Quantum Robustness via Sandwiched Rényi Divergences

Luc Rey-Bellet

November 9, 2026

Abstract

We develop the theory of distributional robustness for quantum states: if a reference state is known only approximately, how much can the expectation of an observable change? Uncertainty is measured by the sandwiched Rényi divergence of order $\alpha \in [\frac{1}{2}, \infty]$, and the distributionally robust value is the worst-case expectation over the states whose divergence from the reference does not exceed a budget q. Convex duality expresses this value in closed form as a one-dimensional optimization over a sandwiched Schatten norm, running from a fidelity risk at $\alpha = \frac{1}{2}$, through the entropic (Gibbs) risk at $\alpha = 1$, to a quantum conditional value-at-risk at $\alpha = \infty$. Our main structural result is a classicalization theorem: for every α the quantum problem coincides exactly with the classical problem for a probability measure ω_α on the spectral interval of the observable, constructed from Heinävaara's tracial joint spectral measure, given by Stahl's theorem (the Bessis–Moussa–Villani conjecture) at $\alpha = 1$, and reducing to the Born measure only at $\alpha = \frac{1}{2}$. For small budgets the value obeys a square-root law whose variance is the inverse of the sandwiched Rényi Fisher metric and, by classicalization, the variance of ω_α. Classicalization also transfers the Hoeffding, Bernstein and Markov–Krein bounds of classical probability to the quantum problem; read on the Helstrom observable they become Pinsker-type inequalities, recovering Pinsker and Fuchs–van de Graaf and adding Bernstein-type refinements. The overall message is a quantum deflation: for $\alpha > \frac{1}{2}$, noncommutativity of the observable with the state strictly reduces the variance of ω_α below the Born variance, and with it every variance-sensitive bound; at $\alpha = 1$ this gives a Bernstein–Pinsker inequality deflated by the Kubo–Mori variance.

1 Introduction

How much can the expectation of an observable change when the state is known only approximately? We take an information theoretic approach: fix a reference state σ, an observable L, and a divergence $\mathcal{D}$; the *distributionally robust* value is

$$V(q) \;=\; \sup\{\mathrm{tr}(L\rho) :\; \mathcal{D}(\rho\|\sigma) \le q\},$$

the worst case over all states within a divergence ball of radius q around σ. Classically — σ a probability distribution, L a random variable — this is the basic object of distributionally robust optimization (DRO), which goes back to Scarf's min–max inventory problem [Scarf, 1958]. The divergence-ball form used here arose in robust control and in economics, with the relative-entropy constraint of Petersen, James and Dupuis [Petersen et al., 2000] and the robustness program of Hansen and Sargent [Hansen and Sargent, 2008], and became the φ-divergence framework of Ben-Tal et al. [Ben-Tal et al., 2013] and the variance-regularized learning of Duchi and Namkoong [Duchi and Namkoong, 2019, 2021]; the other families of ambiguity sets, fixed by moments [Bertsimas and Popescu, 2005, Delage and Ye, 2010, Wiesemann et al., 2014] or by optimal transport [Pflug and Wozabal, 2007, Mohajerin Esfahani and Kuhn, 2018, Blanchet and Murthy, 2019, Gao and Kleywegt, 2023], are outside the scope of this paper, and the survey of Kuhn, Shafiee and Wiesemann [Kuhn

et al., 2025] gives an in-depth account of the subject. The quantum theory developed here continues a line of work on divergence-based robustness and uncertainty quantification via variational formulas for Rényi and related divergences [Dupuis et al., 2020, Birrell et al., 2021, 2019, Katsoulakis et al., 2017].

In the quantum problem σ is a density matrix, L is a Hermitian matrix, and $\mathcal{D}$ a quantum divergence. In this work $\mathcal{D}$ is the sandwiched Rényi divergence of order $\alpha \in [\frac{1}{2}, \infty]$ [Müller-Lennert et al., 2013, Wilde et al., 2014], a quantum extension of the classical Rényi divergence [van Erven and Harremoës, 2014]. We use the Hellinger normalization H_α in Section 3 (classically the Cressie–Read power divergences [Cressie and Read, 1984]): at $\alpha = \frac{1}{2}$ the squared Bures distance (closely related to the fidelity), at $\alpha = 1$ the relative entropy, at $\alpha = \infty$ the max-relative entropy [Datta, 2009]. It is one of several quantum Rényi families that agree on commuting pairs and differ elsewhere [Tomamichel, 2016]. Depending on the parameter regime, the sandwiched Rényi divergence characterizes optimal operational thresholds: for $\alpha > 1$, it determines the exact strong converse exponents for binary quantum hypothesis testing [Mosonyi and Ogawa, 2015] and for $\alpha \in (\frac{1}{2}, 1)$, it yields the exact strong converse exponents for privacy amplification and state smoothing when measured by fidelity [Li and Yao, 2024]. At $\alpha = \frac{1}{2}$ the ball is a fidelity ball, the form in which state-estimation guarantees are stated. On the whole range the divergence is monotone under quantum channels [Frank and Lieb, 2013, Beigi, 2013] and in α [Müller-Lennert et al., 2013], so that the balls are nested and the robust value decreases as the order increases (Remark 3.1), its dual closes in a sandwiched Schatten norm, and it classicalizes. It is also the smallest of the quantum Rényi divergences with the natural properties — additivity, data processing and the classical value on commuting pairs — on the whole range $\alpha \in [\frac{1}{2}, \infty]$, by Tomamichel's pinching argument [Tomamichel, 2016, §4.2.2]; for the robust problem this means that its ball is the largest and its value the most conservative: every upper bound proved below for V_α holds for the DRO value of any such divergence at the same order and budget (Remark 3.2). The DRO problem for other quantum Rényi divergences, such as Petz's family [Petz, 1986, Hiai and Mosonyi, 2017] (which governs the quantum Chernoff and Hoeffding bounds [Nussbaum and Szkoła, 2009, Audenaert et al., 2008, Hayashi, 2007]) or the measured Rényi divergence [Berta et al., 2017], and for the f-divergences built from hockey-stick divergences [Hirche and Tomamichel, 2024, Liu et al., 2025], is an open problem (Section 7).

By convex duality $V(q)$ can be written as a one-dimensional Legendre transform of the free energy $G_\alpha(s) = \sup_\rho\{\mathrm{tr}(L\rho) - s\, H_\alpha(\rho\|\sigma)\}$; at $\alpha = 1$ this is the Gibbs variational principle [Petz, 1988, Ohya and Petz, 2004], and for general α using the sandwiched structure $\sigma^\beta(\,\cdot\,)\sigma^\beta$, $\beta = (\alpha-1)/2\alpha$, we prove, Theorem 3.4,

$$V_\alpha(q) \;=\; \inf_{c\in\mathbb{R}}\Big\{c + \big(1+(\alpha-1)q\big)^{1/\alpha}\big\|(\sigma^\beta(L-cI)\sigma^\beta)_+\big\|_{\alpha'}\Big\}, \qquad \alpha > 1,$$

the quantum form of the optimized certainty equivalent of Ben-Tal and Teboulle [Ben-Tal and Teboulle, 2007], with the classical $L^{\alpha'}$-norm of $(L-c)_+$ replaced by a sandwiched Schatten norm, and, for $\alpha < 1$, with the positive part replaced by a resolvent and the norm by its dual exponent,

$$V_\alpha(q) \;=\; \inf_{c>\lambda_{\max}(L)}\Big\{c - \frac{\big(1+(\alpha-1)q\big)^{1/\alpha}}{\big\|\sigma^\gamma(cI-L)^{-1}\sigma^\gamma\big\|_p}\Big\}, \qquad \tfrac{1}{2} \le \alpha < 1,$$

with $\gamma = -\beta$ and $p = \alpha/(1-\alpha)$, for q below an explicit threshold: a Stieltjes transform of order p, which at $\alpha = \frac{1}{2}$ ($p = 1$, the norm a trace) is a fidelity DRO (Corollary 3.3). The worst-case state is explicit in terms of the dual minimizer (Corollary 3.1); at $\alpha = 2$ the infimum closes in an exact square-root law, with a sandwiched linear tilt of σ as worst case (Corollary 3.2); at $\alpha = \infty$ the

value is a semidefinite program whose optimizer is a Holevo–Helstrom test, a quantum conditional value-at-risk (Theorem 3.5).

The dual formula has a second reading, which is the main structural result of the paper. Theorem 4.1 (*classicalization*) states that for every $\alpha \in [\frac{1}{2}, \infty]$ there is a probability measure ω_α on the spectral interval $[\lambda_{\min}(L), \lambda_{\max}(L)]$, depending on (σ, L, α), such that

$$V_\alpha(q) \;=\; V_\alpha^{\mathrm{cl}}(q;\, \omega_\alpha) \qquad \text{for all } q \geq 0 :$$

the quantum DRO *is* the classical DRO of a scalar random variable with law ω_α. The measure is characterized by one transform per regime — the stop-loss transform of order α' for $\alpha > 1$, the Laplace transform at $\alpha = 1$, the Stieltjes transform of order p for $\alpha < 1$ — each evaluated exactly by the sandwiched trace of the dual. At $\alpha = 1$ this is Stahl's theorem [Stahl, 2013], the former Bessis–Moussa–Villani conjecture [Bessis et al., 1975]; for $\alpha > 1$ it rests on Heinävaara's tracial joint spectral measure [Heinävaara, 2025] for the pencil $(\sigma^\beta L \sigma^\beta, \sigma^{2\beta})$, and for $\alpha < 1$ on a Stieltjes representation derived from it in the companion note [Rey-Bellet, 2026]; at $\alpha = \frac{1}{2}$ the measure is the Born measure of L in the state σ, and at $\alpha = \infty$ the layer-cake measure recalled in the related work below. The measures consist of atoms at the eigenvalues of L and an absolutely continuous part that vanishes if and only if $[\sigma, L] = 0$ or $\alpha = \frac{1}{2}$, with a density given by an integral of Heinävaara's density (Proposition 4.1).

