Schrödinger Mechanisms: Optimal Differential Privacy Mechanisms for Small Sensitivity

Wael Alghamdi*[1], Shahab Asoodeh[2], Flavio P. Calmon[1], Juan Felipe Gomez[1], Oliver Kosut[3] and Lalitha Sankar[3]
[1] School of Engineering and Applied Sciences, Harvard University
(emails: alghamdi@g.harvard.edu, flavio@seas.harvard.edu, juangomez@g.harvard.edu)
[2] Department of Computing and Software, McMaster University (email: asoodehs@mcmaster.ca)
[3] School of Electrical, Computer, and Energy Engineering, Arizona State University (emails: {okosut, lsankar}@asu.edu)

***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 additive noise mechanism has probability density function 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 Airy mechanism.**

I. Introduction

Differential privacy (DP) [1] provides provable privacy guarantees for queries computed over sensitive data. Currently, DP is the standard definition used in privacy-preserving machine learning (ML) deployed in practice by, for example, Google [2], Apple [3], and Facebook [4]. 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., [5]–[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 has the *optimal* privacy guarantee among all possible mechanisms. We investigate this problem under two assumptions: (1) large number of compositions n, and (2) small values of the query sensitivity s. The first assumption is inspired by iterative training procedures for ML models such as stochastic gradient descent, where a dataset is queried many times (often in the thousands) in order to update model parameters (e.g., weights of a neural network). Thus, it is a natural assumption for the privacy analysis of private ML algorithms (see, e.g., [11], [15], [16]). The second assumption holds, for example, when we are interested in counting queries over large datasets, because in this case the sensitivity is inversely proportional to the size of the dataset.

Optimal DP mechanisms under the first assumption (i.e., in the large composition regime) have been recently characterized in [16]. The main technical result validating the approach of [16] is that the privacy guarantee of the n-fold composition of a mechanism $P_{Y|X}$ scales as [15]

$$n \cdot \sup_{|x-x'|\le s} D(P_{Y|X=x} \,\|\, P_{Y|X=x'}), \tag{1}$$

where $D(\,\cdot\,\|\,\cdot\,)$ denotes the KL-divergence. It follows from this observation that the optimal mechanism is the one that solves the following optimization problem:

$$\inf_{\mathbb{E}[c(Y-x)\,|\,X=x]\le C,\ \forall x\in\mathbb{R}} \ \sup_{|x-x'|\le s} \ D(P_{Y|X=x} \,\|\, P_{Y|X=x'}), \tag{2}$$

where the outer infimum is taken over all mechanisms $P_{Y|X}$ that satisfy the prescribed cost constraint dictated by a cost function c and a cost bound C (e.g., a bounded variance). Not surprisingly, the optimal mechanism is additive and continuous (see [16, Theorem 1]), thus (2) reduces to the following optimization over probability density functions (PDFs) p:

$$\inf_{p:\ \mathbb{E}_p[c]\le C} \ \sup_{|a|\le s} \ D(p \,\|\, T_a p), \tag{3}$$

where $(T_a p)(x) := p(x-a)$ denotes the shift operator. The so-called *cactus mechanisms*, were shown by [16] to achieve the optimal value in (3) to arbitrary accuracy and for *fixed* sensitivity $s > 0$. In this paper, we seek to solve (3) with vanishing sensitivity, i.e., $s \to 0^+$. We achieve this goal by a sequence of reductions: from KL-divergence to Fisher information and then to the Schrödinger equation. Thus, we name the ensuing family of optimal mechanisms the *Schrödinger mechanisms*.

We use the folklore expansion [17, Section 2.6]

$$D(p \,\|\, T_a p) = \frac{a^2}{2} I(p) + o(a^2) \quad \text{as } a \to 0, \tag{4}$$

where $I(p)$ is the Fisher information. Consequently, the minimax optimization of KL-divergence in (3) reduces to finding

*Corresponding author, and the remaining authors are in alphabetical order.

This material is based upon work supported by the National Science Foundation under Grant Nos. CAREER-1845852, FAI-2040880, CIF-1900750, SCH-2312666, CIF-1922971, and CIF-1901243, and the U.S. Department of Energy Award No. DE-SC0022158.

the *unique* minimizer of $I(p)$ over all PDFs p satisfying the utility constraint. This reduced formulation reveals a remarkable characterization of the optimizer $p^\star$: it is the square of the ground-state eigenfunction of a Schrödinger operator (Theorem 3). 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 PDF for the L^2 cost function (Proposition 3), 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* (see Definition 4).

A. Related Work and Contributions

Several optimal mechanisms in DP settings are known, e.g., stair-case mechanism [18]–[20], geometric mechanism [21], discrete Laplace mechanism [22], truncated Laplace mechanism [23], and uniform mechanism [24], 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, we focus on characterizing optimal mechanisms under large composition when the query's sensitivity is rather small. Although optimal mechanisms for the large-composition regime are treated in [16], the work therein considers *fixed* sensitivity.

Compared to existing literature on the problem of minimizing the Fisher information, we:

1) work with a larger class of cost functions,
2) do not restrict the support of the PDFs we optimize over,
3) do not require any regularity assumptions whatsoever on the PDFs we optimize over.

We go beyond existing literature by introducing a novel proof technique that does not depend on the calculus of variation, and also by deriving an estimate of the logarithmic derivative of the ground-state eigenfunction of the Schrödinger operator.

The statistics literature is rife with results on Fisher-information-minimizing distributions. The Cramér-Rao bound implies that Gaussian measures are optimal for a given variance. The minimizer over compactly-supported distributions or over those supported on $\mathbb{R}^+$ were characterized in [25] and [26], respectively. Kagan [27] studied the same problem for densities on $\mathbb{R}$ with fixed first and second moments, which was later extended to other moments by Ernst [28]. A connection between minimizing Fisher information and the Schrödinger equation has been established in [29, Example 5.1]. Formulating a privacy problem in terms of minimizing Fisher information has appeared in [30], [31], but not in a DP sense; rather, the analyses therein pertain to privacy-preserving battery charging methods to obfuscate household information, and the Fisher information itself is proposed as a privacy metric. Fisher information minimization in [30] is done for PDFs of compact support, and that is extended to unbounded support in [31] but for only a quadratic cost. Further, the PDFs considered in [30], [31] are assumed *a priori* to be twice continuously differentiable. Therefore, none of these previous works has a setup encompassing ours, namely, they minimize Fisher information: over PDFs supported over a compact set [25], [30] or over $\mathbb{R}^+$ [26]; assuming regularity of the PDFs [29]–[31]; or under a strictly smaller or different class of constraint functions [27], [28], [31].

We discuss in more detail how our work differs from the existing literature closest to ours [28]–[31] in Appendix A.

B. Notation and Assumptions

We let λ denote the Lebesgue measure on $\mathbb{R}$. The set of all probability density functions (PDFs) on $\mathbb{R}$ is denoted by $\mathcal{P}(\mathbb{R})$. For $p \in \mathcal{P}(\mathbb{R})$ and $c : \mathbb{R} \to \mathbb{R}$, the expectation is denoted by $\mathbb{E}_p[c] := \int_{\mathbb{R}} c(x)p(x)\,dx$. The shift operator is denoted by T_x, i.e., $(T_x r)(A) := r(A - x)$.

