = e^\epsilon, \forall k \geq 0.$$

Therefore,

$$V^* = \inf_{\mathcal{P} \in \cup_{i=1}^\infty \mathcal{SP}_{i,\mathrm{pd}}} \int_{x \in \mathbb{R}} \mathcal{L}(x) \mathcal{P}(dx).$$

□

Due to Lemma A.10, we only need to consider probability distribution with symmetric, monotonically decreasing (for $x \geq 0$), and periodically decaying, piecewise constant probability density function. Because of the properties of symmetry and periodic decay, for this class of probability distributions, the probability density function over $\mathbb{R}$ is completely determined by the probability density function over the interval $[0, \Delta)$.

Next, we study what the optimal probability density function should be over the interval $[0, \Delta)$. It turns out that the optimal probability density function over the interval $[0, \Delta)$ is a step function. We use the following three steps to prove this result.

A.2.7 Step 6

Lemma A.11. *Consider a probability distribution $\mathcal{P}_a \in \mathcal{SP}_{i,pd}$ ($i \geq 2$) with density sequence $\{a_0, a_1, \ldots, a_n, \ldots\}$, and $\frac{a_0}{a_{i-1}} < e^\epsilon$. Then there exists a probability distribution $\mathcal{P}_b \in \mathcal{SP}_{i,pd}$ with density sequence $\{b_0, b_1, \ldots, b_n, \ldots\}$such that $\frac{b_0}{b_{i-1}} = e^\epsilon$, and*

$$V(\mathcal{P}_b) \leq V(\mathcal{P}_a).$$

Proof. For each $0 \leq k \leq (i-1)$, define

$$w_k \triangleq \sum_{j=0}^{+\infty} e^{-j\epsilon} \int_{(j+\frac{k}{i})\Delta}^{(j+\frac{k+1}{i})\Delta} \mathcal{L}(x)dx. \tag{A.11}$$

Since $\mathcal{L}(cdot)$ satisfies Property 2.2 and $V^* < \infty$, it is easy to show that the sum of the series in (A.11) exists and is finite, and thus w_k is well defined for all $0 \leq k \leq (i-1)$. In addition, it is easy to see

$$w_0 \leq w_1 \leq w_2 \leq \cdots \leq w_{i-1},$$

since $\mathcal{L}(x)$ is a monotonically increasing function when $x \geq 0$.

Then

$$V(\mathcal{P}_a) = \int_{x\in\mathbb{R}} \mathcal{L}(x)\mathcal{P}_a(dx) = 2\sum_{k=0}^{i-1} w_k a_k.$$

Since $\frac{a_0}{a_{i-1}} < e^\epsilon$, we can scale a_0 up and scale $\{a_1, \ldots, a_{i-1}\}$ down to derive a new valid probability density function with smaller cost. More precisely, define a new probability measure $\mathcal{P}_b \in \mathcal{SP}_{i,\mathrm{pd}}$ with density sequence $\{b_0, b_1, \ldots, b_n, \ldots\}$ via

$$\begin{aligned} b_0 &\triangleq \gamma a_0, \\ b_k &\triangleq \gamma' a_k, \forall 1 \leq k \leq i-1, \end{aligned}$$

for some $\gamma > 1$ and $\gamma' < 1$ such that

$$\frac{b_0}{b_{i-1}} = e^\epsilon.$$

To make $\{b_0, b_1, \ldots, b_n, \ldots\}$ be a valid density sequence, i.e., to make the integral of the

corresponding probability density function over $\mathbb{R}$ be 1, we have

$$\sum_{k=0}^{i-1} b_k = \sum_{k=0}^{i-1} a_k = \frac{1-e^{-\epsilon}}{2}\frac{i}{\Delta}.$$

Define $t \triangleq \frac{1-e^{-\epsilon}}{2}\frac{i}{\Delta}$, then we have two linear equations on γ and γ':

$$\gamma a_0 = e^{\epsilon}\gamma' \tag{A.12}$$

$$\gamma a_0 + \gamma'(t - a_0) = t. \tag{A.13}$$

From (A.12) and (A.13), we can easily get

$$\gamma = \frac{e^{\epsilon} t a_{i-1}}{a_0(t - a_0 + e^{\epsilon} a_{i-1})} > 1$$
$$\gamma' = \frac{t}{t - a_0 + e^{\epsilon} a_{i-1}} < 1.$$

Then we can verify that the $V(\mathcal{P}_a) \geq V(\mathcal{P}_a)$. Indeed,

$$\begin{aligned}
& V(\mathcal{P}_a) - V(\mathcal{P}_b) \\
&= \int_{x\in\mathbb{R}} \mathcal{L}(x)\mathcal{P}_a(dx) - \int_{x\in\mathbb{R}} \mathcal{L}(x)\mathcal{P}_b(dx) \\
&= 2\sum_{k=0}^{i-1} w_k a_k - 2\sum_{k=0}^{i-1} w_k b_k \\
&= 2\left((1-\gamma)w_0 a_0 + (1-\gamma')\sum_{k=1}^{i-1} w_k a_k\right) \\
&\geq 2\left((1-\gamma)w_0 a_0 + (1-\gamma')\sum_{k=1}^{i-1} w_0 a_k\right) \\
&= 2\left((1-\gamma)w_0 a_0 + (1-\gamma')w_0(t-a_0)\right) \\
&= 2w_0\left(a_0 - \frac{a_{i-1}e^{\epsilon}t}{t - a_0 + e^{\epsilon}a_{i-1}} + (t-a_0)\frac{-a_0 + e^{\epsilon}a_{i-1}}{t - a_0 + e^{\epsilon}a_{i-1}}\right) \\
&= 0.
\end{aligned}$$

This completes the proof.

□

Therefore, due to Lemma A.11, for all $i \geq 2$, we only need to consider probability distributions $\mathcal{P} \in \mathcal{SP}_{i,\mathrm{pd}}$ with density sequence $\{a_0, a_1, \ldots, a_n, \ldots\}$ satisfying $\frac{a_0}{a_{i-1}} = e^{\epsilon}$.

More precisely, define

$$\mathcal{SP}_{i,\text{fr}} = \{\mathcal{P} \in \mathcal{SP}_{i,\text{pd}} | \mathcal{P} \text{ has density sequence } \{a_0, a_1, \dots, a_n, \dots\} \text{ satisfying } \frac{a_0}{a_{i-1}} = e^{\epsilon}\}.$$

Then due to Lemma A.11,

Lemma A.12.

$$V^* = \inf_{\mathcal{P} \in \cup_{i=3}^{\infty} \mathcal{SP}_{i,fr}} \int_{x \in \mathbb{R}} \mathcal{L}(x) \mathcal{P}(dx).$$

A.2.8 Step 7

Next, we argue that for each probability distribution $\mathcal{P} \in \mathcal{SP}_{i,\text{fr}}$ $(i \geq 3)$ with density sequence $\{a_0, a_1, \dots, a_n, \dots\}$, we can assume that there exists an integer $1 \leq k \leq (i-2)$, such that

$$a_j = a_0, \forall 0 \leq j < k, \tag{A.14}$$

$$a_j = a_{i-1}, \forall k < j < i. \tag{A.15}$$

More precisely,

Lemma A.13. *Consider a probability distribution $\mathcal{P}_a \in \mathcal{SP}_{i,fr}$ $(i \geq 3)$ with density sequence $\{a_0, a_1, \dots, a_n, \dots\}$. Then there exists a probability distribution $\mathcal{P}_b \in \mathcal{SP}_{i,fr}$ with density sequence $\{b_0, b_1, \dots, b_n, \dots\}$ such that there exists an integer $1 \leq k \leq (i-2)$ with*

$$b_j = a_0, \forall\ 0 \leq j < k, \tag{A.16}$$

$$b_j = a_{i-1}, \forall\ k < j < i, \tag{A.17}$$

and

$$V(\mathcal{P}_b) \leq V(\mathcal{P}_a). \tag{A.18}$$

Proof. If there exists an integer $1 \leq k \leq (i-2)$ such that

$$a_j = a_0, \forall\ 0 \leq j < k,$$

$$a_j = a_{i-1}, \forall\ k < j < i,$$

then we can set $\mathcal{P}_b = \mathcal{P}_a$.

Otherwise, let k_1 be the smallest integer in $\{0, 1, 2, \ldots, i-1\}$ such that

$$a_{k_1} \neq a_0,$$

and let k_2 be the biggest integer in $\{0, 1, 2, \ldots, i-1\}$ such that

$$a_{k_2} \neq a_{i-1}.$$

It is easy to see that $k_1 \neq k_2$. Then we can increase a_{k_1} and decrease a_{k_2} simultaneously by the same amount to derive a new probability distribution $\mathcal{P}_b \in \mathcal{SP}_{i,\text{fr}}$ with smaller cost. Indeed, if

$$a_0 - a_{k_1} \leq a_{k_2} - a_{i-1},$$

then consider a probability distribution $\mathcal{P}_b \in \mathcal{SP}_{i,\text{fr}}$ with density sequence $\{b_0, b_1, \ldots, b_{i-1}, \ldots\}$ defined as

$$\begin{aligned} b_j &= a_0, \forall 0 \leq j \leq k_1, \\ b_j &= a_j, \forall k_1 < j \leq k_2 - 1, \\ b_{k_2} &= a_{k_2} - (a_0 - a_{k_1}), \\ b_j &= a_j, \forall k_2 < j \leq i - 1. \end{aligned}$$

We can verify that $V(\mathcal{P}_a) \geq V(\mathcal{P}_b)$ via

$$\begin{aligned} & V(\mathcal{P}_a) - V(\mathcal{P}_b) \\ &= \int_{x \in \mathbb{R}} \mathcal{L}(x) \mathcal{P}_a(dx) - \int_{x \in \mathbb{R}} \mathcal{L}(x) \mathcal{P}_b(dx) \\ &= 2(w_{k_1} b_{k_1} + w_{k_2} b_{k_2}) - 2(w_{k_1} a_{k_1} + w_{k_2} a_{k_2}) \\ &= 2w_{k_1}(a_0 - a_{k_1}) + 2w_{k_2}(a_{k_2} - (a_0 - a_{k_1}) - a_{k_2}) \\ &= 2(a_0 - a_{k_1})(w_{k_1} - w_{k_2}) \\ &\leq 0, \end{aligned}$$

where w_i is defined in (A.11).

If $a_0 - a_{k_1} \geq a_{k_2} - a_{i-1}$, then accordingly we can construct $\mathcal{P}_b \in \mathcal{SP}_{i,\text{fr}}$ by setting

$$\begin{aligned} b_j &= a_0, \forall 0 \leq j < k_1, \\ b_{k_1} &= a_{k_1} + (a_{k_2} - a_{i-1}), \end{aligned}$$

$$b_j = a_j, \forall k_1 < j \leq k_2 - 1,$$
$$b_j = a_{i-1}, \forall k_2 \leq j \leq i - 1.$$

And similarly, it is easy to verify that $V(\mathcal{P}_a) \geq V(\mathcal{P}_b)$.

Therefore, continue in this way, and finally we will obtain a probability distribution $\mathcal{P}_b \in \mathcal{SP}_{i,\text{fr}}$ with density sequence $\{b_0, b_1, \dots, b_n, \dots\}$ such that (A.16), (A.17) and (A.18) hold.

This completes the proof.

□

Define

$$\mathcal{SP}_{i,\text{step}} = \{\mathcal{P} \in \mathcal{SP}_{i,\text{fr}} \mid \mathcal{P} \text{ has density sequence } \{a_0, a_1, \dots, a_n, \dots\} \text{ satisfying(A.16) and (A.17) for some } 1 \leq k \leq (i-2)\}.$$

Then due to Lemma A.13,

Lemma A.14.

$$V^* = \inf_{\mathcal{P} \in \cup_{i=3}^{\infty} \mathcal{SP}_{i,step}} \int_{x \in \mathbb{R}} \mathcal{L}(x) \mathcal{P}(dx).$$

A.2.9 Step 8

Proof of Theorem 2.4. Since $\{\mathcal{P}_\gamma | \gamma \in [0,1]\} \subseteq \mathcal{SP}$, we have

$$V^* = \inf_{\mathcal{P} \in \mathcal{SP}} \int_{x \in \mathbb{R}} \mathcal{L}(x) \mathcal{P}(dx) \leq \inf_{\gamma \in [0,1]} \int_{x \in \mathbb{R}} \mathcal{L}(x) \mathcal{P}_\gamma(dx).$$

We prove the reverse direction in the following.

We first prove that for any $\mathcal{P} \in \mathcal{SP}_{i,\text{step}}$ ($i \geq 3$), there exists $\gamma \in [0,1]$ such that

$$\int_{x \in \mathbb{R}} \mathcal{L}(x) \mathcal{P}_\gamma(dx) \leq \int_{x \in \mathbb{R}} \mathcal{L}(x) \mathcal{P}(dx).$$

Consider the density sequence $\{a_0, a_1, \dots, a_n, \dots\}$ of $\mathcal{P}$. Since $\mathcal{P} \in \mathcal{SP}_{i,\text{step}}$, there exists an integer $0 \leq k \leq i-2$ such that

$$a_j = a_0, \forall 0 \leq j < k,$$
$$a_j = a_0 e^{-\epsilon}, \forall k < j \leq i - 1.$$

Let

$$\gamma' \triangleq \frac{\frac{1-e^{-\epsilon}}{2\Delta} - a_0 e^{-\epsilon}}{a_0(1-e^{-\epsilon})} \in [0,1].$$

Then $a(\gamma') = a_0$.

It is easy to verify that

$$k\frac{\Delta}{i} \leq \gamma'\Delta \leq (k+1)\frac{\Delta}{i}.$$

The probability density functions of $\mathcal{P}$ and $\mathcal{P}_{\gamma'}$ are the same when $x \in [0, \frac{k}{i}\Delta) \cup [\frac{k+1}{i}\Delta, \Delta)$. Because they periodically decay, the integral of probability density functions over $[0, \Delta)$ is $\frac{1-e^{-\epsilon}}{2}$. Hence, we have

$$a_k\frac{\Delta}{i} = a_0(\gamma' - \frac{k}{i})\Delta + e^{-\epsilon}a_0(\frac{k+1}{i} - \gamma')\Delta.$$

Define $\beta \triangleq i(\gamma' - \frac{k}{i}) \in [0,1]$. Then

$$a_k = \beta a_0 + (1-\beta)e^{-\epsilon}a_0.$$

Define

$$w_k^{(1)} \triangleq \sum_{j=0}^{+\infty} e^{-j\epsilon} \int_{(j+\frac{k}{i})\Delta}^{(j+\gamma')\Delta} \mathcal{L}(x)dx, \tag{A.19}$$

$$w_k^{(2)} \triangleq \sum_{j=0}^{+\infty} e^{-j\epsilon} \int_{(j+\gamma')\Delta}^{(j+\frac{k+1}{i})\Delta} \mathcal{L}(x)dx,. \tag{A.20}$$

Note that $w_k = w_k^{(1)} + w_k^{(2)}$. Since $\mathcal{L}(x)$ is a monotonically increasing function when $x \geq 0$, we have

$$\frac{w_k^{(2)}}{w_k^{(1)}} \geq \frac{(j+\frac{k+1}{i})\Delta - (j+\gamma')\Delta}{(j+\gamma')\Delta - (j+\frac{k}{i})\Delta} = \frac{\frac{k+1}{i} - \gamma'}{\gamma' - \frac{k}{i}}.$$

Therefore,

$$\begin{aligned}&\int_{x\in\mathbb{R}} \mathcal{L}(x)\mathcal{P}(dx) - \int_{x\in\mathbb{R}} \mathcal{L}(x)\mathcal{P}_{\gamma'}(dx)\\ =&2w_k a_k - 2\left(w_k^{(1)}a_0 + w_k^{(2)}a_0 e^{-\epsilon}\right)\end{aligned}$$

$$
\begin{aligned}
=&2\left(w_k^{(1)}+w_k^{(2)}\right)a_k-2\left(w_k^{(1)}a_0+w_k^{(2)}a_0e^{-\epsilon}\right)\\
=&2(a_k-a_0e^{-\epsilon})w_k^{(2)}-2(a_0-a_k)w_k^{(1)}.
\end{aligned}
$$

Since

$$
\begin{aligned}
\frac{a_k-a_0e^{-\epsilon}}{a_0-a_k}&=\frac{\beta(a_0-a_0e^{-\epsilon})}{(1-\beta)(a_0-a_0e^{-\epsilon})}\\
&=\frac{\beta}{1-\beta}\\
&=\frac{\gamma'-\frac{k}{i}}{\frac{k+1}{i}-\gamma'}\\
&\geq\frac{w_k^{(1)}}{w_k^{(2)}},
\end{aligned}
$$

we have

$$
\begin{aligned}
&\int_{x\in\mathbb{R}}\mathcal{L}(x)\mathcal{P}(dx)-\int_{x\in\mathbb{R}}\mathcal{L}(x)\mathcal{P}_{\gamma'}(dx)\\
=&2(a_k-a_0e^{-\epsilon})w_k^{(2)}-2(a_0-a_k)w_k^{(1)}\\
\geq&0.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
V^*&=\inf_{\mathcal{P}\in\cup_{i=3}^{\infty}\mathcal{SP}_{i,\text{step}}}\int_{x\in\mathbb{R}}\mathcal{L}(x)\mathcal{P}(dx)\\
&\geq\inf_{\gamma\in[0,1]}\int_{x\in\mathbb{R}}\mathcal{L}(x)\mathcal{P}_{\gamma}(dx).
\end{aligned}
$$

We conclude

$$
V^*=\inf_{\mathcal{P}\in\mathcal{SP}}\int_{x\in\mathbb{R}}\mathcal{L}(x)\mathcal{P}(dx)=\inf_{\gamma\in[0,1]}\int_{x\in\mathbb{R}}\mathcal{L}(x)\mathcal{P}_{\gamma}(dx)=\inf_{\gamma\in[0,1]}\int_{x\in\mathbb{R}}\mathcal{L}(x)f_{\gamma}(x)dx.
$$

This completes the proof of Theorem 2.4. □

A.3 Proof of Theorem 2.5

Proof of Theorem 2.5. Recall $b \triangleq e^{-\epsilon}$, and $\mathcal{L}(x) = |x|$. We can compute $V(\mathcal{P}_\gamma)$ via

$$\begin{aligned}
V(\mathcal{P}_\gamma) &= \int_{x\in\mathbb{R}} |x| f^\gamma(x) dx \\
&= 2\int_0^{+\infty} x f^\gamma(x) dx \\
&= 2\sum_{k=0}^{+\infty} \left(\int_0^{\gamma\Delta} (x + k\Delta) a(\gamma) e^{-k\epsilon} dx + \int_{\gamma\Delta}^{\Delta} (x + k\Delta) a(\gamma) e^{-\epsilon} e^{-k\epsilon} dx \right) \\
&= 2\Delta^2 a(\gamma) \sum_{k=0}^{+\infty} \left(e^{-k\epsilon} \frac{(k+\gamma)^2 - k^2}{2} + e^{-(k+1)\epsilon} \frac{(k+1)^2 - (k+\gamma)^2}{2} \right) \\
&= 2\Delta^2 a(\gamma) \sum_{k=0}^{+\infty} \left(e^{-k\epsilon} \frac{\gamma^2 + 2k\gamma}{2} + e^{-(k+1)\epsilon} \frac{2k + 1 - 2k\gamma - \gamma^2}{2} \right) \\
&= 2\Delta^2 a(\gamma) \sum_{k=0}^{+\infty} \left((b + (1-b)\gamma) k e^{-k\epsilon} + \frac{b + (1-b)\gamma^2}{2} e^{-k\epsilon} \right) \\
&= 2\Delta^2 a(\gamma) \left((b + (1-b)\gamma) \frac{b}{(1-b)^2} + \frac{b + (1-b)\gamma^2}{2} \frac{1}{1-b} \right) \qquad \text{(A.21)} \\
&= 2\Delta^2 \frac{1-b}{2\Delta(b + (1-b)\gamma)} \left((b + (1-b)\gamma) \frac{b}{(1-b)^2} + \frac{b + (1-b)\gamma^2}{2} \frac{1}{1-b} \right) \\
&= \Delta \left(\frac{b}{1-b} + \frac{1}{2} \frac{b + (1-b)\gamma^2}{b + (1-b)\gamma} \right),
\end{aligned}$$

where in (A.21) we use the formulas

$$\sum_{k=1}^{+\infty} b^k = \frac{1}{1-b}, \qquad \text{(A.22)}$$

$$\sum_{k=1}^{+\infty} k b^k = \frac{b}{(1-b)^2}. \qquad \text{(A.23)}$$

Note that the first term $\frac{b}{1-b}$ is independent of γ. Define

$$g(\gamma) \triangleq \frac{b + (1-b)\gamma^2}{b + (1-b)\gamma},$$

and thus to minimize $V(\mathcal{P}_\gamma)$ over $\gamma \in [0,1]$, we only need to minimize $g(\gamma)$ over $\gamma \in [0,1]$.

