Schrödinger Mechanisms: Optimal Differential Privacy Mechanisms for Small Sensitivity

Wael Alghamdi**, Shahab Asoodeh‡, Flavio P. Calmon*, Oliver Kosut†, Lalitha Sankar†, and Fei Wei†

***Abstract*—We consider the problem of designing optimal differential privacy mechanisms with a favorable privacy-utility tradeoff in the limit of a large number n of compositions (i.e., sequential queries). Here, utility is measured by the average distance between the mechanism's input and output, evaluated by a cost function c. We show that if n is sufficiently large and the sensitivities of all queries are small, then the optimal mechanism is additive and the noise probability density function is fully characterized by the ground-state eigenfunction of the Schrödinger operator with potential c. This leads to a family of optimal mechanisms, dubbed the Schrödinger mechanisms, depending on the choice of the cost function. Instantiating this result, we demonstrate that for $c(x) = x^2$ the Gaussian mechanism is optimal, and for $c(x) = |x|$, the optimal mechanism is obtained by the Airy function, thereby leading to the introduction of the Airy mechanism.**

This paper is Part II of a pair of papers, where Part I is [1]. The full proofs can be found in the appendices.

I. Introduction

Differential privacy (DP) [2] provides provable privacy guarantees when releasing the outcome of a query computed on a sensitive dataset. DP has gained traction in the machine learning (ML) community, and it has led to several privacy-preserving mechanisms implemented in practice by Google [3], Apple [4], and Facebook [5]. The parameters of these mechanisms are determined by the desired level of privacy and the query's sensitivity, denoted by s. When incorporating DP into ML algorithms, one fundamental challenge is to accurately characterize the privacy loss in iterative algorithms. To address this challenge, numerous composition results have been proved in the literature, e.g., [6]–[15].

In this paper, we view composition problems from a different angle: Instead of assuming access to constituent mechanisms, we seek to *construct* a DP mechanism whose n-fold composition exhibits the *optimal* privacy guarantee among all possible mechanisms. We investigate this problem under two assumptions: (1) large number of compositions n, and (2) small query's sensitivity s. The first assumption is inspired by iterative training procedures for machine learning models such as stochastic gradient descent, where a dataset is queried many (often thousands) times in order to update model parameters (e.g., weights in of a neural network). The second assumption holds when queries are computed over large datasets since a query's sensitivity is usually inversely proportional to the size of the dataset.

In a companion paper [1], we demonstrate that for sufficiently large n, *additive* mechanisms are in fact optimal, *i.e.*, we may restrict attention to mechanisms of the form $P_{Y|X=x} = T_x P$ for some Borel probability measure P on $\mathbb{R}$, where T_x denotes the shift operator defined as $(T_x P)(\,\cdot\,) := P(\,\cdot - x)$. Moreover, we show in [1] that an optimal P is obtained by solving the following minimax problem

$$\inf_{\mathbb{E}_P[c]\le C} \sup_{|a|\le s} D(P\|T_a P) \tag{1}$$

where the outer minimization is done over all Borel probability measures P on $\mathbb{R}$ satisfying $\int_{\mathbb{R}} c\,\mathrm{d}P \le C$. By solving this minimax optimization problem in [1], we introduce therein a family of distributions (dubbed the cactus distribution) and show that they can get arbitrarily close to optimal under large compositions for any *fixed* $s > 0$. In this work, we focus on the small-sensitivity regime, i.e., where $s \ll 1$, in which case we derive closed-form optimal distributions that we call the Schrödinger mechanism.

Our approach relies on the folklore approximation $D(T_x P\|T_{x'} P) \approx \frac{1}{2} I(P)|x - x'|^2$ when x and x' are sufficiently close [16, Section 2.6], where $I(P)$ is the Fisher information. Consequently, the minimax optimization of KL-divergence in (1) reduces to finding a *unique* minimizer of $I(P)$ (for the location parameter) over all probability measures P satisfying the utility constraint (cf. Section III). This reduced formulation reveals a remarkable characterization of the optimizer $P^\star$: its square root is the eigenfunction of the Schrödinger operator corresponding to the smallest eigenvalue (cf. Theorem 1). This general characterization provides a powerful tool to identify closed-form DP mechanisms with the optimal privacy-utility trade-off where the utility is measured via the cost function c. In particular, we show that $P^\star$ is the Gaussian measure for the L^2 cost function (cf. Theorem 1), thereby proving that the Gaussian mechanism is optimal in this sense in the small-sensitivity regime. Our results also show that $P^\star$ for the L^1 cost is given by the Airy function, leading to the introduction of a new optimal DP mechanism, which we call the *Airy mechanism* (cf. Definition 4).

*Corresponding author, remaining authors in alphabetical order. *W. Alghamdi and F. P. Calmon are with the School of Engineering and Applied Science, Harvard University (emails: alghamdi@g.harvard.edu, flavio@seas.harvard.edu)

‡ S. Asoodeh is with the Department of Computing and Software, McMaster University (email: asoodehs@mcmaster.ca)

† O. Kosut, L. Sankar, and F. Wei are with the School of Electrical, Computer, and Energy Engineering, Arizona State University (emails: {okosut,lsankar,fwei16}@asu.edu)

This material is based upon work supported by the National Science Foundation under Grant Nos. CAREER-1845852, CIF-1900750, CIF-1815361, CIF-1901243, CIF-1908725, CIF-2007688, CIF-2134256, and CIF-2031799.

A. Related Work

Characterizing optimal mechanisms has been a central problem in the DP literature. Several optimal mechanisms in DP settings are known, e.g., stair-case mechanism [17]–[19], geometric mechanism [20], discrete Laplace mechanism [21], truncated Laplace mechanism [22], and uniform mechanism [23], to name a few. All these works assume a query with a given sensitivity in a single-shot setting (i.e., no compositions). Unlike these works, our work focuses on characterizing optimal mechanisms under large composition when the query's sensitivity is rather small.

The connection between DP and the Schrödinger equation, to the best of our knowledge, is new; however, the component connections have been noted in some form in the literature. Nevertheless, our work serves to make the existing results more complete and rigorous. For instance, the statistics literature is rife with results on Fisher-information-minimizing distributions. The Cramér-Rao bound implies that Gaussian measures have the smallest Fisher information among all densities with a given variance. The minimizer over compactly-supported distributions or over those supported on $\mathbb{R}^+$ were characterized in [24] and [25], respectively. Kagan [26] studied the same problem for densities on $\mathbb{R}$ with fixed first and second moments, which was later extended to other moments by Ernst [27]. A connection between minimizing Fisher information and the Schrödinger equation has been established in [28, Example 5.1]. We discuss how our work differs from the existing literature in more detail in Remark 4.

B. Notation and Assumptions

We let λ denote the Lebesgue measure on $\mathbb{R}$. The set of all regular conditional distributions from $\mathbb{R}$ into $\mathbb{R}$ is denoted by $\mathscr{R}$. The set of all Borel probability measures on $\mathbb{R}$ is denoted by $\mathscr{B}$. The set of all probability density functions (PDFs) on $\mathbb{R}$ is denoted by $\mathcal{P}$. For $P \in \mathscr{B}$ and $c : \mathbb{R} \to \mathbb{R}$, the expectation is denoted by $\mathbb{E}_P[c] := \int_{\mathbb{R}} c \, \mathrm{d}P$. If $P \ll \lambda$ with density p, we denote $\mathbb{E}_p[c] := \mathbb{E}_P[c]$. The Fisher information of $p \in \mathcal{P}$ is denoted by $I(p)$, i.e., if p is absolutely continuous then

$$I(p) := \int_{\{x \in \mathbb{R} \,;\, p(x) > 0\}} \frac{p'(x)^2}{p(x)} \, dx, \tag{2}$$

and $I(p) = \infty$ otherwise. The shift operator is denoted by T_x, i.e., $(T_x r)(\,\cdot\,) := r(\,\cdot - x)$ for $r \in \mathcal{P} \cup \mathscr{B}$ and $x \in \mathbb{R}$. The KL-divergence is denoted by $D(P\|Q)$ or $D(p\|q)$ if $P, Q \in \mathscr{B}$ or $p, q \in \mathcal{P}$. It is well-known (see, e.g., [16, Section 2.6]) that, under mild regularity conditions on a PDF p, one has the expansion

$$D(p\|T_a p) = \frac{a^2}{2} I(p) + o(a^2) \quad \text{as } a \to 0. \tag{3}$$

Let $\mathcal{F} \subset \mathcal{P}$ denote the subset of PDFs satisfying (3).

As a particular case of the assumptions imposed in [1], we require throughout this paper that the cost function is

$$c(x) = |x|^\alpha, \tag{4}$$

where $\alpha > 0$ is a fixed constant. We note that our results hold under more general settings (see Remark 5).