The Fisher information of $p \in \mathcal{P}(\mathbb{R})$ is denoted by $I(p)$, i.e., if p is absolutely continuous then

$$I(p) := \int_{\{x\in\mathbb{R}\,;\,p(x)>0\}} \frac{p'(x)^2}{p(x)}\,dx, \tag{5}$$

and $I(p) = \infty$ otherwise. The KL-divergence is denoted by $D(p\,\|\,q)$ if $p, q \in \mathcal{P}(\mathbb{R})$. The variance of the information density is denoted by (for $p, q \in \mathcal{P}(\mathbb{R})$)

$$\mathrm{V}(p \parallel q) := \mathbb{E}_p\left[\left(\log\frac{p}{q} - D(p\|q)\right)^2\right]. \tag{6}$$

It is well-known (see, e.g., [17, Section 2.6]) that, under mild regularity conditions on a PDF p, one has the expansion in (4). Define the subset of PDFs $\mathcal{F} \subset \mathcal{P}(\mathbb{R})$ by

$$\mathcal{F} := \Bigg\{p \in \mathcal{P}(\mathbb{R}) : (4) \text{ holds, } \sup_{|a|\le s} D(p\|T_a p) < \infty \text{ and } \sup_{|a|\le s} \mathrm{V}(p\|T_a p) < \infty \text{ for some } s > 0\Bigg\}. \tag{7}$$

The minimization problem we solve for Fisher information is global, i.e., over all of $\mathcal{P}(\mathbb{R})$, while the DP optimization we solve will be over the set $\mathcal{F}$ defined in (7).

The results of this paper hold for the following class of cost functions c. We note that this class includes functions such as $c(x) = \beta|x|^\alpha$ and $c(x) = \beta\log(|x|+1)^\alpha$ for any $\alpha, \beta > 0$.

Assumption 1. *The cost function $c : \mathbb{R} \to \mathbb{R}$ satisfies:*

1) Positivity: $c(x) \ge 0$ *for all* $x \in \mathbb{R}$, *and* $c(0) = 0$.
2) Symmetry: $c(x) = c(-x)$ *for all* $x \in \mathbb{R}$.
3) Monotonicity: $c(x_1) \le c(x_2)$ *if* $|x_1| \le |x_2|$.
4) Continuity: *c is continuous over* $\mathbb{R}$.
5) Unbounded: $c(x) \to \infty$ *as* $x \to \infty$,
6) Controlled derivative: $c'(x) = o\big(c(x)^{3/2}\big)$ *as* $x \to \infty$,
7) Tail regularity: $\int_{x_0}^\infty |c'|^2/|c|^{5/2}, \int_{x_0}^\infty |c''|/|c|^{3/2} < \infty$ *for some* $x_0 \in \mathbb{R}$,
8) Moderate growth: $x \mapsto \sqrt{c(x)}/\exp(\gamma\int_0^{|x|}\sqrt{c(t)}\,dt)$ *is integrable for all* $\gamma > 0$,
9) Additive/Multiplicative regularity: *there is a locally bounded strictly positive function ρ on $\mathbb{R}$ such that* $c(x-t), c(tx) \le \rho(t)(c(x)+1)$ *for all* $x, t \in \mathbb{R}$.

In the assumptions involving c' or c'', it is to be understood that c is required to be differentiable (or twice differentiable) *only* at large enough values.

II. From DP to KL-Divergence to Fisher Information

Let $\mathcal{D}$ be the collection of datasets, each of which contains sensitive data of several individuals, and $f : \mathcal{D} \to \mathbb{R}$ be a query function. The quantity of interest is $x = f(d)$, which is the outcome of the query f on the dataset $d \in \mathcal{D}$ (e.g., $f(d)$ could be the percentage of individuals in d falling inside a certain income bracket). To protect the privacy of individuals against membership and inference attacks, a typical approach is to perturb $f(d)$ using a channel (or mechanism) $P_{Y|X=f(d)}$ so that Y cannot be used to distinguish d from a neighboring dataset d' that differs from d in one entry. This approach, known as *differential privacy* [1], 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

$$\sup_{d\sim d'} \sup_{A\subset\mathcal{Y}} \left[P_{Y|X=f(d)}(A) - e^{\varepsilon} P_{Y|X=f(d')}(A)\right] \leq \delta, \tag{8}$$

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 d from an arbitrary neighboring d', thus providing a tunable privacy guarantee for each individual in d. A popular family of such DP mechanisms includes *additive* ones, that is, $Y = f(d) + Z$ where $Z \sim P$ is a noise variable drawn from a distribution P.[1]

We note that the DP definition in (8) can be more compactly expressed using the E_γ-divergence [32] defined for $\gamma \geq 0$ as

$$\mathsf{E}_\gamma(P \,\|\, Q) := \sup_{A \text{ Borel}} P(A) - \gamma Q(A). \tag{9}$$

With this definition at hand, we can say $P_{Y|X}$ is (ε,δ)-DP if

$$\sup_{|x-x'|\leq s} \mathsf{E}_{e^\varepsilon}(P_{Y|X=x} \,\|\, P_{Y|X=x'}) \leq \delta, \tag{10}$$

where s denote the sensitivity of the query f, defined as $s := \sup_{d\sim d'} |f(d) - f(d')|$.

Next, consider a typical composition setting where a dataset d is queried n times sequentially with query functions f_j, $1 \leq j \leq n$, and a mechanism $P_{Y|X}$ is used n times to generate the n-tuple $Y^n = (Y_1, \ldots, Y_n)$ as a private version of the n-tuple $(f_1(d), \cdots, f_n(d))$. For simplicity, we assume that each f_j has the same sensitivity s. Therefore, this n-fold composition $P_{Y|X}^{\otimes n}$ is $(\varepsilon, \delta_{P_{Y|X}^{\otimes n},s}(\varepsilon))$-DP for any $\varepsilon \geq 0$, where[2]

$$\delta_{P_{Y|X}^{\otimes n},s}(\varepsilon) := \sup_{\substack{|u_j - v_j|\leq s\\ 1\leq j\leq n}} \mathsf{E}_{e^\varepsilon}\left(\prod_{j=1}^n P_{Y|X=u_j} \,\middle\|\, \prod_{j=1}^n P_{Y|X=v_j}\right). \tag{11}$$

Equivalently, $P_{Y|X}^{\otimes n}$ is $(\varepsilon_{P_{Y|X}^{\otimes n},s}(\delta), \delta)$-DP for $\delta \in [0,1]$, where

$$\varepsilon_{P_{Y|X}^{\otimes n},s}(\delta) := \inf\left\{\varepsilon \geq 0 \;:\; \delta_{P_{Y|X}^{\otimes n},s}(\varepsilon) \leq \delta\right\}. \tag{12}$$

Since additive continuous channels were shown to be optimal in [16], we henceforth consider only channels of the form $P_{Y|X=x} = T_x P$ with P being absolutely continuous with respect to the Lebesgue measure λ, for which we use the simplified notation $\varepsilon_{p^{\otimes n},s}(\delta)$ where $p = dP/d\lambda$ is the PDF. We derive the following asymptotic formula for $\varepsilon_{P_{Y|X}^{\otimes n},s}(\delta)$.[3]

Theorem 1. *For any PDF $p \in \mathcal{P}(\mathbb{R})$ and $s > 0$ satisfying $\sup_{|a|\leq s} \mathrm{V}(p\|T_a p) < \infty$, and for any $\delta \in (0, 1/2)$, we have*

$$\lim_{n\to\infty} \frac{\varepsilon_{p^{\otimes n},s}(\delta)}{n} = \sup_{|a|\leq s} D(p\|T_a p). \tag{13}$$

According to this theorem, characterizing $\varepsilon_{p^{\otimes n},s}(\delta)$ for sufficiently large n boils down to computing the maximum of $D(p \,\|\, T_a p)$ over all $|a| \leq s$.