Since $\gamma \in [0,1]$, $g(\gamma) \leq 1$. Also note that $g(0) = g(1) = 1$. So the optimal γ^* which minimizes $g(\gamma)$ lies in $(0,1)$.

Compute the derivative of $g(\gamma)$ via

$$\begin{aligned} g'(\gamma) &= \frac{2\gamma(1-b)(b+(1-b)\gamma) - (b+(1-b)\gamma^2)(1-b)}{(b+(1-b)\gamma)^2} \\ &= (1-b)\frac{(1-b)\gamma^2 + 2b\gamma - b}{(b+(1-b)\gamma)^2}. \end{aligned}$$

Set $g'(\gamma^*) = 0$ and we get

$$\begin{aligned} \gamma* &= \frac{\sqrt{b} - b}{1-b} \\ &= \frac{e^{-\frac{1}{2}\epsilon} - e^{-\epsilon}}{1 - e^{-\epsilon}} \\ &= \frac{1}{1 + e^{\frac{\epsilon}{2}}}. \end{aligned}$$

Therefore,

$$\begin{aligned} V(\mathcal{P}_{\gamma^*}) &= \Delta\left(\frac{b}{1-b} + \frac{1}{2}\frac{b+(1-b)\gamma^{*2}}{b+(1-b)\gamma^*}\right) \\ &= \Delta\frac{e^{\frac{\epsilon}{2}}}{e^{\epsilon}-1}. \end{aligned}$$

Due to Theorem 2.4, the minimum expectation of noise amplitude is $V(\mathcal{P}_{\gamma^*}) = \Delta\frac{e^{\frac{\epsilon}{2}}}{e^{\epsilon}-1}$. □

A.4 Proof of Theorem 2.7

Proof of Theorem 2.7. Recall $b \triangleq e^{-\epsilon}$. Then we compute $V(\mathcal{P}_\gamma)$ for the cost function $\mathcal{L}(x) = x^2$ via

$$\begin{aligned} & V(\mathcal{P}_\gamma) \\ &= \int_{x\in\mathbb{R}} x^2 f^\gamma(x)dx \\ &= 2\int_0^{+\infty} x^2 f^\gamma(x)dx \\ &= 2\sum_{k=0}^{+\infty}\left(\int_0^{\gamma\Delta}(x+k\Delta)^2 a(\gamma)e^{-k\epsilon}dx + \int_{\gamma\Delta}^{\Delta}(x+k\Delta)^2 a(\gamma)e^{-\epsilon}e^{-k\epsilon}dx\right) \end{aligned}$$

$$
\begin{aligned}
&= 2\Delta^3 a(\gamma) \sum_{k=0}^{+\infty} \left(e^{-k\epsilon} \frac{(k+\gamma)^3 - k^3}{3} + e^{-(k+1)\epsilon} \frac{(k+1)^3 - (k+\gamma)^3}{3} \right) \\
&= 2\Delta^3 a(\gamma) \sum_{k=0}^{+\infty} \left(e^{-k\epsilon} \frac{\gamma^3 + 3k\gamma^2 + 3k^2\gamma}{3} + e^{-(k+1)\epsilon} \frac{3k^2 + 3k + 1 - 3k^2\gamma - 3k\gamma^2 - \gamma^3}{3} \right) \\
&= 2\Delta^3 a(\gamma) \sum_{k=0}^{+\infty} \left((\frac{1-\gamma^3}{3} b + \frac{\gamma^3}{3}) e^{-k\epsilon} + (\gamma^2 + (1-\gamma^2)b) k e^{-k\epsilon} + (\gamma + (1-\gamma)b) k^2 e^{-k\epsilon} \right) \\
&= 2\Delta^3 a(\gamma) \left((\frac{1-\gamma^3}{3} b + \frac{\gamma^3}{3}) \frac{1}{1-b} + (\gamma^2 + (1-\gamma^2)b) \frac{b}{(1-b)^2} + (\gamma + (1-\gamma)b) \frac{b^2+b}{(1-b)^3} \right)
\end{aligned} \tag{A.24}
$$

$$
\begin{aligned}
&= 2\Delta^3 \frac{1-b}{2\Delta(b + (1-b)\gamma)} \\
&\left((\frac{1-\gamma^3}{3} b + \frac{\gamma^3}{3}) \frac{1}{1-b} + (\gamma^2 + (1-\gamma^2)b) \frac{b}{(1-b)^2} + (\gamma + (1-\gamma)b) \frac{b^2+b}{(1-b)^3} \right) \\
&= \Delta^2 \left(\frac{b^2+b}{(1-b)^2} + \frac{b + (1-b)\gamma^2}{b + (1-b)\gamma} \frac{b}{1-b} + \frac{1}{3} \frac{b + (1-b)\gamma^3}{b + (1-b)\gamma} \right),
\end{aligned} \tag{A.25}
$$

where in (A.24) we use formulas (A.22), (A.23) and

$$
\sum_{k=1}^{+\infty} k^2 b^k = \frac{(b^2+b)}{(1-b)^3}. \tag{A.26}
$$

Note that the first term $\frac{b^2+b}{(1-b)^2}$ is independent of γ. Define

$$
\begin{aligned}
h(\gamma) &\triangleq \frac{b + (1-b)\gamma^2}{b + (1-b)\gamma} \frac{b}{1-b} + \frac{1}{3} \frac{b + (1-b)\gamma^3}{b + (1-b)\gamma} \\
&= \frac{\frac{(1-b)\gamma^3}{3} + b\gamma^2 + \frac{b^2}{1-b} + \frac{b}{3}}{b + (1-b)\gamma},
\end{aligned}
$$

and thus to minimize $V(\mathcal{P}_\gamma)$ over $\gamma \in [0,1]$, we only need to minimize $h(\gamma)$ over $\gamma \in [0,1]$.

Since $\gamma \in [0,1]$, $h(\gamma) \leq \frac{b}{1-b} + \frac{1}{3}$. Also note that $h(0) = h(1) = \frac{b}{1-b} + \frac{1}{3}$. So the optimal γ^* which minimizes $h(\gamma)$ lies in $(0,1)$.

Compute the derivative of $h(\gamma)$ via

$$
\begin{aligned}
h'(\gamma) &= \frac{((1-b)\gamma^2 + 2b\gamma)(b + (1-b)\gamma) - (\frac{1-b}{3}\gamma^3 + b\gamma^2 + \frac{b^2}{1-b} + \frac{b}{3})(1-b)}{(b + (1-b)\gamma)^2} \\
&= \frac{\frac{2}{3}(1-b)^2\gamma^3 + 2b(1-b)\gamma^2 + 2b^2\gamma - \frac{2b^2+b}{3}}{(b + (1-b)\gamma)^2}
\end{aligned}
$$

.

Set $h'(\gamma^*) = 0$ and we get

$$\frac{2}{3}(1-b)^2\gamma^{*3} + 2b(1-b)\gamma^{*2} + 2b^2\gamma^* - \frac{2b^2+b}{3} = 0. \tag{A.27}$$

Therefore, the optimal γ^* is the real-valued root of the cubic equation (A.27), which is

$$\gamma^* = -\frac{b}{1-b} + \frac{(b - 2b^2 + 2b^4 - b^5)^{1/3}}{2^{1/3}(1-b)^2}. \tag{A.28}$$

We plot γ^* as a function of b in Figure 2.5, and we can see $\gamma^* \to \frac{1}{2}$ as $\epsilon \to 0$, and $\gamma^* \to 0$ as $\epsilon \to +\infty$. This also holds in the case $\mathcal{L}(x) = |x|$.

Plug (A.28) into (A.25), and we get the minimum noise power

$$\begin{aligned} V(\mathcal{P}_{\gamma^*}) &= \Delta^2 \left(\frac{b^2+b}{(1-b)^2} + \frac{b + (1-b)\gamma^{*2}}{b + (1-b)\gamma^*} \frac{b}{1-b} + \frac{1}{3} \frac{b + (1-b)\gamma^{*3}}{b + (1-b)\gamma^*} \right) \\ &= \Delta^2 \frac{2^{-2/3} b^{2/3} (1+b)^{2/3} + b}{(1-b)^2}. \end{aligned}$$

Due to Theorem 2.4, the minimum expectation of noise power is

$$V(\mathcal{P}_{\gamma^*}) = \Delta^2 \frac{2^{-2/3} b^{2/3} (1+b)^{2/3} + b}{(1-b)^2}$$

.

□

A.5 Proof of Theorem 2.9

Proof of Theorem 2.9. Let $n = m + 1$, and define

$$c_i \triangleq \sum_{k=0}^{+\infty} b^k k^i, \tag{A.29}$$

for nonnegative integer i.

First we compute $V(\mathcal{P}_\gamma)$ via

$$V(\mathcal{P}_\gamma) = 2 \sum_{k=0}^{+\infty} \left(\int_0^{\gamma\Delta} (x + k\Delta)^m a(\gamma) e^{-k\epsilon} dx + \int_{\gamma\Delta}^{\Delta} (x + k\Delta)^m a(\gamma) e^{-(k+1)\epsilon} dx \right)$$

$$
\begin{aligned}
&= 2a(\gamma)\Delta^{m+1}\sum_{k=0}^{+\infty}\left(b^k\frac{(k+\gamma)^{m+1}-k^{m+1}}{m+1}+b^{k+1}\frac{(k+1)^{m+1}-(k+\gamma)^{m+1}}{m+1}\right)\\
&= 2\Delta^n a(\gamma)\sum_{k=0}^{+\infty}\left(b^k\frac{\sum_{i=1}^n\binom{n}{i}\gamma^i k^{n-i}}{n}+bb^k\frac{\sum_{i=1}^n\binom{n}{i}(1-\gamma^i)k^{n-i}}{n}\right)\\
&= 2\Delta^n a(\gamma)\left(\sum_{i=1}^n\frac{\binom{n}{i}\gamma^i c_{n-i}}{n}+b\sum_{i=1}^n\frac{\binom{n}{i}(1-\gamma^i)c_{n-i}}{n}\right)\\
&= 2\Delta^n a(\gamma)\sum_{i=1}^n\frac{\binom{n}{i}c_{n-i}(\gamma^i(1-b)+b)}{n}\\
&= \frac{2\Delta^n(1-b)}{2\Delta n}\frac{\sum_{i=1}^n\binom{n}{i}c_{n-i}(\gamma^i(1-b)+b)}{\gamma(1-b)+b}.
\end{aligned}
$$

Let $h_i(\gamma)\triangleq\frac{\gamma^i(1-b)+b}{\gamma(1-b)+b}$ for $i\geq 2$. Since $h_i(0)=h_i(1)=1$ and $h_i(\gamma)<1$ for $\gamma\in(0,1)$, $h_i(\gamma)$ achieves the minimum value in the open interval $(0,1)$.

Therefore, if we define $h(\gamma)\triangleq\frac{\sum_{i=1}^n\binom{n}{i}c_{n-i}(\gamma^i(1-b)+b)}{\gamma(1-b)+b}$, the optimal $\gamma^*\in[0,1]$, which minimizes $V(\mathcal{P}_\gamma)$, should satisfy

$$h'(\gamma^*)=0,$$

where $h'(\cdot)$ denotes the first order derivative of $h(\cdot)$.

It is straightforward to derive the expression for $h'(\cdot)$:

$$
\begin{aligned}
h'(\gamma) &= \frac{(\sum_{i=1}^n\binom{n}{i}c_{n-i}i\gamma^{i-1}(1-b))(\gamma(1-b)+b)-(1-b)\sum_{i=1}^n\binom{n}{i}c_{n-i}(\gamma^i(1-b)+b)}{(\gamma(1-b)+b)^2}\\
&= \frac{\sum_{i=1}^n\binom{n}{i}c_{n-i}(i-1)\gamma^i(1-b)^2+\sum_{i=1}^n\binom{n}{i}c_{n-i}i\gamma^{i-1}(1-b)b-\sum_{i=1}^n\binom{n}{i}c_{n-i}b(1-b)}{(\gamma(1-b)+b)^2}.
\end{aligned}
\tag{A.30}
$$

Therefore, γ^* should make the numerator of (A.30) be zero, i.e., γ^* satisfies

$$\sum_{i=1}^n\binom{n}{i}c_{n-i}(i-1)\gamma^i(1-b)^2+\sum_{i=1}^n\binom{n}{i}c_{n-i}i\gamma^{i-1}(1-b)b-\sum_{i=1}^n\binom{n}{i}c_{n-i}b(1-b)=0.$$

Since

$$
\begin{aligned}
&\sum_{i=1}^n\binom{n}{i}c_{n-i}(i-1)\gamma^i(1-b)^2+\sum_{i=1}^n\binom{n}{i}c_{n-i}i\gamma^{i-1}(1-b)b-\sum_{i=1}^n\binom{n}{i}c_{n-i}b(1-b)\\
=&\sum_{i=1}^n\binom{n}{i}c_{n-i}(i-1)\gamma^i(1-b)^2+\sum_{i=0}^{n-1}\binom{n}{i+1}c_{n-(i+1)}(i+1)\gamma^i(1-b)b
\end{aligned}
$$

$$
\begin{aligned}
&- \sum_{i=1}^{n} \binom{n}{i} c_{n-i} b(1-b) \\
=& c_0(n-1)\gamma^n(1-b)^2 + \sum_{i=1}^{n-1} \left(\binom{n}{i} c_{n-i}(i-1)(1-b)^2 + \binom{n}{i+1} c_{n-(i+1)}(i+1)(1-b)b \right) \gamma^i \\
&+ nc_{n-1}(1-b)b - \sum_{i=1}^{n} \binom{n}{i} c_{n-i} b(1-b) \\
=& c_0(n-1)\gamma^n(1-b)^2 + \sum_{i=1}^{n-1} \left(\binom{n}{i} c_{n-i}(i-1)(1-b)^2 + \binom{n}{i+1} c_{n-(i+1)}(i+1)(1-b)b \right) \gamma^i \\
&- \sum_{i=2}^{n} \binom{n}{i} c_{n-i} b(1-b),
\end{aligned}
$$

γ^* satisfies

$$
\begin{aligned}
&c_0(n-1)\gamma^{*n}(1-b)^2 + \sum_{i=1}^{n-1} \left(\binom{n}{i} c_{n-i}(i-1)(1-b)^2 + \binom{n}{i+1} c_{n-(i+1)}(i+1)(1-b)b \right) \gamma^{*i} \\
&- \sum_{i=2}^{n} \binom{n}{i} c_{n-i} b(1-b) = 0. \qquad \text{(A.31)}
\end{aligned}
$$

We can derive the asymptotic properties of γ^* from (A.31). Before deriving the properties of γ^*, we first study the asymptotic properties of c_i, which are functions of b.

There are closed-form formulas for c_i (i=0,1,2,3):

$$
\begin{aligned}
c_0 &= \sum_{k=0}^{+\infty} b^k = \frac{1}{1-b}, \\
c_1 &= \sum_{k=0}^{+\infty} b^k k = \frac{b}{(1-b)^2}, \\
c_2 &= \sum_{k=0}^{+\infty} b^k k^2 = \frac{b^2+b}{(1-b)^3}, \\
c_3 &= \sum_{k=0}^{+\infty} b^k k^3 = \frac{b^3+4b^2+b}{(1-b)^4}.
\end{aligned}
$$

In general, for $i \geq 1$,

$$
c_{i+1} = \sum_{k=0}^{+\infty} b^k k^{i+1} = \sum_{k=1}^{+\infty} b^k k^{i+1} = b + \sum_{k=1}^{+\infty} b^{k+1}(k+1)^{i+1},
$$

$$bc_{i+1} = \sum_{k=0}^{+\infty} b^{k+1}k^{i+1} = \sum_{k=1}^{+\infty} b^{k+1}k^{i+1}.$$

Therefore,

$$\begin{aligned}
c_{i+1} - bc_{i+1} &= b + \sum_{k=1}^{+\infty} b^{k+1}((k+1)^{i+1} - k^{i+1}) \\
&= b + \sum_{k=1}^{+\infty} b^{k+1} \sum_{j=0}^{i} \binom{i+1}{j} k^j \\
&= b + b\sum_{j=0}^{i} \binom{i+1}{j} \sum_{k=1}^{+\infty} k^j b^k \\
&= b + b(\frac{b}{1-b} + \sum_{j=1}^{i} \binom{i+1}{j} c_j) \\
&= \frac{b}{1-b} + b\sum_{j=1}^{i} \binom{i+1}{j} c_j,
\end{aligned}$$

and thus

$$c_{i+1} = \frac{b}{(1-b)^2} + \frac{b}{1-b}\sum_{j=1}^{i} \binom{i+1}{j} c_j. \tag{A.32}$$

From (A.32), by induction we can easily prove that

- as $b \to 0$, $c_i \to 0, \forall i \geq 1$;
- as $b \to 1$, $\forall i \geq 0, c_i \to +\infty, c_i = \Omega(\frac{i!}{(1-b)^{i+1}})$ and

$$\lim_{b\to 1} \frac{c_{i+1}}{c_i}(1-b) = i+1.$$

As $b \to 0$, since $c_i \to 0$ for $i \geq 1$ and $c_0 = 1$, the last two terms of (A.31) go to zero, and thus from (A.31) we can see that γ^* goes to zero as well.

As $b \to 1$, since $c_i = \Omega(\frac{1}{(1-b)^{i+1}})$ and γ^* is bounded by 1, the first term of (A.31) goes to zero, and the dominated terms in (A.31) are

$$\binom{n}{2} c_{n-2} 2(1-b)b\gamma^* - \binom{n}{2} c_{n-2} b(1-b) = 0.$$

Thus, in the limit we have $\gamma^* = \frac{1}{2}$. Therefore, as $b \to 1$, $\gamma^* \to \frac{1}{2}$.

This completes the proof. □

A.6 Proof of Theorem 2.12 and Theorem 2.13

In this section, we prove Theorem 2.12 and Theorem 2.13, which give the optimal noise-adding mechanisms in the discrete setting.

A.6.1 Outline of Proof

The proof technique is very similar to the proof in the continuous settings in Appendix A.2. The proof consists of 5 steps in total, and in each step we narrow down the set of probability distributions where the optimal probability distribution should lie:

- Step 1 proves that we only need to consider probability mass functions which are monotonically increasing for $i \leq 0$ and monotonically decreasing for $i \geq 0$.
- Step 2 proves that we only need to consider symmetric probability mass functions.
- Step 3 proves that we only need to consider symmetric probability mass functions which have periodic and geometric decay for $i \geq 0$, and this proves Theorem 2.12.
- Step 4 and Step 5 prove that the optimal probability mass function over the interval $[0, \Delta)$ is a discrete step function, and they conclude the proof of Theorem 2.13.

A.6.2 Step 1

Recall $\mathcal{SP}$ denotes the set of all probability mass functions which satisfy the ϵ-differential privacy constraint (2.16). Define

$$V^* \triangleq \inf_{\mathcal{P} \in \mathcal{SP}} \sum_{i=-\infty}^{+\infty} \mathcal{L}(i)\mathcal{P}(i).$$

First we prove that we only need to consider probability mass functions which are monotonically increasing for $i \leq 0$ and monotonically decreasing for $i \geq 0$.

Define

$$\mathcal{SP}_{\text{mono}} \triangleq \{\mathcal{P} \in \mathcal{SP} | \mathcal{P}(i) \leq \mathcal{P}(j), \mathcal{P}(m) \geq \mathcal{P}(n), \forall i \leq j \leq 0, 0 \leq m \leq n\}.$$

Lemma A.15.

$$V^* = \inf_{\mathcal{P} \in \mathcal{SP}_{mono}} \sum_{i=-\infty}^{+\infty} \mathcal{L}(i)\mathcal{P}(i).$$

Proof. We will prove that given a probability mass function $\mathcal{P}_a \in \mathcal{SP}$, we can construct a new probability mass function $\mathcal{P}_b \in \mathcal{SP}_{\text{mono}}$ such that

$$\sum_{i=-\infty}^{+\infty} \mathcal{L}(i)\mathcal{P}_a(i) \geq \sum_{i=-\infty}^{+\infty} \mathcal{L}(i)\mathcal{P}_b(i).$$

Given $\mathcal{P}_a \in \mathcal{SP}$, consider the sequence $sa = \{\mathcal{P}_a(0), \mathcal{P}_a(1), \mathcal{P}_a(-1), \mathcal{P}_a(2), \mathcal{P}_a(-2), \dots\}$. Use the same argument in Lemma A.5 and we can show $\mathcal{P}_a(i) > 0, \forall\, i \in \mathbb{Z}$. Let the sequence $sb = \{b_0, b_1, b_{-1}, b_2, b_{-2}, \dots\}$ be a permutation of the sequence sa in descending order. Since $\sum_{i=-\infty}^{+\infty} \mathcal{P}_a(i) = 1$, $\lim_{i\to-\infty} \mathcal{P}_a(i) = \lim_{i\to+\infty} \mathcal{P}_a(i) = 0$, and thus sb is well defined. Let π be the corresponding permutation mapping, i.e., $\pi : \mathbb{Z} \to \mathbb{Z}$, and

$$b_i = \mathcal{P}_a(\pi(i)).$$

Since $\mathcal{L}(\cdot)$ is a symmetric function and monotonically decreasing for $i \geq 0$, we have

$$\begin{aligned}\mathcal{L}(0) \leq \mathcal{L}(1) \leq \mathcal{L}(-1) \leq \mathcal{L}(2) \leq \mathcal{L}(-2) \leq \cdots \\ \leq \mathcal{L}(i) \leq \mathcal{L}(-i) \leq \mathcal{L}(i+1) \leq \mathcal{L}(-(i+1)) \leq \cdots .\end{aligned}$$

Therefore, if we define a probability mass function $\mathcal{P}_b$ with

$$\mathcal{P}_b(i) = b_i, \forall i \in \mathbb{Z},$$

then

$$\sum_{i=-\infty}^{+\infty} \mathcal{L}(i)\mathcal{P}_a(i) \geq \sum_{i=-\infty}^{+\infty} \mathcal{L}(i)\mathcal{P}_b(i).$$

Next, we only need to prove $\mathcal{P}_b \in \mathcal{SP}_{\text{mono}}$, i.e., we need to show that $\mathcal{P}_b$ satisfies the differential privacy constraint (2.16).