II. From Differential Privacy to KL-Divergence

Let $f : \mathcal{D} \to \mathbb{R}$ be a query function, and $d \in \mathcal{D}$ a dataset containing sensitive data of several individuals. The quantity of interest is $f(d)$, the outcome of the query from dataset d. In order to protect the privacy of individuals against membership and inference attacks, a typical approach is to perturb $f(d)$ using a channel (or mechanism) $P_{Y|X}$ so that Y cannot be used to distinguish d from a neighboring dataset d' that differs from d in one entry. This approach, widely known as *differential privacy* [2], is formalized as follows. Given $\varepsilon \geq 0$ and $\delta \in [0, 1]$, a mechanism $P_{Y|X}$ is said to be (ε, δ)-differentially private (or (ε, δ)-DP for short) if

$$\sup_{d \sim d'} \sup_{A \subset \mathcal{Y}} \left[P_{Y|X=f(d)}(A) - e^{\varepsilon} P_{Y|X=f(d')}(A) \right] \leq \delta, \tag{5}$$

where the outer supremum is taken over all pairs of neighboring datasets d and d', denoted by $d \sim d'$, and the inner supremum is taken over all measurable subsets A of the support $\mathcal{Y}$ of Y. If a mechanism $P_{Y|X}$ is (ε, δ)-DP for sufficiently small ε and δ, then an adversary observing Y cannot accurately distinguish between small changes in d (e.g., if a certain user belongs to the original dataset d), thus providing a tunable privacy guarantee for each individual in d. A popular family of such DP mechanisms are *additive* ones, that is, $Y = f(d) + Z$ where $Z \sim P$ is a noise variable drawn from a distribution P. Alternatively, one can express additive mechanisms by $P_{Y|X=x} = T_x P$, where T_x denotes the shift operator defined as $(T_x P)(A) := P(A - x)$.

Now, consider a typical composition setting where a dataset d is queried n times with a query function f and a mechanism $P_{Y|X}$ is used n times to generate i.i.d. private copies[1] $Y^n = (Y_1, \ldots, Y_n)$ of $f(d)$. The mechanism in this setting can be viewed as $P_{Y^n|X^n=x^n} = \prod_i P_{Y|X=x_i}$ (known as n-fold composition of $P_{Y|X}$) with input $x^n = (f(d), \ldots, f(d))$ and output Y^n. We wish $P_{Y|X}$, on the one hand, to yield as private an n-fold composition as possible and, on the other hand, to generate each Y_i with minimal randomization (i.e., better utility). To formalize the latter goal, we impose the bound $\mathbb{E}[c(Y - x) \mid X = x] \leq C$ for all $x \in \mathbb{R}$ and a given $C \geq 0$, where $c : \mathbb{R} \to \mathbb{R}^+$ is a measurable cost function. Taking c to be the L^1 cost function, this constraint ensures that each Y_i is not too different from $f(d)$. Given the query f, we say that $P_{Y|X}$ is optimal if its n-fold composition is (ε, δ)-DP with the *smallest* ε and δ among n-fold compositions of any other mechanism $Q_{Y|X}$ satisfying $\mathbb{E}[c(Y - x) \mid X = x] \leq C$ for all $x \in \mathbb{R}$. Notice that for additive mechanisms $P_{Y|X=x} = T_x P$, this constraint reduces to $\mathbb{E}_P[c] \leq C$.

In [1], we study characterizations of optimal mechanisms $P_{Y|X}$ under the assumptions that n is sufficiently large and the cost function satisfies some regularity conditions (see Assumption 1 in [1]). Specifically, it was shown that additive mechanisms are optimal. Furthermore, optimal noise distributions P (where $P_{Y|X=x} = T_x P$) are the solutions of the

[1]This setting is usually referred to as non-adaptive composition. One can instead consider n different mechanisms to adaptively generate independent (but not necessarily i.i.d.) $Y_1, \ldots, Y_n$. However, the non-adaptive case is in a sense a worst-case setting, and we focus on it.

following minimax optimization problem

$$\begin{aligned} \inf_{\substack{P\in\mathscr{B}\\ P\ll\lambda}} \quad & \sup_{|a|\le s} D(P\|T_aP) \\ \text{subject to} \quad & \mathbb{E}_P[c]\le C. \end{aligned} \tag{6}$$

In the next section, we develop machinery for solving this optimization problem when operating in the small sensitivity regime, i.e., $s \ll 1$.

III. From KL-Divergence to Fisher Information: The Small-Sensitivity Regime

It is well-known (see, e.g., [16, Section 2.6]) that, under mild regularity conditions on a PDF p, one has the expansion

$$D(p\|T_ap) = \frac{a^2}{2}I(p) + o(a^2) \quad \text{as } a\to 0. \tag{7}$$

We restrict attention in this section to PDFs satisfying this expansion. The precise definition of the optimality of a noise PDF for queries with small sensitivities is as follows.

Definition 1. Let $\mathcal{F}$ denote the set of PDFs satisfying expansion (7). We say that a PDF $p\in\mathcal{F}$ is *optimal in the small-sensitivity regime* for the cost function c and the cost bound C if $\mathbb{E}_p[c]\le C$, and for every other PDF $q\in\mathcal{F}$ (i.e., $\lambda(\{p=q\})=0$) satisfying $\mathbb{E}_q[c]\le C$ there is a constant $s=s(q)>0$ such that

$$D(p\|T_ap) < D(q\|T_aq) \quad \text{for every } 0<|a|<s. \tag{8}$$

If such a p exists, then it is unique, and we denote it by $p^\star_{c,C}$.

Remark 1. We note that for the Gaussian density $\varphi^\sigma(x) = e^{-x^2/(2\sigma^2)}/\sqrt{2\pi\sigma^2}$, we have $D(\varphi^\sigma\|T_a\varphi^\sigma) = a^2/(2\sigma^2)$. Thus, if one insists that the PDF p satisfies $D(p\|T_ap)\le D(\varphi^\sigma\|T_a\varphi^\sigma)$ for all small a, then the mapping $a\mapsto D(p\|T_ap)$ is necessarily differentiable at $a=0$ with vanishing derivative. In particular, one reasonably expects that desirable PDFs for the small-shift regime to satisfy the expansion (7).

In light of (7), the optimal PDF in the small sensitivity region $p^\star_{c,C}$ is the unique minimizer of the Fisher information

$$I(p^\star_{c,C}) = \inf_{\substack{p\in\mathcal{P}\\ \mathbb{E}_p[c]\le C}} I(p), \tag{9}$$

if such a unique minimizer exists. We will formally state the equivalence between $p^\star_{c,C}$ and the minimizer of the optimization problem (9) in Theorem 1.

IV. From Fisher Information to the Schrödinger Equation

Solving (the dual of) the Fisher information minimization problem reveals a bridge between differential privacy and the celebrated Schrödinger operator. This connection enables us to borrow tools from the rich theory of the Schrödinger equation and show that $p^\star_{c,C}$ is fully characterized by the minimal-eigenvalue eigenfunction of the Schrödinger operator (see Theorem 1) with the potential given by the cost function c. More specifically, $p^\star_{c,C}$ is identical to the distribution of a particle that is subjected to an energy potential given by c and is in the ground state.

A. The Schrödinger Equation

We begin by recalling the definition of the Schrödinger operator and some of its known properties.

Definition 2. Given a measurable $v:\mathbb{R}\to\mathbb{R}$, the Schrödinger operator $\mathcal{H}_v$ on $L^2(\mathbb{R})$ with potential v is defined as[2]

$$\mathcal{H}_v(y) := -y'' + vy. \tag{10}$$

As in [29, Section 2.4], we say $y\in L^2(\mathbb{R})$ is an eigenfunction of $\mathcal{H}_v$ if y is differentiable, y' is absolutely continuous, and there exists a constant E such that $\mathcal{H}_v(y)=Ey$ holds a.e.

The spectrum of $\mathcal{H}_v$ is discrete: if v is locally bounded and $\lim_{|x|\to\infty} v(x)=\infty$ then $L^2(\mathbb{R})$ has an orthonormal complete set consisting of eigenfunctions of $\mathcal{H}_v$ with eigenvalues $\{E_k\}_{k\in\mathbb{N}}$ such that $E_k\to\infty$ (see [29, Chapter 2, Theorem 3.1]). Moreover, one may order the E_k in an increasing fashion, and then the eigenfunction associated to E_k has exactly k zeros (see [29, Chapter 2, Theorem 3.5]). We are interested in the smallest eigenvalue E_0 and the associated eigenfunction, i.e., the ground-state eigenfunction.