At small budget the value linearizes, $V_\alpha(q) = \mathrm{tr}(L\sigma) + s_\alpha\sqrt{2q} + o(\sqrt{q})$, and the *DRO variance* s_α^2 has two forms (Theorem 5.1): intrinsically it is the inverse Hessian $\langle \hat{L}, \mathcal{H}_\alpha^{-1}\hat{L}\rangle$ of $H_\alpha(\cdot\|\sigma)$ at σ, a generalized variance in the sense of Petz dual to the sandwiched Rényi metric of Takahashi and Fujiwara, and extrinsically it is $\mathrm{Var}_{\omega_\alpha}(\ell)/\alpha$, the variance of the classicalizing measure. At matched curvature this variance decreases from the Born variance at $\alpha = \frac{1}{2}$ to that of the conditional-value-at-risk measure at $\alpha = \infty$, strictly when $[\sigma, L] \neq 0$ (Theorem 5.3).

Finally, classicalization makes the concentration bounds of classical probability available for *non-asymptotic* statements, and the comparison carries one message, a *quantum deflation*: noncommutativity reduces the worst case. Every classical bound evaluated on the Born measure ν_L of L in the state σ is a valid quantum bound, using Golden–Thompson and Araki–Lieb–Thirring inequalities (Proposition 6.1). The spectral measure ω_α shares mean and support with ν_L but has a strictly smaller variance for every $\alpha > \frac{1}{2}$ whenever $[\sigma, L] \neq 0$, so the Bernstein bound built on ω_α, sharp over the moment class, lies strictly below the same bound built on ν_L whenever the measured certificate is below saturation (Theorem 6.1, Proposition 6.3, Figure 6), and explicit bounds follow at every order from three anchors, the fidelity at $\alpha = \frac{1}{2}$, Bernstein's inequality at $\alpha = 1$ and Cauchy–Schwarz at $\alpha = 2$ (Proposition 6.4). Read on the Helstrom observable, the same bounds are Pinsker-type inequalities: Pinsker, its optimal form and Fuchs–van de Graaf are recovered with no quantum penalty, the optimal trace-distance versus H_α inequality being the classical one (Remark 6.3); the deflated tier gives a Bernstein–Pinsker inequality with the Kubo–Mori variance (Corollary 6.1); and at $\alpha = \frac{1}{2}$ the fidelity robustness bound of Weber et al., read on the Helstrom observable, is Fuchs–van de Graaf times the cosine of the mismatch between the Helstrom weight and the extremal weight, with a spectrum-dependent refinement (Theorem 6.3, Proposition 6.5).

Related work. Guarantees of this form are what quantum state estimation delivers: tomography and certification protocols return a reconstructed state with a confidence region in fidelity or trace distance [Christandl and Renner, 2012, Faist and Renner, 2016, Flammia and Liu, 2011, Haah et al., 2017, O'Donnell and Wright, 2016], classical shadows predict expectation values from few measurements with error bars of the same kind [Huang et al., 2020], and the robust value converts such a guarantee on the state into a guarantee on any observable. The same worst-case-over-a-ball structure underlies the robustness certification of quantum classifiers [Lu et al., 2020, Weber et al.,

2021, Guan et al., 2021], the smoothing of one-shot entropies [Tomamichel, 2016] and, at $\alpha = 1$, the sensitivity of Gibbs-state expectations to misspecification of the Hamiltonian, the quantum form of the model-bias question of [Gourgoulias et al., 2020]. The problem itself has been met at the endpoints of the family: at $\alpha = 1$ in quantum risk-sensitive control, where Yamamoto and Bouten [Yamamoto and Bouten, 2009] and Vladimirov, Petersen and James [Vladimirov et al., 2018] bound the worst case over a relative-entropy ball through the Golden–Thompson inequality, the quantum counterpart of [Petersen et al., 2000]; at $\alpha = \frac{1}{2}$, where Weber et al. [Weber et al., 2022, Thms. 1 and 2] bound the expectation of an observable over a fidelity ball by its mean and range, or its mean and variance, in the reference state, the mean-only bound being implicit, for pure states, in [Weber et al., 2021]; and at $\alpha = \infty$, where for $L \geq 0$ the value is, up to normalization, the quantum relative Lorenz curve of Buscemi and Gour [Buscemi and Gour, 2017], represented by the layer-cake measure of Gour [Gour, 2026] and of Liu, Hirche and Cheng [Liu et al., 2025], which Schmitt [Schmitt, 2026] obtains independently and concurrently as a push-forward of Heinävaara's measure. What the paper adds at each endpoint is stated at the corresponding results (Proposition 6.1, Theorem 6.2, Theorem 4.1). Neighbouring lines of work are the thermodynamic uncertainty relations for Petz f-divergences of [Salazar, 2026], mean–variance inequalities for a different family, and the use of CVaR [Barkoutsos et al., 2020], of the optimized certainty equivalent [Zhu and Simeone, 2026] and of the tilted loss [Qiu et al., 2026] in variational and learning settings, for the classical outcome of a measurement.

The paper is organized as follows. Section 2 sets up the DRO problem and its duality for an abstract divergence. Section 3 computes the dual for the sandwiched Rényi family. Section 4 proves the classicalization theorem and constructs the measures ω_α. Section 5 treats the linearization and the quantum Fisher metrics, Section 6 the concentration, comparison and state-discrimination bounds. Section 7 collects open problems, and Appendix A the classical proofs.

Notation

M_n^{sa}: Hermitian $n \times n$ matrices. $\mathcal{S}_n \subset M_n^{\mathrm{sa}}$: density matrices (trace-one, positive semidefinite). $\sigma \in \mathcal{S}_n$, $\sigma > 0$: the fixed reference state, eigenvalues $p_1, \dots, p_n$, ρ: a generic state. $L = L^* \in M_n^{\mathrm{sa}}$: the observable, eigenvalues $l_1 \leq \cdots \leq l_n$. $\mu := \operatorname{tr}(L\sigma)$, $\hat{L} := L - \mu I$, I is the identity; $\langle A, B\rangle = \operatorname{tr}(AB)$.

$F(t), G(s), R(x), V(q)$: the abstract CGF, free energy, rate function, and value function (Section 2). $q_{\max}$: the saturation budget, $V(q) = \lambda_{\max}(L)$ iff $q \geq q_{\max}$ (7).

Rényi data ($\alpha \in [\frac{1}{2}, \infty]$, Section 3): $\alpha' = \alpha/(\alpha-1)$; $\beta = (\alpha-1)/(2\alpha)$ for $\alpha > 1$, $\gamma := -\beta$ for $\alpha < 1$ (9); $p := -\alpha' = \alpha/(1-\alpha) \geq 1$, the Stieltjes order on $\alpha < 1$. Q_α: sandwiched Rényi power (10). $H_\alpha = (Q_\alpha - 1)/(\alpha - 1)$: the divergence dualized throughout (Definition 3.1); $D_\alpha = \log(Q_\alpha)/(\alpha-1)$: its logarithmic form (12); q: the budget, $\kappa_\alpha(q) := (1 + (\alpha-1)q)^{1/\alpha}$. $\mathrm{Fid}(\rho, \sigma) := \operatorname{tr}|\sigma^{1/2}\rho^{1/2}|$: root fidelity, $H_{1/2} = 2(1 - \mathrm{Fid})$. $T(\rho, \sigma) := \frac{1}{2}\|\rho - \sigma\|_1$: trace distance.

Spectral measures (Section 4): ω_α: the quantum spectral measure, $V_\alpha(q) = V_\alpha^{\mathrm{cl}}(q; \omega_\alpha)$. ω_1: the Stahl/BMV measure. $\nu_L = \sum_i \langle i|\sigma|i\rangle\, \delta_{l_i} = \omega_{1/2}$: the Born measure. ω_∞: the CVaR endpoint measure. $V_\alpha^{\mathrm{cl}}(q; \nu)$: the classical DRO of a law ν (Lemma 4.1); $V_\alpha^{\mathrm{meas}} := V_\alpha^{\mathrm{cl}}(\cdot; \nu_L)$: the measured DRO.

$\mathrm{Var}_\sigma(L) := \operatorname{tr}(\sigma \hat{L}^2)$: the Born variance.

Variance and metric (Section 5): $s_\alpha^2(\hat{L}) = \langle \hat{L}, \mathcal{H}_\alpha^{-1}\hat{L}\rangle = \mathrm{Var}_{\omega_\alpha}(\ell)/\alpha$: the DRO variance for $\mathcal{D} = H_\alpha$ (Corollary 5.3), specializing the curvature $s_{\mathcal{D}}^2$ of Proposition 2.3. $\mathcal{H}_\alpha$: the Hessian of $H_\alpha(\cdot\|\sigma)$ at σ, with kernel $h_\alpha(p_i, p_j)$ (60). Symbols used within a single section — the tracial measure $\mu_{A,B}$ and its density (Section 4), the mean functions m_α and representing functions $\mathfrak{f}_\alpha$ (Section 5),

the data (a, b, Δ, θ), transforms, profiles and certificates of Section 6 — are defined where they are used.

2 The DRO framework

This section isolates the convex-duality skeleton common to all divergences — strong duality, the constraint–objective exchange, the four functions F, G, R, V and their interrelations — for any divergence $\mathcal{D}$ satisfying (D1)–(D3) below. The formalism applies equally to classical and quantum state spaces, with the trace pairing replacing integration. Nothing in this section is new: the results are standard for divergence balls of probability measures [Ben-Tal et al., 2013, Shapiro, 2017, Rahimian and Mehrotra, 2022, Kuhn et al., 2025], and references are given at the statements.

2.1 The primal problem

Let M_n^{sa} be the Hermitian $n \times n$ matrices, let $\mathcal{S}_n = \{\rho \in M_n^{\mathrm{sa}} : \rho \geq 0, \operatorname{tr}\rho = 1\}$ be the set of states (density matrices) on $\mathbb{C}^n$, let $\sigma \in \mathcal{S}_n$ with $\sigma > 0$ be a faithful nominal state, and let $\mathcal{D}(\cdot\|\sigma) : \mathcal{S}_n \to [0, +\infty]$ be a divergence satisfying:

(D1) **Non-negativity.** $\mathcal{D}(\rho\|\sigma) \geq 0$ for all $\rho \in \mathcal{S}_n$.