Due to the way how we construct the sequence sb, we have

$$\begin{aligned}b_0 \geq b_1 \geq b_2 \geq b_3 \geq \cdots , \\ b_0 \geq b_{-1} \geq b_{-2} \geq b_{-3} \geq \cdots .\end{aligned}$$

Therefore, it is both sufficient and necessary to prove that

$$\frac{b_i}{b_{i+\Delta}} \leq e^{\epsilon}, \forall i \geq 0,$$
$$\frac{b_i}{b_{i-\Delta}} \leq e^{\epsilon}, \forall i \leq 0.$$

Since $\mathcal{P}_a \in \mathcal{SP}$, $\forall\, i \in \{\pi(0) - \Delta, \pi(0) - \Delta + 1, \pi(0) - \Delta + 2, \ldots, \pi(0) + \Delta\}$,

$$\frac{\mathcal{P}_a(\pi(0))}{\mathcal{P}_a(i)} \leq e^{\epsilon}.$$

Therefore, in the sequence sb there exist at least 2Δ elements which are no smaller than $b_0 e^{-\epsilon}$. Since $b_{-\Delta}$ and b_{Δ} are the 2Δth and $(2\Delta - 1)$th largest elements in the sequence sb other than b_0, we have $\frac{b_0}{b_{-\Delta}} \leq e^{\epsilon}$ and $\frac{b_0}{b_{\Delta}} \leq e^{\epsilon}$.

In general, given $i \in \mathbb{Z}$, we can use Algorithm 3 to find at least 2Δ elements in the sequence sb which are no bigger than b_i and no smaller than $b_i e^{-\epsilon}$.

More precisely, given $i \in \mathbb{Z}$, let j_R^* and j_L^* be the output of Algorithm 3. Note that since the while loops in Algorithm 3 can take only at most $2(|i| + 1)$ steps, the algorithm will always terminate. For all integers $j \in [\pi(j_L^*) - \Delta, \pi(j_L^*) - 1]$, $\mathcal{P}_a(j)$ is no bigger than b_i and is no smaller than $\mathcal{P}_a(j_L^*)e^{-\epsilon}$; and for all integers $j \in [\pi(j_R^*) + 1, \pi(j_R^*) + \Delta]$, $\mathcal{P}_a(j)$ is no bigger than b_i and is no smaller than $\mathcal{P}_a(j_R^*)e^{-\epsilon}$. Since $\mathcal{P}_a(j_R^*), \mathcal{P}_a(j_L^*) \geq b_i$, for all $j \in [\pi(j_L^*) - \Delta, \pi(j_L^*) - 1] \cup [\pi(j_R^*) + 1, \pi(j_R^*) + \Delta]$, $\mathcal{P}_a(j)$ is no bigger than b_i and is no smaller than $b_i e^{-\epsilon}$. Therefore, there exist at least 2Δ elements in the sequence sb which are no bigger than b_i and no smaller than $b_i e^{-\epsilon}$.

If $i \leq 0$, then $b_{i-\Delta}$ is the 2Δth largest element in the sequence sb which is no bigger than b_i and no smaller than $b_i e^{-\epsilon}$; and if $i \geq 0$, then $b_{i+\Delta}$ is the $(2\Delta - 1)$th largest element in the sequence sb which is no bigger than b_i and no smaller than $b_i e^{-\epsilon}$. Therefore, we have

$$\frac{b_i}{b_{i+\Delta}} \leq e^{\epsilon}, \forall i \geq 0,$$
$$\frac{b_i}{b_{i-\Delta}} \leq e^{\epsilon}, \forall i \leq 0.$$

This completes the proof of Lemma A.15.

□

Algorithm 3

$j_R^* \leftarrow i$
while there exists some j which appears before i in the sequence $\{0, 1, -1, 2, -2, \dots\}$ and $\pi(j) \in [\pi(j_R^*) + 1, \pi(j_R^*) + \Delta]$ **do**
  $j_R^* \leftarrow j$
end while

$j_L^* \leftarrow i$
while there exists some j which appears before i in the sequence $\{0, 1, -1, 2, -2, \dots\}$ and $\pi(j) \in [\pi(j_L^*) - \Delta, \pi(j_L^*) - 1]$ **do**
  $j_L^* \leftarrow j$
end while

Output j_R^* and j_L^*.

A.6.3 Step 2

Next we prove that we only need to consider symmetric probability mass functions which are monotonically decreasing when $i \geq 0$.

Define

$$\mathcal{SP}_{\text{sym}} \triangleq \{\mathcal{P} \in \mathcal{SP}_{\text{mono}} |\ \mathcal{P}(i) = \mathcal{P}(-i), \forall\, i \in \mathbb{Z}\}.$$

Lemma A.16.

$$V^* = \inf_{\mathcal{P} \in \mathcal{SP}_{sym}} \sum_{i=-\infty}^{+\infty} \mathcal{L}(i)\mathcal{P}(i).$$

Proof. The proof is essentially the same as the proof of Lemma A.3.

Given $\mathcal{P}_a \in \mathcal{SP}_{\text{mono}}$, define a new probability mass function $\mathcal{P}_b$ with

$$\mathcal{P}_b(i) \triangleq \frac{\mathcal{P}_a(i) + \mathcal{P}_a(-i)}{2}, \forall i \in \mathbb{Z}.$$

It is easy to see $\mathcal{P}_b$ is a valid probability mass function and symmetric. Since the cost function $\mathcal{L}(\cdot)$ is symmetric,

$$\sum_{i=-\infty}^{+\infty} \mathcal{L}(i)\mathcal{P}_a(i) = \sum_{i=-\infty}^{+\infty} \mathcal{L}(i)\mathcal{P}_b(i).$$

Next we show that $\mathcal{P}_b$ also satisfies the differential privacy constraint (2.16). For any

$i \in \mathbb{Z}$ and $|d| \leq \Delta$, since $\mathcal{P}_a(i) \leq e^\epsilon \mathcal{P}_a(i+d)$ and $\mathcal{P}_a(-i) \leq e^\epsilon \mathcal{P}_a(-i-d)$, we have

$$\begin{aligned}\mathcal{P}_b(i) &= \frac{\mathcal{P}_a(i) + \mathcal{P}_a(-i)}{2} \\ &\leq \frac{e^\epsilon \mathcal{P}_a(i+d) + e^\epsilon \mathcal{P}_a(-i-d)}{2} \\ &= e^\epsilon \mathcal{P}_b(i+d).\end{aligned}$$

Therefore, $\mathcal{P}_b$ satisfies (2.16).

Finally, for any $0 \leq i \leq j$,

$$\begin{aligned}\mathcal{P}_b(i) &= \frac{\mathcal{P}_a(i) + \mathcal{P}_a(-i)}{2} \\ &\geq \frac{\mathcal{P}_a(j) + \mathcal{P}_a(-j)}{2} \\ &= \mathcal{P}_b(j).\end{aligned}$$

So $\mathcal{P}_b \in \mathcal{SP}_{\text{mono}}$, and thus $\mathcal{P}_b \in \mathcal{SP}_{\text{sym}}$. We conclude

$$V^* = \inf_{\mathcal{P} \in \mathcal{SP}_{\text{sym}}} \sum_{i=-\infty}^{+\infty} \mathcal{L}(i)\mathcal{P}(i).$$

□

A.6.4 Step 3

Next we show that among all symmetric and monotonically decreasing (for $i \geq 0$) probability mass functions, we only need to consider those which are periodically and geometrically decaying.

More precisely, define

$$\mathcal{SP}_{\text{pd}} \triangleq \{\mathcal{P} \in \mathcal{SP}_{\text{sym}} | \frac{\mathcal{P}(i)}{\mathcal{P}(i+\Delta)} = e^\epsilon, \forall\, i \in \mathbb{N}\}.$$

Then

Lemma A.17.

$$V^* = \inf_{\mathcal{P} \in \mathcal{SP}_{pd}} V(\mathcal{P}).$$

Proof. Due to Lemma A.16, we only need to consider probability mass functions which are symmetric and monotonically decreasing for $i \geq 0$.

We first show that given $\mathcal{P}_a \in \mathcal{SP}_{\text{sym}}$, if $\frac{\mathcal{P}_a 0}{\mathcal{P}_a \Delta} < e^\epsilon$, then we can construct a probability mass function $\mathcal{P}_b \in \mathcal{SP}_{\text{sym}}$ such that $\frac{\mathcal{P}_b 0}{\mathcal{P}_b \Delta} = e^\epsilon$ and

$$V(\mathcal{P}_a) \geq V(\mathcal{P}_b).$$

Since $\mathcal{P}_a$ is symmetric,

$$V(\mathcal{P}_a) = \mathcal{L}(0)\mathcal{P}_a(0) + 2\sum_{i=1}^{+\infty} \mathcal{L}(i)\mathcal{P}_a(i).$$

Suppose $\frac{\mathcal{P}_a 0}{\mathcal{P}_a \Delta} < e^\epsilon$, then define a new symmetric probability mass function $\mathcal{P}_b$ with

$$\begin{aligned}
\mathcal{P}_b(0) &\triangleq (1+\delta)\mathcal{P}_a(0), \\
\mathcal{P}_b(i) &\triangleq (1-\delta')\mathcal{P}_a(i), \forall i \in \mathbb{Z}\backslash\{0\},
\end{aligned}$$

where

$$\begin{aligned}
\delta &= \frac{e^\epsilon \frac{\mathcal{P}_a(\Delta)}{\mathcal{P}_a(0)} - 1}{1 + e^\epsilon \frac{\mathcal{P}_a(\Delta)}{1-\mathcal{P}_a(0)}} > 0, \\
\delta' &= \frac{e^\epsilon \frac{\mathcal{P}_a(\Delta)}{\mathcal{P}_a(0)} - 1}{\frac{1}{\mathcal{P}_a(0)} + e^\epsilon \frac{\mathcal{P}_a(\Delta)}{\mathcal{P}_a(0)} - 1} > 0,
\end{aligned}$$

so that $\frac{\mathcal{P}_b(0)}{\mathcal{P}_b(\Delta)} = e^\epsilon$.

It is easy to see $\mathcal{P}_b \in \mathcal{SP}_{\text{sym}}$, and

$$\begin{aligned}
& V(\mathcal{P}_b) - V(\mathcal{P}_a) \\
=& \delta\mathcal{L}(0)\mathcal{P}_a(0) - 2\delta' \sum_{i=1}^{+\infty} \mathcal{L}(i)\mathcal{P}_a(i) \\
\leq& \delta\mathcal{L}(0)\mathcal{P}_a(0) - 2\delta' \sum_{i=1}^{+\infty} \mathcal{L}(0)\mathcal{P}_a(i) \\
\leq& \delta\mathcal{L}(0)\mathcal{P}_a(0) - \delta'\mathcal{L}(0)(1 - \mathcal{P}_a(0)) \\
=& 0.
\end{aligned}$$

Therefore, we only need to consider $\mathcal{P} \in \mathcal{SP}_{\text{sym}}$ satisfying $\frac{\mathcal{P}(0)}{\mathcal{P}(\Delta)} = e^\epsilon$.

By using the same argument as in the proof of Lemma A.10, one can conclude that we only need to consider $\mathcal{P} \in \mathcal{SP}_{\text{sym}}$ satisfying

$$\frac{\mathcal{P}(i)}{\mathcal{P}(i+\Delta)} = e^{\epsilon}, \forall i \in \mathbb{N}. \tag{A.33}$$

Therefore, $V^* = \inf_{\mathcal{P} \in \mathcal{SP}_{\text{pd}}} V(\mathcal{P})$.

□

Proof of Theorem 2.12. In the case that $\Delta = 1$, due to Lemma A.17, the symmetry property and (A.33) completely characterize the optimal noise probability mass function, which is the geometric mechanism. □

A.6.5 Step 4

Due to Lemma A.17, the optimal probability mass function $\mathcal{P}$ is completely characterized by $\mathcal{P}(0), \mathcal{P}(1), \ldots, \mathcal{P}(\Delta-1)$. Next we derive the properties of optimal probability mass function in the domain $\{0, 1, 2, \ldots, \Delta - 1\}$.

Since Lemma A.17 solves the case $\Delta = 1$, in the remaining of this section, we assume $\Delta \geq 2$.

Define

$$\begin{aligned}\mathcal{SP}_{\text{step}_\lambda} \triangleq \{\, \mathcal{P} \in \mathcal{SP}_{\text{pd}} \mid \exists\, k \in \{0, 1, \ldots, \Delta - 2\}, \mathcal{P}(i) = \mathcal{P}(0), \forall i \in \{0, 1, \ldots, k\},\\ \mathcal{P}(j) = \lambda \mathcal{P}(0), \forall j \in \{k+1, k+2, \ldots, \Delta - 1\}\}.\end{aligned}$$

Lemma A.18.

$$V^* = \inf_{\mathcal{P} \in \cup_{\lambda \in [e^{-\epsilon}, 1]} \mathcal{SP}_{step_\lambda}} V(\mathcal{P}).$$

Proof. If $\Delta = 2$, then for any $\mathcal{P} \in \mathcal{SP}_{\text{pd}}$, we can set $k = 0$, and $\mathcal{P} \in \mathcal{SP}_{\text{step}\, \frac{\mathcal{P}(\Delta-1)}{\mathcal{P}(0)}}$. Therefore, Lemma A.18 holds for $\Delta = 2$.

Assume $\Delta \geq 3$. First, we prove that we only need to consider probability mass function $\mathcal{P} \in \mathcal{SP}_{\text{pd}}$ such that there exists $k \in \{1, 2, \ldots, \Delta - 2\}$ with

$$\mathcal{P}(i) = \mathcal{P}(0), \forall i \in \{0, 1, \ldots, k-1\} \tag{A.34}$$

$$\mathcal{P}(j) = \mathcal{P}(\Delta - 1), \forall i \in \{k+1, k+2, \ldots, \Delta - 1\}. \tag{A.35}$$

More precisely, let $\mathcal{P}_a \in \mathcal{SP}_{\text{pd}}$, we can construct a probability mass function $\mathcal{P}_b \in \mathcal{SP}_{\text{pd}}$ such that there exists k satisfying (A.34) and (A.35), and $V(\mathcal{P}_b) \geq V(\mathcal{P}_a)$.

The proof technique is very similar to proof of Lemma A.13. Suppose there does not exists such k for $\mathcal{P}_a$, then let k_1 be the smallest integer in $\{1, 2, \ldots, \Delta - 1\}$ such that

$$\mathcal{P}_a(k_1) \neq \mathcal{P}_a(0),$$

and let k_2 be the biggest integer in $\{0, 1, \ldots, \Delta - 2\}$ such that

$$\mathcal{P}_a(k_2) \neq \mathcal{P}_a(\Delta - 1).$$

It is easy to see that $k_1 < k_2$, and $k_1 \neq 0$. Then we can increase $\mathcal{P}_a(k_1)$ and decrease $\mathcal{P}_a(k_2)$ simultaneously by the same amount to derive a new probability mass function $\mathcal{P}_b \in \mathcal{SP}_{\text{pd}}$ with smaller cost. Indeed, if

$$\mathcal{P}_a(0) - \mathcal{P}_a(k_1) \leq \mathcal{P}_a(k_2) - \mathcal{P}_a(\Delta - 1),$$

then consider a probability mass function $\mathcal{P}_b \in \mathcal{SP}_{\text{pd}}$ with

$$\begin{aligned}
\mathcal{P}_b(i) &= \mathcal{P}_a(0), \forall 0 \leq i \leq k_1, \\
\mathcal{P}_b(i) &= \mathcal{P}_a(i), \forall k_1 < i < k_2, \\
\mathcal{P}_b(k_2) &= \mathcal{P}_a(k_2) - (\mathcal{P}_a(0) - \mathcal{P}_a(k_1)), \\
\mathcal{P}_b(i) &= \mathcal{P}_a(i), \forall k_2 < i \leq \Delta - 1.
\end{aligned}$$

Define

$$\begin{aligned}
w_0 &\triangleq \mathcal{L}(0) + 2\sum_{k=1}^{\infty} \mathcal{L}(k\Delta)e^{-k\epsilon}, \\
w_i &\triangleq 2\sum_{k=0}^{\infty} \mathcal{L}(i + k\Delta)e^{-k\epsilon}, \forall i \in \{1, 2, \ldots, \Delta - 1\}.
\end{aligned}$$

Note that since $\mathcal{L}(\cdot)$ is a monotonically decreasing function when $i \geq 0$, we have $w_0 \leq w_1 \leq \cdots \leq w_{\Delta-1}$.

Then we can verify that $V(\mathcal{P}_b) \leq V(\mathcal{P}_a)$ via

$$V(\mathcal{P}_b) - V(\mathcal{P}_a)$$

$$
\begin{aligned}
&= \sum_{i=0}^{\Delta-1} \mathcal{P}_b(i) w_i - \sum_{i=0}^{\Delta-1} \mathcal{P}_a(i) w_i \\
&= (\mathcal{P}_a(0) - \mathcal{P}_a(k_1))(w_{k_1} - w_{k_2}) \\
&\leq 0.
\end{aligned}
$$

If

$$
\mathcal{P}_a(0) - \mathcal{P}_a(k_1) \geq \mathcal{P}_a(k_2) - \mathcal{P}_a(\Delta - 1),
$$

then we can define $\mathcal{P}_b \in \mathcal{SP}_{\text{pd}}$ by setting

$$
\begin{aligned}
\mathcal{P}_b(i) &= \mathcal{P}_a(0), \forall 0 \leq i < k_1, \\
\mathcal{P}_b(k_1) &= \mathcal{P}_a(k_1) + (\mathcal{P}_a(k_2) - \mathcal{P}_a(\Delta - 1)), \\
\mathcal{P}_b(i) &= \mathcal{P}_a(i), \forall k_1 < i < k_2, \\
\mathcal{P}_b(i) &= \mathcal{P}_a(\Delta - 1), \forall k_2 \leq i \leq \Delta - 1.
\end{aligned}
$$

And similarly, we have

$$
V(\mathcal{P}_b) - V(\mathcal{P}_a) = (\mathcal{P}_a(k_2) - \mathcal{P}_a(\Delta - 1))(w_{k_1} - w_{k_2}) \leq 0.
$$

Therefore, continue in this way, and finally we will obtain a probability mass function $\mathcal{P}_b \in \mathcal{SP}_{\text{pd}}$ such that there exists k to satisfy (A.34) and (A.35) and $V(\mathcal{P}_b) \leq V(\mathcal{P}_a)$.

From the above argument, we can see that in the optimal solution $\mathcal{P}^* \in \mathcal{SP}_{\text{pd}}$, the probability mass function can only take at most three distinct values for all $i \in \{0, 1, \ldots, \Delta - 1\}$, which are $\mathcal{P}^*(0), \mathcal{P}^*(k)$, and $\mathcal{P}^*(\Delta - 1)$. Next we show that indeed either $\mathcal{P}^*(k) = \mathcal{P}^*(0)$ and $\mathcal{P}^*(k) = \mathcal{P}^*(\Delta - 1)$, and this will complete the proof of Lemma A.18.

The optimal probability mass function $\mathcal{P} \in \mathcal{SP}_{\text{pd}}$ can be specified by four parameters $\mathcal{P}(0), \lambda \in [e^{-\epsilon}, 1]$, $k \in \{1, 2, \ldots, \Delta - 2\}$, and $\mathcal{P}(k)$. We will show that when k and λ are fixed, to minimize the cost, we have either $\mathcal{P}(k) = \mathcal{P}(0)$ or $\mathcal{P}(k) = \mathcal{P}(\Delta - 1) = \lambda \mathcal{P}(0)$.