Lemma 1. *For any $\theta>0$, there exists a unique unit-L^2-norm eigenfunction $y_{\theta,c}$ of $\mathcal{H}_{\theta c}$ satisfying $y_{\theta,c}(x)>0$ for all $x\in\mathbb{R}$. Further, $y_{\theta,c}$ is even, and its eigenvalue is the smallest eigenvalue of $\mathcal{H}_{\theta c}$.*

B. The Schrödinger Mechanism

Since $y_{\theta,c}$ is a Borel function satisfying $\|y_{\theta,c}\|_2=1$, we get $y^2_{\theta,c}\in\mathcal{P}$. We call $y^2_{\theta,c}$ the Schrödinger mechanism.

Definition 3. The *Schrödinger mechanism* given the cost function c and parameter $\theta>0$ is defined by $Y=X+Z$ for Z having the PDF y^2 where y is the unique unit-L^2-norm and strictly positive solution to the Schrödinger equation

$$y'' = (\theta c - E)y, \tag{11}$$

with E an arbitrary constant.

Remark 2. By Lemma 1, there is a unique E for which the ODE (11) is solvable, and the solution then is $y_{\theta,c}$.

The main result of this section is that the Schrödinger mechanism is optimal in the small sensitivity regime.

Theorem 1. *If $\theta>0$ satisfies $C=\mathbb{E}_{y^2_{\theta,c}}[c]$, then*

$$p^\star_{c,C} = y^2_{\theta,c}, \tag{12}$$

i.e., the Schrödinger mechanism given the cost function c and parameter θ is optimal in the small-sensitivity regime for the cost function c and cost bound C (see Definitions 1 and 3). Furthermore, $p^\star_{c,C}$ uniquely *minimizes the Fisher information over all possible PDFs:*

$$I\left(p^\star_{c,C}\right) = \min_{\substack{p\in\mathcal{P}\\ \mathbb{E}_p[c]\le C}} I(p). \tag{13}$$

[2]One may define $\mathcal{H}_v$ initially on compactly-supported $\mathcal{C}^\infty$ functions, then show that its closure is self-adjoint if v satisfies mild conditions (see [29, Chapter 2, Theorem 1.1]). In particular, this extension goes through if v is nonnegative (and measurable).

Remark 3. For the two examples we discuss in the next section, we give a reversing procedure producing θ given C that takes the form $\theta = aC^{-b}$ for absolute constants a and b.

Remark 4. The connection between Fisher information minimization and the Schrödinger equation has been previously noted in [28, Example 5.1] and [27]. Although there is no general statement (e.g., a theorem) in [28] showing a result similar to Theorem 1, one can distill from Section 4.5 in [28] a claim that roughly translates as follows. For a PDF p to uniquely minimize the Fisher information over all PDFs satisfying $\mathbb{E}_p[c] \le C$, it suffices to satisfy the following: p is strictly positive, absolutely continuous, and twice differentiable, the following integration by parts holds for the ratio $\psi = p'/p$

$$\int_{\mathbb{R}} \psi(x)(q'(x) - p'(x))\,dx = -\int_{\mathbb{R}} \psi'(x)(q(x) - p(x))\,dx \tag{14}$$

for *every* PDF q with $I(q) < \infty$ and $\mathbb{E}_q[c] \le C$, and there is a $\theta > 0$ such that $y = \sqrt{p}$ uniquely solves the Schrödinger equation $y'' = (\theta c - E)y$ with E being the smallest possible constant. Example 5.1 of [28] gives full details for the special case when $c(x) = -a \cdot 1_{|x|\le 1} + b \cdot 1_{|x|>1}$ (and notes the well-known case $c(x) = x^2$). On the other hand, the derivations in [27] assume without proof some of the above mentioned properties, such as positivity, smoothness, and the validity of the integration by parts in (14); there are no worked examples in [27]. We note that equation (14) should not be expected to hold for arbitrary cost c (which is what the expositions in [27], [28] require or assume).

Since Theorem 1 gives a general unconditional result, our work can be seen as a way to fill the gaps in [27], [28]. In the next section, we also provide a new *explicit* solution for the absolute-value cost case. Our method of proof deviates from those in [27], [28], where we borrow results from the quantum mechanics literature (such as [29]) to show that the needed properties for p can be derived instead of assumed. For instance, we show that the unqiue eigenfunction $y_{\theta,c}$ as given by Lemma 1 satisfies the following bound.

Lemma 2. *For any $\theta > 0$, we have that*

$$\limsup_{|x|\to\infty} \left| \frac{y'_{\theta,c}(x)}{y_{\theta,c}(x)\sqrt{c(x)}} \right| \le \sqrt{\theta}. \tag{15}$$

Remark 5. While we assume throughout this paper that $c(x) = |x|^\alpha$, examining the proofs of Lemma 2 and Theorem 1, we see that the derived results hold under weaker conditions on c. Specifically, the results hold if we allow c to be any function that is non-negative, even, monotonic in $|x|$, and satisfies the following: $c(x_0)$, $\int_{x_0}^{\infty} |c'|^2/|c|^{5/2}$, $\int_{x_0}^{\infty} |c''|/|c|^{3/2} \to \infty$ as $x_0 \to \infty$, and

$$c'(x) = o(c^{3/2}(x)),\ \ \sqrt{c(x)}e^{-\gamma\int_0^x \sqrt{c(t)}\,dt} \in L^1(\mathbb{R}) \tag{16}$$

for all $\gamma > 0$. Note that $c(x) = |x|^\alpha$ and $c(x) = \log(|x| + 1)$ satisfy these relaxed conditions.

V. From the Schrödinger Equation to the Gaussian and Airy Mechanisms

Next, we instantiate Theorem 1 for two different cost functions, namely the quadratic and absolute-value cost functions.

A. Quadratic cost: optimality of Gaussian

Consider first the quadratic cost function $c(x) = x^2$. By particularizing Theorem 1 to this case, we show that the Gaussian distribution is optimal in the small-sensitivity regime in the sense of Definition 1. This is a direct consequence of the Cramér-Rao bound, but we derive it here using Theorem 1. The Schrödinger to be solved becomes

$$y''(x) = \left(\theta x^2 - E\right) y(x). \tag{17}$$

Proposition 1. *Let $c(x) = x^2$. For any $C > 0$, we have*

$$p^\star_{c,C}(x) = \frac{1}{\sqrt{2\pi C}} e^{-x^2/(2C)}, \tag{18}$$

i.e., the Gaussian distribution is optimal in the small-sensitivity regime under a variance cost (see Definition 1).

B. Absolute value cost: optimality of Airy

We next consider the absolute value cost function $c(x) = |x|$. In this case, the eigenvalue problem $\mathcal{H}_{\theta c}(y) = Ey$ becomes

$$y''(x) = (\theta|x| - E)\, y(x), \tag{19}$$

for some $\theta > 0$. It will be useful to recall the definition of the Airy functions. The differential equation

$$y''(x) = xy(x) \tag{20}$$

has two linearly independent solutions, called the Airy functions [30, Chapter 9]. They are denoted by Ai and Bi, where Ai is the solution such that $\mathsf{Ai}(x) \to 0$ as $x \to \infty$; specifically, Ai is approximated as $\mathsf{Ai}(x) \sim e^{-2x^{3/2}/3}/(2\sqrt{\pi}x^{1/4})$. This function can be expressed by the improper Riemann integral

$$\mathsf{Ai}(x) = \frac{1}{\pi} \lim_{N\to\infty} \int_0^N \cos\left(\frac{t^3}{3} + xt\right) dt. \tag{21}$$

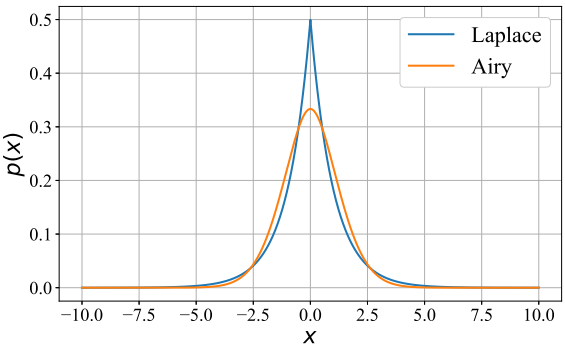
This function is analytic, and there are countably many zeros of Ai and Ai' all falling on the negative half-line. As is customary, the zeros of Ai and Ai' are denoted by $a_1 > a_2 > \cdots$ and $a'_1 > a'_2 > \cdots$, respectively. It is also known that approximately

$$a_1 = -2.33810, \quad a'_1 = -1.01879, \quad \text{and} \quad \mathsf{Ai}(a'_1) = 0.53565.$$

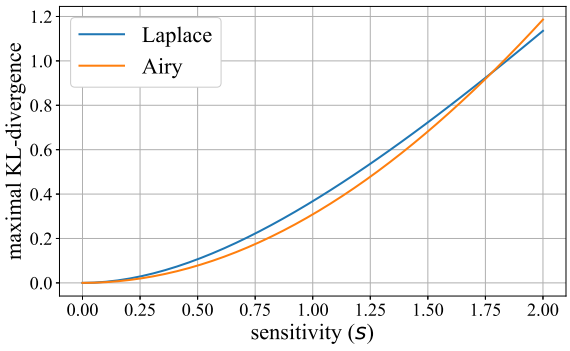
In particular, the function Ai is strictly positive and strictly decreasing over $[a'_1, \infty)$. We use the Airy function to construct the following density, which we show afterwards to be equal to $p^\star_{c,C}$.