(D2) **Faithfulness.** $\mathcal{D}(\rho\|\sigma) = 0$ if and only if $\rho = \sigma$.

(D3) **Convexity and lower-semicontinuity.** $\rho \mapsto \mathcal{D}(\rho\|\sigma)$ is convex and lower semicontinuous on $\mathcal{S}_n$.

Remark. (D1) and (D3) suffice for the strong duality of Proposition 2.1; (D2) is used only for the normalization $V(0) = \operatorname{tr}(L\sigma)$. All divergences in this paper satisfy (D1)–(D3) and are finite on $\mathcal{S}_n$ because $\sigma > 0$; sublevel sets $\{\mathcal{D}(\cdot\|\sigma) \leq q\}$ are compact, so the extrema over them below are attained.

Definition 2.1 (Distribution model risk)**.** For a Hermitian observable $L \in M_n^{\mathrm{sa}}$ and an uncertainty level $q \geq 0$, the **distribution model risk** is

$$V(q; L) = \sup\{\operatorname{tr}(L\,\rho) : \rho \in \mathcal{S}_n,\ \mathcal{D}(\rho\|\sigma) \leq q\}. \tag{1}$$

This asks: what is the worst-case expected value of the observable L over all states within a $\mathcal{D}$-ball of radius q around σ? The name follows Breuer and Csiszár [Breuer and Csiszár, 2013, 2016], who define distribution model risk classically as exactly this quantity; the same object is the robust risk measurement of Glasserman and Xu [Glasserman and Xu, 2014] in finance and, for the relative entropy, the model-misspecification problem of Petersen, James and Dupuis [Petersen et al., 2000] and of Hansen and Sargent [Hansen and Sargent, 2008].

Remark 2.1. The infimum $U(q; L) = \inf\{\operatorname{tr}(L\rho) : \rho \in \mathcal{S}_n, \mathcal{D}(\rho\|\sigma) \leq q\}$ satisfies $U(q; L) = -V(q; -L)$, so it suffices to study V. Translation equivariance $V(q; L + cI) = V(q; L) + c$ allows us to work with the **centered observable** $\hat{L} = L - \operatorname{tr}(L\sigma)\, I$, for which $V(q; \hat{L}) = \sup\{\operatorname{tr}(\hat{L}(\rho - \sigma)) : \mathcal{D}(\rho\|\sigma) \leq q\}$ is the maximal deviation from the baseline expectation.

2.2 Strong duality

Fix an observable $L \in M_n^{\mathrm{sa}}$ and consider the **value function** $V(q) = V(q; L)$ as a function of the uncertainty level q. The following result is the standard duality for divergence balls [Ben-Tal et al., 2013, Shapiro, 2017, Breuer and Csiszár, 2016], whose penalty form goes back to [Ben-Tal et al., 1991], stated for the trace pairing: strong duality reduces the computation of the value function — a priori an optimization over $\mathcal{S}_n$ — to a one-dimensional optimization.

Proposition 2.1 (Properties of the value function and strong duality)**.** *For a fixed observable $L \in M_n^{sa}$, the value function $V(q) = V(q; L)$ satisfies the following properties.*

(i) ***Monotonicity.*** *$V(q)$ is nondecreasing on $[0, \infty)$.*

(ii) ***Concavity.*** *$V(q)$ is concave on $[0, \infty)$.*

(iii) ***Bounds.*** *$\mathrm{tr}(L\sigma) \le V(q) \le \lambda_{\max}(L)$ for all $q \ge 0$.*

(iv) ***Strong duality.*** *We have*

$$V(q) = \inf_{s \ge 0} \{ s\, q + G(s) \}, \tag{2}$$

where the ***free energy*** *G is*

$$G(s) = \sup_{\rho \in \mathcal{S}_n} \{ \mathrm{tr}(L\rho) - s\, \mathcal{D}(\rho \| \sigma) \},$$

the supremum being over the effective domain $\{\mathcal{D}(\cdot\|\sigma) < \infty\}$. G is convex and nonincreasing, with $G(0) = \sup\{\mathrm{tr}(L\rho) : \mathcal{D}(\rho\|\sigma) < \infty\}$, which equals $\lambda_{\max}(L)$ whenever $\mathcal{D}(\cdot\|\sigma)$ is finite on the top eigenspace of L.

Proof. (i) Larger q gives a larger feasible set. (ii) If ρ_i is optimal for $V(q_i)$ and $\lambda \in [0,1]$, convexity of $\mathcal{D}(\cdot\|\sigma)$ makes $\lambda\rho_1 + (1-\lambda)\rho_2$ feasible for $V(\lambda q_1 + (1-\lambda)q_2)$, with objective $\lambda V(q_1) + (1-\lambda)V(q_2)$. (iii) $\rho = \sigma$ is feasible for every $q \ge 0$ by (D2), and $\mathrm{tr}(L\rho) \le \lambda_{\max}(L)$ for every state.

(iv) Define $g(q) := -V(q)$ for $q \ge 0$ and $g := +\infty$ on $(-\infty, 0)$: proper by (iii), convex by (ii), and lower semicontinuous because V is continuous on $[0,\infty)$ — on $(0,\infty)$ as a finite concave function, and at 0 because if $q_k \downarrow 0$ and ρ_k is optimal for $V(q_k)$, every limit point ρ satisfies $\mathcal{D}(\rho\|\sigma) \le \liminf_k \mathcal{D}(\rho_k\|\sigma) = 0$, so ρ is feasible at $q = 0$ and $\lim_k V(q_k) = \mathrm{tr}(L\rho) \le V(0) \le \lim_k V(q_k)$ by (i). Hence $g = g^{**}$ (Fenchel–Moreau). The conjugate $g^*(\mu) = \sup_{q \ge 0}\{\mu q + V(q)\}$ is $+\infty$ for $\mu > 0$ by (iii); for $\mu = -s < 0$, writing $V(q)$ as a supremum over $\{\mathcal{D}(\rho\|\sigma) \le q\}$ and exchanging the two suprema, the decreasing map $q \mapsto -sq$ is maximized at $q = \mathcal{D}(\rho\|\sigma)$, so $g^*(-s) = \sup_\rho\{\mathrm{tr}(L\rho) - s\mathcal{D}(\rho\|\sigma)\} = G(s)$; and $g^*(0) = \sup_q V(q) = G(0)$, a state being feasible for some finite q iff its divergence is finite. Then $V(q) = -g^{**}(q) = \inf_{\mu \le 0}\{g^*(\mu) - \mu q\} = \inf_{s \ge 0}\{sq + G(s)\}$. □ □

2.3 The rate function

The value function $V(q)$ answers the question "given a divergence budget q, how large can $\mathrm{tr}(\hat{L}\,\rho)$ be?" We introduce next the **rate function** which asks the reverse: given a target expectation x, what is the minimal divergence cost to achieve it?

Definition 2.2 (Rate function)**.** Given $L \in M_n^{\mathrm{sa}}$, $x \in \mathbb{R}$, and a baseline model σ, the rate function is

$$R(x, L) := \inf\{\mathcal{D}(\rho\|\sigma) : \rho \in \mathcal{S}_n,\ \mathrm{tr}(L\,\rho) \ge x\}. \tag{3}$$

The names follow the large-deviation analogy — for the classical relative entropy R is the rate function of Cramér's theorem — whose convexity structure survives even where the probabilistic interpretation does not. Alongside the **free energy** G of Proposition 2.1 we introduce the **abstract cumulant generating function**

$$F(t) := \sup_{\rho \in \mathcal{S}_n} \{ t\,\mathrm{tr}(L\rho) - \mathcal{D}(\rho\|\sigma) \}, \qquad t \in \mathbb{R}, \tag{4}$$

so that G is the perspective of F: $G(s) = s\,F(1/s)$ for $s > 0$, equivalently $F(t) = t\,G(1/t)$. F is convex, $F(0) = 0$, $F(t) \ge t\,\mathrm{tr}(L\sigma)$, and $F(t; L) = F(-t; -L)$. The following proposition connects the rate function to the value function and identifies R with the Fenchel–Legendre transform

$F^*(x) = \sup_{t\geq 0}\{xt - F(t)\}$ of F. Throughout, L is fixed and $R(x) = R(x, L)$. Part (iv) and the perspective formula (6) in its proof are the standard inversion of a Legendre transform [Boucheron et al., 2013]; for the relative entropy see [Ahmadi-Javid, 2012, Breuer and Csiszár, 2013, Chowdhary and Dupuis, 2013], for Rényi divergences [Atar et al., 2015], and for the same inversion as a goal-oriented uncertainty bound [Dupuis et al., 2016].

Proposition 2.2 (Constraint-objective exchange)**.** *For all $q \geq 0$ and $x \in \mathbb{R}$,*

$$V(q) \geq x \quad \Longleftrightarrow \quad R(x) \leq q. \tag{5}$$

Consequently, V and R are generalized inverses:

$$V(q) = \sup\{x \in \mathbb{R} : R(x) \leq q\} \quad \text{and} \quad R(x) = \inf\{q \geq 0 : V(q) \geq x\}.$$

Furthermore:

(i) ***Monotonicity.*** *R is nondecreasing on $[\mathrm{tr}(L\sigma), \infty)$.*

(ii) ***Convexity.*** *R is convex on $[\mathrm{tr}(L\sigma), \infty)$.*

(iii) ***Bounds.*** *$R(\mathrm{tr}(L\sigma)) = 0$ and $R(x) = \infty$ for $x > \lambda_{\max}(L)$.*

(iv) ***Strong duality.*** *$R(x) = F^*(x)$ for all $x \geq \mathrm{tr}(L\sigma)$.*

Proof. The exchange. Both extrema are attained (the feasible sets are compact and the objectives are continuous, resp. lower semicontinuous), so $V(q) \geq x$ and $R(x) \leq q$ each say that some $\rho \in \mathcal{S}_n$ satisfies both $\mathcal{D}(\rho\|\sigma) \leq q$ and $\mathrm{tr}(L\rho) \geq x$. This proves (5) and the generalized-inverse relation.