Since $\sum_{i=-\infty}^{+\infty} \mathcal{P}(i) = 1$,

$$
2\frac{k\mathcal{P}(0) + \mathcal{P}(k) + (\Delta - k - 1)\lambda \mathcal{P}(0)}{1 - b} - \mathcal{P}(0) = 1,
$$

and thus $\mathcal{P}(k) = \frac{(1+\mathcal{P}(0))(1-b) - 2\mathcal{P}(0)k - 2\lambda\mathcal{P}(0)(\Delta - k - 1)}{2}$.

The cost for $\mathcal{P}$ is

$$
\begin{aligned}
&V(\mathcal{P}) \\
&= \mathcal{P}(0)\sum_{i=0}^{k-1} w_i + \mathcal{P}(\Delta-1)\sum_{i=k+1}^{\Delta-1} w_i + \mathcal{P}(k)w_k \\
&= \mathcal{P}(0)\sum_{i=0}^{k-1} w_i + \lambda\mathcal{P}(0)\sum_{i=k+1}^{\Delta-1} w_i + (\frac{(1+\mathcal{P}(0))(1-b) - 2\mathcal{P}(0)k - 2\lambda\mathcal{P}(0)(\Delta-k-1)}{2})w_k,
\end{aligned}
$$

which is a linear function of the parameter $\mathcal{P}(0)$.

Since $\mathcal{P}(k) \geq \lambda\mathcal{P}(0)$ and $\mathcal{P}(k) \leq \mathcal{P}(0)$, we have

$$
\begin{aligned}
1 &= 2\frac{k\mathcal{P}(0) + \mathcal{P}(k) + (\Delta-k-1)\lambda\mathcal{P}(0)}{1-b} - \mathcal{P}(0) \\
&\leq 2\frac{k\mathcal{P}(0) + \mathcal{P}(0) + (\Delta-k-1)\lambda\mathcal{P}(0)}{1-b} - \mathcal{P}(0), \\
1 &= 2\frac{k\mathcal{P}(0) + \mathcal{P}(k) + (\Delta-k-1)\lambda\mathcal{P}(0)}{1-b} - \mathcal{P}(0) \\
&\geq 2\frac{k\mathcal{P}(0) + \lambda\mathcal{P}(0) + (\Delta-k-1)\lambda\mathcal{P}(0)}{1-b} - \mathcal{P}(0),
\end{aligned}
$$

and thus the constraints on $\mathcal{P}(0)$ are

$$
\frac{1-b}{2k+2+2\lambda(\Delta-k-1)-1+b} \leq \mathcal{P}(0) \leq \frac{1-b}{2k+2\lambda(\Delta-k)-1+b}. \tag{A.36}
$$

Since $V(\mathcal{P})$ is a linear function of $\mathcal{P}(0)$, to minimize the cost $V(\mathcal{P})$, either $\mathcal{P}(0) = \frac{1-b}{2k+2+2\lambda(\Delta-k-1)-1+b}$ or $\mathcal{P}(0) = \frac{1-b}{2k+2\lambda(\Delta-k)-1+b}$, i.e., $\mathcal{P}(0)$ should take one of the two extreme points of (A.36). To get these two extreme points, we have either $\mathcal{P}(k) = \mathcal{P}(0)$ or $\mathcal{P}(k) = \lambda\mathcal{P}(0) = \mathcal{P}(\Delta-1)$.

Therefore, in the optimal probability mass function $\mathcal{P} \in \mathcal{SP}_{\text{pd}}$, there exists $k \leq \Delta-2$ such that

$$
\begin{aligned}
&\mathcal{P}(i) = \mathcal{P}(0), \forall i \in \{0,1,\ldots,k\} \\
&\mathcal{P}(i) = \mathcal{P}(\Delta-1), \forall i \in \{k+1, k+2, \ldots, \Delta-1\}.
\end{aligned}
$$

This completes the proof of Lemma A.18.

□

A.6.6 Step 5

In the last step, we prove that although $\lambda \in [e^{-\epsilon}, 1]$, in the optimal probability mass function, λ is either $e^{-\epsilon}$ or 1, and this will complete the proof of Theorem 2.13.

Proof. For fixed $k \in \{0, 1, \ldots, \Delta - 2\}$, consider $\mathcal{P} \in \mathcal{SP}_{\text{pd}}$ with

$$\begin{aligned}
&\mathcal{P}(i) = \mathcal{P}(0), \forall i \in \{0, 1, \ldots, k\}, \\
&\mathcal{P}(i) = \lambda \mathcal{P}(0), \forall i \in \{k+1, k+2, \ldots, \Delta - 1\}.
\end{aligned}$$

Since $\sum_{i=-\infty}^{+\infty} \mathcal{P}(i) = 1$,

$$2\frac{(k+1)\mathcal{P}(0) + (\Delta - k - 1)\lambda \mathcal{P}(0)}{1-b} - \mathcal{P}(0) = 1,$$

and thus

$$\mathcal{P}(0) = \frac{1-b}{2(k+1) + 2(\Delta - k - 1)\lambda - 1 + b}.$$

Hence, $\mathcal{P}$ is specified by only one parameter λ.

The cost of $\mathcal{P}$ is

$$\begin{aligned}
V(\mathcal{P}) &= \sum_{i=0}^{\Delta-1} \mathcal{P}(i) w_i \\
&= \mathcal{P}(0) \sum_{i=0}^{k} w_i + \lambda \mathcal{P}(0) \sum_{k+1}^{\Delta-1} w_i \\
&= \frac{(1-b)(\sum_{i=0}^{k} w_i + \lambda \sum_{i=k+1}^{\Delta-1} w_i)}{2(k+1) + 2(\Delta - k - 1)\lambda - 1 + b} \\
&= (1-b)(C_1 + \frac{C_2}{2(k+1) + 2(\Delta - k - 1)\lambda - 1 + b}),
\end{aligned}$$

where C_1 and C_2 are constant terms independent of λ. Therefore, to minimize $V(\mathcal{P})$ over $\lambda \in [e^{-\epsilon}, 1]$, λ should take the extreme points, either $e^{-\epsilon}$ or 1, depending on whether C_2 is negative or positive.

When $\lambda = 1$, then the probability mass function is uniquely determined, which is $\mathcal{P} \in \mathcal{SP}_{\text{pd}}$ with

$$\mathcal{P}(i) = \frac{1-b}{2\Delta - 1 + b}, \forall i \in \{0, 1, \ldots, \Delta - 1\},$$

which is exactly $\mathcal{P}_r$ defined in (2.17) with $r = \Delta$.

When $\lambda = e^{-\epsilon}$, the probability mass function is exactly $\mathcal{P}_r$ with $r = k + 1$.

Therefore, we conclude that

$$V^* = \min_{\{r \in \mathbb{N} | 1 \leq r \leq \Delta\}} \sum_{i=-\infty}^{+\infty} \mathcal{L}(i) \mathcal{P}_r(i).$$

□

APPENDIX B

PROOFS FOR CHAPTER 3

B.1 Proof of Theorem 3.1

In this section, we give detailed and rigorous proof of Theorem 3.1.

B.1.1 Outline of Proof

The key idea of the proof is to use a sequence of probability distributions with piecewise constant probability density functions to approximate any probability distribution satisfying the differential privacy constraint (3.6). The proof consists of 4 steps in total, and in each step we narrow down the set of probability distributions where the optimal probability distribution should lie:

- Step 1 proves that we only need to consider probability distributions which have symmetric and piecewise constant probability density functions.
- Step 2 proves that we only need to consider those symmetric and piecewise constant probability density functions which are monotonically decreasing.
- Step 3 proves that the optimal probability density function should periodically decay.
- Step 4 proves that the optimal probability density function is staircase-shaped in the multiple dimensional setting, and it concludes the proof of Theorem 3.1.

B.1.2 Step 1

Given $\mathcal{P} \in \mathcal{SP}$, define

$$V(\mathcal{P}) \triangleq \int \int \dots \int_{\mathbb{R}^d} \mathcal{L}(\mathbf{x}) \mathcal{P}(dx_1 dx_2 \dots dx_d).$$

Define

$$V^* \triangleq \inf_{\mathcal{P}\in\mathcal{SP}} V(\mathcal{P}).$$

Our goal is to prove that $V^* = \inf_{\gamma\in[0,1]} \int\int\ldots\int_{\mathbb{R}^d} \mathcal{L}(\mathbf{x}) f_\gamma(\mathbf{x}) dx_1 dx_2 \ldots dx_d$.

If $V^* = +\infty$, then due to the definition of V^*, we have

$$\inf_{\gamma\in[0,1]} \int\int\ldots\int_{\mathbb{R}^d} \mathcal{L}(\mathbf{x}) f_\gamma(\mathbf{x}) dx_1 dx_2 \ldots dx_d \geq V^* = +\infty,$$

and thus $\inf_{\gamma\in[0,1]} \int\int\ldots\int_{\mathbb{R}^d} \mathcal{L}(\mathbf{x}) f_\gamma(\mathbf{x}) dx_1 dx_2 \ldots dx_d = V^* = +\infty$. So we only need to consider the case $V^* < +\infty$, i.e., V^* is finite. Therefore, in the rest of the proof, we assume V^* is finite.

First we show that given any probability measure $\mathcal{P} \in \mathcal{SP}$, we can use a sequence of probability measures with multiple dimensionally piecewise constant probability density functions to approximate $\mathcal{P}$.

Given $i \in \mathbb{N}$ and $k \in \mathbb{N}$, define

$$A_i(k) = \{\mathbf{x} \in \mathbb{R}^d | k\frac{\Delta}{i} \leq \|\mathbf{x}\|_1 < (k+1)\frac{\Delta}{i}\} \subset \mathbb{R}^d.$$

It is easy to calculate the volumn of $A_i(k)$, which is

$$\mathrm{Vol}(A_i(k)) = \frac{2^d}{d!}\left((k+1)^d - k^d\right)\frac{\Delta^d}{i^d}.$$

.

Lemma B.1. *Given $\mathcal{P} \in \mathcal{SP}$ with $V(\mathcal{P}) < +\infty$, any positive integer $i \in \mathbb{N}$, define $\mathcal{P}_i$ as the probability distribution with probability density function $f_i(\mathbf{x})$ defined as*

$$f_i(\mathbf{x}) = a_i(k) \triangleq \frac{\mathcal{P}(A_i(k))}{Vol(A_i(k))}\mathbf{x} \in A_i(k) \text{ for } k \in \mathbb{N}. \tag{B.1}$$

Then $\mathcal{P}_i \in \mathcal{SP}$ and

$$\lim_{i\to+\infty} V(\mathcal{P}_i) = V(\mathcal{P}).$$

Before proving Lemma B.1, we prove an auxiliary lemma which shows that for probability mass function over $\mathbb{Z}^2$ satisfying the ϵ-differential privacy constraint, we can construct a new probability mass function by averaging the old probability mass function over each ℓ^1-ball and the new probability mass function still satisfies the ϵ-differential privacy constraint.

Lemma B.2. *For any given probability mass function $\mathcal{P}$ defined over the set $\mathbb{Z}^2$ satisfying that*

$$\mathcal{P}(i_1, j_1) \leq e^{\epsilon}\mathcal{P}(i_2, j_2), \forall |i_1 - i_2| + |j_1 - j_2| \leq \Delta, \tag{B.2}$$

define the probability mass function $\tilde{\mathcal{P}}$ via

$$\tilde{\mathcal{P}}(i, j) = \begin{cases} \mathcal{P}(0, 0) & (i, j) = (0, 0) \\ p_{|i|+|j|} & (i, j) \neq (0, 0) \end{cases}$$

where $p_k \triangleq \frac{\sum_{(i',j') \in \mathbb{Z}^2 : |i'|+|j'|=k} \mathcal{P}(i',j')}{4k}, \forall k \geq 1$.

Then $\tilde{\mathcal{P}}$ is also a probability mass function satisfying the differential privacy constraint, i.e.,

$$\tilde{\mathcal{P}}(i_1, j_1) \leq e^{\epsilon}\tilde{\mathcal{P}}(i_2, j_2), \forall |i_1 - i_2| + |j_1 - j_2| \leq \Delta. \tag{B.3}$$

Proof. Due to the way we define $\tilde{\mathcal{P}}$, we have

$$\sum_{(i,j) \in \mathbb{Z}^2} \tilde{\mathcal{P}}(i, j) = \sum_{(i,j) \in \mathbb{Z}^2} \mathcal{P}(i, j) = 1,$$

and thus $\tilde{\mathcal{P}}$ is a valid probability mass function defined over $\mathbb{Z}^2$.

Next we prove that $\tilde{\mathcal{P}}$ satisfies (B.3). To simplify notation, define $p_0 \triangleq \mathcal{P}(0, 0)$. Then we only need to prove that for any $k_1, k_2 \in \mathbb{N}$ such that $|k_1 - k_2| \leq \Delta$, we have

$$p_{k_1} \leq e^{\epsilon} p_{k_2}.$$

Due to the symmetry property, without loss of generality, we can assume $k_1 < k_2$.

The easiest case is $k_1 = 0$. When $k_1 = 0$, we have $k_2 \leq \Delta$ and

$$\mathcal{P}(0, 0) \leq e^{\epsilon}\mathcal{P}(i, j), \forall |i| + |j| = k_2. \tag{B.4}$$

The number of distinct pairs (i, j) satisfying $|i| + |j| = k$ is $4k$ for $k \geq 1$. Sum up all inequalities in (B.4), and we get

$$\begin{aligned} 4k_2\mathcal{P}(0, 0) &\leq e^{\epsilon} \sum_{(i,j) \in \mathbb{Z}^2 : |i|+|j|=k_2} \mathcal{P}(i, j) \\ \Leftrightarrow \mathcal{P}(0, 0) &\leq e^{\epsilon} \frac{\sum_{(i,j) \in \mathbb{Z}^2 : |i|+|j|=k_2} \mathcal{P}(i, j)}{4k_2} \end{aligned}$$

$$\Leftrightarrow p_0 \leq e^{\epsilon} p_{k_2}.$$

For general $0 < k_1 < k_2$, let $\Delta' \triangleq k_2 - k_1 \leq \Delta$. Define B_k via

$$B_k \triangleq \{(i,j) \in \mathbb{Z}^2 || i| + |j| = k\}, \forall k \in \mathbb{N}.$$

Then the differential privacy constraint (B.2) implies that

$$\mathcal{P}(i_1, j_1) \leq e^{\epsilon} \mathcal{P}(i_2, j_2), \forall (i_1, j_1) \in B_{k_1}, (i_2, j_2) \in B_{k_2}, |i_1 - i_2| + |j_1 - j_2| = \Delta'. \tag{B.5}$$

The set of points in B_k forms a rectangle, which has 4 corner points and $4(k-1)$ interior points on the edges. For each corner point in B_{k_1}, which appears in the left side of (B.5), there are $(2\Delta' + 1)$ points in B_{k_2} close to it with an ℓ^1 distance of Δ'. And for each interior point in B_{k_1}, there are $(\Delta' + 1)$ points in B_{k_2} close to it with an ℓ^1 distance of Δ'. Therefore, there are in total $4(2\Delta' + 1) + 4(k_1 - 1)(\Delta' + 1)$ distinct inequalities in (B.5).

If we can find certain nonnegative coefficients such that multiplying each inequality in (B.5) by these nonnegative coefficients and summing them up gives us

$$\frac{\sum_{(i',j') \in \mathbb{Z}^2 : |i'| + |j'| = k_1} \mathcal{P}(i', j')}{4k_1} \leq e^{\epsilon} \frac{\sum_{(i',j') \in \mathbb{Z}^2 : |i'| + |j'| = k_2} \mathcal{P}(i', j')}{4k_2},$$

then (B.3) holds. Therefore, our goal is to find the "right" coefficients associated with each inequality in (B.5). We formulate it as a matrix filling-in problem in which we need to choose nonnegative coefficients for certain entries in a matrix such that the sum of each row is $\frac{k_1 + \Delta'}{k_1}$, and the sum of each column is 1.

More precisely, label the $4k_1$ points in B_{k_1} by $\{I_1, I_2, I_3, \ldots, I_{4k_1}\}$, where we label the topmost point by 1 and sequentially label other points clockwise. Similarly, we label the $4k_2$ points in B_{k_2} by $\{O_1, O_2, O_3, \ldots, O_{4k_2}\}$, where we label the topmost point by 1 and sequentially label other points clockwise.

Consider the following $4k_1$ by $4k_2$ matrix M, where each row corresponds to the point in B_{k_1} and each column corresponds to the point in B_{k_2}, and the entry M_{ij} in the ith row and jth column is the coefficient corresponds to inequality involved with the points I_i and O_j. If there is no inequality associated with the points I_i and O_j, then $M_{ij} = 0$.

In the case $k_1 = 2$ and $\Delta' = 3$, the zeros/nonzeros pattern of M has the following form:

$$\begin{pmatrix} x & x & x & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & x & x \\ 0 & x & x & x & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & x & x & x & x & x & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & x & x & x & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & x & x & x & x & x & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & x & x & x & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & x & x & x & x & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & x & x \end{pmatrix},$$

where x denotes an entry which can take any nonnegative coefficient.

For general k_1 and k_2, the pattern of M is that the first, (k_1+1)th, $(2k_1+1)$th and $(3k_1+1)$th rows can have $2\Delta'+1$ nonzero entries, and all other rows can have $\Delta'+1$ nonzero entries.

We want to show that

$$\frac{\sum_{(i',j')\in\mathbb{Z}^2:|i'|+|j'|=k_1}\mathcal{P}(i',j')}{4k_1} \leq e^{\epsilon}\frac{\sum_{(i',j')\in\mathbb{Z}^2:|i'|+|j'|=k_2}\mathcal{P}(i',j')}{4k_2},$$

or equivalently,

$$(1+\frac{\Delta'}{k_1})\sum_{(i',j')\in\mathbb{Z}^2:|i'|+|j'|=k_1}\mathcal{P}(i',j') \leq e^{\epsilon}\sum_{(i',j')\in\mathbb{Z}^2:|i'|+|j'|=k_2}\mathcal{P}(i',j').$$

Therefore, our goal is to find nonnegative coefficients to substitute each x in the matrix such that the sum of each column is 1 and the sum of each column is $(1+\frac{\Delta'}{k_1})$. We will give explicit formulas on how to choose the coefficients.

The case $k_1 = 1$ is trivial. Indeed, one can set all diagonal entries to be 1, and set all other nonzero entries to be $\frac{1}{2}$. Therefore, we can assume $k_1 > 1$.

Consider two different cases: $k_1 \leq \Delta'$ and $k_1 \geq \Delta'+1$.

We first consider the case $k_1 \leq \Delta'$. Due to the periodic patterns in M, we only need to consider rows from 1 to k_1+1. Set all entries to be zero except that we set

$$\begin{aligned} M_{11} = M_{22} = \cdots = M_{k_1k_1} &= 1, \\ M_{2,\Delta'+2} = M_{3,\Delta'+3} = \cdots = M_{k_1+1,k_1+\Delta'+1} &= 1 \\ M_{1,j} = \frac{\Delta'}{2k_1(\Delta'-k_1+1)}&, j \in [k_1+1,\Delta'+1]\cup[4k_1-\Delta'+1,4k_1] \end{aligned}$$

$$M_{k_1+1,j} = \frac{\Delta'}{2k_1(\Delta' - k_1 + 1)}, j \in [k_1 + 1, \Delta' + 1] \cup [2k_1 + 1 + \Delta', k_1 + 1 + 2\Delta']$$
$$M_{i,j} = \frac{1 - \frac{\Delta'}{k_1(\Delta' - k_1 + 1)}}{k_1 - 1}.$$

It is straightforward to verify that the above matrix M satisfies the properties that the sum of each column is 1 and the sum of each row is $(1 + \frac{\Delta'}{k_1})$. Therefore, we have

$$p_{k_1} \leq e^{\epsilon} p_{k_2}, \forall 0 < k_1 < k_2, k_1 \leq k_2 - k_1 \leq \Delta.$$

Next we solve the case $k_1 \geq \Delta' + 1$. Again due to the periodic patterns in M, we only need to consider the nonzero entries in rows from 1 to k_1+1. We use the following procedures to construct M:

1. For the first row, set $M_{11} = 1$ and set all other $2\Delta'$ nonzero entries to be $\frac{1}{2k_1}$.

2. For the second row, M_{22} is uniquely determined to be $1 - \frac{1}{2k_1}$. Set the next $\Delta' - 1$ nonzero entries in the second row to be $\frac{1}{k_1}$, i.e., $M_{2j} = \frac{1}{k_1}$ for $j \in [3, \Delta' + 1]$. The last nonzero entry $M_{2,\Delta'+2}$ is uniquely determined to be

$$(1 + \frac{\Delta'}{k_1}) - (1 - \frac{1}{2k_1}) - \frac{\Delta' - 1}{k_1} = \frac{3}{2k_1}.$$

3. For the third row, the first nonzero entry M_{33} is uniquely determined to be $1 - \frac{1}{2k_1} - \frac{1}{k_1} = 1 - \frac{3}{2k_1}$. Set the next $\Delta' - 1$ nonzero entries to be $\frac{1}{k_1}$, i.e., $M_{3j} = \frac{1}{k_1}$ for $j \in [4, \Delta' + 2]$. The last nonzero entry $M_{3,\Delta'+3}$ is uniquely determined to be

$$(1 + \frac{\Delta'}{k_1}) - (1 - \frac{3}{2k_1}) - \frac{\Delta' - 1}{k_1} = \frac{5}{2k_1}.$$

4. In general, for the ith row ($i \in [2, k_1 - 1]$), the first nonzero entry M_{ii} is set to be $M_{ii} = 1 - \frac{2i-3}{2k_1}$, and the next $\Delta' - 1$ nonzero entries are $\frac{1}{k_1}$, and the last nonzero entry $M_{i,i+\Delta'} = \frac{2i-1}{2k_1}$.