Definition 4. For $C > 0$, we define the *Airy distribution* with first absolute moment C as the probability measure whose PDF $p_{\mathsf{Ai},C}$ is given by

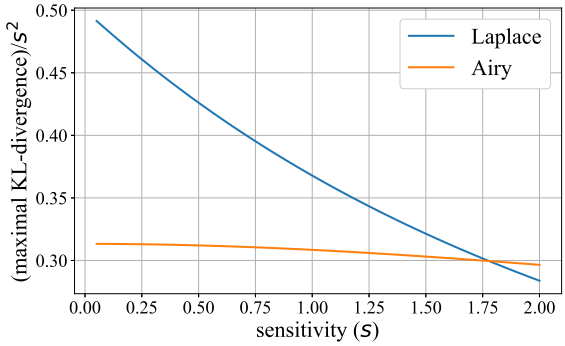
$$p_{\mathsf{Ai},C}(x) := \frac{1}{3C\mathsf{Ai}(a'_1)^2} \mathsf{Ai}\left(\frac{-2a'_1}{3C}|x| + a'_1\right)^2. \tag{22}$$



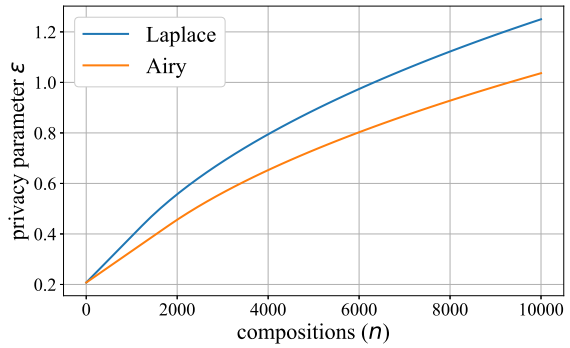
(a) The densities of the Laplace distribution and the Airy distribution, $p_{\mathsf{Ai},C}(x)$ (introduced in Definition 4). Both of these densities have absolute first moment equal to one.



(b) The achieved $\sup_{|a|\le s} D(p\|T_a p)$ (the objective function of (6)) versus the sensitivity s, with L^1 cost constraint $\mathbb{E}_p[|Z|] \le 1$.



(c) A zoomed-in version of Fig. 1b, showing $\sup_{|a|\le s} D(p\|T_a p)/s^2$ versus sensitivity (s), with L^1 cost constraint $\mathbb{E}_p[|Z|] \le 1$.



(d) The privacy parameter ε versus the number of compositions, for L^1 constraint $\mathbb{E}_p[|Z|] \le 1$, $\delta = 10^{-8}$, sensitivity $s = 1$, subsampling rate $q = 0.01$. Curves computed using the moments accountant.

Fig. 1: Comparisons between the Laplace and Airy mechanisms.

Remark 6. It can be verified with some algebra that $p_{\mathsf{Ai},C}$ is a valid PDF and that its first absolute moment is indeed C.

In Fig. 1a, we illustrate the Airy distribution and compare it with the Laplace distribution. As per Theorem 1, the Airy distribution uniquely minimizes the Fisher information subject to an absolute value cost. If the cost bound is set to $C = 1$, then we obtain the minimal value

$$I(p_{\mathsf{Ai},C}) \approx 0.6266 < 1 = I(q), \tag{23}$$

where $q(x) := e^{-|x|}/2$ is the Laplace distribution. We note that the Airy distribution has a lighter tail than that of the Laplace distribution, where the exponential decay of the former is $e^{-2x^{3/2}/3}$ and that of the latter is e^{-x}. Further, since the Airy function Ai is strictly positive and strictly decreasing over $[a'_1, \infty)$, we see that the Airy PDF $p_{\mathsf{Ai},C}$ is even, strictly positive everywhere, and strictly decreasing over $[0, \infty)$.

Combining Definition 3 and Theorem 1, we now show that the Airy distribution is optimal in the small-sensitivity regime for the absolute-value cost.

Proposition 2. *Let $c(x) = |x|$. For any $C > 0$, we have that*

$$p^\star_{c,C} = p_{\mathsf{Ai},C}, \tag{24}$$

i.e., the Airy distribution is optimal in the small-sensitivity regime for the absolute-value cost constraint (see Definition 1).

VI. Numerical Results

We compare the Airy mechanism with the Laplace mechanism in several ways. In Figure 1b we compare the objective function in the optimization problem (6) (i.e., $\sup_{|a|\le s} D(p\|T_a p)$) between the two mechanisms. Figure 1c shows the same objective, but normalized by s^2, which is known to be approximated by Fisher information $I(p)$ for small s (see (7)). Normalizing by s^2 makes the behavior at small s more pronounced. As expected, the Airy mechanism outperforms the Laplace distribution over the small sensitivity regime, namely s less than about 1.75.

For a comparison of (ε, δ)-DP performance, we use the moments accountant method [8], based on the Rényi-divergence, to quantify the privacy achieved over n compositions. In Figure 1d, we fix $\delta = 10^{-8}$ and compute the achieved ε as a function of the number of compositions for both mechanisms. The privacy is computed with a subsampling rate[3] of $q = 0.01$. The two mechanisms satisfy the L^1 constraint $\mathbb{E}_p[|Z|] \le 1$, and we set the sensitivity $s = 1$. As seen in the figure, the Airy mechanism achieves better privacy (i.e., smaller ε for the same δ) in comparison to the Laplace mechanism.

[3]The Airy mechanism and, more generally, mechanisms derived from solving (1) are not optimized for subsampling. Nevertheless, we present results with subsampling since subsampling data prior to computing a query and adding noise is a standard practice in machine learning for amplifying the privacy guarantee [8], [31], [32].

REFERENCES

[1] W. Alghamdi, S. Asoodeh, F. Calmon, O. Kosut, L. Sankar, and F. Wei, "Cactus mechanisms: Optimal differential privacy mechanisms in the large-composition regime," 2022. [Online]. Available: https://github.com/WaelAlghamdi/DP-Cactus

[2] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Proc. Theory of Cryptography (TCC)*, Berlin, Heidelberg, 2006, pp. 265–284.

[3] Ú. Erlingsson, V. Pihur, and A. Korolova, "Rappor: Randomized aggregatable privacy-preserving ordinal response," in *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*. ACM, 2014, pp. 1054–1067.

[4] Differential privacy team Apple, "Learning with privacy at scale," 2017.

[5] D. Kifer, S. Messing, A. Roth, A. Thakurta, and D. Zhang, "Guidelines for implementing and auditing differentially private systems," *ArXiv*, vol. abs/2002.04049, 2020.

[6] C. Dwork, G. N. Rothblum, and S. Vadhan, "Boosting and differential privacy," in *51st Annual Symposium on Foundations of Computer Science*. IEEE, 2010, pp. 51–60.

[7] J. Murtagh and S. Vadhan, "The complexity of computing the optimal composition of differential privacy," in *Proc. Int. Conf. Theory of Cryptography*, 2016, pp. 157–175.

[8] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep learning with differential privacy," in *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, 2016, pp. 308–318.

[9] S. Asoodeh, J. Liao, F. P. Calmon, O. Kosut, and L. Sankar, "Three variants of differential privacy: Lossless conversion and applications," *IEEE Journal on Selected Areas in Information Theory*, vol. 2, no. 1, pp. 208–222, 2021.

[10] P. Kairouz, S. Oh, and P. Viswanath, "The composition theorem for differential privacy," in *Proceedings of the 32nd International Conference on Machine Learning*, F. Bach and D. Blei, Eds., vol. 37, 2015, pp. 1376–1385.

[11] S. Meiser and E. Mohammadi, "Tight on budget? tight bounds for r-fold approximate differential privacy," in *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '18, 2018, p. 247–264.

[12] J. Dong, A. Roth, and W. J. Su, "Gaussian differential privacy," *CoRR*, vol. abs/1905.02383, 2019. [Online]. Available: http://arxiv.org/abs/1905.02383

[13] A. Koskela, J. Jälkö, and A. Honkela, "Computing tight differential privacy guarantees using fft," in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2020, pp. 2560–2569.