Properties (i)–(iii). Through the exchange, monotonicity of R is monotonicity of V, and convexity of R is concavity of V: if $R(x_i) \leq q_i$ then $V(\lambda q_1 + (1-\lambda)q_2) \geq \lambda V(q_1) + (1-\lambda)V(q_2) \geq \lambda x_1 + (1-\lambda)x_2$. The bounds follow from (D2) and $\mathrm{tr}(L\rho) \leq \lambda_{\max}(L)$.

Strong duality: $R = F^*$. By Proposition 2.1 and the substitution $s = 1/t$ (the $s = 0$ term never wins, G being nonincreasing), the perspective formula holds:

$$V(q) = \inf_{s>0}\{s\,q + G(s)\} = \inf_{t>0}\left\{\frac{q + F(t)}{t}\right\}. \tag{6}$$

Now, by the exchange (5),

$$R(x) = \inf\{q \geq 0 : V(q) \geq x\}.$$

Using (6): $V(q) \geq x$ if and only if $(q + F(t))/t \geq x$ for all $t > 0$, i.e., $q \geq xt - F(t)$ for all $t > 0$. Therefore

$$R(x) = \inf\{q \geq 0 : q \geq xt - F(t) \text{ for all } t > 0\} = \sup_{t>0}\{xt - F(t)\} = F^*(x),$$

the last equality because the $t = 0$ term of F^* is $0 \leq \sup_{t>0}\{xt - F(t)\}$ when $x \geq \mathrm{tr}(L\sigma)$ (let $t \to 0^+$). □ □

The saturation budget. Since $\mathrm{tr}(L\rho) \leq \lambda_{\max}(L)$ with equality iff ρ is supported on the top eigenspace of L, we set

$$q_{\max} \;:=\; R(\lambda_{\max}(L)) \;=\; \inf\{\mathcal{D}(\rho\|\sigma) : \rho \in \mathcal{S}_n,\ \mathrm{tr}(L\rho) = \lambda_{\max}(L)\}, \tag{7}$$

attained by compactness of the feasible set and lower semicontinuity of $\mathcal{D}$, and finite for every divergence in this paper since $\sigma > 0$; by (5), $V(q) = \lambda_{\max}(L)$ iff $q \geq q_{\max}$.

The four functions form the commutative square of Figure 1, and the entire DRO strategy reduces to a single question: **can you compute $F(t)$ or $G(s)$ explicitly?** This is the only step that depends on $\mathcal{D}$; the rest — concavity and strong duality — is automatic.

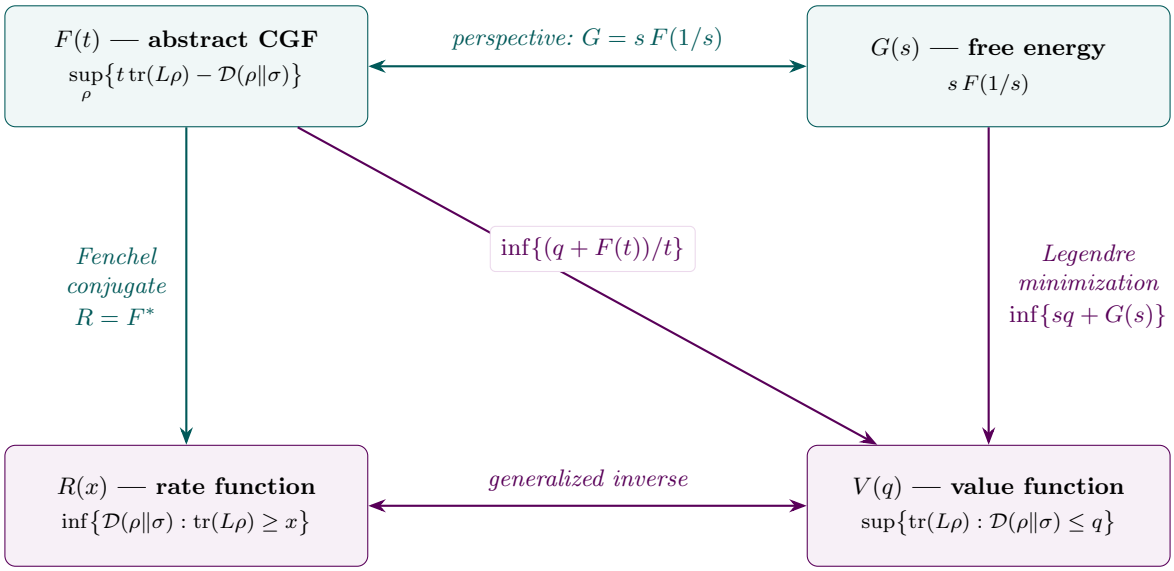


Figure 1: The DRO square: F and G are related by the perspective $G(s) = sF(1/s)$, R and V are generalized inverses, and the two paths from F to V — Fenchel conjugation $R = F^*$ then inversion, or perspective then $V(q) = \inf_s\{sq + G(s)\}$ — commute (Proposition 2.2); the diagonal is $V(q) = \inf_{t>0}\{(q + F(t))/t\}$.

Remark 2.2 (Illustration: relative entropy and the BMV bridge). For the classical relative entropy the Gibbs variational principle gives $F(t) = \log \mathbb{E}_Q[e^{tL}]$, the classical cumulant generating function, and the square is the classical duality between risk-sensitive integrals and relative entropy [Petersen et al., 2000, Hansen and Sargent, 2008, Chowdhary and Dupuis, 2013]. For the Umegaki relative entropy it gives $F(t) = \log \mathrm{tr}\, e^{\log\sigma + tL}$, which is *not* the log-moment generating function $\log \mathrm{tr}[\sigma e^{tL}]$ when $[\sigma, L] \neq 0$ — yet, by Stahl's theorem [Stahl, 2013], is the cumulant generating function of a probability measure ω_1 on the spectral interval of L (Section 4.3.3): the **BMV bridge** that carries classical large-deviation tools to the quantum $\alpha = 1$ problem.

2.4 The rate function picture

Applied to $\hat{L}$ and to $-\hat{L}$, the perspective formula (6) reads both branches of F: the line through $(0, -q)$ tangent to the graph of $F(\cdot; \hat{L})$ on the right has slope $x_+ = V(q; \hat{L})$, the tangent on the left has slope $x_- = -V(q; -\hat{L})$, and $\{x : R(x; \hat{L}) \leq q,\ R(-x; -\hat{L}) \leq q\} = [x_-, x_+]$ is the **two-sided uncertainty interval**:

$$-V(q; -\hat{L}) \ \leq\ \mathrm{tr}(\hat{L}(\rho - \sigma)) \ \leq\ V(q; \hat{L}) \qquad \text{for all } \rho \text{ with } \mathcal{D}(\rho\|\sigma) \leq q, \tag{8}$$

asymmetric in general.

The full convex duality chain $F \to R \to V$ is illustrated in Figure 2.

Read as the inverse of R, V has three regimes: a **square-root onset** as $q \to 0^+$, governed by the curvature of R at the origin; a **strict increase** on $(0, q_{\max})$, where the constraint is active; and **saturation** at $\lambda_{\max}(L)$ for $q \geq q_{\max}$. The onset is the classical square-root law [Dupuis et al., 2016, Lam, 2016, Gotoh et al., 2018, Duchi and Namkoong, 2019], governed by a single number, here called the **DRO variance** $s^2_{\mathcal{D}}(\hat{L})$:

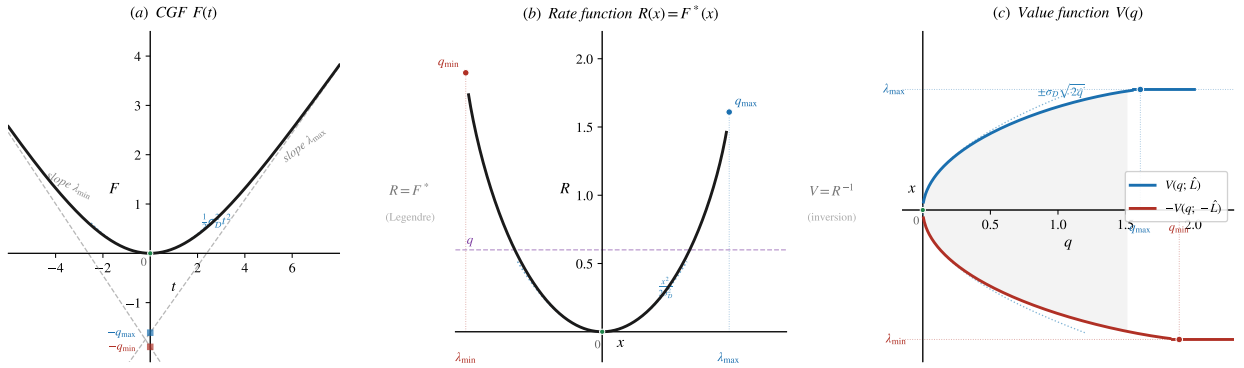


Figure 2: The convex duality chain. *(a)* $F(t)$, convex with $F(0) = F'(0) = 0$, linear asymptotes of slopes $\lambda_{\min}(\hat{L})$, $\lambda_{\max}(\hat{L})$ and tangent parabola $\frac{1}{2}s_{\mathcal{D}}^2 t^2$ at the origin. *(b)* $R = F^*$ on $[\lambda_{\min}, \lambda_{\max}]$, with $R(\lambda_{\max}(\hat{L})) = q_{\max}$. *(c)* The two branches $V(q; \hat{L})$ and $-V(q; -\hat{L})$, diverging as $\pm s_{\mathcal{D}}\sqrt{2q}$ and saturating at $\lambda_{\max}$, $\lambda_{\min}$; the shaded region is (8).