5. For $(k_1 + 1)$th row, by symmetry, we set $M_{k_1+1,k_1+1} = 1$ and set other $2\Delta'$ nonzero entries to be $\frac{1}{2k_1}$.

6. The nonzero entries in the k_1th row are uniquely determined. Indeed, we have

$$M_{k_1,k_1} = 1 - \frac{2k_1 - 3}{2k_1},$$

$$M_{k_1,k_1+\Delta'} = 1 - \frac{1}{2k_1},$$
$$M_{k_1,k_1+j} = \frac{1}{k_1}, j \in [2, \Delta' - 1].$$

It is straightforward to verify that each entry in M is nonnegative and M satisfies the properties that the sum of each column is 1 and the sum of each row is $(1 + \frac{\Delta'}{k_1})$. Therefore, we have

$$p_{k_1} \leq e^{\epsilon} p_{k_2}, \forall 0 < k_1 < k_2, k_1 \geq \Delta' + 1 = k_2 - k_1 + 1.$$

Therefore, for all $k_1, k_2 \in \mathbb{N}$ such that $|k_2 - k_1| \leq \Delta$, we have

$$p_{k_1} \leq e^{\epsilon} p_{k_2}.$$

This completes the proof of Lemma B.2.

□

Proof of Lemma B.1. First we prove that $\mathcal{P}_i \in \mathcal{SP}$, i.e., $\mathcal{P}_i$ satisfies the differential privacy constraint (3.6).

By the definition of $f_i(\mathbf{x})$, $f_i(\mathbf{x})$ is a nonnegative function, and

$$\begin{aligned}
&\int \int \dots \int_{\mathbb{R}^d} f_i(\mathbf{x}) dx_1 dx_2 \dots dx_d \\
=&\sum_{k=0}^{+\infty} a_i(k) \text{Vol}(A_i(k)) \\
=&\sum_{k=0}^{+\infty} \mathcal{P}(A_i(k)) \\
=&\mathcal{P}(\mathbb{R}^d) = 1.
\end{aligned}$$

So $\mathcal{P}_i$ is a valid probability distribution.

Next we show that $f_i(\mathbf{x})$ satisfies the differential privacy constraint. For fixed i, on the $x_1 - x_2$ plane, we can use the lines $x_2 = x_1 + \frac{k}{i}\Delta$ and $x_2 = -x_1 + \frac{k}{i}\Delta$ for all $k \in \mathbb{Z}$ to divide each $A_i(k)$ into distinct squares with the same size (each $A_i(k)$ will be divided into $8k + 4$ squares). By taking the average of the probability density function over each square, we reduce the probability density function to a discrete probability mass function over $\mathbb{Z}^2$ satisfying ϵ-differential privacy constraint. Then apply Lemma B.2, and we have

$$a_i(k_1) \leq e^{\epsilon} a_i(k_2), \forall k_1, k_2 \in \mathbb{N} \text{ with } |k_1 - k_2| \leq i.$$

Given $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ such that $\|\mathbf{x} - \mathbf{y}\|_1 \leq \Delta$, let k_1, k_2 be the integers such that

$$\mathbf{x} \in A_i(k_1),$$
$$\mathbf{y} \in A_i(k_2).$$

Then $|k_1 - k_2| \leq i$. Therefore,

$$f_i(\mathbf{x}) \leq e^{\epsilon} f_i(\mathbf{y}),$$

which implies that the probability distribution $\mathcal{P}_i$ satisfies the differential privacy constraint (3.6).

Therefore, for any integer $i \geq 1$, $\mathcal{P}_i \in \mathcal{SP}$.

Next we show that

$$\lim_{i \to +\infty} V(\mathcal{P}_i) = V(\mathcal{P}).$$

To simplify notation, we use $d\mathbf{x}$ to denote $dx_1 dx_2 \dots dx_d$.

Given $\delta > 0$, since $V(\mathcal{P})$ is finite, there exists $T^* = m\Delta > 1$ for some $m \in \mathbb{N}$ such that

$$\int \int \dots \int_{\{\mathbf{x} \in \mathbb{R}^d | \|\mathbf{x}\|_1 \geq T^*\}} \mathcal{L}(\mathbf{x}) \mathcal{P}(d\mathbf{x}) < \frac{\delta}{2}.$$

For each $A_i(k)$ we have

$$\begin{aligned}
\int \int \dots \int_{A_i(k)} \mathcal{L}(\mathbf{x}) \mathcal{P}_i(d\mathbf{x}) &= \int \int \dots \int_{A_i(k)} \|\mathbf{x}\|_1 \mathcal{P}_i(d\mathbf{x}) \\
&\leq \mathcal{P}_i(A_i(k))(k+1)\frac{\Delta}{i} \\
&= \mathcal{P}(A_i(k))(k+1)\frac{\Delta}{i} \\
&\leq 2\mathcal{P}(A_i(k))k\frac{\Delta}{i} \\
&\leq 2 \int \int \dots \int_{A_i(k)} \mathcal{L}(\mathbf{x}) \mathcal{P}(d\mathbf{x}).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\int \int \dots \int_{\{\mathbf{x} \in \mathbb{R}^d | \|\mathbf{x}\|_1 \geq T^*\}} \mathcal{L}(\mathbf{x}) \mathcal{P}_i(d\mathbf{x}) &\leq 2 \int \int \dots \int_{\{\mathbf{x} \in \mathbb{R}^d | \|\mathbf{x}\|_1 \geq T^*\}} \mathcal{L}(\mathbf{x}) \mathcal{P}(d\mathbf{x}) \\
&\leq 2\frac{\delta}{2} = \delta.
\end{aligned}$$

$\mathcal{L}(\mathbf{x})$ is a bounded function when $\|\mathbf{x}\|_1 \leq T^*$, and thus by the definition of Riemann-Stieltjes integral, we have

$$\lim_{i\to\infty} \int\int \dots \int_{\{\mathbf{x}\in\mathbb{R}^d | \|\mathbf{x}\|_1 < T^*\}} \mathcal{L}(\mathbf{x}) \mathcal{P}_i(d\mathbf{x}) = \int\int \dots \int_{\{\mathbf{x}\in\mathbb{R}^d | \|\mathbf{x}\|_1 < T^*\}} \mathcal{L}(\mathbf{x}) \mathcal{P}(d\mathbf{x}).$$

So there exists a sufficiently large integer i^* such that for all $i \geq i^*$

$$\left| \int\int \dots \int_{\{\mathbf{x}\in\mathbb{R}^d | \|\mathbf{x}\|_1 < T^*\}} \mathcal{L}(\mathbf{x}) \mathcal{P}_i(d\mathbf{x}) - \int\int \dots \int_{\{\mathbf{x}\in\mathbb{R}^d | \|\mathbf{x}\|_1 < T^*\}} \mathcal{L}(\mathbf{x}) \mathcal{P}(d\mathbf{x}) \right| \leq \delta.$$

Hence, for all $i \geq i^*$

$$\begin{aligned}
& |V(\mathcal{P}_i) - V(\mathcal{P})| \\
= & \left| \int_{\mathbb{R}^d} \mathcal{L}(\mathbf{x}) \mathcal{P}_i(d\mathbf{x}) - \int_{\mathbb{R}^d} \mathcal{L}(\mathbf{x}) \mathcal{P}(d\mathbf{x}) \right| \\
= & | \int_{\{\mathbf{x}\in\mathbb{R}^d | \|\mathbf{x}\|_1 < T^*\}} \mathcal{L}(\mathbf{x}) \mathcal{P}_i(d\mathbf{x}) - \int_{\{\mathbf{x}\in\mathbb{R}^d | \|\mathbf{x}\|_1 < T^*\}} \mathcal{L}(\mathbf{x}) \mathcal{P}(d\mathbf{x}) \\
& + \int_{\{\mathbf{x}\in\mathbb{R}^d | \|\mathbf{x}\|_1 \geq T^*\}} \mathcal{L}(\mathbf{x}) \mathcal{P}_i(d\mathbf{x}) - \int_{\{\mathbf{x}\in\mathbb{R}^d | \|\mathbf{x}\|_1 \geq T^*\}} \mathcal{L}(\mathbf{x}) \mathcal{P}(d\mathbf{x}) | \\
\leq & \left| \int_{\{\mathbf{x}\in\mathbb{R}^d | \|\mathbf{x}\|_1 < T^*\}} \mathcal{L}(\mathbf{x}) \mathcal{P}_i(d\mathbf{x}) - \int_{\{\mathbf{x}\in\mathbb{R}^d | \|\mathbf{x}\|_1 < T^*\}} \mathcal{L}(\mathbf{x}) \mathcal{P}(d\mathbf{x}) \right| \\
& + \int_{\{\mathbf{x}\in\mathbb{R}^d | \|\mathbf{x}\|_1 \geq T^*\}} \mathcal{L}(\mathbf{x}) \mathcal{P}_i(d\mathbf{x}) + \int_{\{\mathbf{x}\in\mathbb{R}^d | \|\mathbf{x}\|_1 \geq T^*\}} \mathcal{L}(\mathbf{x}) \mathcal{P}(d\mathbf{x}) \\
\leq & (\delta + \delta + \frac{\delta}{2}) \\
\leq & \frac{5}{2}\delta.
\end{aligned}$$

Therefore,

$$\lim_{i\to+\infty} V(\mathcal{P}_i) = V(\mathcal{P}).$$

□

Define $\mathcal{SP}_{i,\text{sym}} \triangleq \{\mathcal{P}_i | \mathcal{P} \in \mathcal{SP}\}$ for $i \geq 1$, i.e., $\mathcal{SP}_{i,\text{sym}}$ is the set of probability distributions satisfying differential privacy constraint (3.6) and having symmetric and piecewise constant (over $A_i(k)$ $\forall k \in \mathbb{N}$) probability density functions.

Due to Lemma B.1,

Lemma B.3.

$$V^* = \inf_{\mathcal{P} \in \cup_{i=1}^{\infty} \mathcal{SP}_{i,sym}} V(\mathcal{P}).$$

Therefore, to characterize V^*, we only need to study probability distributions with symmetric and piecewise constant probability density functions.

B.1.3 Step 2

Given $\mathcal{P} \in \mathcal{P}_{\text{sym}}$, we call $\{a_i(0), a_i(1), a_i(2), \dots\}$ the density sequence of $\mathcal{P}_i \in \mathcal{SP}_{i,\text{sym}}$, where $a_i(k)$ is defined in (B.1) $\forall k \in \mathbb{N}$.

Next we show that indeed we only need to consider those probability distributions with symmetric and piecewise constant probability density functions the density sequences of which are *monotonically decreasing*.

Define

$$\mathcal{SP}_{i,\text{md}} \triangleq \{\mathcal{P} | \mathcal{P} \in \mathcal{SP}_{i,\text{sym}}, \text{ and the density sequence of } \mathcal{P} \text{ is monotonically decreasing}\}.$$

Then

Lemma B.4.

$$V^* = \inf_{\mathcal{P} \in \cup_{i=1}^{\infty} \mathcal{SP}_{i,md}} V(\mathcal{P}).$$

Proof. We first show that among $\mathcal{SP}_{i,\text{sym}}$, to minimize the cost we only need to consider these probability distributions with density sequences $\{a_0, a_1, a_2, \dots\}$ satisfying that $a_0 \geq a_1$. Indeed, given $\mathcal{P}_a \in \mathcal{SP}_{i,\text{sym}}$ with density sequence $\{a_0, a_1, a_2, \dots\}$ such that $a_0 < a_1$, there exists $\mathcal{P}_b \in \mathcal{SP}_{i,\text{sym}}$ with density sequence $\{b_0, b_1, b_2, \dots\}$ such that $b_0 \geq b_1$ and

$$V(\mathcal{P}_b) \leq V(\mathcal{P}_a).$$

Consider the probability distribution $\mathcal{P}_b \in \mathcal{SP}_{i,\text{sym}}$ with density sequence $\{b_0, b_1, b_2, , \dots\}$ defined as

$$\begin{aligned} b_0 &= (1+\delta)a_0, \\ b_k &= (1-\delta')a_k, \forall k \geq 1, \end{aligned}$$

where we choose $\delta > 0$ and $0 < \delta' < 1$ such that

$$b_0 = b_1, \tag{B.6}$$

$$\sum_{k=0}^{+\infty} b_k \mathrm{Vol}(A_i(k)) = \sum_{k=0}^{+\infty} a_k \mathrm{Vol}(A_i(k)) = 1. \tag{B.7}$$

Equation (B.7) makes $\mathcal{P}_b$ be a valid probability distribution. One can easily solve (B.6) and (B.7), and write down the explicit expression for δ, δ'. The density sequence $\{b_0, b_1, b_2, \dots\}$ satisfies $b_0 \geq b_1$ (indeed, we have $b_0 = b_1$), and it is easy to verify that it satisfies the differential privacy constraint, i.e.,

$$b_{k_1} \leq e^{\epsilon} b_{k_2}, \forall k_1, k_2 \in \mathbb{N} \text{ with } |k_1 - k_2| \leq i.$$

Note that $\mathcal{C}(\|\mathbf{x}\|_1)$ is a monotonically increasing function of $\|\mathbf{x}\|_1$, and compared to $\mathcal{P}_a$, $\mathcal{P}_b$ moves some probability of $\mathcal{SP}_{i,\mathrm{md}}$ from the (higher cost) area $\{\mathbf{x} | \|bx\| \geq \frac{\Delta}{i}\}$ to the (lower cost) area $\{\mathbf{x} | \|bx\| \leq \frac{\Delta}{i}\}$, and thus we have

$$V(\mathcal{P}_b) \leq V(\mathcal{P}_a).$$

Therefore, among $\mathcal{SP}_{i,\mathrm{sym}}$, to minimize the cost we only need to consider these probability distributions with density sequences $\{a_1, a_2, a_3, \dots\}$ satisfying that $a_0 \geq a_1$.

Next we show that among $\mathcal{SP}_{i,\mathrm{sym}}$ with density sequences $\{a_1, a_2, a_3, \dots\}$ satisfying $a_0 \geq a_1$, to minimize the cost we only need to consider these probability distributions with density sequences also satisfying that $a_1 \geq a_2$.

Given $\mathcal{P}_a \in \mathcal{SP}_{i,\mathrm{sym}}$ with density sequence $\{a_1, a_2, a_3, \dots\}$ such that $a_0 \geq a_1$ and $a_1 < a_2$, there exists $\mathcal{P}_b \in \mathcal{SP}_{i,\mathrm{sym}}$ with density sequence $\{b_1, b_2, b_3, \dots\}$ such that $b_0 \geq b_1$ and

$$b_1 \geq b_2.$$

If $i \leq 2$, we can construct $\mathcal{P}_b$ by scaling up a_0, a_1 and scale down a_k for all $k \geq 2$. More precisely, define $\mathcal{P}_b$ with density sequence $\{b_0, b_1, b_2, \dots\}$ via

$$b_k = (1 + \delta) a_k, k \leq 1,$$
$$b_k = (1 - \delta') a_k, k \geq 2,$$

for some $\delta > 0$ and $0 < \delta' < 1$ such that

$$b_2 = b_1,$$

$$\sum_{k=0}^{+\infty} b_k \mathrm{Vol}(A_i(k)) = \sum_{k=0}^{+\infty} a_k \mathrm{Vol}(A_i(k)) = 1.$$

So we have $b_0 \geq b_1 \geq b_2$. It is easy to check that $\mathcal{P}_b$ satisfies the differential privacy constraint, and $V(\mathcal{P}_b) \leq V(\mathcal{P}_a)$ using the fact that $\mathcal{C}(\|\mathbf{x}\|_1)$ is a monotonically decreasing function in terms of $\|\mathbf{x}\|_1$.

If $i \geq 3$, then without loss of generality we can assume $a_2 \leq a_0$. Indeed, if $a_2 > a_0$, we can scale up a_0, a_1 and scale down a_k for all $k \geq 2$ to make $a_2 = a_0$, and this operation will preserve the differential privacy constraint and decrease the cost. Note that in this case we cannot use the same scaling operation to make $a_2 \leq a_0$, because it is possible that after the scaling operation $\frac{a_0}{a_k} > e^\epsilon$ for some $3 \leq k \geq i$, which violates the differential privacy constraint. Hence, we can assume $a_0 \geq a_2 > a_1$. Let $a_{k'}$ be the largest value in $\{a_3, \ldots, a_{2+i}\}$. If $\frac{a_{k'}}{a_2} < e^\epsilon$, we can scale up a_1 and scale down a_2 until $a_1 = a_2$ or $\frac{a_{k'}}{a_2} = e^\epsilon$. It is easy to see this scaling operation will preserve differential privacy and decrease the cost. If after this scaling operation we have $a_2 = a_1$, then we are done. Suppose a_1 is still bigger than a_2. Then a_2 is the smallest element in $\{a_2, a_3, \ldots, a_{2+i}\}$. Therefore, we have $\max_{2 \leq k \leq i} \frac{a_0}{a_k} = \frac{a_0}{a_2}$. Then we can scale up a_0, a_1 and scale down a_k for $k \geq 2$ until $a_1 = a_2$. This operation will preserve the differential privacy constraint and decrease the cost. If we call the final probability distribution we obtained $\mathcal{P}_b$, we have $\mathcal{P}_b \in \mathcal{SP}_{i,\mathrm{sym}}$, and the density sequence satisfying $b_0 \geq b_1 \geq b_2$ (indeed, $b_1 = b_2$), and $V(\mathcal{P}_b) \leq V(\mathcal{P}_a)$.

By induction, we can show that among all probability distributions in $\mathcal{SP}_{i,\mathrm{sym}}$, to minimize the cost we only need to consider probability distributions with monotonically decreasing density sequence.

Suppose among $\mathcal{SP}_{i,\mathrm{sym}}$ to minimize the cost we only need to consider probability distribution with density sequence $\{a_0, a_1, a_2, \ldots\}$ satisfying $a_0 \geq a_1 \geq a_2 \geq \cdots \geq a_n$. Then we can show that among $\mathcal{SP}_{i,\mathrm{sym}}$ to minimize the cost we only need to consider probability distribution with density sequence $\{a_0, a_1, a_2, \ldots\}$ satisfying $a_0 \geq a_1 \geq a_2 \geq \cdots \geq a_n \geq a_{n+1}$.

Indeed, given $\mathcal{P}_a \in \mathcal{SP}_{i,\mathrm{sym}}$ with density sequence $\{a_0, a_1, a_2, \ldots\}$ satisfying $a_0 \geq a_1 \geq a_2 \geq \cdots \geq a_n$, we can construct $\mathcal{P}_b \in \mathcal{SP}_{i,\mathrm{sym}}$ with density sequence $\{b_0, b_1, b_2, \ldots\}$ satisfying

$$b_0 \geq b_1 \geq b_2 \geq \cdots \geq b_n \geq b_{n+1},$$

and

$$V(\mathcal{P}_b) \leq V(\mathcal{P}_a).$$

If $a_{n+1} \leq a_n$, then we can choose $\mathcal{P}_b = \mathcal{P}_a$.

Suppose $a_{n+1} > a_n$. Without loss of generality, we can assume

$$a_{n+1} \leq a_k, \text{ for } k \leq n+2-i. \tag{B.8}$$

If $a_{n+1} > a_{n+2-i}$, then we can scale up $\{a_0, a_1, \dots, a_n\}$ and scale down $\{a_{n+1}, a_{n+2}, \dots\}$ until $a_{n+1} = a_k$. It is easy to verify that this scaling operation will preserve the differential privacy constraint and decrease the cost.