[14] S. Gopi, Y. T. Lee, and L. Wutschitz, "Numerical composition of differential privacy," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

[15] A. Koskela, J. Jälkö, L. Prediger, and A. Honkela, "Tight differential privacy for discrete-valued mechanisms and for the subsampled gaussian mechanism using fft," in *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, A. Banerjee and K. Fukumizu, Eds., vol. 130. PMLR, 13–15 Apr 2021, pp. 3358–3366. [Online]. Available: https://proceedings.mlr.press/v130/koskela21a.html

[16] S. Kullback, *Information Theory and Statistics*. Wiley, 1959.

[17] Q. Geng and P. Viswanath, "The optimal noise-adding mechanism in differential privacy," *IEEE Transactions on Information Theory*, vol. 62, no. 2, pp. 925–951, 2015.

[18] Q. Geng, P. Kairouz, S. Oh, and P. Viswanath, "The staircase mechanism in differential privacy," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 7, pp. 1176–1184, 2015.

[19] J. Soria-Comas and J. Domingo-Ferrer, "Optimal data-independent noise for differential privacy," *Information Sciences*, vol. 250, no. Complete, pp. 200–214, 2013.

[20] A. Ghosh, T. Roughgarden, and M. Sundararajan, "Universally utility-maximizing privacy mechanisms," *SIAM Journal on Computing*, vol. 41, no. 6, pp. 1673–1693, 2012.

[21] Q. Geng and P. Viswanath, "Optimal noise adding mechanisms for approximate differential privacy," *IEEE Transactions on Information Theory*, vol. 62, no. 2, pp. 952–969, 2016.

[22] Q. Geng, W. Ding, R. Guo, and S. Kumar, "Tight analysis of privacy and utility tradeoff in approximate differential privacy," in *Proc. International Conference on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, S. Chiappa and R. Calandra, Eds., vol. 108, 2020, pp. 89–99.

[23] ——, "Optimal noise-adding mechanism in additive differential privacy," in *Proc. International Conference on Artificial Intelligence and Statistics*, K. Chaudhuri and M. Sugiyama, Eds., vol. 89, 2019, pp. 11–20.

[24] E. Uhrmann-Klingen, "Minimal fisher information distributions with compact-supports," *Sankhyā: The Indian Journal of Statistics*, vol. 57, no. 3, pp. 360–374, 1995.

[25] J. F. Bercher and C. Vignat, "On minimum fisher information distributions with restricted support and fixed variance," *Inf. Sci.*, vol. 179, no. 22, p. 3832–3842, 2009.

[26] A. M. Kagan, "Information property of exponential families," *Theory of Probability & Its Applications,*, vol. 30, no. 4, pp. 831–835, 1986.

[27] P. A. Ernst, "Minimizing fisher information with absolute moment constraints," *Statistics & Probability Letters*, vol. 129, pp. 167–170, 2017.

[28] P. J. Huber and E. M. Ronchetti, *Robust Statistics, Second Edition*. Wiley, 2009.

[29] F. A. Berezin and M. Shubin, *The Schrödinger Equation*. Dordrecht: Springer, 1991.

[30] "*NIST Digital Library of Mathematical Functions*," http://dlmf.nist.gov/, Release 1.1.4 of 2022-01-15, f. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, eds. [Online]. Available: http://dlmf.nist.gov/

[31] 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, p. 793–826, jun 2011.

[32] A. Beimel, K. Nissim, and U. Stemmer, "Characterizing the sample complexity of private learners," in *Proceedings of the 4th Conference on Innovations in Theoretical Computer Science*, 2013, p. 97–110.

[33] V. I. Bogachev, *Measure theory*. Berlin: Springer, 2007.

APPENDIX A
AUXILIARY LEMMAS

We prove in this appendix Lemmas 1 and 2, and we also introduce and prove the following lemma, which will be useful in the proof of Theorem 1 in the next appendix.

Lemma 3. *With $\mathcal{P}_0 \subset \mathcal{P}$ denoting the set of strictly positive PDFs, we have that*

$$\inf_{\substack{p\in\mathcal{P}_0\\ \mathbb{E}_p[c]\le C}} I(p) = \inf_{\substack{p\in\mathcal{P}\\ \mathbb{E}_p[c]\le C}} I(p). \tag{25}$$

A. Proof of Lemma 1

By [29, Chapter 2, Theorems 3.1 and 3.5], there is a minimal eigenvalue E_0 of $\mathcal{H}_{\theta c}$, which corresponds to a 1-dimensional eigenspace $\{\gamma y\}_{\gamma\in\mathbb{R}} \subset L^2(\mathbb{R})$ where $y \in L^2(\mathbb{R})$ has no zeros. Then, there is a unique $\gamma \in \mathbb{R}$ such that $\|\gamma y\|_2 = 1$ and $\gamma y(x) > 0$ for all $x \in \mathbb{R}$, namely, $\gamma := \operatorname{sgn}(y(0))/\|y\|_2$. Setting $y_{\theta,c} = \gamma y$ yields the desired uniqueness result. Further, this uniqueness yields that $y_{\theta,c}$ is even since $y_{\theta,c}(-x)$ also satisfies the same differential equation, so a normalized version of $y_{\theta,c}(x) + y_{\theta,c}(-x)$ does too.

B. Proof of Lemma 2

We will use the following asymptotic of $y_{\theta,c}$.

Theorem 2 ([29, Chapter 2, Theorem 4.6]). *Fix $\theta > 0$, and let E_0 be the eigenvalue associated with $y_{\theta,c}$. As $x_1, x - x_1 \to \infty$ or $x_1, x - x_1 \to -\infty$, we have the asymptotic*

$$y_{\theta,c}(x) \sim \frac{\exp\left(-\int_{x_1}^{x} \sqrt{\theta c(t) - E_0}\, dt\right)}{(\theta c(x))^{1/4}}. \tag{26}$$

We denote $y = y_{\theta,c}$ for readability. Denote $f = -y'/y$ and $g = \theta c - E_0$, and note that f satisfies the Riccati equation

$$-f' + f^2 = g. \tag{27}$$

With this notation, the eigenvalue equation for y is $y'' = gy$. Since c grows without bound, g is eventually strictly positive. Since y is strictly positive and $y'' = gy$, we conclude that y'' is eventually positive a.e., i.e., there is an N such that $\lambda(\{x \in (N,\infty)\ ;\ y''(x) < 0\}) = 0$. Since y' is absolutely continuous,

$$y'(t_1) - y'(t_2) = \int_{t_2}^{t_1} y''(t)\,dt \geq 0 \tag{28}$$

for all large t_1 and t_2 with $t_1 > t_2$, i.e., y' is eventually increasing. As y decays to zero at infinity, and as y' eventually increases, we infer that y' is eventually negative. Thus, f is eventually positive. We will show that, for all large x,

$$f(x) \leq \sqrt{2g(x)}, \tag{29}$$

which is equivalent to

$$\left|\frac{y'(x)}{y(x)}\right| \leq \sqrt{2\left(\theta c(x) - E_0\right)}. \tag{30}$$

This is enough to finish the proof of the lemma by evenness of y and c. Now, we show that (29) holds.

Set $h = \sqrt{2g}$, so we want to show that $f \leq h$ is eventually satisfied. Denote $z = f - h$. Differentiating and using $-f' + f^2 = g$ (see (27)), we obtain

$$-z' + z^2 + 2zh = h' - g. \tag{31}$$

Now, we note that $h' < g$ eventually holds. Indeed, as $x \to \infty$, we have that

$$\frac{h'(x)}{g(x)} = \frac{1}{\sqrt{2}}\frac{g'(x)}{g(x)^{3/2}} = \frac{\alpha\theta x^{\alpha-1}}{\sqrt{2}\left(\theta x^\alpha - E_0\right)^{3/2}} \sim \frac{\alpha x^{-(\alpha/2+1)}}{\sqrt{2\theta}}. \tag{32}$$

Thus, by (31), we eventually have $-z' < 0$, i.e., z is eventually strictly increasing. In fact, by the above argument, we have that z is strictly increasing over (x_0,∞) where

$$x_0 = \max\left(\left(\frac{4\alpha}{\sqrt{\theta}}\right)^{1/(1+\alpha/2)}, \left(\frac{2\max(E_0,0)}{\theta}\right)^{1/\alpha}\right). \tag{33}$$