Analogous to [16], we address the utility of the mechanism $P_{Y|X}$ by imposing 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. Notice that for additive mechanisms $P_{Y|X=x} = T_x P$, this utility constraint reduces to $\mathbb{E}_P[c] \leq C$. Motivated by the asymptotic given in Theorem 1, we consider the following optimality in the small sensitivity regime.

Definition 1. Let $\mathcal{F} \subset \mathcal{P}(\mathbb{R})$ be as defined in (7).[4] 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 $\mathbb{E}_p[c] \leq C$, and for every other PDF $q \in \mathcal{F}$ (i.e., $\lambda(\{p = q\}) = 0$) satisfying $\mathbb{E}_q[c] \leq C$ there is a constant $s(q) > 0$ such that $0 < s < s(q)$ implies

$$\sup_{0<\delta<\frac{1}{2}} \lim_{n\to\infty} \frac{\varepsilon_{p^{\otimes n},s}(\delta)}{\varepsilon_{q^{\otimes n},s}(\delta)} < 1. \tag{14}$$

An immediate corollary of Theorem 1 is that the unique minimizer of the Fisher information is automatically the optimal PDF in the small-sensitivity regime.

Corollary 1. *If $p \in \mathcal{F}$ is the unique minimizer*

$$p = \operatorname*{argmin}_{\substack{q\in\mathcal{F}\\ \mathbb{E}_q[c]\leq C}} I(q), \tag{15}$$

[1]Alternatively, one can express additive mechanisms by $P_{Y|X=x} = T_x P$, where T_x denotes the shift operator defined as $(T_x P)(A) := P(A - x)$.

[2]While the sensitivity s is usually suppressed from the notation of δ and ε in the literature, we include it here since we consider a variable sensitivity.

[3]A similar result appears in [15] under additional third-moment constraints, and also under the assumption of existence of "worst-case shifts." Thus, our result can be seen as a generalization of the asymptotic formula in [15].

[4]For the Gaussian density φ^σ, we have $D(\varphi^\sigma\|T_a\varphi^\sigma) = a^2/(2\sigma^2)$. Thus, if one insists that the PDF p satisfy $D(p\|T_a p) \leq D(\varphi^\sigma\|T_a\varphi^\sigma)$ for all small a, then the mapping $a \mapsto D(p\|T_a p)$ is necessarily differentiable at $a = 0$ with vanishing derivative. In particular, one reasonably expects that desirable PDFs for the small-sensitivity regime to belong to $\mathcal{F}$.

then p is the optimal PDF in the small-sensitivity regime for the cost function c and the cost bound C.

We derive in the next section minimizers of Fisher information over *all* PDFs $\mathcal{P}(\mathbb{R})$, then we also show that such minimizers in fact fall within the set $\mathcal{F}$.

III. From Fisher Information to the Schrödinger Equation

Solving the Fisher information minimization problem reveals a bridge between DP and the Schrödinger operator. This connection enables us to show that the global minimizers of Fisher information are fully characterized by the minimal-eigenvalue eigenfunctions of the Schrödinger operator (see Theorem 2) with the potential given by the cost function c.

A. The Schrödinger Equation

We begin by recalling the setup of the Schrödinger operator eigen-problem and some of its known properties.

Definition 2 ([33, Section 2.4]). For 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[5]

$$\mathcal{H}_v(y) := -y'' + vy. \tag{16}$$

We say that $y \in L^2(\mathbb{R})$ 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 [33, 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 [33, Chapter 2, Theorem 3.5]). We are interested in the smallest eigenvalue E_0 and the associated eigenfunction, i.e., the ground-state eigenfunction. An easy consequence of known properties of the ground-state eigenfunction is as follows.

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 $\mathcal{H}_{\theta c}$.*

The notation $y_{\theta,c}$ as given by Lemma 1 will be used in the remainder of the paper.

B. Global Minimization of Fisher Information

Recall the recipe we provide in Section II for finding optimal DP mechanisms in the small-sensitivity regime:

1) globally minimize Fisher information (i.e., over $\mathcal{P}(\mathbb{R})$),
2) show that the solution in fact falls within $\mathcal{F}$,
3) use Theorem 1 to conclude that the Fisher information global minimizer is the optimal DP mechanism.

We carry out step 1 in Theorem 2 below, where we show that $y_{\theta,c}^2$ is the unique global minimizer of the Fisher information. After that, we complete our general derivations in Proposition 2 by showing that step 2 holds, i.e., $y_{\theta,c}^2 \in \mathcal{F}$.

Theorem 2. *Suppose c satisfies Assumption 1, fix $\theta > 0$, consider the PDF $p = y_{\theta,c}^2$, and set $C = \mathbb{E}_p[c]$. Then, the PDF p uniquely minimizes the Fisher information among all PDFs $q \in \mathcal{P}(\mathbb{R})$ that satisfy $\mathbb{E}_q[c] \le C$, i.e.,*

$$p = \underset{\substack{q\in\mathcal{P}(\mathbb{R}) \\ \mathbb{E}_q[c]\le C}}{\operatorname{argmin}}\ I(q). \tag{17}$$

Since Theorem 2 gives a general unconditional result, our work can be seen as a way to fill the gaps in [28]–[31]. 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 [28]–[31], where we borrow results from the quantum mechanics literature (such as [33]) to show that the needed properties for p can be derived instead of assumed. For instance, we show that the unique eigenfunction $y_{\theta,c}$ as given by Lemma 1 satisfies the following bound.

Proposition 1. *For c satisfying Assumption 1 and any $\theta > 0$, we have the bound*

$$\limsup_{|x|\to\infty} \left| \frac{y_{\theta,c}'(x)}{y_{\theta,c}(x)\sqrt{c(x)}} \right| \le \sqrt{\theta}. \tag{18}$$

Finally, we show in the following result that the PDF $y_{\theta,c}^2$ falls within the set $\mathcal{F}$ defined in (7).

Proposition 2. *For any c satisfying Assumption 1 and any $\theta > 0$, we have that $y_{\theta,c}^2 \in \mathcal{F}$.*

Next, we combine Theorems 1–2 and Proposition 2 to show in Theorem 3 that the PDF $y_{\theta,c}^2$ is the optimal DP mechanism in the sense of Definition 1.

C. The Schrödinger Mechanism

Since $y_{\theta,c}$ is a Borel function satisfying $\|y_{\theta,c}\|_2 = 1$, we get that $y_{\theta,c}^2$ is a PDF. We call $y_{\theta,c}^2$ 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 $y_{\theta,c}^2$ where $y = y_{\theta,c}$ is the unique unit-L^2-norm and strictly positive solution to the Schrödinger equation

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

with E an arbitrary constant.[6]

Combining our results, we get that the Schrödinger mechanism is optimal in the small-sensitivity regime.

Theorem 3. *If a cost satisfies Assumption 1, the Schrödinger mechanism is optimal in the small-sensitivity regime in the sense of Definition 1.*

[5] 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 [33, Chapter 2, Theorem 1.1]). In particular, this extension goes through if v is nonnegative (and measurable).

[6] By Lemma 1, there is a unique E for which the ODE (19) is solvable with the prescribed properties for the solution y, and the solution then is $y = y_{\theta,c}$.

Remark 1. 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.