Let k^* be the smallest integer such that $a_{k^*} < a_{n+1}$. Note that by (B.8) we have $n+3-i \leq k^* \leq n$. Let a_j be the biggest element in $\{a_{n+2}, a_{n+3}, \dots, a_{n+1+i}\}$. Due to the differential privacy constraint, we have $\frac{a_j}{a_{n+1}} \leq e^\epsilon$. Then we can scale up a_{k^*} and scale down a_{n+1} until $a_{k^*} = a_{n+1}$ or $\frac{a_j}{a_{n+1}} = e^\epsilon$. This operation will preserve the differential privacy constraint and decrease the cost. If after this scaling operation a_{k^*} is still bigger than a_{n+1}, then we can scale up $\{a_0, a_1, \dots, a_n\}$ and scale down $\{a_{n+1}, a_{n+2}, \dots\}$ until $a_{k^*} = a_{n+1}$. Due to the fact that a_{n+1} is the smallest element in $\{a_{n+1}, a_{n+2}, \dots, a_{n+1+i}\}$, this scaling operation will preserve the differential privacy constraint and decrease the cost. Therefore, we will have $a_{n+1} \leq a_{k^*}$.

Repeat the above steps for each $k \in k^*+1, k^*+2, \dots, n$ such that $a_k < a_{n+1}$. If we call the final probability distribution we obtained $\mathcal{P}_b$, we have $\mathcal{P}_b \in \mathcal{SP}_{i,\text{sym}}$, and the density sequence satisfying

$$b_0 \geq b_1 \geq b_2 \geq \cdots \geq b_n,$$

and $V(\mathcal{P}_b) \leq V(\mathcal{P}_a)$.

Hence, among $\mathcal{SP}_{i,\text{sym}}$ to minimize the cost we only need to consider probability distribution with density sequence $\{a_0, a_1, a_2, \dots\}$ satisfying $a_0 \geq a_1 \geq a_2 \geq \cdots \geq a_n \geq a_{n+1}$.

Therefore, among all probability distributions in $\mathcal{SP}_{i,\text{sym}}$, to minimize the cost we only need to consider probability distributions with monotonically decreasing density sequence.

We conclude that

$$V^* = \inf_{\mathcal{P} \in \cup_{i=1}^{\infty} \mathcal{SP}_{i,\text{md}}} V(\mathcal{P}).$$

This completes the proof of Lemma B.4. □

B.1.4 Step 3

Next we show that among all symmetric and piecewise constant probability density functions, we only need to consider those which are geometrically decaying.

More precisely, given positive integer i,

$$\mathcal{SP}_{i,\text{pd}} \triangleq \{\ \mathcal{P} \mid \mathcal{P} \in \mathcal{SP}_{i,\text{md}},\ \text{and } \mathcal{P} \text{ has density sequence } \{a_0, a_1, \ldots, a_n, \ldots,\} \\ \text{satisfying} \frac{a_k}{a_{k+i}} = e^\epsilon, \forall k \in \mathbb{N}\},$$

then

Lemma B.5.

$$V^* = \inf_{\mathcal{P} \in \cup_{i=1}^{\infty} \mathcal{SP}_{i,pd}} V(\mathcal{P}).$$

Proof. Due to Lemma B.4, we only need to consider probability distributions with symmetric and piecewise constant probability density functions which are monotonically decreasing.

We first show that given $\mathcal{P}_a \in \mathcal{SP}_{i,\text{md}}$ with density sequence $\{a_0, a_1, \ldots, a_n, \ldots,\}$, if $\frac{a_0}{a_i} < e^\epsilon$, then we can construct a probability distributions $\mathcal{P}_b \in \mathcal{SP}_{i,\text{md}}$ with density sequence $\{b_0, b_1, \ldots, b_n, \ldots,\}$ such that $\frac{b_0}{b_i} = e^\epsilon$ and

$$V(\mathcal{P}_b) \leq V(\mathcal{P}_a).$$

Define a new sequence $\{b_0, b_1, \ldots, b_n, \ldots\}$ by scaling up a_0 and scaling down $\{a_1, a_2, \ldots\}$. More precisely, define $\{b_0, b_1, \ldots, b_n, \ldots\}$ via

$$b_0 = a_0(1+\delta),$$
$$b_k = a_k(1-\delta'), \forall\ k \geq 1,$$

for some $\delta > 0$ and $0 < \delta' < 1$ such that

$$\frac{b_0}{b_i} = e^\epsilon,$$
$$\sum_{k=0}^{+\infty} b_k \text{Vol}(A_i(k)) = \sum_{k=0}^{+\infty} a_k \text{Vol}(A_i(k)) = 1.$$

So $\{b_0, b_1, \ldots, b_n, \ldots\}$ is a valid probability density sequence. Let $\mathcal{P}_b$ be the corresponding probability distribution. It is easy to check that $\mathcal{P}_b$ satisfies the differential privacy constraint, i.e.,

$$\frac{b_k}{b_{k+i}} \leq e^\epsilon, \forall k \geq 0.$$

Hence, $\mathcal{P}_b \in \mathcal{SP}_{i,\text{md}}$. Since $\mathcal{C}(\|bx\|_1)$ is a monotonically increasing function of $\|\mathbf{x}\|_1$, we have

$V(\mathcal{P}_b) \leq V(\mathcal{P}_a)$.

Therefore, for given $i \in \mathbb{N}$, we only need to consider $\mathcal{P} \in \mathcal{SP}_{i,\mathrm{md}}$ with density sequence $\{a_0, a_1, \ldots, a_n, \ldots\}$ satisfying $\frac{a_0}{a_i} = e^\epsilon$.

Next, we argue that among all probability distributions $\mathcal{P} \in \mathcal{SP}_{i,\mathrm{md}}$ with density sequence $\{a_0, a_1, \ldots, a_n, \ldots,\}$ satisfying $\frac{a_0}{a_i} = e^\epsilon$, we only need to consider those probability distributions with density sequence also satisfying $\frac{a_1}{a_{i+1}} = e^\epsilon$.

Given $\mathcal{P}_a \in \mathcal{SP}_{i,\mathrm{md}}$ with density sequence $\{a_0, a_1, \ldots, a_n, \ldots\}$ satisfying $\frac{a_0}{a_i} = e^\epsilon$ and $\frac{a_1}{a_{i+1}} < e^\epsilon$, we can construct a new probability distribution $\mathcal{P}_b \in \mathcal{SP}_{i,\mathrm{md}}$ with density sequence $\{b_0, b_1, \ldots, b_n, \ldots\}$ satisfying

$$\frac{b_0}{b_i} = e^\epsilon,$$
$$\frac{b_1}{b_{i+1}} = e^\epsilon,$$

and $V(\mathcal{P}_a) \geq V(\mathcal{P}_b)$.

First, it is easy to see a_1 is strictly less than a_0, since if $a_0 = a_1$, then $\frac{a_1}{a_{i+1}} = \frac{a_0}{a_{i+1}} \geq \frac{a_0}{a_i} = e^\epsilon$. We can construct a new density sequence by increasing a_1 and decreasing a_{i+1} to make $\frac{a_1}{a_{i+1}}$. More precisely, we define a new sequence $\{b_0, b_1, \ldots, b_n, \ldots\}$ as

$$\begin{aligned} b_k &= a_k, \forall k \neq 1, k \neq i+1, \\ b_1 &= a_1(1+\delta), \\ b_{i+1} &= a_{i+1}(1-\delta'), \end{aligned}$$

where $\delta > 0$ and $\delta' > 0$ are chosen such that $\frac{b_1}{b_{i+1}} = e^\epsilon$ and

$$\sum_{k=0}^{+\infty} b_k \mathrm{Vol}(A_i(k)) = \sum_{k=0}^{+\infty} a_k \mathrm{Vol}(A_i(k)) = 1.$$

It is easy to verify that $\{b_0, b_1, \ldots, b_n, \ldots\}$ is a valid probability density sequence and the corresponding probability distribution $\mathcal{P}_b$ satisfies the differential privacy constraint (3.6). Moreover, $V(\mathcal{P}_b) \leq V(\mathcal{P}_a)$. Therefore, we only need to consider $\mathcal{P} \in \mathcal{SP}_{i,\mathrm{md}}$ with density sequences $\{a_0, a_1, \ldots, a_n, \ldots\}$ satisfying $\frac{a_0}{a_i} = e^\epsilon$ and $\frac{a_1}{a_{i+1}} = e^\epsilon$.

Use the same argument, we can show that we only need to consider $\mathcal{P} \in \mathcal{SP}_{i,\mathrm{md}}$ with density sequences $\{a_0, a_1, \ldots, a_n, \ldots\}$ satisfying

$$\frac{a_k}{a_{i+k}} = e^\epsilon, \forall k \geq 0.$$

Therefore,

$$V^* = \inf_{\mathcal{P} \in \cup_{i=1}^{\infty} \mathcal{SP}_{i,\mathrm{pd}}} V(\mathcal{P}).$$

□

Due to Lemma B.5, we only need to consider probability distribution with a symmetric, monotonically decreasing, and geometrically decaying piecewise constant probability density function. Because of the properties of symmetry and periodic (geometric) decay, for this class of probability distributions, the probability density function over $\mathbb{R}^d$ is completely determined by the probability density function over the set $\{\mathbf{x} \in \mathbb{R}^d | \|\mathbf{x}\|_1 < \Delta\}$.

Next, we study what the optimal probability density function should be over the set $\{\mathbf{x} \in \mathbb{R}^d | \|\mathbf{x}\|_1 < \Delta\}$. It turns out that the optimal probability density function over the set $\{\mathbf{x} \in \mathbb{R}^d | \|\mathbf{x}\|_1 < \Delta\}$ is a step function. We use the following three steps to prove this result.

B.1.5 Step 4

Lemma B.6. *Consider a probability distribution $\mathcal{P}_a \in \mathcal{SP}_{i,pd}$ ($i \geq 2$) with density sequence $\{a_0, a_1, \ldots, a_n, \ldots\}$. Then there exists an integer $k(i)$ and a probability distribution $\mathcal{P}_b \in \mathcal{SP}_{i,pd}$ with density sequence $\{b_0, b_1, \ldots, b_n, \ldots\}$ such that*

$$b_0 = b_1 = b_2 = \cdots = b_{k(i)},$$
$$\frac{b_0}{b_{i-1}} = e^{\epsilon},$$

and

$$V(\mathcal{P}_b) \leq V(\mathcal{P}_a).$$

Proof. For $0 \leq k \leq i-1$, define

$$w_k \triangleq \sum_{j=0}^{+\infty} e^{-j\epsilon} \int \int \cdots \int_{(j+\frac{k}{i})\Delta \leq \|\mathbf{x}\|_1 < (j+\frac{k}{i})\Delta} \mathcal{C}(\mathbf{x}) dx_1 dx_2 \ldots dx_d,$$

and

$$u_k \triangleq \sum_{j=0}^{+\infty} e^{-j\epsilon} \mathrm{Vol}(A_i(ji+k).$$

Then the cost $V(\mathcal{P}_a) = \sum_{k=0}^{i-1} w_k a_k$, and the constraint on a_k is that

$$\begin{aligned} & a_0 \geq a_1 \geq \cdots \geq a_{i-1}, \\ & a_0 \leq a_{i-1} e^{\epsilon}, \\ & \sum_{k=0}^{+\infty} u_k a_k = 1. \end{aligned}$$

Therefore, to minimize $V(\mathcal{P})$ among all probability distributions $\mathcal{P} \in \mathcal{SP}_{i,\mathrm{pd}}$, we need to solve the following linear programming problem

$$\begin{aligned} \underset{a_0,a_1,\ldots,a_{i-1}}{\text{minimize}} \quad & \sum_{k=0}^{i-1} w_k a_k, \\ \text{subject to} \quad & a_0 \geq a_1 \geq \cdots \geq a_{i-1}, \\ & a_0 \leq a_{i-1} e^{\epsilon}, \\ & \sum_{k=0}^{+\infty} u_k a_k = 1. \end{aligned}$$

Let

$$h_k \triangleq \frac{w_k}{u_k}.$$

In the following we show that when $d = 2$, there exists an integer $k(i)$ such that

$$h_0 \geq h_1 \geq \cdots \geq h_{k(i)}, \tag{B.9}$$

$$h_{k(i)} \leq h_{k(i)+1} \leq \cdots \leq h_{i-1}, \tag{B.10}$$

$$h_0 \leq h_{i-1}. \tag{B.11}$$

When $d = 2$,

$$\begin{aligned} h_k &= \frac{w_k}{u_k} \\ &= \frac{\frac{4}{3}\frac{\Delta^3}{i^3} \sum_{j=0}^{+\infty} e^{-j\epsilon}(1 + 3(ji+k) + 3(ij+k)^2}{2\frac{\Delta^2}{i^2} \sum_{j=0}^{+\infty} e^{-j\epsilon}(1 + 2(ji+k))} \\ &= \frac{2}{3}\frac{\Delta}{i} \frac{3i^2 c_2 + (6ik+3i)c_1 + (1+3k+3k^2)c_0}{(1+2k)c_0 + 2ic_1}. \end{aligned}$$

Let $g(k) =\triangleq \frac{3i^2 c_2 + (6ik+3i)c_1 + (1+3k+3k^2)c_0}{(1+2k)c_0 + 2ic_1}$. It is easy to compute the derivative of $g(k)$ with

respect to k:

$$g'(k) = \frac{6c_0^2k^2 + 6c_0^2k + c_0^2 + 12c_0c_1ik + 6c_0c_1i - 6c_2c_0i^2 + 12c_1^2i^2}{((1+2k)c_0 + 2ic_1)^2}.$$

Note that the numerator of $g'(k)$ is an increasing function of k, and

$$\begin{aligned} g'(0) &= c_0^2 + 6c_0c_1i - 6c_2c_0i^2 + 12c_1^2i^2 \\ &= \frac{b(6i^2 - 6i + 1) - 1}{(b-1)^3} < 0, \end{aligned}$$

for sufficiently large i, and

$$g'(i-1) = \frac{6i^2 - 6i + 1 - b}{(1-b)^3} > 0.$$

Therefore, h_k first increases as k increases, and then decreases as k increases to $i-1$. Hence, there exists an integer $k(i)$ such that (B.9) and (B.10) hold.

Next we compare h_{i-1} and h_0:

$$\begin{aligned} h_{i-1} - h_0 &= \frac{w_{i-1}}{u_{i-1}} - \frac{w_0}{u_0} \\ &= \frac{2}{3}\frac{\Delta}{i}\frac{(3i-2)(b-1)^2(i-1)}{(2bi - b + 1)(b + 2i - 1)} > 0. \end{aligned}$$

Hence, (B.11) also holds.

Now we are ready to prove Lemma B.6.

Suppose $a_{k(i)} < a_{k(i)-1}$. We can scale up $a_{k(i)}$ and scale down $a_{k(i)-1}$ to make $a_{k(i)} = a_{k(i)-1}$. Since $h_{k(i)} \le h_{k(i)-1}$, i.e., $\frac{w_{k(i)}}{u_{k(i)}} \le \frac{w_{k(i)-1}}{u_{k(i)-1}}$, this scaling operation will not increase the cost $V(\mathcal{P}_a)$. Now we have $a_{k(i)} = a_{k(i)-1}$.

Suppose $a_{k(i)} = a_{k(i)-1} < a_{k(i)-2}$. Then we can scale up $a_{k(i)}$ and $a_{k(i)-1}$, and scale down $a_{k(i)-2}$ to make $a_{k(i)} = a_{k(i)-1} = a_{k(i)-2}$. Since $h_{k(i)} \le h_{k(i)-1} \le h_{k(i)-2}$, this scaling operation will not increase the cost $V(\mathcal{P}_a)$. Now we have $a_{k(i)} = a_{k(i)-1} = a_{k(i)-2}$.

After $k(i)$ steps of these scaling operations, we can make $a_0 = a_1 = \cdots = a_{k(i)}$, and this will not increase the cost $V(\mathcal{P}_a)$.

Finally, if $\frac{a_0}{a_{i-1}} < e^\epsilon$, we can scale up $a_0, a_1, \ldots, a_{k(i)}$, and scale down a_{i-1} to make $\frac{a_0}{a_{i-1}} = e^\epsilon$. Since $h_{i-1} \ge h_0 \ge h_1 \ge \cdots \ge h_{k(i)}$, this scaling operation will not increase the cost $V(\mathcal{P}_a)$.

Let $\mathcal{P}_b$ be the probability distribution we obtained after the $k(i)+1$ steps of scaling

operations. Then $\mathcal{P}_b \in \mathcal{SP}_{i,\mathrm{pd}}$, and its density sequence $\{b_0, b_1, \ldots, b_n, \ldots\}$ satisfies

$$b_0 = b_1 = b_2 = \cdots = b_{k(i)},$$
$$\frac{b_0}{b_{i-1}} = e^{\epsilon},$$

and

$$V(\mathcal{P}_b) \leq V(\mathcal{P}_a).$$

This completes the proof of Lemma B.6.

□

Therefore, due to Lemma B.6, for sufficiently large i, we only need to consider probability distributions $\mathcal{P} \in \mathcal{SP}_{i,\mathrm{pd}}$ with density sequence $\{a_0, a_1, \ldots, a_n, \ldots\}$ satisfying

$$a_0 = a_1 = a_2 = \cdots = a_{k(i)}, \tag{B.12}$$
$$\frac{b_0}{b_{i-1}} = e^{\epsilon}. \tag{B.13}$$

More precisely, define

$$\mathcal{SP}_{i,\mathrm{fr}} = \{\ \mathcal{P} \in \mathcal{SP}_{i,\mathrm{pd}} \mid \mathcal{P} \text{ has density sequence } \{a_0, a_1, \ldots, a_n, \ldots\} \text{ satisfying (B.12) and (B.13) }\}.$$

Then due to Lemma B.6,

Lemma B.7.

$$V^* = \inf_{\mathcal{P} \in \cup_{i=3}^{\infty} \mathcal{SP}_{i,fr}} V(\mathcal{P}).$$

Next, we argue that for each probability distribution $\mathcal{P} \in \mathcal{SP}_{i,\mathrm{fr}}$ $(i \geq 3)$ with density sequence $\{a_0, a_1, \ldots, a_n, \ldots\}$, we can assume that there exists an integer $k(i)+1 \leq k \leq (i-2)$, such that

$$a_j = a_0, \forall 0 \leq j < k, \tag{B.14}$$
$$a_j = a_{i-1}, \forall k < j < i. \tag{B.15}$$

More precisely,

Lemma B.8. *Consider a probability distribution $\mathcal{P}_a \in \mathcal{SP}_{i,fr}$ ($i \geq 3$) with density sequence $\{a_0, a_1, \ldots, a_n, \ldots\}$. Then there exists a probability distribution $\mathcal{P}_b \in \mathcal{SP}_{i,fr}$ with density sequence $\{b_0, b_1, \ldots, b_n, \ldots\}$ such that there exists an integer $k(i)+1 \leq k \leq (i-2)$ with*

$$b_j = a_0, \forall\ 0 \leq j < k, \tag{B.16}$$

$$b_j = a_{i-1}, \forall\ k < j < i, \tag{B.17}$$

and

$$V(\mathcal{P}_b) \leq V(\mathcal{P}_a). \tag{B.18}$$

Proof. If there exists an integer $k(i)+1 \leq k \leq (i-2)$ such that

$$a_j = a_0, \forall\ 0 \leq j < k,$$
$$a_j = a_{i-1}, \forall\ k < j < i,$$

then we can let $\mathcal{P}_b = \mathcal{P}_a$.

Otherwise, let k_1 be the smallest integer in $\{k(i)+1, k(i)+2, \ldots, i-1\}$ such that

$$a_{k_1} \neq a_0,$$

and let k_2 be the biggest integer in $\{k(i)+1, k(i)+2, \ldots, i-1\}$ such that

$$a_{k_2} \neq a_{i-1}.$$

It is easy to see that $k_1 \neq k_2$. Then we can scale up a_{k_1} and scale down a_{k_2} simultaneously until either $a_{k_1} = a_0$ or $a_{k_2} = a_{i-1}$. Since $h_k \triangleq \frac{w_k}{u_k}$ is an increasing function of k when $k > k(i)$, and $k(i) < k_1 < k_2$, this scaling operation will not increase the cost.

After this scaling operation we can update k_1 and k_2, and either k_1 is increased by one or k_2 is decreased by one.

Therefore, continue in this way, and finally we will obtain a probability distribution $\mathcal{P}_b \in \mathcal{SP}_{i,\text{fr}}$ with density sequence $\{b_0, b_1, \ldots, b_n, \ldots\}$ such that (B.16), (B.17), and (B.18) hold.

This completes the proof.