Proposition 2.3 (Square-root law)**.** *Suppose that* $F(t) = \frac{1}{2}s_{\mathcal{D}}^2 t^2 + o(t^2)$ *as* $t \to 0^+$ *for some* $s_{\mathcal{D}} = s_{\mathcal{D}}(\hat{L}) > 0$*. Then*

$$V(q; \hat{L}) = s_{\mathcal{D}}(\hat{L})\sqrt{2q} + o(\sqrt{q}) \qquad \text{as } q \to 0^+.$$

Proof. Write $s := s_{\mathcal{D}}$ and use the perspective formula $V(q; \hat{L}) = \inf_{t>0}(q + F(t))/t$ of (6) for the centered observable. *Upper bound.* Taking $t = \sqrt{2q}/s$ gives $V(q; \hat{L}) \le s\sqrt{2q} + o(\sqrt{q})$. *Lower bound.* Fix $\varepsilon \in (0, s^2)$ and $\delta > 0$ with $F(t) \ge \frac{1}{2}(s^2 - \varepsilon)t^2$ on $(0, \delta]$. For $t \le \delta$, $(q + F(t))/t \ge \inf_{t>0}(q + \frac{1}{2}(s^2 - \varepsilon)t^2)/t = \sqrt{2q(s^2 - \varepsilon)}$. For $t > \delta$: F is convex with $F(0) = 0$, so $F(t)/t$ is nondecreasing and $(q + F(t))/t \ge F(\delta)/\delta \ge \frac{1}{2}(s^2 - \varepsilon)\delta$, which exceeds $\sqrt{2q(s^2 - \varepsilon)}$ once q is small. Hence $V(q; \hat{L}) \ge \sqrt{2q(s^2 - \varepsilon)}$ for small q, and ε is arbitrary. □ □

For $\mathcal{D} = H_\alpha$, $\alpha \in [1/2, \infty)$, the hypothesis is verified in Section 5 for every L that is not a multiple of the identity, with $s_{\mathcal{D}}^2(\hat{L}) = \langle \hat{L}, \mathcal{H}_\alpha^{-1}\hat{L}\rangle$ the inverse Hessian of $H_\alpha(\cdot\|\sigma)$ at σ (Theorem 5.2, Corollary 5.2).

Remark 2.3 (Connection to risk measures). For fixed $q \ge 0$, the map $L \mapsto V(q; L)$ is a **coherent risk measure** in the sense of Artzner, Delbaen, Eber, and Heath [Artzner et al., 1999] — monotone, translation equivariant, positively homogeneous and subadditive, all immediate from the supremum form; at $\alpha = 1$ it is the quantum counterpart of the entropic value-at-risk [Ahmadi-Javid, 2012], with the Rényi extension of [Pichler and Schlotter, 2020] at the other orders, at $\alpha = \infty$ of the conditional value-at-risk of Rockafellar and Uryasev [Rockafellar and Uryasev, 2000] (Section 3.4), and in the dual form of Theorem 3.4 an optimized certainty equivalent [Ben-Tal and Teboulle, 2007].

3 DRO for α-divergences

3.1 Definitions and notations

This section computes $V_\alpha(q)$ on the whole range $\alpha \in [\frac{1}{2}, \infty]$. The answer (Theorem 3.4) is a one-dimensional dual, the quantum form of the optimized certainty equivalent of Ben-Tal and Teboulle [Ben-Tal and Teboulle, 2007], with the worst-case state explicit in terms of the dual minimizer (Corollary 3.1); $\alpha = 1$ is the Gibbs variational principle and $\alpha = \infty$ a quantum conditional

value-at-risk. Everything rests on convexity of $H_\alpha(\cdot\|\sigma)$ in ρ, so the duality of Section 2 applies verbatim.

Quantum refinements of the classical Rényi divergence [Rényi, 1961, van Erven and Harremoës, 2014] are not unique. Alongside the sandwiched family, Petz's quasi-entropies [Petz, 1986, Ohya and Petz, 2004], the α-z divergences [Jakšić et al., 2012, Audenaert and Datta, 2015], and the measured and maximal f-divergences [Hiai et al., 2011, Hiai and Mosonyi, 2017] all agree with it on the commuting locus but differ elsewhere; the measured Rényi divergences, the largest classical divergence obtainable by a measurement, have Donsker–Varadhan-type variational formulas of their own [Berta et al., 2017]. The sandwiched family is monotone under quantum channels for $\alpha \geq \frac{1}{2}$ [Frank and Lieb, 2013] (for $\alpha > 1$ also [Beigi, 2013]) and not below [Berta et al., 2017], and monotone in α [Müller-Lennert et al., 2013]; its $\alpha = 1$ member is the Umegaki relative entropy, monotone under channels by [Lindblad, 1975, Uhlmann, 1977], and its $\alpha = \infty$ endpoint the max-relative entropy [Datta, 2009].

Throughout this section, $\sigma > 0$ denotes a faithful state on $\mathbb{C}^n$, $L = L^*$ a Hermitian observable, and $\alpha \in [\frac{1}{2}, \infty] \setminus \{1\}$ carries the conjugate exponent and sandwiching data

$$\alpha' = \frac{\alpha}{\alpha - 1}, \qquad \beta = \frac{\alpha - 1}{2\alpha} = \frac{1}{2\alpha'}, \qquad \gamma := -\beta = \frac{1 - \alpha}{2\alpha}, \qquad p := -\alpha'\big|_{\alpha<1} = \frac{\alpha}{1 - \alpha} \geq 1. \tag{9}$$

So $\beta > 0$ for $\alpha > 1$ and $\gamma > 0$ for $\alpha \in [\frac{1}{2}, 1)$, where $\alpha' = -p$: β is used on the first range and $\gamma = -\beta$ on the second, so that the sandwiching exponents σ^β, σ^γ read as positive powers. The **sandwiched Rényi power** is given for $\alpha \neq 1$ by

$$Q_\alpha(\rho\|\sigma) := \operatorname{tr}\Big[\big(\sigma^{\frac{1-\alpha}{2\alpha}}\,\rho\,\sigma^{\frac{1-\alpha}{2\alpha}}\big)^\alpha\Big] = \|\sigma^{\frac{1-\alpha}{2\alpha}}\,\rho\,\sigma^{\frac{1-\alpha}{2\alpha}}\|_\alpha^\alpha. \tag{10}$$

The **sandwiched α-divergence** is then defined by

Definition 3.1 (Sandwiched α-divergence)**.**

$$H_\alpha(\rho\|\sigma) := \begin{cases} \frac{Q_\alpha(\rho\|\sigma)-1}{\alpha-1} & \alpha \in [\frac{1}{2}, 1) \cup (1, \infty) \\ \operatorname{tr}(\rho(\log\rho - \log\sigma)) & \alpha = 1 \\ \|\sigma^{-1/2}\rho\,\sigma^{-1/2}\|_\infty - 1 & \alpha = \infty \end{cases}$$

Only convexity of $H_\alpha(\cdot\|\sigma)$ in ρ is used below; at $\alpha = \infty$ it is the only convexity available.

The **DRO primal** with H_α-budget $q \geq 0$ is

$$V_\alpha(q) = \sup\{\operatorname{tr}(L\rho) : \rho \in \mathcal{S}_n,\ H_\alpha(\rho\|\sigma) \leq q\}. \tag{11}$$

The logarithmic sandwiched Rényi divergence is the monotone reparametrization $D_\alpha = \frac{1}{\alpha-1}\log Q_\alpha = \frac{1}{\alpha-1}\log(1 + (\alpha-1)H_\alpha)$ (with $D_1 = H_1$ and $D_\infty = \log(1 + H_\infty)$), so the H_α-ball of radius q *is* the D_α-ball of radius

$$r = \frac{\log(1 + (\alpha-1)q)}{\alpha - 1}, \qquad \text{equivalently} \qquad q = \frac{e^{(\alpha-1)r} - 1}{\alpha - 1} \tag{12}$$

($r = q$ at $\alpha = 1$, $r = \log(1 + q)$ at $\alpha = \infty$), and at each fixed order every result below transfers verbatim to the logarithmic convention through this dictionary. We dualize H_α rather than D_α for a convexity reason: $H_\alpha(\cdot\|\sigma)$ is convex (axiom (D3) of Section 2), which is what the DRO duality rests on, whereas D_α is jointly convex only for $\alpha \in [\frac{1}{2}, 1]$ (as in the classical case [van Erven and Harremoës, 2014]) and for $\alpha > 1$, including $\alpha = \infty$, is not even convex in ρ — already for commuting pairs.

Remark 3.1 (Monotonicity in α and the limit $\alpha \to \infty$). For every pair, $\alpha \mapsto D_\alpha(\rho\|\sigma)$ is nondecreasing on $[\frac{1}{2}, \infty)$ [Müller-Lennert et al., 2013, Thm. 7; Beigi, 2013], and $x \mapsto (e^{(\alpha-1)x} - 1)/(\alpha - 1)$ is increasing in both α and x; hence $\alpha \mapsto H_\alpha(\rho\|\sigma)$ is nondecreasing as well, the H_α-balls of a fixed radius are nested downward in α, and

$$\alpha \mapsto V_\alpha(q) \;\text{ is nonincreasing on } [\tfrac{1}{2}, \infty) \text{ for every } q \geq 0$$

(visible in Figure 3). The same holds classically for H_α^{cl} [van Erven and Harremoës, 2014, Thm. 3], which is how it is used in Proposition 6.4.

The limit $\alpha \to \infty$, however, belongs to the Rényi scale. At fixed q the radius r of (12) tends to 0, so the balls shrink to $\{\sigma\}$ and $V_\alpha(q) \downarrow \mathrm{tr}(L\sigma)$. At fixed r the balls $\{D_\alpha \leq r\}$ decrease to $\{D_\infty \leq r\} = \{H_\infty \leq e^r - 1\}$, and the values $V_\alpha(q)$ at $q = (e^{(\alpha-1)r} - 1)/(\alpha - 1)$ decrease to $V_\infty(e^r - 1)$. Thus H_∞ in Definition 3.1 is the endpoint of the Rényi scale, not a limit of the H_α.