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

Next, we instantiate Theorem 3 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 3 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 3. The Schrödinger equation to be solved becomes

$$y''(x) = \left(\theta x^2 - E\right) y(x). \tag{20}$$

Proposition 3. *For a quadratic cost $c(x) = x^2$, the Gaussian mechanism is optimal in the small-sensitivity regime.*

B. Absolute value cost: optimality of Airy

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

$$y''(x) = (\theta|x| - E)y(x), \tag{21}$$

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

$$y''(x) = xy(x) \tag{22}$$

has two linearly independent solutions, called the Airy functions [34, Chapter 9]. They are denoted by Ai and Bi, where Ai is the solution such that $\mathsf{Ai}(x) \to 0$ as $x \to \infty$. This function can be expressed by the improper Riemann integral

$$\mathsf{Ai}(x) = \frac{1}{\pi} \lim_{N\to\infty} \int_0^N \cos\left(\frac{t^3}{3} + xt\right) dt. \tag{23}$$

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 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 optimal.

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

$$p_{\mathsf{Ai},C}(x) := \frac{1}{3C\mathsf{Ai}(a_1')^2} \mathsf{Ai}\left(\frac{-2a_1'}{3C}|x| + a_1'\right)^2. \tag{24}$$

Remark 2. 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.

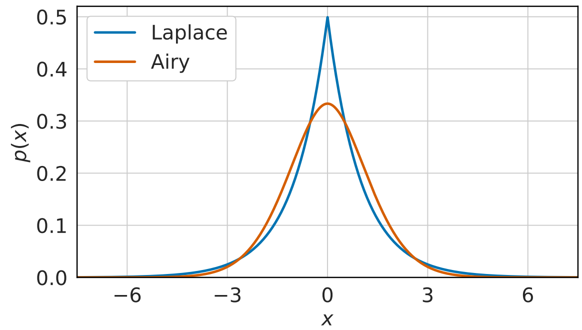

Fig. 1: The densities of the Laplace and Airy distributions ($p_{\mathsf{Ai},C}(x)$, introduced in Definition 4), with $C = \mathbb{E}[|X|] = 1$.

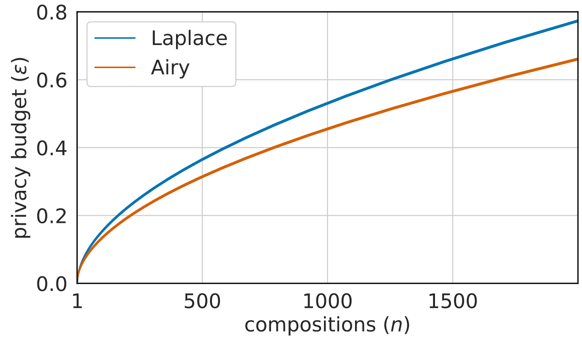

Fig. 2: The privacy budget ε versus the number of the compositions n, for the constraint $C = \mathbb{E}[|X|] = 2$, $s = 1$, fixed privacy parameter $\delta = 10^{-8}$, and subsampling rate $q = 0.01$.

Proposition 4. *For an absolute-value cost $c(x) = |x|$, the Airy mechanism is optimal in the small-sensitivity regime.*

In Figure 1, we illustrate the Airy distribution and compare it with 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}.

Experiments: Finally, we demonstrate that the Airy mechanism can achieve better DP parameters than the Laplace mechanism for the same fixed absolute-value cost. In particular, we subsample both mechanisms, following standard practice in the DP machine learning community for amplifying privacy [7], [35], [36]. We also use the arbitrary-accuracy FFT-based numerical accountant introduced in [13] to compute tight privacy bounds for finite compositions. In Figure 2, we fix $\delta = 10^{-8}$ and estimate ε under a varying number of compositions. Under this construction, the accountant in [13] computes both upper and lower bounds on ε. We choose $\varepsilon_{\text{error}} = 0.002$ and $\delta_{\text{error}} = 10^{-10}$, effectively making the upper and lower bounds indistinguishable (they are both plotted in Figure 2). The Airy mechanism provides stronger privacy guarantees for all values of compositions ($1 \le n \le 2000$), and the gap increases with composition.

REFERENCES

[1] 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.

[2] 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.

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

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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, pp. 247–264.

[11] 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

[12] 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.

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

[14] 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

[15] W. Alghamdi, S. Asoodeh, F. P. Calmon, J. F. Gomez, O. Kosut, L. Sankar, and F. Wei, "The saddle-point accountant for differential privacy," *arXiv preprint arXiv:2208.09595*, 2022.

[16] W. Alghamdi, S. Asoodeh, F. P. Calmon, O. Kosut, L. Sankar, and F. Wei, "Cactus mechanisms: Optimal differential privacy mechanisms in the large-composition regime," in *2022 IEEE International Symposium on Information Theory (ISIT)*, 2022, pp. 1838–1843.

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

[18] 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.

[19] 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.

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

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

[22] 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.

[23] 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.

[24] ——, "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.

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

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

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

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

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

[30] F. Farokhi and H. Sandberg, "Fisher information as a measure of privacy: Preserving privacy of households with smart meters using batteries," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 4726–4734, 2018.

[31] ——, "Ensuring privacy with constrained additive noise by minimizing fisher information," *Automatica*, vol. 99, pp. 275–288, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0005109818304862

[32] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.

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

[34] "*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/

[35] 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.

[36] 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.

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

Appendix A
Further Comparisons with the Literature

We contrast in this appendix our Fisher information minimization contribution with the relevant literature.

- [29, Example 5.1]: Although there is no general statement (e.g., a theorem) in [29] showing a result similar to our result in Theorem 2, one can distill from Section 4.5 in [29] 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: **(i)** p is strictly positive, absolutely continuous, and twice differentiable, **(ii)** the following integration by parts holds[7] for the ratio $\psi = p'/p$

$$\int_{\mathbb{R}} \psi(x)(q'(x)-p'(x))\,dx = -\int_{\mathbb{R}} \psi'(x)(q(x)-p(x))\,dx \tag{25}$$

for *every* PDF q with $I(q) < \infty$ and $\mathbb{E}_q[c] \le C$, and **(iii)** 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 [29] 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 $c(x) = x^2$). In contrast, our results on Fisher information minimization assumes none of the assumptions made in [29]; rather, we *derive* similar results that are required for our proof technique to follow through (e.g., via proving Proposition 1).
- [28]: The derivations therein assume without proof some of the above mentioned properties regarding [29], such as positivity, smoothness, and the validity of the integration by parts in (25); there are no worked examples in [28]. Similarly to our comparison with [29], we derive rather than assume the required properties.
- The use of Fisher information for optimizing privacy has appeared in [30], [31]. However, in these papers, rather than connecting DP to Fisher information, the authors set up the privacy problem as one where Fisher information is to be minimized. Then, the problem of minimizing Fisher information is connected to the Schrödinger equation. However, we note that the mathematical setup for the Fisher-information minimization problems in [30], [31] is different from, and less general than, what we consider herein. Recall that we derive the unique minimizers of the Fisher information $I(p)$ for $p \in \mathcal{P}(\mathbb{R})$, i.e., over *all* possible PDFs, subject to the constraint $\mathbb{E}_p[c] \le C$ where c satisfies Assumption 1. In contrast, [30] considers only bounded-support PDFs that are also twice continuously differentiable. The work in [30] is extended in [31] to consider unbounded-support PDFs, but subject to two restrictions: the PDF must be twice continuously differentiable, and the cost constraint is the variance cost constraint. Again, we do not assume these properties *a priori*, but derive whatever properties are necessary for our approach.

[7]We note that the integration by parts in equation (25) should not be expected to hold for arbitrary cost c.