□

Define

$$\mathcal{SP}_{i,\text{step}} = \{\ \mathcal{P} \in \mathcal{SP}_{i,\text{fr}} \mid \mathcal{P} \text{ has density sequence } \{a_0, a_1, \ldots, a_n, \ldots\} \\ \text{satisfying (B.16) and (B.17) for some } k(i) < k \leq (i-2)\ \}.$$

Then due to Lemma B.8,

Lemma B.9.

$$V^* = \inf_{\mathcal{P} \in \cup_{i=3}^{\infty} \mathcal{SP}_{i,step}} V(\mathcal{P}).$$

As $i \to \infty$, the probability density function of $\mathcal{P} \in \mathcal{SP}_{i,\text{fr}}$ will converge to a multiple dimensional staircase function. Therefore, for $d = 2$ and the cost function $\mathcal{L}(\mathbf{x}) = \|\mathbf{x}\|_1, \forall \mathbf{x} \in \mathbb{R}^2$, then

$$\inf_{\mathcal{P} \in \mathcal{SP}} \int\int_{\mathbb{R}^2} \mathcal{L}(\mathbf{x}) \mathcal{P}(dx_1 dx_2) = \inf_{\gamma \in [0,1]} \int\int_{\mathbb{R}^2} \mathcal{L}(\mathbf{x}) f_\gamma(\mathbf{x}) dx_1 dx_2.$$

This completes the proof of Theorem 3.1.

APPENDIX C

PROOFS FOR CHAPTER 4

C.1 Proof of Theorem 4.2

Proof of Theorem 4.2 . Consider a feasible solution to the optimization problem (4.8) with primal variables

$$p_k = \begin{cases} \delta & k = 1 + i\Delta, \text{for} \quad i = 0, 1, 2, \ldots, \frac{1}{2\delta} - 1 \\ 0 & \text{otherwise} \end{cases}.$$

The corresponding value of the objective function is

$$2\delta \sum_{i=0}^{\frac{1}{2\delta}-1} \mathcal{L}(1 + i\Delta).$$

Therefore,

$$V_{LB} \leq 2\delta \sum_{i=0}^{\frac{1}{2\delta}-1} \mathcal{L}(1 + i\Delta). \tag{C.1}$$

We claim that the above primal variables are the optimal solution. We prove this claim by constructing the corresponding dual variables.

Associating dual variables μ with the constraint in (4.9), y_k with the constraint in (4.10), we have the dual linear program:

$$\begin{aligned} V_{LB} = \max \quad & \mu - 2\delta \sum_{k=0}^{\infty} y_k \\ \text{such that} \quad & \mu \geq 0, y_k \geq 0, \forall k \in \mathbb{N}, && \text{(C.2)} \\ & \frac{1}{2}\mu - y_0 \leq 0, && \text{(C.3)} \end{aligned}$$

$$\mu - \sum_{i=\max(0,k-\Delta+1)}^{k} y_k \leq \mathcal{L}(k), \forall k \geq 1. \tag{C.4}$$

The complementary slackness conditions require that

$$\begin{aligned}
\mu - y_0 - y_1 &= \mathcal{L}(1), \\
\mu - \sum_{i=2+(k-1)\Delta}^{1+k\Delta} y_k &= \mathcal{L}(1+k\Delta), \text{for } k = 1, 2, \ldots, \frac{1}{2\delta} - 1, \\
y_k &= 0, \forall k \geq (\frac{1}{2\delta} - 1)\Delta + 2.
\end{aligned}$$

Consider the following dual variables:

$$\begin{aligned}
\mu &= \mathcal{L}(1 + \frac{\Delta}{2\delta}), \\
y_k &= 0, \forall k \geq (\frac{1}{2\delta} - 1)\Delta + 2, \\
y_k &= \mathcal{L}(k+\Delta) - \mathcal{L}(k+\Delta-1) + y(k+\Delta), \forall 2 \leq k \leq (\frac{1}{2\delta} - 1)\Delta + 1, \\
y_1 &= \sum_{i=1}^{\frac{1}{2\delta}} (\mathcal{L}(1+i\Delta) - \mathcal{L}(i\Delta)) \geq 0, \\
y_0 &= \mu - \mathcal{L}(1) - y_1 = \mathcal{L}(1 + \frac{\Delta}{2\delta}) - \mathcal{L}(1) - \sum_{i=1}^{\frac{1}{2\delta}} (\mathcal{L}(1+i\Delta) - \mathcal{L}(i\Delta)) \geq 0.
\end{aligned}$$

It is easy to verify that these dual variables satisfy the constraints of the dual linear program, and the value of the objective function is

$$\begin{aligned}
\mu - 2\delta \sum_{k=0}^{+\infty} y_k =& \mu - 2\delta \sum_{i=0}^{\frac{1}{2\delta}-1} (\mu - \mathcal{L}(1+i\Delta)) \\
=& 2\delta \sum_{i=0}^{\frac{1}{2\delta}-1} \mathcal{L}(1+i\Delta).
\end{aligned}$$

Therefore, by weak duality we have

$$V_{LB} \geq 2\delta \sum_{i=0}^{\frac{1}{2\delta}-1} \mathcal{L}(1+i\Delta).$$

Due to (C.1), we conclude

$$V_{LB} = 2\delta \sum_{i=0}^{\frac{1}{2\delta}-1} \mathcal{L}(1+i\Delta).$$

□

C.2 Proof of Corollary 4.5

Proof of Corollary 4.5. First we compute the lower bound V_{LB} via

$$\begin{aligned}
V_{LB} &= 2\sum_{i=0}^{\frac{1}{2\delta}-1} \delta\mathcal{L}(1+i\Delta) \\
&= 2\delta \sum_{i=0}^{\frac{1}{2\delta}-1} (1+i\Delta)^2 \\
&= 2\delta \sum_{i=0}^{\frac{1}{2\delta}-1} (1+2i\Delta+i^2\Delta^2) \\
&= 2\delta(\frac{1}{2\delta} + 2\Delta\frac{\frac{1}{2\delta}(\frac{1}{2\delta}-1)}{2} + \Delta^2\frac{(\frac{1}{2\delta}-1)\frac{1}{2\delta}(2\frac{1}{2\delta}-1)}{6}) \\
&= 1 + \Delta(\frac{1}{2\delta}-1) + \frac{\Delta^2}{12\delta^2} + \frac{\Delta^2}{6} - \frac{\Delta^2}{4\delta} \\
&= \Theta(\frac{\Delta^2}{12\delta^2}).
\end{aligned}$$

The upper bound is

$$\begin{aligned}
V_{UB} &= 2\sum_{i=1}^{\frac{\Delta}{2\delta}-1} \frac{\delta}{\Delta}\mathcal{L}(i) + \frac{\delta}{\Delta}\mathcal{L}(\frac{\Delta}{2\delta}) \\
&= 2\frac{\delta}{\Delta}\frac{(\frac{\Delta}{2\delta}-1)\frac{\Delta}{2\delta}(\frac{\Delta}{\delta}-1)}{6} + \frac{\delta}{\Delta}\frac{\Delta^2}{4\delta^2} \\
&= \frac{1}{6}(\frac{\Delta^2}{2\delta^2} + 1 - \frac{3\Delta}{2\delta}) + \frac{\Delta}{4\delta} \\
&= \frac{\Delta^2}{12\delta^2} + \frac{1}{6} \\
&= \Theta(\frac{\Delta^2}{12\delta^2}).
\end{aligned}$$

Therefore, the multiplicative gap goes to one as $\delta \to 0$, i.e.,

$$\lim_{\delta \to 0} \frac{V_{UB}}{V_{LB}} = 1.$$

□

C.3 Proof of Corollary 4.7

Proof of Corollary 4.7. Using the fact that $\mathcal{L}(\cdot)$ is a monotonically increasing function for $k \geq 0$, we have

$$\begin{aligned} V_{UB} - V_{LB} &= 2 \sum_{i=1}^{\frac{\Delta}{2\delta}-1} \frac{\delta}{\Delta} \mathcal{L}(i) + \frac{\delta}{\Delta} \mathcal{L}(\frac{\Delta}{2\delta}) - 2\delta \sum_{i=0}^{\frac{1}{2\delta}-1} \mathcal{L}(1+i\Delta) \\ &\leq -2\delta \mathcal{L}(1) + \frac{\delta}{\Delta} \mathcal{L}(\frac{\Delta}{2\delta}) + 2\delta \mathcal{L}(\frac{\Delta}{2\delta} - 1) \\ &\leq (2 + \frac{1}{\Delta}) \delta \mathcal{L}(\frac{\Delta}{2\delta}). \end{aligned}$$

Therefore,

$$\begin{aligned} \frac{V_{UB}}{V_{LB}} &= 1 + \frac{V_{UB} - V_{LB}}{V_{LB}} \\ &\leq 1 + \frac{(2 + \frac{1}{\Delta})\delta \mathcal{L}(\frac{\Delta}{2\delta})}{2\delta \sum_{i=0}^{\frac{1}{2\delta}-1} \mathcal{L}(1+i\Delta)} \\ &\leq 1 + \frac{(2 + \frac{1}{\Delta})\delta \mathcal{L}(\frac{\Delta}{2\delta})}{2\delta \mathcal{L}(1 + (\frac{1}{2\delta} - 1)\Delta)}, \end{aligned}$$

and thus

$$\lim_{\delta \to 0} \frac{V_{UB}}{V_{LB}} \leq 1 + (1 + \frac{1}{2\Delta})C.$$

□

C.4 Proof of Theorem 4.8

Proof of Theorem 4.8. Consider the feasible primal variables $\{p_k\}_{k\in\mathbb{N}}$ defined as

$$\mathcal{P}_k = \begin{cases} ab^i & \text{for } k = 1 + i\Delta, 0 \le i \le n-1 \\ 0 & \text{otherwise} \end{cases}. \tag{C.5}$$

It is straightforward to verify that the above primal variables satisfy the constraints of the relaxed linear program, and the corresponding value of the objective function is

$$2\sum_{k=0}^{n-1} ab^k \mathcal{L}(1+k\Delta).$$

We prove it is also the optimal value by constructing the optimal dual variables for the corresponding dual linear program.

Associating dual variables μ, y_0, y_1, y_i with the primal constraints in (4.19), (4.20), (4.21), and (4.22), respectively, we have the dual linear program:

$$V_{LB} := \min \quad \mu - (2\delta + e^\epsilon - 1)\sum_{k=0}^{+\infty} y_k \tag{C.6}$$

$$\text{such that} \quad \mu \ge 0, y_k \ge 0 \quad \forall k \in N \tag{C.7}$$

$$\frac{1}{2}\mu - \frac{1+e^\epsilon}{2} y_0 - \frac{e^\epsilon - 1}{2} y_1 - \frac{e^\epsilon - 1}{2}\sum_{k=2}^{+\infty} y_k \le 0 \tag{C.8}$$

$$\mu - e^\epsilon y_0 - e^\epsilon y_1 - (e^\epsilon - 1)\sum_{k=2}^{+\infty} y_k \le \mathcal{L}(1) \tag{C.9}$$

$$\mu - e^\epsilon \sum_{l=\max(0,k-\Delta+1)}^{k} y_l - (e^\epsilon - 1)\sum_{l=k+1}^{+\infty} y_l \le \mathcal{L}(k), \forall k \ge 2. \tag{C.10}$$

If the primal variables defined in (C.5) are the optimal solution, the complementary slackness conditions require that the corresponding dual variables satisfy that

$$\mu = \mathcal{L}(1) + e^\epsilon(y_0 + y_1) + (e^\epsilon - 1)\sum_{l=2}^{+\infty} y_l$$

$$\mu = \mathcal{L}(1+\Delta) + e^\epsilon \sum_{l=2}^{1+\Delta} y_l + (e^\epsilon - 1)\sum_{l=2+\Delta}^{+\infty} y_l$$

$$\mu = \mathcal{L}(1+k\Delta) + e^{\epsilon} \sum_{l=2+(k-1)\Delta}^{1+k\Delta} y_l + (e^{\epsilon}-1) \sum_{l=2+k\Delta}^{+\infty} y_l, \forall 1 \le k \le n-1,$$
$$y_l = 0, \forall l \ge 2 + (n-1)\Delta.$$

Consider the following dual variables defined via

$$\begin{aligned}
\mu &= \mathcal{L}(1+(n-1)\Delta), \\
y_k &= 0, \forall k \ge 2 + (n-2)\Delta, \\
y_k &= b(y_{k+\Delta} + \mathcal{L}(k+\Delta) - \mathcal{L}(k+\Delta-1)), \forall 2 \le k \le 1 + (n-2)\Delta, \\
y_1 &= \sum_{i=1}^{n-1} b^i (\mathcal{L}(1+i\Delta) - \mathcal{L}(i\Delta)), \\
y_0 &= \sum_{i=1}^{n-1} b^i (\mathcal{L}(i\Delta) - \mathcal{L}(1+(i-1)\Delta)).
\end{aligned}$$

We verify that the above dual variables satisfy the inequality (C.8) in the following

$$\begin{aligned}
&(1+e^{\epsilon})y_0 + (e^{\epsilon}-1)y_1 + (e^{\epsilon}-1)\sum_{k=2}^{+\infty} y_k - \mu \ge 0 \\
\Leftrightarrow &y_0 - y_1 + e^{\epsilon}(y_0+y_1) + (e^{\epsilon}-1)\sum_{k=2}^{+\infty} y_k - \mu \ge 0 \\
\Leftrightarrow &y_0 - y_1 + \mu - \mathcal{L}(1) - \mu \ge 0 \\
\Leftrightarrow &y_0 - y_1 - \mathcal{L}(1) \ge 0 \\
\Leftrightarrow &\sum_{i=1}^{n-1} b^i (2\mathcal{L}(i\Delta) - \mathcal{L}(1+(i-1)\Delta) - \mathcal{L}(1+i\Delta)) \ge \mathcal{L}(1).
\end{aligned}$$

It is easy to verify that the dual variables satisfy the constraints (C.7), (C.8), (C.9), and (C.10) in the dual linear program. Next we compute the corresponding value of the objective function

$$\begin{aligned}
&\mu - (2\delta + e^{\epsilon} - 1)\sum_{k=0}^{+\infty} y_k \\
=&\mu - (2\delta + e^{\epsilon} - 1)(y_0 + y_1 + \frac{\mu - \mathcal{L}(1) - e^{\epsilon}(y_0+y_1)}{e^{\epsilon}-1}) \\
=&\mu - \frac{2\delta + e^{\epsilon} - 1}{e^{\epsilon}-1}(\mu - \mathcal{L}(1) - y_0 - y_1)
\end{aligned}$$

$$
\begin{aligned}
=&\mathcal{L}(1+(n-1)\Delta) - \frac{2\delta + e^{\epsilon} - 1}{e^{\epsilon} - 1}(\mathcal{L}(1+(n-1)\Delta) - \mathcal{L}(1) - \\
&\sum_{i=1}^{n-1} b^i(\mathcal{L}(1+i\Delta) - \mathcal{L}(1+(i-1)\Delta))) \\
=&2\sum_{k=0}^{n-1} ab^k \mathcal{L}(1+k\Delta),
\end{aligned}
$$

which is also the value of the objective function in the primal problem achieved by the primal variables defined in (C.5). Therefore, we conclude that

$$
V_{LB} = 2\sum_{k=0}^{n-1} ab^k \mathcal{L}(1+k\Delta).
$$

□

C.5 Proof of Corollary 4.12

Proof of Corollary 4.12. For the cost function $\mathcal{L}(k) = |k|$,

$$
\begin{aligned}
V_{LB} &= 2\sum_{k=0}^{n-1} ab^k \mathcal{L}(1+k\Delta) \\
&= 2\sum_{k=0}^{n-1} ab^k (1+k\Delta) \\
&= 1 + 2a\Delta \sum_{k=0}^{n-1} b^k k \\
&= 1 + 2a\Delta\left(\frac{b - b^n}{(1-b)^2} - \frac{(n-1)b^n}{1-b}\right).
\end{aligned}
$$

Given $\delta > 0$, V_{LB} is a decreasing function of ϵ. Therefore, to lower bound $\frac{V_{UB}^{\text{uniform}}}{V_{LB}}$ in the regime $\epsilon \leq \delta$, we only need to consider the case $\epsilon = \delta$. Thus, in the following we set $\epsilon = \delta$.

Since $\sum_{k=0}^{n-1} ab^k = \frac{1}{2}$, we have

$$
\begin{aligned}
a\frac{1-b^n}{1-b} &= \frac{1}{2} \\
\Leftrightarrow b^n &= 1 - \frac{1-b}{2a}.
\end{aligned}
$$

As $\delta \to 0$, $\frac{1-b}{2a} = \frac{1-e^{-\epsilon}}{2\frac{\delta+\frac{e^\epsilon-1}{e^\epsilon}}{2}} \to \frac{1}{3}$, and thus

$$\lim_{\delta\to 0} b^n = 1 - \frac{1}{3} = \frac{2}{3},$$
$$n = \Theta(\frac{\log(\frac{3}{2})}{\epsilon}).$$

Note that $a = \Theta(\frac{3}{2}\delta)$ as $\delta \to 0$.
Therefore, as $\delta \to 0$,

$$\begin{aligned} V_{LB} &\approx 2\Delta a(\frac{1-\frac{2}{3}}{\epsilon^2} - \frac{\frac{\log(\frac{3}{2})}{\epsilon}\frac{2}{3}}{\epsilon}) \\ &\approx 2\Delta\frac{3}{2}\delta(\frac{1}{3\delta^2} - \frac{\frac{2}{3}\log(\frac{3}{2})}{\delta^2}) \\ &= \frac{\Delta}{\delta}(1 - 2\log\frac{3}{2}) \\ &\approx 0.19\frac{\Delta}{\delta}. \end{aligned}$$

Recall $V_{UB}^{\text{uniform}} = \frac{\Delta}{4\delta}$.
Therefore,

$$\lim_{\epsilon=\delta\to 0} \frac{V_{UB}}{V_{LB}} = \frac{1}{4(1-2\log\frac{3}{2})} \approx 1.32,$$

and thus

$$\lim_{\epsilon\le\delta\to 0} \frac{V_{UB}}{V_{LB}} \le \frac{1}{4(1-2\log\frac{3}{2})} \approx 1.32.$$

□

C.6 Proof of Corollary 4.13

Proof of Corollary 4.13. Using the same argument in the proof of Corollary 4.12, we can set $\epsilon = \delta$.

For the cost function $\mathcal{L}(k) = k^2$,

$$V_{LB} = 2\sum_{k=0}^{n-1} ab^k\mathcal{L}(1+k\Delta)$$

$$
\begin{aligned}
&= 2\sum_{k=0}^{n-1} ab^k(1+k\Delta)^2 \\
&= 1 + 4a\Delta\sum_{k=0}^{n-1} b^k k + 2a\Delta^2\sum_{k=0}^{n-1} b^k k^2 \\
&\approx 2a\Delta^2\sum_{k=0}^{n-1} b^k k^2 \\
&= 2\Delta^2\frac{\delta+\frac{e^\epsilon-1}{2}}{e^\epsilon}\frac{-b+2(\frac{b(1-b^{n-1})}{(1-b)^2}-\frac{(n-1)b^n}{1-b})-\frac{b^2(1-b^{n-2})}{1-b}-(n-1)^2b^n}{1-b} \\
&\approx 2\Delta^2\frac{3}{2}\epsilon\frac{2(\frac{1-\frac{2}{3}}{\epsilon^2}-\frac{\frac{2}{3}\log(\frac{3}{2})}{\epsilon^2})-\frac{1}{3\epsilon}-\frac{2}{3}\frac{(\log(\frac{3}{2}))^2}{\epsilon^2}}{\epsilon} \\
&\approx \frac{3\Delta^2}{\epsilon^2}(\frac{2}{3}-\frac{4}{3}\log(\frac{3}{2})-\frac{2}{3}(\log(\frac{3}{2}))^2) \\
&= \frac{\Delta^2}{\epsilon^2}(2-4\log(\frac{3}{2})-2(\log(\frac{3}{2}))^2) \\
&\approx \frac{\Delta^2}{20\epsilon^2} \\
&= \frac{\Delta^2}{20\delta^2}
\end{aligned}
$$

Recall $V_{UB}^{\text{uniform}} = \frac{\Delta^2}{12\delta^2}$.

Therefore,

$$\lim_{\epsilon=\delta\to 0}\frac{V_{UB}^{\text{uniform}}}{V_{LB}} = \frac{1}{12(2-4\log(\frac{3}{2})-2(\log(\frac{3}{2}))^2)} \approx \frac{5}{3},$$

and thus

$$\lim_{\epsilon\le\delta\to 0}\frac{V_{UB}^{\text{uniform}}}{V_{LB}} \le \frac{1}{12(2-4\log(\frac{3}{2})-2(\log(\frac{3}{2}))^2)} \approx \frac{5}{3}.$$

□

C.7 Proof of Corollary 4.14

Proof of Corollary 4.14. For the cost function $\mathcal{L}(k) = |k|$,

$$V_{LB} = 2\sum_{k=0}^{n-1} ab^k\mathcal{L}(1+k\Delta)$$

$$
\begin{aligned}
&= 2\sum_{k=0}^{n-1} ab^k(1+k\Delta) \\
&= 1 + 2a\Delta\sum_{k=0}^{n-1} b^k k \\
&= 1 + 2a\Delta(\frac{b-b^n}{(1-b)^2} - \frac{(n-1)b^n}{1-b}).
\end{aligned}
$$

Given $\epsilon > 0$, V_{LB} is a decreasing function of δ. Therefore, to lower bound $\frac{V_{UB}^{\text{Lap}}}{V_{LB}}$ in the regime $\delta \leq \epsilon$, we only need to consider the case $\delta = \epsilon$. Thus, in the following we set $\delta = \epsilon$.