Remark 3.2 (Minimality of the sandwiched family). Any quantum Rényi divergence D_α that is additive under tensor products, monotone under quantum channels and equal to the classical Rényi divergence on commuting pairs dominates the sandwiched one: pinching $\rho^{\otimes n}$ in the eigenbasis of $\sigma^{\otimes n}$ gives $D_\alpha(\rho\|\sigma) \geq \frac{1}{n} D_\alpha(\mathcal{P}_{\sigma^{\otimes n}}(\rho^{\otimes n}) \,\|\, \sigma^{\otimes n})$, a classical quantity whose limit as $n \to \infty$ is the sandwiched divergence [Tomamichel, 2016, §4.2.2 and Prop. 4.12]; for this reason Tomamichel calls the latter the *minimal* quantum Rényi divergence. Since the ball of such a D_α is contained in the sandwiched ball of the same radius, in either normalization, its robust value is at most $V_\alpha(q)$: the value of Theorem 3.4 is the largest, and every upper bound of Section 6 remains valid, for every admissible quantum Rényi divergence at the same order and budget. Petz's family dominates it at every order, by the Araki–Lieb–Thirring inequality; the measured Rényi divergence, which lies below the sandwiched one [Berta et al., 2017], is not additive and is not covered.

3.2 Relative entropy DRO ($\alpha = 1$)

At $\alpha = 1$ the generalized cumulant is the Gibbs free energy $\log \mathrm{tr}\, e^{\log\sigma + tL}$, in closed form, whereas for $\alpha \neq 1$ it carries a one-dimensional optimization. The quantum Gibbs variational principle below is a classical result [Petz, 1988, Ohya and Petz, 2004, Donald, 1986]; we include a short, self-contained proof.

Theorem 3.1 (Gibbs Variational Principle). *For $\sigma > 0$ and $L = L^*$,*

$$\sup_{\rho \in \mathcal{S}_n} \{\mathrm{tr}(L\rho) - H_1(\rho\|\sigma)\} = \log \mathrm{tr}(e^{\log\sigma + L}),$$

attained uniquely at the Gibbs state $\tau := e^{\log\sigma + L}/\mathrm{tr}(e^{\log\sigma + L})$.

Proof. A direct computation gives the key identity

$$H_1(\rho\|\tau) = H_1(\rho\|\sigma) - \mathrm{tr}(L\rho) + \log \mathrm{tr}(e^{\log\sigma + L}),$$

so $\mathrm{tr}(L\rho) - H_1(\rho\|\sigma) = \log \mathrm{tr}(e^{\log\sigma + L}) - H_1(\rho\|\tau)$. Non-negativity and faithfulness of H_1 give the supremum and the uniqueness of the maximizer $\rho = \tau$. □ □

The Gibbs cumulant $F_1(t) := \log \mathrm{tr}(e^{\log\sigma + tL})$, $t > 0$, identifies a one-parameter family of tilted Gibbs states

$$\sigma_t := \frac{e^{\log\sigma + tL}}{\mathrm{tr}(e^{\log\sigma + tL})}, \qquad F_1'(t) = \mathrm{tr}(L\sigma_t),$$

and the Legendre identity $H_1(\sigma_t\|\sigma) = tF_1'(t) - F_1(t)$ follows by the same computation as in the proof of Theorem 3.1.

The next theorem is the quantum counterpart of the classical relative-entropy DRO [Petersen et al., 2000, Ahmadi-Javid, 2012, Breuer and Csiszár, 2013]. In quantum risk-sensitive control, the expectation under a state at finite relative entropy from a nominal state is bounded through the Gibbs variational principle and the Golden–Thompson inequality, that is with $\log \operatorname{tr}(\sigma e^{tL})$ in place of $F_1(t)$ [Yamamoto and Bouten, 2009, Lemma 2.1 and eq. (13)], and the worst case over a relative-entropy ball is bounded by the corresponding infimum over the risk parameter [Vladimirov et al., 2018, Thm. 4]. The theorem gives the exact value, with F_1, and the worst-case state.

Theorem 3.2 (Relative entropy robustness)**.** *Let $\sigma > 0$ and let $L = L^*$ not be a multiple of the identity. For every $q \geq 0$,*

$$V_1(q) \;=\; \inf_{t>0} \frac{q + F_1(t)}{t}. \tag{13}$$

For $q \in (0, q_{\max})$ ($q_{\max}$ as in (7)) the infimum is attained at the unique $t^* > 0$ satisfying $H_1(\sigma_{t^*}\|\sigma) = q$, and

$$V_1(q) \;=\; F_1'(t^*) \;=\; \operatorname{tr}(L\,\sigma_{t^*}). \tag{14}$$

Proof. Equation (13) is the perspective formula (6), since by Theorem 3.1 applied to tL the abstract CGF of (4) is $F = F_1$. Setting the derivative in t to zero,

$$\frac{d}{dt}\frac{q + F_1(t)}{t} = \frac{tF_1'(t) - F_1(t) - q}{t^2},$$

gives the stationarity condition $tF_1'(t) - F_1(t) = q$, which by the Legendre identity above is exactly $H_1(\sigma_t\|\sigma) = q$. Existence and uniqueness of t^* follow from strict convexity of F_1 — its second derivative $F_1''(t)$ is the Kubo–Mori (Bogoliubov) variance of L in the tilted state σ_t, hence strictly positive since L is not a multiple of the identity [Petz and Toth, 1993, Hiai and Petz, 2014] — as follows: $\frac{d}{dt}H_1(\sigma_t\|\sigma) = tF_1''(t) > 0$, so $t \mapsto H_1(\sigma_t\|\sigma)$ is continuous and strictly increasing on $(0,\infty)$, starting from 0; and its limit as $t \to \infty$ is $\lim_t\{\lambda_{\max}(L)\,t - F_1(t)\} = F_1^*(\lambda_{\max}(L)) = q_{\max}$ by Proposition 2.2(iv) and (7). More explicitly, if P_n is the spectral projection of L at $l_n = \lambda_{\max}(L)$, finite-dimensional analytic perturbation of $L + t^{-1}\log\sigma$ gives

$$F_1(t) = tl_n + \log \operatorname{tr}_{\operatorname{Ran} P_n} \exp(P_n \log \sigma P_n) + O(t^{-1}),$$

and differentiation (or the same perturbation expansion one order further) yields $t(l_n - F_1'(t)) \to 0$. Notice that in the noncommuting case the spectral subspace itself need only approach $\operatorname{Ran} P_n$ at order t^{-1}; exponential convergence of σ_t to that fixed subspace is not required and is generally false. At t^*, (13) gives $V_1(q) = (q + F_1(t^*))/t^* = t^*F_1'(t^*)/t^* = F_1'(t^*)$, which is (14). □ □

3.3 DRO for finite $\alpha \neq 1$

For $\alpha \in [\frac{1}{2}, \infty)\backslash\{1\}$ the sandwiched divergence H_α is convex in its first argument — for $\alpha > 1$ because Q_α itself is convex in ρ, and for $\alpha \in [\frac{1}{2}, 1)$ because Q_α is *concave* and $H_\alpha = (Q_\alpha - 1)/(\alpha - 1)$ divides by the negative constant $\alpha - 1$ — so the abstract DRO machinery of Section 2 applies throughout. The entire calculation reduces to computing the generalized cumulant $G_\alpha(s) = \sup_\rho\{\operatorname{tr}(L\rho) - sH_\alpha(\rho\|\sigma)\}$, which Theorem 3.3 below obtains uniformly on both sides of $\alpha = 1$ — via the signed variable $s = 1/(\alpha - 1)$ tracking the convexity/concavity of Q_α — from the convex duality of trace functions, a matrix Fenchel–Young inequality; see [Carlen, 2010] for this circle of ideas. (The formulas below hold for all $\alpha \in (0, 1)$; the range $\alpha \geq \frac{1}{2}$ is where H_α is monotone under quantum channels and where the classicalizing measure of Section 4 is positive.)

Frank and Lieb [Frank and Lieb, 2013, Lemma 4] write Q_α as a supremum, respectively an infimum, of functions affine in ρ; the variational formula below is the dual companion of that representation, derived here directly from the Fenchel–Young inequality for trace functions.

3.3.1 Variational formula and its consequences

Theorem 3.3 (Quantum variational formula for H_α)**.** *For $\sigma > 0$, $L = L^*$ and $\alpha \in [\frac{1}{2}, \infty) \setminus \{1\}$, let $\widehat{G}(s) := \sup_{\rho \in \mathcal{S}_n} \{\mathrm{tr}(L\rho) - s\, Q_\alpha(\rho\|\sigma)\}$.*
Case (i), *$\alpha > 1$: for every $s > 0$,*

$$\widehat{G}(s) = \inf_{c \in \mathbb{R}} \left\{ c + \frac{\alpha - 1}{\alpha^{\alpha'}\, s^{\alpha'-1}} \left\| (\sigma^\beta (L - cI) \sigma^\beta)_+ \right\|_{\alpha'}^{\alpha'} \right\}. \tag{15}$$

Case (ii), *$\alpha \in [\frac{1}{2}, 1)$: for every $\zeta > 0$,*

$$\widehat{G}(-\zeta) = \inf_{c > \lambda_{\max}(L)} \left\{ c + (1 - \alpha)\, \alpha^p\, \zeta^{1+p}\, \|\sigma^\gamma (cI - L)^{-1} \sigma^\gamma\|_p^p \right\}. \tag{16}$$

Proof. Both cases are proved by the same four steps, for s with $\mathrm{sign}(s) = \mathrm{sign}(\alpha - 1)$ — the sign, verified in Step 1, that makes $\widehat{G}$ the supremum of a concave function; only Step 3 depends on it. Recall $Q_\alpha(\rho\|\sigma) = \|\sigma^{-\beta}\rho\sigma^{-\beta}\|_\alpha^\alpha$.
Step 1 — Infimal-convolution split. On M_n^{sa} with pairing $\langle A, \rho\rangle = \mathrm{tr}(A\rho)$, define

$$M(\rho) = \begin{cases} s\, Q_\alpha(\rho\|\sigma) & \rho \geq 0 \\ +\infty & \text{else} \end{cases}, \qquad N(\rho) = \begin{cases} 0 & \mathrm{tr}\,\rho = 1 \\ +\infty & \text{else} \end{cases}.$$