Appendix B
Proof of Theorem 1

For $a \in [-s,s]$, define $L_a = \log \frac{p(X)}{p(X-a)}$ where $X \sim p$. For $\boldsymbol{a} = (a_1, \cdots, a_n)$, let $L_{\boldsymbol{a}} = \sum_{i=1}^n L_{a_i}$ for independent $L_{a_1}, \cdots, L_{a_n}$. We rewrite $\delta_{p^{\circ n},s}$ (see (11)) as

$$\delta_{p^{\circ n},s}(\varepsilon) = \sup_{\boldsymbol{a} \in \mathbb{R}^n, \|\boldsymbol{a}\|_\infty \le s} f(\boldsymbol{a}, \varepsilon) \tag{26}$$

for $f(\boldsymbol{a},\varepsilon) := \mathbb{E}[(1-e^{\varepsilon - L_{\boldsymbol{a}}})^+]$ with $x^+ := \max(0,x)$. We derive bounds on f. Denote $\mathrm{KL}_a = \mathbb{E}[L_a]$ and $\mathrm{V}_a = \mathrm{Var}(L_a)$. We may restrict attention to $\mathrm{V}_{\boldsymbol{a}} > 0$, for otherwise we get $f(\boldsymbol{a},\varepsilon) = 0$. Let $W_{\boldsymbol{a}} \sim \mathcal{N}(\mathrm{KL}_{\boldsymbol{a}}, \mathrm{V}_{\boldsymbol{a}})$. Expressing the f as the integral of the complementary CDF, we get from the CLT that

$$f(\boldsymbol{a},\varepsilon) = \mathbb{E}\left[\left(1-e^{\varepsilon - W_{\boldsymbol{a}}}\right)^+\right] + r(n), \tag{27}$$

where $r(n) = o(1)$ uniformly in $\boldsymbol{a}$ and ε by the assumption of uniformly bounded variances. Further, the Gaussian expectation above evaluates to (using the Gaussian Q-function)

$$Q\left(\frac{\varepsilon - \mathrm{KL}_{\boldsymbol{a}}}{\sqrt{\mathrm{V}_{\boldsymbol{a}}}}\right) - e^{\varepsilon - \mathrm{KL}_{\boldsymbol{a}} + \frac{\mathrm{V}_{\boldsymbol{a}}}{2}} Q\left(\frac{\varepsilon - \mathrm{KL}_{\boldsymbol{a}} + \mathrm{V}_{\boldsymbol{a}}}{\sqrt{\mathrm{V}_{\boldsymbol{a}}}}\right). \tag{28}$$

Let $\mathrm{KL}_{\max} = \sup_{|a|\le s} \mathrm{KL}_a$ and $\mathrm{V}_{\max} = \sup_{|a|\le s} \mathrm{V}_a$, and note that $\mathrm{KL}_{\boldsymbol{a}} \le n \cdot \mathrm{KL}_{\max}$ and $\mathrm{V}_{\boldsymbol{a}} \le n \cdot \mathrm{V}_{\max}$. Now, fix $0 < \underline{\tau} < \mathrm{KL}_{\max} < \overline{\tau}$. Then,

$$\lim_{n\to\infty} \sup_{\boldsymbol{a} \in \mathbb{R}^n, \|\boldsymbol{a}\|_\infty \le s} Q\left(\frac{\overline{\tau} n - \mathrm{KL}_{\boldsymbol{a}}}{\sqrt{\mathrm{V}_{\boldsymbol{a}}}}\right) = 0. \tag{29}$$

In particular, $\delta_{p^{\circ n},s}(\overline{\tau} n) < \delta$ for large n. Hence,

$$\limsup_{n\to\infty} \frac{\varepsilon_{p^{\circ n},s}(\delta)}{n} \le \overline{\tau}, \tag{30}$$

which, upon taking $\overline{\tau} \searrow \mathrm{KL}_{\max}$, yields

$$\limsup_{n\to\infty} \frac{\varepsilon_{p^{\circ n},s}(\delta)}{n} \le \mathrm{KL}_{\max}. \tag{31}$$

Similarly, using $\underline{\tau} n$ instrad, the reverse inequality is obtained from the well-known bound $Q(z) \le \frac{1}{\sqrt{2\pi}\cdot z} e^{-z^2/2}$, so

$$\liminf_{n\to\infty} \frac{\varepsilon_{p^{\circ n},s}(\delta)}{n} \ge \mathrm{KL}_{\max}, \tag{32}$$

completing the proof of the theorem.

Appendix C
Proof of Lemma 1

By [33, 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, $\gamma := \mathrm{sgn}(y(0))/\|y\|_2$. 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.

Appendix D
Proof of Proposition 1

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

Theorem 4 ([33, 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*

$$y_{\theta,c}(x) \sim \frac{\exp\left(-\int_{x_1}^{x}\sqrt{\theta c(t) - E_0}\,dt\right)}{(\theta c(x))^{1/4}}. \tag{33}$$

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

$$-f' + f^2 = g. \tag{34}$$

The eigenvalue equation for y becomes $y'' = gy$. Since c is unbound, g is eventually strictly positive. Since y is strictly positive and $y'' = gy$, we see 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,

$$y'(t_1) - y'(t_2) = \int_{t_2}^{t_1} y''(t)\,dt \geq 0 \tag{35}$$

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,

$$f(x) \leq \sqrt{2g(x)}, \tag{36}$$

which is equivalent to $\left|\frac{y'(x)}{y(x)}\right| \leq \sqrt{2(\theta c(x) - E_0)}$. This is enough to finish the proof by evenness of y and c.

Set $h = \sqrt{2g}$. Denote $z = f - h$. Differentiating and using $-f' + f^2 = g$ (see (34)), we obtain

$$-z' + z^2 + 2zh = h' - g. \tag{37}$$

Note that $h' < g$ eventually holds. Indeed, as $x \to \infty$,

$$\frac{h'(x)}{g(x)} = \frac{1}{\sqrt{2}}\frac{g'(x)}{g(x)^{3/2}} \propto \frac{c'(x)}{c(x)^{3/2}} \to 0 \tag{38}$$

by assumption on c. Thus, by (37), we eventually have $-z' < 0$, i.e., z is strictly increasing over (x_0,∞) for some $x_0 > 0$.

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, $-\frac{y'(x)}{y(x)} > \sqrt{2g(x)}$ for all $x \geq x_1$. Increase x_1 if necessary so that $y(x) < 1$ for $x \geq x_1$. Then,

$$y(x) \leq \exp\left(-\int_{x_2}^{x}\sqrt{2g(t)}\,dt\right) \tag{39}$$

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

$$y(x) \geq \frac{1}{2g(x)^{1/4}}\exp\left(-\int_{x_3}^{x}\sqrt{g(t)}\,dt\right) \tag{40}$$

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

$$(\sqrt{2}-1)\int_{x_3}^{x}\sqrt{g(t)}\,dt \leq \log\left(2g(x)^{1/4}\right). \tag{41}$$

Denote $w(t) = \sqrt{(\sqrt{2}-1)\sqrt{g(t)}}$, so (41) can be rewritten as

$$\frac{\int_{x_3}^{x} w(t)^2\,dt}{\log(\gamma\cdot w(x))} \leq 1, \tag{42}$$

where $\gamma := 2\sqrt{1+\sqrt{2}}$ is an absolute constant. To arrive at a contradiction, we take $x \to \infty$ and use L'Hôpital's rule:

$$\lim_{x\to\infty}\frac{\int_{x_3}^{x} w(t)^2\,dt}{\log(\gamma\cdot w(x))} = \lim_{x\to\infty}\frac{w(x)^3}{w'(x)} = \infty. \tag{43}$$

To see that the last limit diverges, note that

$$\frac{w(x)^3}{w'(x)} \propto \frac{c(x)^{3/2}}{c'(x)} \to \infty. \tag{44}$$

The limit in (43) contradicts inequality (42). Thus, there is no $x_1 > x_0$ such that $f(x_1) > h(x_1)$. Hence, (36) eventually holds, and the proof is complete.