Suppose, for the sake of contradiction, that there is an $x_1 > x_0$ such that $f(x_1) > h(x_1)$, i.e., $z(x_1) > 0$. Then, as z is strictly increasing over (x_0,∞), we have that $z(x) > 0$ for all $x \geq x_1$. In other words,

$$-\frac{y'(x)}{y(x)} > \sqrt{2g(x)} \tag{34}$$

for all $x \geq x_1$. Increase x_1 if necessary so that $y(x) < 1$ for $x \geq x_1$. Then,

$$y(x) \leq \exp\left(-\int_{x_2}^{x} \sqrt{2g(t)}\,dt\right) \tag{35}$$

for all $x > x_2 \geq x_1$. Let x_3 and x_4 satisfying $x_4 > x_3 > x_1$ be such that

$$y(x) \geq \frac{1}{2g(x)^{1/4}} \exp\left(-\int_{x_3}^{x} \sqrt{g(t)}\,dt\right) \tag{36}$$

for every $x > x_4$. Then, for all $x > x_4$,

$$(\sqrt{2}-1)\int_{x_3}^{x} \sqrt{g(t)}\,dt \leq \log\left(2g(x)^{1/4}\right). \tag{37}$$

Denote

$$w(t) = \sqrt{(\sqrt{2}-1)\sqrt{g(t)}}, \tag{38}$$

so (37) can be rewritten as

$$\frac{\int_{x_3}^{x} w(t)^2\,dt}{\log\left(\gamma \cdot w(x)\right)} \leq 1, \tag{39}$$

where $\gamma := 2\sqrt{1+\sqrt{2}}$ is an absolute constant. To see that this leads to a contradiction, we take $x \to \infty$. By L'Hôpital's rule,

$$\lim_{x\to\infty} \frac{\int_{x_3}^{x} w(t)^2\,dt}{\log\left(\gamma \cdot w(x)\right)} = \lim_{x\to\infty} \frac{w(x)^3}{w'(x)} = \infty. \tag{40}$$

To see the last limit diverges, note that

$$\frac{w(x)^3}{w'(x)} \sim \eta x^{1+\alpha/2} \tag{41}$$

for some constant $\eta > 0$ depending only on α, θ, and E_0. The limit in (40) contradicts inequality (39). Thus, there is no $x_1 > x_0$ such that $f(x_1) > h(x_1)$. Hence, (29) eventually holds, and the proof is complete.

C. *Proof of Lemma 3*

For each $p \in \mathcal{P}$ and $\sigma > 0$, denote $p^\sigma(x) = p(x/\sigma)/\sigma$. Let ϕ denote the Gaussian density $\phi(x) := e^{-x^2/2}/\sqrt{2\pi}$.

We begin by showing that the limit

$$\lim_{\sigma\to 0^+} \mathbb{E}_{p*\phi^\sigma}[c] = \mathbb{E}_p[c] \tag{42}$$

holds for every PDF p that satisfies $\mathbb{E}_p[c] < \infty$.

Let (Ω,Σ,P) be a probability space and $Z, V : \Omega \to \mathbb{R}$ be independent random variables with PDFs p and ϕ, respectively, with respect to λ, i.e., with $P_Z(B) := P(Z^{-1}(B))$ and $P_V(B) := P(V^{-1}(B))$ we have

$$\frac{dP_Z}{d\lambda} = p, \qquad \frac{dP_V}{d\lambda} = \phi. \tag{43}$$

Then, for any $\sigma > 0$, the random variable $Z_\sigma := Z + \sigma V$ has PDF $p * \phi^\sigma$. Denote integration against P by $\mathbb{E}$; in particular,

$$\mathbb{E}[f(Z,V)] := \int_\Omega f(Z(\omega), V(\omega))\,dP(\omega) \tag{44}$$

for any Borel function $f : \mathbb{R}^2 \to \mathbb{R}$ whose integral is well-defined.

By Slutsky's theorem, we have that $Z_\sigma \to Z$ in distribution. By the continuous mapping theorem, we also have that $c(Z_\sigma) \to c(Z)$ in distribution. Thus, by the Lebesgue-Vitali theorem [33, Theorem 4.5.4], to conclude that (42) holds, it suffices to show uniform integrability of $\{c(Z_\sigma)\}_{0<\sigma\leq 1}$, i.e.,

it suffices to show that

$$\lim_{K\to\infty} \sup_{0<\sigma\le 1} \mathbb{E}\left[c(Z_\sigma)\cdot 1_{(K,\infty)}(c(Z_\sigma))\right] = 0. \tag{45}$$

To establish (45), it suffices to uniformly upper bound the $c(Z_\sigma)$ (for $\sigma \in (0,1]$) by an integrable random variable. To see this, note that if

$$\sup_{0<\sigma\le 1} c(Z_\sigma) \le U \tag{46}$$

for some random variable $U : \Omega \to \mathbb{R}$ with $\mathbb{E}[U] < \infty$, then we have the inequality

$$\sup_{0<\sigma\le 1} \mathbb{E}\left[c(Z_\sigma)\cdot 1_{(K,\infty)}(c(Z_\sigma))\right] \le \mathbb{E}\left[U\cdot 1_{(K,\infty)}(U)\right], \tag{47}$$

and the limit

$$\lim_{K\to\infty} \mathbb{E}\left[U\cdot 1_{(K,\infty)}(U)\right] = 0 \tag{48}$$

follows by absolute continuity of the Lebesgue integral in view of $\mathbb{E}[U] < \infty$.

Now, we show that a uniform bound as in (46) holds. Recall that for any $(z,v) \in \mathbb{R}^2$ and $0 < w < t$, denoting $\|(z,v)\|_w := (|z|^w + |v|^w)^{1/w}$, one has from Hölder's inequality that

$$\|(z,v)\|_t \le \|(z,v)\|_w \le 2^{\frac{1}{w}-\frac{1}{t}} \|(z,v)\|_t. \tag{49}$$

In particular, for any $r > 0$, denoting $\ell_r := \max(1, 2^{r-1})$, one has that

$$(|z| + |v|)^r \le \ell_r(|z|^r + |v|^r). \tag{50}$$

Therefore,

$$c(Z_\sigma) \le \ell_\alpha(|Z|^\alpha + |V|^\alpha) =: U. \tag{51}$$

As $\mathbb{E}[|Z|^\alpha] < \infty$ by assumption, and as $\mathbb{E}[|V|^\alpha] < \infty$, we see that U is an integrable upper bound on $\{c(Z_\sigma)\}_{0<\sigma\le 1}$. Hence, by absolute continuity of the Lebesgue integral, we obtain uniform integrability of the set $\{c(Z_\sigma)\}_{0<\sigma\le 1}$, so (42) follows by the Lebesgue-Vitali theorem.

Next, we show that the function $I_0^\star : \mathbb{R} \to [0,\infty]$ defined by

$$I_0^\star(C) := \inf_{\substack{p\in\mathcal{P}_0 \\ \mathbb{E}_p[c]\le C}} I(p) \tag{52}$$

is continuous at C. We may write

$$I_0^\star(C) = \inf_{p\in\mathcal{P}_0} I(p) + \mathbb{I}_{(-\infty,C]}\left(\mathbb{E}_p[c]\right). \tag{53}$$

Being the infimum of a jointly convex function over a convex set, $I_0^\star$ is convex. Further, this function is finite over $(0,\infty)$. To see that $I_0^\star(C)$ is finite, we only need to take p Gaussian with α-th moment less than C. Hence, being convex and finite, $I_0^\star$ is continuous over $(0,\infty)$.

Define

$$I^\star(C) := \inf_{\substack{p\in\mathcal{P} \\ \mathbb{E}_p[c]\le C}} I(p) \tag{54}$$

Now, fix $\varepsilon, \eta > 0$, and let $p \in \mathcal{P}$ be a PDF such that $\mathbb{E}_p[c] \le C$ and

$$I(p) \le I^\star(C) + \varepsilon. \tag{55}$$

Since the Fisher information satisfies the convolution inequality, we have

$$I(p * \phi^\sigma) \le I(p) \tag{56}$$

for every $\sigma > 0$. By the limit in (42), there is a $\sigma = \sigma(\eta)$ such that

$$\mathbb{E}_{p*\phi^\sigma}[c] \le \mathbb{E}_p[c] + \eta \le C + \eta. \tag{57}$$

Note that $p * \phi^\sigma \in \mathcal{P}_0$ by strict positivity of ϕ. Therefore,

$$I_0^\star(C+\eta) \le I(p*\phi^\sigma) \le I(p) \le I^\star(C) + \varepsilon. \tag{58}$$

By continuity of $I_0^\star$ at C, we may take $\eta \to 0^+$ to obtain

$$I_0^\star(C) \le I^\star(C) + \varepsilon. \tag{59}$$

By arbitrariness of ε, we deduce

$$I_0^\star(C) \le I^\star(C). \tag{60}$$

But the reverse inequality is trivial, thus equality is attained in (60), completing the proof of the lemma.