N is proper, convex, and lsc. So is M: $\rho \mapsto Q_\alpha(\rho\|\sigma)$ is convex on the positive cone for $\alpha > 1$ and concave for $\alpha \in [\frac{1}{2}, 1)$ — since $X \mapsto \mathrm{tr}\, X^\alpha$ is convex, respectively concave, on positive matrices, as $\mathrm{tr}\, f$ inherits the convexity of the scalar function f [Carlen, 2010] — so sQ_α is convex exactly when $\mathrm{sign}(s) = \mathrm{sign}(\alpha - 1)$, the case at hand. Since $I/n \in \mathrm{ri}(\mathrm{dom}\, M) \cap \mathrm{ri}(\mathrm{dom}\, N)$, the Moreau–Rockafellar / Fenchel duality theorem [Rockafellar, 1970] gives

$$\widehat{G}(s) = (M + N)^*(L) = \inf_A \{M^*(L - A) + N^*(A)\} = \inf_{c \in \mathbb{R}} \{c + M^*(L - cI)\},$$

the last step because $N^*(A) = c$ for $A = cI$ and $+\infty$ otherwise, collapsing the A-infimum onto that line; here $M^*(B) = \sup_{\rho \geq 0} \{\mathrm{tr}(B\rho) - sQ_\alpha(\rho\|\sigma)\}$.
Step 2 — Change of variables. Set $X = \sigma^{-\beta}\rho\sigma^{-\beta}$ (so $\rho = \sigma^\beta X \sigma^\beta$), a bijection between positive cones — the same substitution on both sides of $\alpha = 1$. By cyclicity, $\mathrm{tr}(B\rho) = \mathrm{tr}(\widetilde{B}X)$ with $\widetilde{B} := \sigma^\beta B \sigma^\beta$, and $M^*(B) = \Psi(\widetilde{B})$ where

$$\Psi(A) := \sup_{X \geq 0} \{\mathrm{tr}(AX) - s\|X\|_\alpha^\alpha\}.$$

Step 3 — Computation of $\Psi(A)$: the sign split. Let $a_1 \geq \cdots \geq a_n$ and $x_1 \geq \cdots \geq x_n \geq 0$ be the eigenvalues of A, X. By von Neumann's trace inequality [Bhatia, 1997], $\mathrm{tr}(AX) - s\|X\|_\alpha^\alpha \leq \sum_i (a_i x_i - s x_i^\alpha)$, with equality when A, X share a co-ordered eigenbasis, reducing $\Psi(A)$ to n copies of the scalar problem $\sup_{x \geq 0}(ax - sx^\alpha)$.
(a) $\alpha > 1$ ($s > 0$). The optimum is 0 at $x^* = 0$ for $a \leq 0$; for $a > 0$, $x^*(a) = (a/(s\alpha))^{1/(\alpha-1)}$, with value $\frac{\alpha-1}{\alpha^{\alpha'} s^{\alpha'-1}} a^{\alpha'}$. Since $a \mapsto x^*(a_+)$ is nondecreasing, the optimal x_i^* inherit the order of a_i, and equality in von Neumann is attained by diagonalizing X in the eigenbasis of A. Summing,

$$\Psi(A) = \frac{\alpha - 1}{\alpha^{\alpha'}\, s^{\alpha'-1}} \|A_+\|_{\alpha'}^{\alpha'}. \tag{17}$$

(b) $\alpha \in [\frac{1}{2}, 1)$ ($s < 0$; write $\zeta := -s > 0$). Here $ax - sx^\alpha = ax + \zeta x^\alpha$; since $x \mapsto x^\alpha$ is sublinear, the supremum is $+\infty$ whenever $a \geq 0$, so finiteness of $\Psi(A)$ forces $A < 0$. For $a < 0$,

$x^*(a) = (\zeta\alpha/|a|)^{1/(1-\alpha)}$, with value $(1-\alpha)\zeta(\zeta\alpha/|a|)^p$; again $a \mapsto x^*(a)$ is nondecreasing on $\{a < 0\}$ (it decreases in $|a|$), so the rearrangement is non-binding and equality is attained by diagonalizing X in the eigenbasis of A. Summing,

$$\Psi(A) = \begin{cases} (1-\alpha)\,\alpha^p\,\zeta^{1+p}\,\|(-A)^{-1}\|_p^p & A < 0 \\ +\infty & \text{otherwise.} \end{cases} \tag{18}$$

Step 4 — Reassembly. Combining Steps 1–3 with $B = L - cI$, $\widetilde{B} = \sigma^\beta(L-cI)\sigma^\beta$: for $s > 0$, (17) gives (15); for $s = -\zeta < 0$, finiteness of $\Psi(\widetilde{B})$ forces $\widetilde{B} < 0$, i.e. $c > \lambda_{\max}(L)$, and $(-\widetilde{B})^{-1} = \sigma^{-\beta}(cI-L)^{-1}\sigma^{-\beta} = \sigma^\gamma(cI-L)^{-1}\sigma^\gamma$ in (18) gives (16). □ □

3.3.2 DRO value

For commuting σ and L, case (i) of the following theorem is the dual formula of Duchi and Namkoong [Duchi and Namkoong, 2021, Lemma 1], constant included, in their normalization of the radius (q/α in place of q); see Lemma 4.1. The norm $X \mapsto \|\sigma^\beta X \sigma^\beta\|_{\alpha'}$ is the σ-weighted noncommutative $L_{\alpha'}$ norm of [Beigi, 2013, §II]; in the theorem it is applied to a positive part taken after sandwiching.

Theorem 3.4 (DRO for sandwiched α-divergence)**.** *Let $\sigma > 0$, $L = L^*$, and $q \geq 0$.*

Case (i), *$\alpha > 1$:*

$$V_\alpha(q) = \inf_{c\in\mathbb{R}} \left\{ c + (1+(\alpha-1)q)^{1/\alpha} \left\| (\sigma^\beta(L-cI)\sigma^\beta)_+ \right\|_{\alpha'} \right\}, \tag{19}$$

and for $q > 0$ the infimum is attained.

Case (ii), *$\alpha \in [\frac{1}{2}, 1)$:*

$$V_\alpha(q) = \inf_{c > \lambda_{\max}(L)} \left\{ c - \frac{(1+(\alpha-1)q)_+^{1/\alpha}}{\left\| \sigma^\gamma(cI-L)^{-1}\sigma^\gamma \right\|_p} \right\}, \tag{20}$$

and for $q \in (0, q_{\max})$ the infimum is attained.

The positive part in the prefactor of (ii) matters only for $q \geq 1/(1-\alpha)$, where $H_\alpha \leq 1/(1-\alpha)$ makes the ball all of $\mathcal{S}_n$ and both sides equal $\lambda_{\max}(L)$.

Proof. Both cases are proved by the same Step 1; the inner minimization and the attainment argument then differ genuinely between the two ranges, so Steps 2–3 are given separately for each.

Step 1 — Rewrite in terms of Q_α. Since $H_\alpha = (Q_\alpha - 1)/(\alpha - 1)$, the substitution $\tau := s/(\alpha-1)$ turns (6) into

$$V_\alpha(q) = \inf_{s>0}\{qs + G_\alpha(s)\} = \inf_{\mathrm{sign}(\tau)=\mathrm{sign}(\alpha-1)} \{\widetilde{q}\,\tau + \widehat{G}(\tau)\}, \qquad \widetilde{q} := 1 + (\alpha-1)q,$$

where $\widehat{G}$ is the signed Q_α-cumulant of Theorem 3.3, given in closed form by (15) ($\tau > 0$) and (16) ($\tau < 0$).

Case (i), $\alpha > 1$ ($\tau = t > 0$).

Step 2 — Inner t-minimization. Fix c, set $\widetilde{L}(c) := \sigma^\beta(L-cI)\sigma^\beta$ and $T := \|\widetilde{L}(c)_+\|_{\alpha'}^{\alpha'}$. Elementary calculus, using $\alpha' - 1 = 1/(\alpha-1)$, gives

$$\inf_{t>0}\left\{ \widetilde{q}\,t + \frac{(\alpha-1)T}{\alpha^{\alpha'}}\,t^{1-\alpha'} \right\} = \widetilde{q}^{\,1/\alpha}\,T^{1/\alpha'} = \widetilde{q}^{\,1/\alpha}\,\|\widetilde{L}(c)_+\|_{\alpha'}.$$

Substituting into Step 1,

$$V_\alpha(q) = \inf_{c\in\mathbb{R}} \left\{ c + \left(1 + (\alpha-1)q\right)^{1/\alpha} \|\widetilde{L}(c)_+\|_{\alpha'} \right\}.$$

Step 3 — Attainment. Denote the objective by $\Phi(c)$. It is convex (composition of the affine map $c \mapsto \widetilde{L}(c) = \sigma^\beta L \sigma^\beta - c\sigma^{2\beta}$ with $A \mapsto \|A_+\|_{\alpha'} = \sup\{\mathrm{tr}(AX) : X \geq 0,\ \|X\|_\alpha \leq 1\}$, a supremum of linear functionals), finite and continuous on $\mathbb{R}$. For $c \geq \lambda_{\max}(L)$ we have $\widetilde{L}(c)_+ = 0$, so $\Phi(c) = c$; in particular $\Phi(\lambda_{\max}(L)) = \lambda_{\max}(L)$, which is the value $V_\alpha(q)$ when $q \geq q_{\max}$, so the infimum is attained there. For $0 < q < q_{\max}$ we have $\widetilde{q} > 1$, and Φ is coercive: $\Phi(c) = c \to +\infty$ as $c \to +\infty$, while for $c < \lambda_{\min}(L)$, $\widetilde{L}(c) \succ 0$, so $\widetilde{L}(c)_+ = \widetilde{L}(c)$ and, using $\|\sigma^{2\beta}\|_{\alpha'} = 1$ (since $\mathrm{tr}\,\sigma = 1$ and $2\beta\alpha' = 1$), the triangle inequality gives $\Phi(c) \geq (\widetilde{q}^{1/\alpha} - 1)\,|c| - \widetilde{q}^{1/\alpha}\|\sigma^\beta L \sigma^\beta\|_{\alpha'} \to +\infty$ as $c \to -\infty$.