In the course of this proof of Proposition 1, we have shown the following useful property of $y_{\theta,c}$ that will be used later.

Lemma 2. *$y_{\theta,c}$ is eventually decreasing.*

Appendix E
An Auxiliary Lemma

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

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

$$\inf_{\substack{p\in\mathcal{P}_0\\ \mathbb{E}_p[c]\leq C}} I(p) = \inf_{\substack{p\in\mathcal{P}\\ \mathbb{E}_p[c]\leq C}} I(p). \tag{45}$$

Proof. 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 noting that the limit

$$\lim_{\sigma\to 0^+}\mathbb{E}_{p*\phi^\sigma}[c] = \mathbb{E}_p[c] \tag{46}$$

holds for every PDF p that satisfies $\mathbb{E}_p[c] < \infty$. Indeed, by the assumed additive and multiplicative regularity of c, it is not hard to see that, for the random variables $Z_\sigma \sim p * \phi^\sigma$, the set $\{c(Z_\sigma)\}_{0<\sigma\leq 1}$ is uniformly bounded by an integrable random variable. In particular, the set $\{c(Z_\sigma)\}_{0<\sigma\leq 1}$ is uniformly integrable, so the Lebesgue-Vitali theorem [37, Theorem 4.5.4] yields that the limit (46) holds.

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

$$I_0^\star(C) := \inf_{\substack{p\in\mathcal{P}_0\\ \mathbb{E}_p[c]\leq C}} I(p) \tag{47}$$

is continuous at C. We may write

$$I_0^\star(C) = \inf_{p\in\mathcal{P}_0} I(p) + \mathbb{I}_{(-\infty,C]}(\mathbb{E}_p[c]). \tag{48}$$

Being the infimum of a jointly convex function over a convex set, $I_0^\star$ is convex. Further, this function is finite over $(0,\infty)$

(e.g., take the Gaussian PDF with small enough variance). Hence, being convex and finite, $I_0^\star$ is continuous over $(0,\infty)$.

Define

$$I^\star(C) := \inf_{\substack{p\in\mathcal{P}\\ \mathbb{E}_p[c]\le C}} I(p) \tag{49}$$

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

$$I(p)\le I^\star(C)+\varepsilon. \tag{50}$$

Since Fisher information satisfies the convolution inequality,

$$I(p*\phi^\sigma)\le I(p) \tag{51}$$

for every $\sigma>0$. By (46), there is a $\sigma=\sigma(\eta)$ such that

$$\mathbb{E}_{p*\phi^\sigma}[c]\le \mathbb{E}_p[c]+\eta\le C+\eta. \tag{52}$$

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

$$I_0^\star(C+\eta)\le I(p*\phi^\sigma)\le I(p)\le I^\star(C)+\varepsilon. \tag{53}$$

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

$$I_0^\star(C)\le I^\star(C)+\varepsilon. \tag{54}$$

By arbitrariness of ε, we deduce

$$I_0^\star(C)\le I^\star(C). \tag{55}$$

But the reverse inequality is trivial, thus equality is attained in (55), completing the proof of the lemma. □

Appendix F
Proof of Theorem 2

We use the integration shorthand $\int_A f := \int_A f(x)\,dx$. Denote $y=y_{\theta,c}$ for short, and set $p=y^2$. First, we note that $C=\|y\|_{2,c}^2$ is indeed finite as can be deduced from the expansion of y in Theorem 4.

Denote the space of absolutely continuous functions on $\mathbb{R}$ by $\mathrm{AC}(\mathbb{R})$, and those that are locally absolutely continuous 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

$$V := L^2(\mathbb{R})\cap L^2(\mathbb{R},c)\cap \mathrm{AC}_{\mathrm{loc}}(\mathbb{R}). \tag{56}$$

Let E be the eigenvalue of y, so

$$y''=(\theta c-E)y. \tag{57}$$

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

$$\mathcal{E}(w) := \|w'\|_2^2+\theta\|w\|_{2,c}^2-E\|w\|_2^2. \tag{58}$$

We start by showing that y is a global minimizer of $\mathcal{E}$, and $0=\mathcal{E}(y)=\inf_{w\in V}\mathcal{E}(w)$. 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 (57) satisfied by y, we obtain that a.e.

$$(w')^2\ge \frac{2y'ww'}{y}-\frac{(y')^2w^2}{y^2}=\left(\frac{y'w^2}{y}\right)'-(\theta c-E)w^2. \tag{59}$$

Note that $y'w^2/y\in\mathrm{AC}_{\mathrm{loc}}(\mathbb{R})$. Thus, integrating (59) over any $[-t,t]$ with $t>0$, we obtain

$$\|w'1_{[-t,t]}\|_2^2\ge \frac{y'w^2}{y}\bigg|_{-t}^{t}-\theta\|w1_{[-t,t]}\|_{2,c}^2+E\|w1_{[-t,t]}\|_2^2. \tag{60}$$

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

$$\liminf_{n\to\infty}\frac{y'w^2}{y}\bigg|_{-t_n}^{t_n}\ge 0. \tag{61}$$

This would readily yield $\mathcal{E}(w)\ge 0$ from inequality (60). By assumption, $w\in L^2(\mathbb{R},c)$, so symmetry of c implies

$$\int_0^\infty (w(x)^2+w(-x)^2)c(x)\,dx=\int_{\mathbb{R}} w^2c<\infty. \tag{62}$$

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

$$\big(w(t_n)^2+w(-t_n)^2\big)c(t_n)\to 0. \tag{63}$$

By (18), there is an $A\in(0,\infty)$ such that $\left|\frac{y'(x)}{y(x)}\right|\le A\cdot c(|x|)$ holds for all large $|x|$. Then, for all large n,

$$\begin{aligned}\frac{y'w^2}{y}\bigg|_{-t_n}^{t_n} &= \frac{y'(t_n)w(t_n)^2}{y(t_n)}-\frac{y'(-t_n)w(t_n)^2}{y(-t_n)} && (64)\\ &\ge -\left|\frac{y'(t_n)}{y(t_n)}\right|w(t_n)^2-\left|\frac{y'(-t_n)}{y(-t_n)}\right|w(-t_n)^2 && (65)\\ &\ge -Ac(t_n)\big(w(t_n)^2+w(-t_n)^2\big). && (66)\end{aligned}$$

Taking the limit inferior in (66), and using (63),

$$\liminf_{n\to\infty}\frac{y'w^2}{y}\bigg|_{-t_n}^{t_n}\ge 0. \tag{67}$$

Since $w\in L^2(\mathbb{R})$, the monotone convergence theorem implies

$$\lim_{n\to\infty}\theta\|w1_{[-t_n,t_n]}\|_{2,c}^2-E\|w1_{[-t_n,t_n]}\|_2^2=\theta\|w\|_{2,c}^2-E\|w\|_2^2. \tag{68}$$

Taking the limit inferior of (60) along the t_n, and using (67) and (68), we conclude that

$$\|w'\|_2^2\ge -\theta\|w\|_{2,c}^2+E\|w\|_2^2. \tag{69}$$

As $w\in L^2(\mathbb{R},c)\cap L^2(\mathbb{R})$, (69) is equivalent to $\mathcal{E}(w)\ge 0$.