Appendix B
Proof of Theorem 1

We use the integration shorthand

$$\int_A f := \int_A f(x)\,dx. \tag{61}$$

Denote $y = y_{\theta,c}$ for short, and set $p = y^2$. We will show that $p^\star_{c,C} = p$. First, we note that C, which is defined by $C = \|y\|^2_{2,c}$, is indeed finite. This can be seen using the expansion of y given in Theorem 2, which yields

$$y(x)^2 c(x) = e^{-\Omega(x^{\alpha/2+1})}. \tag{62}$$

Denote the space of absolutely continuous functions on $\mathbb{R}$ by $\mathrm{AC}(\mathbb{R})$, and those that are locally absolutely continuous over $\mathbb{R}$ by $\mathrm{AC}_{\mathrm{loc}}(\mathbb{R})$. Let $L^2(\mathbb{R},c)$ be the weighted L^2-space of functions square-integrable against c. Consider the vector space

$$V := L^2(\mathbb{R}) \cap L^2(\mathbb{R},c) \cap \mathrm{AC}_{\mathrm{loc}}(\mathbb{R}). \tag{63}$$

Let E be the eigenvalue of y, so

$$y'' = (\theta c - E)y. \tag{64}$$

Consider the modified Dirichlet energy $\mathcal{E} : V \to \mathbb{R}\cup\{\infty\}$ defined by

$$\mathcal{E}(w) := \|w'\|_2^2 + \theta\|w\|^2_{2,c} - E\|w\|_2^2. \tag{65}$$

We start by showing that y is a global minimizer of $\mathcal{E}$, and

$$0 = \mathcal{E}(y) = \inf_{w\in V} \mathcal{E}(w). \tag{66}$$

Note that $y \in V$ since $y \in \mathcal{C}^1(\mathbb{R})$.

Fix an arbitrary $w \in V$, and we will show that $\mathcal{E}(w) \ge 0$. Since w is a.e. differentiable, we have $(y\cdot(w/y)')^2 \ge 0$ a.e. Rearranging this inequality, and noting the eigenvalue equation (64) satisfied by y, we obtain that a.e.

$$(w')^2 \ge \frac{2y'ww'}{y} - \frac{(y')^2w^2}{y^2} \tag{67}$$

$$= \left(\frac{y'w^2}{y}\right)' - \frac{y''w^2}{y} \tag{68}$$

$$= \left(\frac{y'w^2}{y}\right)' - (\theta c - E)w^2. \tag{69}$$

Note that $y'w^2/y \in \mathrm{AC}_{\mathrm{loc}}(\mathbb{R})$. Thus, integrating (69) over any $[-t,t]$ with $t>0$, we obtain

$$\|w'1_{[-t,t]}\|_2^2 \geq \left.\frac{y'w^2}{y}\right|_{-t}^{t} - \theta\|w1_{[-t,t]}\|_{2,c}^2 + E\|w1_{[-t,t]}\|_2^2. \tag{70}$$

Next, we show that there exists a sequence $t_n \nearrow \infty$ such that

$$\liminf_{n\to\infty} \left.\frac{y'w^2}{y}\right|_{-t_n}^{t_n} \geq 0. \tag{71}$$

This would readily yield $\mathcal{E}(w) \geq 0$ from inequality (70). By assumption, $w \in L^2(\mathbb{R},c)$, so symmetry of c implies

$$\int_0^\infty (w(x)^2 + w(-x)^2)c(x)\,dx = \int_{\mathbb{R}} w^2 c < \infty. \tag{72}$$

In particular, there is a sequence $\{t_n\}_{n\in\mathbb{N}} \subset (0,\infty)$ such that, as $n\to\infty$, we have $t_n \nearrow \infty$ and

$$\left(w(t_n)^2 + w(-t_n)^2\right) c(t_n) \to 0. \tag{73}$$

In addition, by the upper bound (15) in Theorem 2, there is an $A \in (0,\infty)$ such that

$$\left|\frac{y'(x)}{y(x)}\right| \leq A\cdot c(|x|) \tag{74}$$

holds for all large $|x|$. Then, for all large n,

$$\left.\frac{y'w^2}{y}\right|_{-t_n}^{t_n} = \frac{y'(t_n)w(t_n)^2}{y(t_n)} - \frac{y'(-t_n)w(t_n)^2}{y(-t_n)} \tag{75}$$

$$\geq -\left|\frac{y'(t_n)}{y(t_n)}\right| w(t_n)^2 - \left|\frac{y'(-t_n)}{y(-t_n)}\right| w(-t_n)^2 \tag{76}$$

$$\geq -Ac(t_n)\left(w(t_n)^2 + w(-t_n)^2\right). \tag{77}$$

Taking the limit inferior in (77) we obtain, in view of (73), that

$$\liminf_{n\to\infty} \left.\frac{y'w^2}{y}\right|_{-t_n}^{t_n} \geq 0. \tag{78}$$

In addition, by the assumption that $w \in L^2(\mathbb{R})$, the monotone convergence theorem implies

$$\lim_{n\to\infty} \theta\|w1_{[-t_n,t_n]}\|_{2,c}^2 - E\|w1_{[-t_n,t_n]}\|_2^2 = \theta\|w\|_{2,c}^2 - E\|w\|_2^2. \tag{79}$$

Taking the limit inferior of (70) along the t_n, and using (78) and (79) we conclude that

$$\|w'\|_2^2 \geq -\theta\|w\|_{2,c}^2 + E\|w\|_2^2. \tag{80}$$

As $w \in L^2(\mathbb{R},c)\cap L^2(\mathbb{R})$, (80) is equivalent to $\mathcal{E}(w)\geq 0$.

We have just shown that

$$\inf_{w\in V} \mathcal{E}(w) \geq 0. \tag{81}$$

On the other hand, we may show that $\mathcal{E}(y) = 0$. Indeed, as $y \in \mathcal{C}^1(\mathbb{R})$ and $y' \in \mathrm{AC}(\mathbb{R})$, we have that $yy' \in \mathrm{AC}_{\mathrm{loc}}(\mathbb{R})$. Note that $yy' = O(y^2 c)$ by the upper bound in Lemma 2. As $y \in L^2(\mathbb{R},c)$, we get $yy' \in L^1(\mathbb{R})$. Thus, there exist sequences $a_n, b_n \nearrow \infty$ such that $y(-a_n)y'(-a_n), y(b_n)y'(b_n) \to 0$. Therefore, we have that

$$\mathcal{E}(y)$$

$$= \|y'\|_2^2 + \theta\|y\|_{2,c}^2 - E\|y\|_2^2 \tag{82}$$

$$= \lim_{n\to\infty} \|y'1_{[-a_n,b_n]}\|_2^2 + \theta\|y1_{[-a_n,b_n]}\|_{2,c}^2 - E\|y1_{[-a_n,b_n]}\|_2^2 \tag{83}$$

$$= \lim_{n\to\infty} \int_{-a_n}^{b_n} (y')^2 + (\theta c - E)y^2 \tag{84}$$

$$= \lim_{n\to\infty} \int_{-a_n}^{b_n} (y')^2 + yy'' \tag{85}$$

$$= \lim_{n\to\infty} \int_{-a_n}^{b_n} (yy')' \tag{86}$$

$$= \lim_{n\to\infty} y(b_n)y'(b_n) - y(-a_n)y'(-a_n) \tag{87}$$

$$= 0, \tag{88}$$

where (83) follows by the monotone convergence theorem as $y \in L^2(\mathbb{R})\cap L^2(\mathbb{R},c)$. Thus, y globally minimizes $\mathcal{E}$ over V.

Next, we show that the already shown properties of y imply that p (which we defined by $p=y^2$ in the beginning of this proof) minimizes the Fisher information. For that, we consider first a couple of important quantities.