Following the same calculations in the proof of Corollary 4.12, we have

$$
\begin{aligned}
V_{LB} &\approx \frac{\Delta}{\delta}(1 - 2\log\frac{3}{2}) \\
&\approx 0.19\frac{\Delta}{\delta} \\
&= 0.19\frac{\Delta}{\epsilon}.
\end{aligned}
$$

On the other hand, we have

$$
\begin{aligned}
V_{UB}^{\text{Lap}} &= 2\sum_{k=1}^{+\infty} \frac{1-\lambda}{1+\lambda}\lambda^k k \\
&= \frac{2e^{-\frac{\epsilon}{\Delta}}}{1 - e^{-2\frac{\epsilon}{\Delta}}} \\
&\approx \frac{\Delta}{\epsilon},
\end{aligned}
$$

as $\epsilon \to 0$.

Therefore,

$$
\lim_{\epsilon=\delta\to 0} \frac{V_{UB}^{\text{Lap}}}{V_{LB}} = \frac{1}{1 - 2\log\frac{3}{2}} \approx 5.29,
$$

and thus

$$
\lim_{\epsilon\leq\delta\to 0} \frac{V_{UB}^{\text{Lap}}}{V_{LB}} \leq \frac{1}{1 - 2\log\frac{3}{2}} \approx 5.29.
$$

□

C.8 Proof of Corollary 4.15

Proof of Corollary 4.15. Using the same argument in the proof of Corollary 4.14, we can set $\epsilon = \delta$.

For the cost function $\mathcal{L}(k) = k^2$, following the same calculations in the proof of Corollary 4.13, we have

$$\begin{aligned} V_{LB} &\approx \frac{\Delta^2}{\epsilon^2}(2 - 4\log(\frac{3}{2}) - 2(\log(\frac{3}{2}))^2) \\ &\approx \frac{\Delta^2}{20\epsilon^2}. \end{aligned}$$

On the other hand, we have

$$\begin{aligned} V_{UB}^{\text{Lap}} &= 2\sum_{k=1}^{+\infty} \frac{1-\lambda}{1+\lambda}\lambda^k k^2 \\ &= \frac{2\lambda}{(1-\lambda)^2} \\ &\approx 2\frac{\Delta^2}{\epsilon^2}, \end{aligned}$$

as $\epsilon \to 0$.

Therefore,

$$\lim_{\epsilon=\delta\to 0} \frac{V_{UB}^{\text{Lap}}}{V_{LB}} = \frac{2}{(2 - 4\log(\frac{3}{2}) - 2(\log(\frac{3}{2}))^2)} \approx 40,$$

and thus

$$\lim_{\epsilon\leq\delta\to 0} \frac{V_{UB}^{\text{Lap}}}{V_{LB}} \leq \frac{2}{(2 - 4\log(\frac{3}{2}) - 2(\log(\frac{3}{2}))^2)} \approx 40.$$

□

C.9 Proof of Theorem 4.16

Proof of Theorem 4.16. Consider the dual program of the linear program (4.27),

$$V_{LB} := \max \quad \mu - \delta \left(\sum_{i_1\in\mathbb{Z}} y_{i_1}^{(1)} + \sum_{i_2\in\mathbb{Z}} y_{i_2}^{(2)} + \cdots + \sum_{i_d\in\mathbb{Z}} y_{i_d}^{(d)} \right)$$

$$\text{such that} \quad y_{i_1}^{(1)}, y_{i_2}^{(2)}, \ldots, y_{i_d}^{(d)} \geq 0, \forall i_1 \in \mathbb{Z}, i_2 \in \mathbb{Z}, \ldots, i_d \in \mathbb{Z}$$
$$\mu - \sum_{i_1 \in [k_1 - \Delta + 1, k_1]} y_{i_1}^{(1)} - \cdots - \sum_{i_d \in [k_d - \Delta + 1, k_d]} y_{i_d}^{(d)} \leq |k_1| + |k_2| + \cdots + |k_d|, \forall (k_1, \ldots, k_d) \in \mathbb{Z}^d.$$

Consider a candidate solution with

$$\mu = \frac{d\Delta}{2\delta}$$

and for all $m \in \{1, 2, \ldots, d\}$,

$$y_i^{(m)} = \begin{cases} \frac{\mu}{d} & i = 0 \\ \max(\frac{\mu}{d} - k\Delta, 0) & i = k\Delta, \text{ for } k \in \mathbb{Z}, k \geq 1 \\ \max(\frac{\mu}{d} - (|k| - 1)\Delta - 1, 0) & i = k\Delta, \text{ for } k \in \mathbb{Z}, k \leq -1 \\ 0 & \text{otherwise} \end{cases}.$$

It is easy to verify that this candidate solution satisfies the constraints, and the corresponding value of the objective function is

$$\begin{aligned} & \mu - \delta \left(\sum_{i_1 \in \mathbb{Z}} y_{i_1}^{(1)} + \sum_{i_2 \in \mathbb{Z}} y_{i_2}^{(2)} + \cdots + \sum_{i_d \in \mathbb{Z}} y_{i_d}^{(d)} \right) \\ =& \mu - \delta d \sum_{i_1 \in \mathbb{Z}} y_{i_1}^{(1)} \\ =& \mu - \delta d \left(\sum_{i=0}^{\frac{\mu}{d\Delta}} (\frac{\mu}{d} - i\Delta) + \sum_{i=0}^{\frac{\mu}{d\Delta} - 1} (\frac{\mu}{d} - i\Delta - 1) \right) \\ =& \mu - \delta d \left(\frac{\frac{\mu}{d}(\frac{\mu}{d\Delta} + 1)}{2} + \frac{(\frac{\mu}{d} + \Delta - 2)\frac{\mu}{d\Delta}}{2} \right) \\ =& \mu - \delta d (\frac{\mu^2}{d^2\Delta} + \frac{\mu}{d} - \frac{\mu}{d\Delta}) \\ =& \mu - \delta (\frac{\mu^2}{d\Delta} + \mu - \frac{\mu}{\Delta}) \\ =& \frac{d\Delta}{4\delta} - \frac{\Delta - 1}{2} d. \end{aligned}$$

Therefore, we have

$$V_{LB} \geq \frac{d\Delta}{4\delta} - \frac{\Delta - 1}{2} d.$$

□

C.10 Proof of Theorem 4.17

Proof of Theorem 4.17. Consider the dual program of the linear program (4.27),

$$V_{LB} := \max \quad \mu - \delta \left(\sum_{i_1 \in \mathbb{Z}} y_{i_1}^{(1)} + \sum_{i_2 \in \mathbb{Z}} y_{i_2}^{(2)} + \cdots + \sum_{i_d \in \mathbb{Z}} y_{i_d}^{(d)} \right)$$

$$\text{such that} \quad y_{i_1}^{(1)}, y_{i_2}^{(2)}, \ldots, y_{i_d}^{(d)} \geq 0, \forall i_1 \in \mathbb{Z}, i_2 \in \mathbb{Z}, \ldots, i_d \in \mathbb{Z}$$

$$\mu - \sum_{i_1 \in [k_1 - \Delta + 1, k_1]} y_{i_1}^{(1)} - \cdots - \sum_{i_d \in [k_d - \Delta + 1, k_d]} y_{i_d}^{(d)} \leq |k_1|^2 + |k_2|^2 + \cdots + |k_d|^2, \forall (k_1, \ldots, k_d) \in \mathbb{Z}^d.$$

To avoid integer-rounding issues, assume that $\frac{1}{2\delta}$ is an integer. Consider a candidate solution with

$$\mu = \frac{d\Delta^2}{4\delta^2}$$

and for all $m \in \{1, 2, \ldots, d\}$,

$$y_i^{(m)} = \begin{cases} \frac{\mu}{d} & i = 0 \\ \frac{\mu}{d} - k^2\Delta^2 & i = k\Delta, \text{ for } 1 \leq k \geq \frac{1}{2\delta} \\ \frac{\mu}{d} - ((|k| - 1)\Delta + 1)^2 & i = k\Delta, \text{ for } -\frac{1}{2\delta} \leq k \leq -1 \\ 0 & \text{otherwise} \end{cases}.$$

It is easy to verify that this candidate solution satisfies the constraints, and the corresponding value of the objective function is

$$\begin{aligned} & \mu - \delta \left(\sum_{i_1 \in \mathbb{Z}} y_{i_1}^{(1)} + \sum_{i_2 \in \mathbb{Z}} y_{i_2}^{(2)} + \cdots + \sum_{i_d \in \mathbb{Z}} y_{i_d}^{(d)} \right) \\ = \quad & \mu - \delta d \sum_{i_1 \in \mathbb{Z}} y_{i_1}^{(1)} \\ = \quad & \mu - \delta d \left(\sum_{i=0}^{\frac{1}{2\delta}} (\frac{\mu}{d} - i^2\Delta^2) + \sum_{i=0}^{\frac{1}{2\delta}-1} (\frac{\mu}{d} - (i\Delta + 1)^2) \right) \end{aligned}$$

$$
\begin{aligned}
&= \mu - \delta d\left((\frac{1}{2\delta}+1)\frac{\mu}{d} - \Delta^2\frac{\frac{1}{2\delta}(\frac{1}{2\delta}+1)(\frac{1}{\delta}+1)}{6} + \frac{1}{2\delta}\frac{\mu}{d} - \frac{1}{2\delta}\right.\\
&\qquad \left. -\Delta^2\frac{(\frac{1}{2\delta}-1)\frac{1}{2\delta}(\frac{1}{\delta}-1)}{6} - \Delta\frac{1}{2\delta}(\frac{1}{2\delta}-1)\right)\\
&= \mu - \delta d\left((\frac{1}{\delta}+1)\frac{\mu}{d} - \frac{\Delta^2\frac{1}{2\delta}(\frac{1}{2\delta^2}+1)}{3} - \frac{1}{2\delta} - \Delta\frac{1}{2\delta}(\frac{1}{2\delta}-1)\right)\\
&= \frac{d\Delta^2}{12\delta^2} + (\frac{1}{\Delta}-1)\frac{d\Delta^2}{4\delta} + \frac{1-\Delta}{2}d + \frac{d\Delta^2}{6}.
\end{aligned}
$$

Therefore, we have

$$
V_{LB} \geq \frac{d\Delta^2}{12\delta^2} + (\frac{1}{\Delta}-1)\frac{d\Delta^2}{4\delta} + \frac{1-\Delta}{2}d + \frac{d\Delta^2}{6}.
$$

□

C.11 Proof of Theorem 4.25

Proof of Theorem 4.25. Consider a candidate solution with $\mu = \frac{d\Delta\log\frac{3}{2}}{\beta}$ (assuming $k \triangleq \frac{\mu}{d\Delta}$ is an integer), and for all $m \in \{1, 2, \ldots, d\}$,

$$
y_i^{(m)} = \begin{cases} 0 & i \leq -k\Delta \\ e^{\beta}y_{i-\Delta}^{(m)} + 1 & i \in [-k\Delta+1, 0] \\ \max(e^{\beta}y_{i-\Delta}^{(m)} - 1, 0) & i \geq 0 \end{cases}.
$$

It is easy to verify that the above candidate solution satisfies the constraints of the dual linear program. We can derive the analytical expression for y_i^m, which is

$$
y_i^{(m)} = \begin{cases} 0 & i \leq -k\Delta \\ \frac{e^{(k-j)\beta}-1}{e^{\beta}-1} & i \in [-(j+1)\Delta+1, -j\Delta], \text{for } j \in [0, k-1] \\ \max(e^{j\beta}\frac{e^{k\beta}-2}{e^{\beta}-1} + \frac{1}{e^{\beta}-1}, 0) & i \in [(j-1)\Delta+1, j\Delta] \end{cases}.
$$

To avoid integer-rounding issues, assume that $n \triangleq \frac{1}{\beta}\log\frac{1}{2-e^{k\beta}} = \frac{\log 2}{\beta}$ is an integer. Then the value of the objective function with this candidate solution is

$$
\mu - \beta\left(\sum_{i_1\in\mathbb{Z}} y_{i_1}^{(1)} + \sum_{i_2\in\mathbb{Z}} y_{i_2}^{(2)} + \cdots + \sum_{i_d\in\mathbb{Z}} y_{i_d}^{(d)}\right)
$$

$$
\begin{aligned}
=&\mu - \beta d \sum_{i_1 \in \mathbb{Z}} y_{i_1}^{(1)} \\
=&\mu - \beta d \Delta \left(\sum_{i=1}^{k} \frac{e^{i\beta} - 1}{e^{\beta} - 1} + \sum_{i=1}^{n} (e^{i\beta} \frac{e^{k\beta} - 2}{e^{\beta} - 1} + \frac{1}{e^{\beta} - 1}) \right) \\
=&\mu - \beta d \Delta \left(\frac{\frac{e^{\beta}(1-e^{k\beta})}{1-e^{\beta}} - k}{e^{\beta} - 1} + \frac{e^{k\beta} - 2}{e^{\beta} - 1} \frac{e^{\beta}(1 - e^{n\beta})}{1 - e^{\beta}} + \frac{n}{e^{\beta} - 1} \right) \\
=&\frac{d\Delta \log \frac{3}{2}}{\beta} - \beta d \Delta \left(\frac{\frac{e^{\beta}(1-\frac{3}{2})}{1-e^{\beta}} - \frac{\log \frac{3}{2}}{\beta}}{e^{\beta} - 1} + \frac{-\frac{1}{2}}{e^{\beta} - 1} \frac{e^{\beta}(1 - 2)}{1 - e^{\beta}} + \frac{\log 2}{\beta(e^{\beta} - 1)} \right) \\
=&\frac{d\Delta \log \frac{3}{2}}{\beta} - \beta d \Delta \left(\frac{e^{\beta}}{2(e^{\beta} - 1)^2} - \frac{\log \frac{3}{2}}{\beta(e^{\beta} - 1)} - \frac{e^{\beta}}{2(e^{\beta} - 1)^2} + \frac{\log 2}{\beta(e^{\beta} - 1)} \right) \\
=&\Theta \left(\frac{d\Delta}{\beta} (\log \frac{3}{2} - \frac{1}{2} + \log \frac{3}{2} + \frac{1}{2} - \log 2) \right) \\
=&\log \frac{9}{8} \Theta \left(\frac{d\Delta}{\beta} \right) \\
\approx&\Theta \left(0.1178 \frac{d\Delta}{\beta} \right),
\end{aligned}
$$

as $\beta \triangleq \max(\epsilon, \delta) \to 0$.

Therefore,

$$
\lim_{\max(\epsilon,\delta) \to 0} \frac{V'_{LB}}{\frac{d\Delta}{\beta}} \geq \log \frac{9}{8} \approx 0.1178.
$$

□

C.12 Proof of Theorem 4.26

Proof of Theorem 4.26. Let $\alpha = \frac{3}{2}$. Consider a candidate solution with $\mu = \frac{d\Delta^2 \log^2 \alpha}{\beta^2}$ (assuming $k \triangleq \frac{\sqrt{\frac{\mu}{d}}}{\Delta} = \frac{\log \alpha}{\beta}$ is an integer), and for all $m \in \{1, 2, \ldots, d\}$,

$$
y_i^{(m)} = \begin{cases} 0 & i \leq -k\Delta \\ e^{\beta} y_{i-\Delta}^{(m)} + 2|i| + 1 & i \in [-k\Delta + 1, 0] \\ \max(e^{\beta} y_{i-\Delta}^{(m)} - (2i + 1), 0) & i \geq 0 \end{cases}.
$$

It is easy to verify that the above candidate solution satisfies the constraints of the dual linear program.

Define

$$
\begin{aligned}
z_1 &= \frac{2}{e^{\beta}-1},\\
z_2 &= \frac{1-\frac{2e^{\beta}\Delta}{e^{\beta}-1}}{e^{\beta}-1},\\
z_3 &= \frac{2}{1-e^{\beta}},\\
z_4 &= \frac{1-\frac{2e^{\beta}\Delta}{1-e^{\beta}}}{1-e^{\beta}}.
\end{aligned}
$$

We can derive the analytical expression for y_i^m, which is

$$
y_i^{(m)} = \begin{cases} 0 & i \le -k\Delta \\ e^{(k-k')\beta}\left(z_1(k\Delta+j)+z_2\right) - z_1(k'\Delta+j) - z_2 & i = -(k'\Delta+j) \end{cases},
$$

where $k' \in [0, k-1], j \in [0, \Delta-1]$, and for $i = (m-1)\Delta + j$, where $j \in [1, \Delta], m \ge 1$,

$$
y_i^{(m)} = \max(a_{m,j}, 0),
$$

where

$$
\begin{aligned}
a_{m,j} \triangleq\, & e^{m\beta}\left(z_1(k\Delta+\Delta-j)+z_2\right) - z_1(\Delta-j) - z_2 - z_3(\Delta-j) + z_4) \\
& - z_4 - z_3((m-1)\Delta+j).
\end{aligned}
$$

For each $j \in [1, \Delta]$, and we are interested in finding the number $m(j)$ such that $a_{m(j),j} = 0$. As $\beta \to 0$, from $a_{m(j),j} = 0$, we get

$$
e^{m(j)\beta} e^{k\beta}\left(\frac{2}{\beta}k\Delta - \frac{2\Delta}{\beta^2}\right) = -\frac{2\Delta}{\beta^2} - \frac{2}{\beta}m(j)\Delta + o\left(\frac{1}{\beta^2}\right).
$$

Therefore,

$$
m(j) = \frac{\log\gamma}{\beta} + o\left(\frac{1}{\beta}\right),
$$

where γ is the solution to

$$
\gamma\alpha(\log\alpha - 1) = -(1 + \log\gamma).
$$

When $\alpha = \frac{3}{2}$, we have $\gamma \approx 1.7468$.

Therefore, the value of the objective function is

$$
\begin{aligned}
&\mu - \beta\left(\sum_{i_1\in\mathbb{Z}} y_{i_1}^{(1)} + \sum_{i_2\in\mathbb{Z}} y_{i_2}^{(2)} + \cdots + \sum_{i_d\in\mathbb{Z}} y_{i_d}^{(d)}\right)\\
=&\mu - \beta d \sum_{i_1\in\mathbb{Z}} y_{i_1}^{(1)}\\
=&\mu - \beta d\left(\sum_{k'=0}^{k-1}\sum_{j=0}^{\Delta-1} y_{-(k'\Delta+j)}^{(1)} + \sum_{j=1}^{\Delta}\sum_{m=1}^{m(j)} y_{(m-1)\Delta+j}^{(1)}\right)\\
=&\frac{d\Delta^2\log^2\alpha}{\beta^2} -\\
&\beta d\left(\frac{1-e^{-k\beta}}{1-e^{-\beta}}e^{k\beta}((z_1k\Delta+z_2)\Delta + z_1\frac{\Delta(\Delta-1)}{2}) - z_1\Delta^2\frac{k(k-1)}{2} - z_1k\frac{\Delta(\Delta-1)}{2} - z_2k\Delta\right)\\
&-\beta d\sum_{j=1}^{\Delta}(\frac{e^{\beta}(1-e^{m(j)\beta})}{1-e^{\beta}}(e^{k\beta}(z_1(k\Delta+\Delta-j)+z_2) - z_1(\Delta-j)\\
&-z_2 - z_3(\Delta-j) + z_4) - z_4m(j) - z_3\Delta\frac{m(j)(m(j)+1)}{2} + z_3(\Delta-j)m')\\
=&\frac{d\Delta^2}{\beta^2}(\log^2\alpha - (\alpha-1)(2\log\alpha-2) + \log^2\alpha - 2\log\alpha + (1-\gamma)\alpha(2\log\alpha-2)\\
&-2\log\gamma - \log^2\gamma) + o(\frac{1}{\beta^2})\\
=&\frac{d\Delta^2}{\beta^2}\left(2\log^2\alpha - 2 - 2\alpha\gamma\log\alpha + 2\alpha\gamma - 2\log\gamma - \log^2\gamma\right) + o(\frac{1}{\beta^2})\\
\approx&0.0177\frac{d\Delta^2}{\beta^2} + o(\frac{1}{\beta^2}),
\end{aligned}
$$

as $\beta \triangleq \max(\epsilon,\delta) \to 0$.

Therefore,

$$\lim_{\max(\epsilon,\delta)\to 0} \frac{V'_{LB}}{\frac{d\Delta^2}{\beta^2}} \geq 0.0177.$$

□