We have just shown that

$$\inf_{w\in V}\mathcal{E}(w)\ge 0. \tag{70}$$

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 $yy'=O(y^2c)$ by Proposition 1. 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,

$$\begin{aligned}\mathcal{E}(y)&=\|y'\|_2^2+\theta\|y\|_{2,c}^2-E\|y\|_2^2 && (71)\\ &=\lim_{n\to\infty}\|y'1_{[-a_n,b_n]}\|_2^2+\theta\|y1_{[-a_n,b_n]}\|_{2,c}^2-E\|y1_{[-a_n,b_n]}\|_2^2 && (72)\\ &=\lim_{n\to\infty}\int_{-a_n}^{b_n}(y')^2+(\theta c-E)y^2 && (73)\end{aligned}$$

$$= \lim_{n\to\infty} \int_{-a_n}^{b_n} (y')^2 + yy'' = \lim_{n\to\infty} \int_{-a_n}^{b_n} (yy')' \tag{74}$$

$$= \lim_{n\to\infty} y(b_n)y'(b_n) - y(-a_n)y'(-a_n) = 0, \tag{75}$$

where (72) 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 $p = y^2$ minimizes the Fisher information.

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

$$I_\gamma^\star := \inf_{\substack{w\in V\\ \|w\|_{2,c}^2\le\gamma,\ \|w\|_2=1}} 4\|w'\|_2^2. \tag{76}$$

It is not hard to see that V is closed under positive dilation, so in particular $u(x) = w(x/\sigma)/\sqrt{\sigma}$ is in V if $w \in V$. This in turn yields (via choosing σ large enough if necessary) that the inequality $\|w\|_{2,c}^2 \le \gamma$ in the definition of $I_\gamma^\star$ can be replaced with an equality, i.e.,

$$I_\gamma^\star := \inf_{\substack{w\in V\\ \|w\|_{2,c}^2=\gamma,\ \|w\|_2=1}} 4\|w'\|_2^2. \tag{77}$$

Our next goal is to show that $E \le E^\star$, where we define

$$E^\star := \inf_{\gamma\in\mathbb{R}} I_\gamma^\star + \theta\gamma. \tag{78}$$

Indeed, by (77), we use (78) to deduce that $E^\star$ satisfies

$$E^\star = \inf_{\substack{w\in V\\ \|w\|_2=1}} 4\|w'\|_2^2 + \theta\|w\|_{2,c}^2 \tag{79}$$

$$= \inf_{w\in V\setminus\{0\}} \frac{4\|w'\|_2^2 + \theta\|w\|_{2,c}^2}{\|w\|_2^2} \tag{80}$$

$$= E + \inf_{w\in V\setminus\{0\}} \frac{\mathcal{E}(w)}{\|w\|_2^2} \ge E, \tag{81}$$

where (80) follows since V is a vector space, and (81) since $\inf_{w\in V} \mathcal{E}(w) \ge 0$ (see (70)).

Next, we show $I(p) = I_C^\star$. We have $(p')^2/p = 4(y')^2$. Thus, $I(p) = 4\|y'\|_2^2$. From $E \le E^\star$ and (78), we get

$$E\|y\|_2^2 = E \le E^\star \le I_C^\star + \theta C = I_C^\star + \theta\|y\|_{2,c}^2. \tag{82}$$

Adding $4\|y'\|_2^2 - E\|y\|_2^2$ to both sides yields $I(p) \le I_C^\star + \mathcal{E}(y)$. As $\mathcal{E}(y) = 0$ (see (75)), we conclude that $I(p) \le I_C^\star$. The reverse inequality also holds as $\|y\|_2 = 1$ and $\|y\|_{2,c}^2 = C$, so

$$I(p) = I_C^\star. \tag{83}$$

We show that p globally minimizes the Fisher information:

$$I(p) = \inf_{\substack{q\in\mathcal{P}\\ \mathbb{E}_q[c]\le C}} I(q). \tag{84}$$

We start by showing that $I(p)$ is minimal among strictly positive PDFs, denoted $\mathcal{P}_0$. 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 $I(q) = 4\|w'\|_2^2$. Thus, we conclude from $I(p) = I_C^\star$ (see (83)) that

$$I(p) = \inf_{\substack{q\in\mathcal{P}_0\\ \mathbb{E}_q[c]\le C}} I(q). \tag{85}$$

However, this argument cannot be applied to a PDF q that has zeros. Instead, we use Lemma 3 to obtain from (85) that

$$I(p) = \inf_{\substack{q\in\mathcal{P}_0\\ \mathbb{E}_q[c]\le C}} I(q) = \inf_{\substack{q\in\mathcal{P}\\ \mathbb{E}_q[c]\le C}} I(q), \tag{86}$$

which is the global optimality of p claimed in (84). Since p is strictly positive, we conclude that it is the unique minimizer of the Fisher information over all possible PDFs [29, Proposition 4.5], and the proof is complete.

Appendix G
Proof of Proposition 2

We need the following well-known differentiation under the integral sign result.

Theorem 5. *Let $U \subset \mathbb{R}$ be open, and (X, Σ, μ) be a measure space. Suppose $f : U \times X \to \mathbb{R}$ satisfies:*

1) *For each $a \in U$, we have $f(a, \cdot) \in L^1(\mu)$.*
2) *For μ-almost every $x \in X$, the function $f(\cdot, x)$ is differentiable over U.*
3) *The function $x \mapsto \sup_{a_0\in U} \left|\frac{\partial f}{\partial a}(a_0, x)\right|$ is μ-integrable.*

Then, over U,

$$\frac{d}{da} \int_X f(a, x)\, d\mu(x) = \int_X \frac{\partial f}{\partial a}(a, x)\, d\mu(x). \tag{87}$$

We use Theorem 5 to differentiate $a \mapsto D(p \,\|\, T_a p)$ twice, then conclude using Taylor's theorem. We have that

$$D(p \,\|\, T_a p) = \int_{\mathbb{R}} p(x) \log \frac{p(x)}{p(x-a)}\, dx. \tag{88}$$

Denote the function $f_1(a, x) := \log \frac{p(x)}{p(x-a)}$. For each $a \in \mathbb{R}$, we have that $f_1(a, \cdot)$ is continuous; indeed, it is differentiable by differentiability and strict positivity of p (recall that $p = y_{\theta,c}^2$). We consider the Borel space $(X, \Sigma, \mu) = (\mathbb{R}, \mathcal{B}, p(x)\, dx)$. Hence, for the sake of showing the integrability $f_1(a, \cdot) \in L^1(p(x)\, dx)$, we may ignore any bounded interval. By the asymptotic expansion of $y_{\theta,c}$ in Theorem 4, we have the following asymptotic formula. Let E denote the eigenvalue of $y_{\theta,c}$. For each $a \in \mathbb{R}$, as $x \to \infty$ we have

$$\frac{p(x)}{p(x-a)} \sim \sqrt{\frac{c(x-a)}{c(x)}} \cdot \exp\left(-2 \int_{x-a}^{x} \sqrt{\theta c(t) - E}\, dt\right). \tag{89}$$

By Assumption 1, we have that for all large x

$$\frac{1}{2\rho(-a)} \le \frac{c(x-a)}{c(x)} \le 2\rho(a). \tag{90}$$

Further, for $|a| < 1$, we have that for all large x

$$\int_{x-a}^{x} \sqrt{\theta c(t) - E}\, dt \le \sqrt{\theta c(x)} \le \sqrt{\theta} \cdot c(x). \tag{91}$$

Thus, we conclude the integrability $f_1(a, \cdot) \in L^1(p(x)\, dx)$.