Define, for $\gamma \in \mathbb{R}$,

$$I_\gamma^\star := \inf_{\substack{w\in V \\ \|w\|_{2,c}^2 \leq \gamma \\ \|w\|_2 = 1}} 4\|w'\|_2^2. \tag{89}$$

The inequality $\|w\|_{2,c}^2 \leq \gamma$ in the definition of $I_\gamma^\star$ can be replaced with an equality. This follows by closedness of V under positive dilation, as we show now. Suppose $\gamma > 0$ (for otherwise the claim we are about to show is trivially satisfied). Consider $w \in V$, and let $u(x) = w(x/\sigma)/\sqrt{\sigma}$ where $\sigma > 0$ is a constant. Then, $u \in L^2(\mathbb{R})\cap \mathrm{AC}_{\mathrm{loc}}(\mathbb{R})$; in fact, $\|u\|_2 = \|w\|_2$. Further, $\|u\|_{2,c}^2 = \sigma^\alpha \|w\|_{2,c}^2$ and $\|u'\|_2^2 = \|w'\|_2^2/\sigma^2$. In particular, $w \in L^2(\mathbb{R},c)$ implies $u \in L^2(\mathbb{R},c)$, hence $u \in V$. Thus, if $\|w\|_{2,c}^2 = \gamma_1 < \gamma$, we may replace w with u where $\sigma = (\gamma/\gamma_1)^{1/\alpha} > 1$ to obtain the lower objective value $\|u'\|_2^2 < \|w'\|_2^2$ with the cost constraint $\|u\|_{2,c}^2 = \gamma$. In other words,

$$I_\gamma^\star := \inf_{\substack{w\in V \\ \|w\|_{2,c}^2 = \gamma \\ \|w\|_2 = 1}} 4\|w'\|_2^2. \tag{90}$$

Next, define

$$E^\star := \inf_{\gamma\in\mathbb{R}} I_\gamma^\star + \theta\gamma. \tag{91}$$

Our next goal is to show that $E \leq E^\star$. By (90), we use (91) to deduce that $E^\star$ satisfies

$$E^\star = \inf_{\substack{w\in V \\ \|w\|_2 = 1}} 4\|w'\|_2^2 + \theta\|w\|_{2,c}^2 \tag{92}$$

$$= \inf_{w\in V\setminus\{0\}} \frac{4\|w'\|_2^2 + \theta\|w\|_{2,c}^2}{\|w\|_2^2} \tag{93}$$

$$= E + \inf_{w\in V\setminus\{0\}} \frac{\mathcal{E}(w)}{\|w\|_2^2} \tag{94}$$

$$\geq E, \tag{95}$$

where (93) follows since V is a vector space, and (95) since $\inf_{w\in V}\mathcal{E}(w)\geq 0$ (see (81)).

Next, we deduce that $I(p)=I_C^\star$. Note that $p=y^2$ implies $(p')^2/p=4(y')^2$. Thus, $I(p)=4\|y'\|_2^2$. From $E\leq E^\star$ and the definition of $E^\star$ in (91), we obtain

$$E\|y\|_2^2=E\leq E^\star\leq I_C^\star+\theta C=I_C^\star+\theta\|y\|_{2,c}^2. \tag{96}$$

Adding $4\|y'\|_2^2-E\|y\|_2^2$ to both sides, we obtain

$$I(p)\leq I_C^\star+\mathcal{E}(y). \tag{97}$$

As $\mathcal{E}(y)=0$ (see (88)), we conclude that $I(p)\leq I_C^\star$. The reverse inequality also holds since $\|y\|_2=1$ and $\|y\|_{2,c}^2=C$, so we conclude that

$$I(p)=I_C^\star. \tag{98}$$

Next, we show that p globally minimizes the Fisher information, i.e., with $\mathcal{P}$ denoting the set of all possible PDFs, we show that

$$I(p)=\inf_{\substack{q\in\mathcal{P}\\ \mathbb{E}_q[c]\leq C}} I(q). \tag{99}$$

We start by showing that $I(p)$ is minimal among strictly positive PDFs. Denote the set of strictly positive PDFs by $\mathcal{P}_0$,

$$\mathcal{P}_0:=\{q\in\mathcal{P}\ ;\ q(x)>0 \text{ for every } x\in\mathbb{R}\}. \tag{100}$$

Note that, by definition of the Fisher information, $q\in\mathrm{AC}(\mathbb{R})$ if $I(q)<\infty$. Further, if $q\in\mathrm{AC}(\mathbb{R})$, then $\sqrt{q}\in\mathrm{AC}_{\mathrm{loc}}(\mathbb{R})$. Then, for every $q\in\mathcal{P}_0$ such that $I(q)<\infty$, setting $w=\sqrt{q}$, we get

$$I(q)=4\|w'\|_2^2. \tag{101}$$

Thus, we conclude from $I(p)=I_C^\star$ (see (98)) that

$$I(p)=\inf_{\substack{q\in\mathcal{P}_0\\ \mathbb{E}_q[c]\leq C}} I(q). \tag{102}$$

However, the same argument cannot be applied to a PDF q that has zeros. For this, we apply Lemma 3, to obtain from (102) that

$$I(p)=\inf_{\substack{q\in\mathcal{P}_0\\ \mathbb{E}_q[c]\leq C}} I(q)=\inf_{\substack{q\in\mathcal{P}\\ \mathbb{E}_q[c]\leq C}} I(q), \tag{103}$$

which is the global optimality of p claimed in (99).

Since p is strictly positive, we conclude that it is the unique minimizer of the Fisher information over all possible PDFs. Finally, note that $p\in\mathcal{F}$, i.e.,

$$D(p\|T_a p)=\frac{a^2}{2}I(p)+o(a^2) \tag{104}$$

as $a\to 0$. If $q\in\mathcal{F}$ is different from p, then $I(p)<I(q)$ and

$$D(q\|T_a q)=\frac{a^2}{2}I(q)+o(a^2). \tag{105}$$

Hence, there is an $s=s(q)>0$ such that $0<|a|<s$ implies

$$D(p\|T_a p)<D(q\|T_a q). \tag{106}$$

As this is true for all $q\in\mathcal{F}$, we deduce that

$$p=p_{c,C}^\star, \tag{107}$$

and the proof of the theorem is complete.

Appendix C
Proofs of Section V

A. Proof of Proposition 1

According to Theorem 1, we need to solve the following differential equation

$$y''(x)=\left(\theta x^2-E\right)y(x), \tag{108}$$

for some $\theta>0$ and eigenvalue $E>0$. It can be easily verified that

$$y_1(x)=\left(\frac{\sqrt{\theta}}{\pi}\right)^{1/4}e^{-x^2\cdot\sqrt{\theta}/2}, \tag{109}$$

solves (108) with the corresponding eigenvalue

$$E_0=\sqrt{\theta}. \tag{110}$$

Thus, by the uniqueness property in Lemma 1, the unit-norm, strictly-positive, ground-state eigenfunction, denoted by $y_{\theta,c}$ in Lemma 1, is given by y_1. Therefore, the optimal density is given by $y_{\theta,c}^2$. The corresponding cost for this density is therefore equal to

$$\left(\frac{\sqrt{\theta}}{\pi}\right)^{1/2}\int_{\mathbb{R}}x^2e^{-x^2\sqrt{\theta}}\,\mathrm{d}x=\frac{1}{2\sqrt{\theta}}. \tag{111}$$

To ensure that the incurred cost is equal to C, we thus need to choose

$$\theta=\frac{1}{4C^2}. \tag{112}$$

Plugging this into (109), we obtain

$$y_{\theta,c}(x)^2=\frac{1}{\sqrt{2\pi C}}e^{-x^2/(2C)}, \tag{113}$$

which, according to Theorem 1, is the optimal density $p_{c,C}^\star$. Thus, Gaussian density is optimal in the small-sensitivity regime.

B. Proof of Proposition 2

First, we notice that, according to Theorem 1, we need to solve the differential equation problem (19). Let

$$y_1=\sqrt{p_{\mathsf{Ai},C}}. \tag{114}$$

Thus, from Definition 4, we have

$$y_1(x)=\gamma\cdot\mathsf{Ai}\left(\theta^{1/3}|x|+a_1'\right), \tag{115}$$

where

$$\gamma:=\frac{1}{\sqrt{3C}\cdot\mathsf{Ai}(a_1')}, \tag{116}$$

and

$$\theta=\left(\frac{-2a_1'}{3C}\right)^3. \tag{117}$$

Differentiating separately for $x < 0$, $x = 0$, and $x > 0$, we obtain

$$y_1'(x) = \theta^{1/3}\gamma \, \mathrm{sgn}(x)\mathsf{Ai}'(\theta^{1/3}|x| + a_1'), \tag{118}$$

for every $x \in \mathbb{R}$ (where $\mathrm{sgn}(x) = x/|x|$ for $x \neq 0$, and $\mathrm{sgn}(0) = 0$). Thus, y_1' is absolutely continuous. Differentiating again, we obtain for every $x \in \mathbb{R}$

$$y_1''(x) = \theta^{2/3}\gamma\mathsf{Ai}''\left(\theta^{1/3}|x| + a_1'\right). \tag{119}$$

Since Ai is a solution of the differential equation of (20), it follows that $\mathsf{Ai}''(z) = z\mathsf{Ai}(z)$ for every $z \in \mathbb{R}$ and thus

$$y_1''(x) = \left(\theta|x| + \theta^{2/3}a_1'\right) y_1(x), \tag{120}$$

and hence y_1 solves the equation (19). Therefore, we conclude from Lemma 1 that $y_{\theta,c} = y_1$ is the ground-state eigenfunction of $\mathcal{H}_{\theta c}$. Moreover, since $\int_{\mathbb{R}} c y_{\theta,c}^2 = \int_{\mathbb{R}} c p_{\mathsf{Ai},C} = C$, Theorem 1 implies that $p^\star_{c,C} = p_{\mathsf{Ai},C}$, as desired.