For each $x \in \mathbb{R}$, $f_1(\cdot, x)$ is differentiable with derivative

$$\frac{\partial f_1}{\partial a}(a_0, x) = \frac{-p'(x-a_0)}{p(x-a_0)} = \frac{-2y_{\theta,c}'(x-a_0)}{y_{\theta,c}(x-a_0)}. \tag{92}$$

We consider $a_0 \in U = (-1, 1)$. From Proposition 1, there is some $z_0 = z_0(\theta, c)$ such that $z > z_0(\theta, c)$ implies

$$\left| \frac{y'_{\theta,c}(z)}{y_{\theta,c}(z)} \right| \le 2\sqrt{2\theta} \cdot c(z). \tag{93}$$

Hence, for $x > z_0 + 1$, we have for all $|a_0| < 1$

$$\left| \frac{y'_{\theta,c}(x - a_0)}{y_{\theta,c}(x - a_0)} \right| \le 2\sqrt{2\theta} \cdot c(x - a_0). \tag{94}$$

Using Assumption 1, we have

$$\sup_{|a_0|<1} c(x - a_0) \le \left(\sup_{|a_0|<1} \rho(a_0) \right) \cdot (c(x) + 1), \tag{95}$$

where $A := \sup_{|a_0|<1} \rho(a_0)$ is finite. Combining these inequalities, we conclude that

$$\sup_{|a_0|<1} \left| \frac{\partial f_1}{\partial a}(a_0, x) \right| \le 4A\sqrt{2\theta} \cdot (c(x) + 1) \tag{96}$$

for all large x. As $(a, x) \mapsto \frac{\partial f_1}{\partial a}(a, x)$ is continuous, we conclude that $\sup_{|a_0|<1} \left| \frac{\partial f_1}{\partial a}(a_0, x) \right| \in L^1(p(x)\, dx)$.

Therefore, we may apply Theorem 5 to differentiate the KL-divergence and obtain

$$\frac{d}{da} D(p \,\|\, T_a p) = -\int_{\mathbb{R}} p(x) \cdot \frac{p'(x-a)}{p(x-a)}\, dx \tag{97}$$

over $|a| < 1$. Performing a change of variable, we obtain that

$$\frac{d}{da} D(p \,\|\, T_a p) = -\int_{\mathbb{R}} p(x+a) \cdot \frac{p'(x)}{p(x)}\, dx. \tag{98}$$

Next, we apply Theorem 5 to differentiate the KL-divergence a second time. This time, we use the usual Lebesgue space $(\mathbb{R}, \mathcal{B}, \lambda)$. Consider the function $f_2(a, x) := p(x+a) \cdot \frac{p'(x)}{p(x)}$. Inequality (96) shows that $f_2(a, \cdot) \in L^1(\lambda)$ for each $a \in (-1, 1)$. Further, for each $x \in \mathbb{R}$, $f_2(\cdot, x)$ is differentiable over $(-1, 1)$ with derivative

$$\frac{\partial f_2}{\partial a}(a, x) = p'(x+a) \cdot \frac{p'(x)}{p(x)}. \tag{99}$$

We write

$$\left| \frac{\partial f_2}{\partial a}(a, x) \right| = \left| \frac{p'(x+a)}{p(x+a)} \right| \cdot \left| \frac{p'(x)}{p(x)} \right| \cdot p(x+a). \tag{100}$$

Via the same derivation of inequality (96), but using the full power of Proposition 1 this time (i.e., $\sqrt{c}$ as an upper bound instead of c), and applying Lemma 2 (i.e., that p is eventually decreasing), we obtain the bound

$$\sup_{|a|<1} \left| \frac{\partial f_2}{\partial a}(a, x) \right| \le 8\sqrt{2A\theta} \cdot c(x) p(x - 1) \tag{101}$$

for all large x. Therefore, $\sup_{|a|<1} \left| \frac{\partial f_2}{\partial a}(a, x) \right| \in L^1(\lambda)$.

Hence, we may apply Theorem 5 again to obtain that

$$\frac{d^2}{da^2} D(p \,\|\, T_a p) = -\int_{\mathbb{R}} p'(x+a) \cdot \frac{p'(x)}{p(x)}\, dx \tag{102}$$

over $(-1, 1)$. Setting $a = 0$ in $D(p \,\|\, T_a p)$ and its first two derivatives, we obtain from Taylor's theorem that

$$D(p \,\|\, T_a p) = \frac{a^2}{2} I(p) + o(a^2) \tag{103}$$

as $a \to 0$, i.e., expansion (4) holds.

Next, we show that $\sup_{|a|\le s} D(p \,\|\, T_a p) < \infty$ and $\sup_{|a|\le s} \mathrm{V}(p \,\|\, T_a p) < \infty$ for some $s > 0$. For this, it suffices to show that

$$\int_{\mathbb{R}} p(x) \cdot \sup_{|a|\le 1} \left| \log \frac{p(x)}{p(x-a)} \right|^2 dx < \infty \tag{104}$$

Using the asymptotic formula for $y_{\theta,c}$ in Theorem 1, we have

$$\frac{p(x)}{p(x-a)} \sim \sqrt{\frac{c(x-a)}{c(x)}} \cdot \exp\left(-2 \int_{x-a}^{x} \sqrt{\theta c(t) - E}\, dt \right). \tag{105}$$

Hence, the same method showing integrability of $f_1(a, \cdot)$ shows the desired result.

Appendix H
Proofs of Section IV

We give the proof of Proposition 4 and omit that of Proposition 3 due to space limitation and since it is well-known that the Gaussian function is the ground-state eigenfunction for the quantum harmonic oscillator.

A. *Proof of Proposition 4*

From Theorem 3, we need to solve the ODE (21). Let $y_1 = \sqrt{p_{\mathsf{Ai},C}}$. Thus, from Definition 4, we have

$$y_1(x) = \gamma \cdot \mathsf{Ai}\left(\theta^{1/3}|x| + a_1' \right), \tag{106}$$

where $\gamma := \frac{1}{\sqrt{3C} \cdot \mathsf{Ai}(a_1')}$ and $\theta = \left(\frac{-2a_1'}{3C} \right)^3$. Differentiating separately for $x < 0$, $x = 0$, and $x > 0$, we obtain

$$y_1'(x) = \theta^{1/3} \gamma \,\mathrm{sgn}(x) \mathsf{Ai}'(\theta^{1/3}|x| + a_1'), \tag{107}$$

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 $x \in \mathbb{R}$

$$y_1''(x) = \theta^{2/3} \gamma \mathsf{Ai}''\left(\theta^{1/3}|x| + a_1' \right). \tag{108}$$

Since Ai is a solution of the differential equation of (22), it follows that $\mathsf{Ai}''(z) = z\mathsf{Ai}(z)$ for every $z \in \mathbb{R}$ and thus

$$y_1''(x) = \left(\theta|x| + \theta^{2/3} a_1' \right) y_1(x), \tag{109}$$

and hence y_1 solves the equation (21). Therefore, we conclude from Lemma 1 that $y_{\theta,c} = y_1$ is the ground-state eigenfunction of $\mathcal{H}_{\theta c}$. Moreover, since $\int_{\mathbb{R}} c y_{\theta,c}^2 = \int_{\mathbb{R}} c p_{\mathsf{Ai},C} = C$, invoking Theorem 3 finishes the proof that the Airy mechanism is optimal in the small-sensitivity regime for the absolute-value cost.