Case (ii), $\alpha \in [\frac{1}{2}, 1)$ ($\tau = -\zeta$, $\zeta > 0$).

Step 2 — Inner ζ-minimization. Fix $c > \lambda_{\max}(L)$ and set $T := \|\sigma^\gamma (cI - L)^{-1} \sigma^\gamma\|_p > 0$. For $\widetilde{q} > 0$, elementary calculus, using $1 + 1/p = 1/\alpha$ and $(1-\alpha)(1+p) = 1$, gives

$$\inf_{\zeta>0} \{ -\widetilde{q}\,\zeta + (1-\alpha)\,\alpha^p\, T^p\, \zeta^{1+p} \} = -\widetilde{q}^{1/\alpha}/T.$$

Substituting into Step 1,

$$V_\alpha(q) = \inf_{c\,>\,\lambda_{\max}(L)} \left\{ c - \widetilde{q}^{1/\alpha}/T(c) \right\}.$$

Step 3 — Attainment. Denote the objective by $\Phi(c)$, continuous on $(\lambda_{\max}(L), +\infty)$. As $c \to \lambda_{\max}(L)^+$, $T(c) \to +\infty$ and $\Phi(c) \to \lambda_{\max}(L)$. As $c \to +\infty$, expanding $(cI - L)^{-1} = c^{-1}(I + L/c + O(c^{-2}))$ and using $\|\sigma^{2\gamma}\|_p = 1$ (since $2\gamma p = 1$) together with the derivative of the Schatten norm, $T(c) = c^{-1}(1 + \mathrm{tr}(\sigma L)/c + O(c^{-2}))$, hence

$$\Phi(c) = (1 - \widetilde{q}^{1/\alpha})\, c + \widetilde{q}^{1/\alpha}\, \mathrm{tr}(\sigma L) + O(1/c),$$

which tends to $+\infty$ exactly when $\widetilde{q} < 1$, i.e. $q > 0$. For $0 < q < q_{\max}$ the infimum $V_\alpha(q)$ is strictly below the left-end limit $\lambda_{\max}(L)$ (by (5)), so it is attained at some $c^* \in (\lambda_{\max}(L), \infty)$. □ □

3.3.3 Optimal state and primal–dual identity

The dual characterization in Theorem 3.4 pins down the worst-case state in each case, by the same stationarity argument, and yields a clean decomposition of the robust value.

Corollary 3.1 (Worst-case state and primal–dual identity)**.** *In the active regime $q \in (0, q_{\max})$, let c^* be a minimizer in the corresponding case of Theorem 3.4 and set $\widetilde{q} := 1 + (\alpha-1)q$.*

Case (i), $\alpha > 1$: *with $\widetilde{L}^* := \sigma^\beta(L - c^*I)\sigma^\beta$, the supremum in $V_\alpha(q)$ is attained at*

$$\rho^* \;=\; \widetilde{q}^{1/\alpha}\, \frac{\sigma^\beta\, (\widetilde{L}^*_+)^{\alpha'-1}\, \sigma^\beta}{\|\widetilde{L}^*_+\|_{\alpha'}^{\alpha'-1}}, \qquad \alpha' - 1 = \tfrac{1}{\alpha-1} > 0, \tag{21}$$

the divergence constraint is tight, $H_\alpha(\rho^*\|\sigma) = q$, and

$$V_\alpha(q) \;=\; \mathrm{tr}(L\rho^*) \;=\; c^* + \widetilde{q}^{1/\alpha}\, \|\widetilde{L}^*_+\|_{\alpha'}. \tag{22}$$

Case (ii), $\alpha \in [\frac{1}{2}, 1)$: *with $\widetilde{M}^* := \sigma^{-\gamma}(c^*I - L)\sigma^{-\gamma} > 0$, the supremum in $V_\alpha(q)$ is attained at*

$$\rho^* \;=\; \widetilde{q}^{1/\alpha}\, \frac{\sigma^{-\gamma}\, (\widetilde{M}^*)^{\alpha'-1}\, \sigma^{-\gamma}}{\|(\widetilde{M}^*)^{-1}\|_p^{1+p}}, \qquad \alpha' - 1 = \tfrac{1}{\alpha-1} = -(1+p) < 0, \tag{23}$$

the divergence constraint is tight, $H_\alpha(\rho^*\|\sigma) = q$*, and*

$$V_\alpha(q) \;=\; \mathrm{tr}(L\rho^*) \;=\; c^* - \frac{\widetilde{q}^{\,1/\alpha}}{\|(\widetilde{M}^*)^{-1}\|_p}. \tag{24}$$

Proof. **Case (i),** $\alpha > 1$**.**

Form of ρ^***.** Define ρ^* by (21), the form read off Step 3(a) of the proof of Theorem 3.3. Since $\|A_+\|_{\alpha'}$ is differentiable in A wherever $A_+ \neq 0$ ($\alpha' > 1$), and $\widetilde{L}^*_+ \neq 0$ because $\Phi(c^*) = V_\alpha(q) < \lambda_{\max}(L) \leq \Phi(c)$ for every $c \geq \lambda_{\max}(L)$, stationarity of Φ at c^* reads

$$0 = \Phi'(c^*) = 1 - \widetilde{q}^{\,1/\alpha}\,\frac{\mathrm{tr}\big[\sigma^{2\beta}(\widetilde{L}^*_+)^{\alpha'-1}\big]}{\|\widetilde{L}^*_+\|_{\alpha'}^{\alpha'-1}},$$

which is exactly $\mathrm{tr}\,\rho^* = 1$ for the state (21); and $\rho^* \geq 0$ is clear.

Feasibility and tightness. With $X^* = \sigma^{-\beta}\rho^*\sigma^{-\beta} = \widetilde{q}^{\,1/\alpha}(\widetilde{L}^*_+)^{\alpha'-1}/\|\widetilde{L}^*_+\|_{\alpha'}^{\alpha'-1}$ and $(\alpha'-1)\alpha = \alpha'$,

$$Q_\alpha(\rho^*\|\sigma) = \|X^*\|_\alpha^\alpha = \widetilde{q}\,\frac{\mathrm{tr}\big[(\widetilde{L}^*_+)^{\alpha'}\big]}{\|\widetilde{L}^*_+\|_{\alpha'}^{\alpha'}} = \widetilde{q},$$

i.e. $H_\alpha(\rho^*\|\sigma) = (\widetilde{q}-1)/(\alpha-1) = q$: ρ^* is feasible and the constraint is tight.

Primal–dual identity and optimality. Cyclicity and (21) give

$$\mathrm{tr}((L - c^*I)\rho^*) = \mathrm{tr}(\widetilde{L}^*X^*) = \widetilde{q}^{\,1/\alpha}\,\frac{\|\widetilde{L}^*_+\|_{\alpha'}^{\alpha'}}{\|\widetilde{L}^*_+\|_{\alpha'}^{\alpha'-1}} = \widetilde{q}^{\,1/\alpha}\,\|\widetilde{L}^*_+\|_{\alpha'},$$

so $\mathrm{tr}(L\rho^*) = c^* + \mathrm{tr}(\widetilde{L}^*X^*) = \Phi(c^*) = V_\alpha(q)$ by Theorem 3.4 (case (i)). Since ρ^* is feasible, it is optimal, which is (22).

Case (ii), $\alpha \in [\frac{1}{2}, 1)$**.**

Form of ρ^***.** Define ρ^* by (23), the form read off Step 3(b) of the proof of Theorem 3.3. Stationarity of Φ at c^*, $0 = \Phi'(c^*) = 1 - \widetilde{q}^{\,1/\alpha}\,\mathrm{tr}[\sigma^{-2\gamma}(\widetilde{M}^*)^{-1-p}]/\|(\widetilde{M}^*)^{-1}\|_p^{1+p}$, reads

$$\mathrm{tr}\big[\sigma^{-\gamma}\,(\widetilde{M}^*)^{\alpha'-1}\,\sigma^{-\gamma}\big] \;=\; \widetilde{q}^{\,-1/\alpha}\,\|(\widetilde{M}^*)^{-1}\|_p^{1+p},$$

which is $\mathrm{tr}\,\rho^* = 1$ for (23); $\rho^* \geq 0$ is clear.

Feasibility and tightness. With $X^* = \sigma^\gamma\rho^*\sigma^\gamma = \widetilde{q}^{\,1/\alpha}(\widetilde{M}^*)^{-(1+p)}/\|(\widetilde{M}^*)^{-1}\|_p^{1+p}$, $(1+p)\alpha = p$ and $\|X^*\|_\alpha^\alpha = \widetilde{q}\,\mathrm{tr}[(\widetilde{M}^*)^{-p}]/\|(\widetilde{M}^*)^{-1}\|_p^p = \widetilde{q}$, so $Q_\alpha(\rho^*\|\sigma) = \widetilde{q}$ and $H_\alpha(\rho^*\|\sigma) = (\widetilde{q}-1)/(\alpha-1) = q$: ρ^* is feasible and the constraint is tight.

Primal–dual identity and optimality. With $\alpha' = -p$, cyclicity and (23) give

$$\mathrm{tr}\big((c^*I - L)\rho^*\big) = \mathrm{tr}(\widetilde{M}^*X^*) = \widetilde{q}^{\,1/\alpha}\,\frac{\mathrm{tr}\big((\widetilde{M}^*)^{\alpha'}\big)}{\|(\widetilde{M}^*)^{-1}\|_p^{1+p}} = \widetilde{q}^{\,1/\alpha}\,\frac{\|(\widetilde{M}^*)^{-1}\|_p^p}{\|(\widetilde{M}^*)^{-1}\|_p^{1+p}} = \frac{\widetilde{q}^{\,1/\alpha}}{\|(\widetilde{M}^*)^{-1}\|_p},$$

so $\mathrm{tr}(L\rho^*) = c^* - \mathrm{tr}\big((c^*I - L)\rho^*\big) = \Phi(c^*) = V_\alpha(q)$ by Theorem 3.4 (case (ii)); since ρ^* is feasible it is optimal, which is (24). □ □

3.3.4 The Hilbert–Schmidt endpoint $\alpha = 2$

At $\alpha = 2$ the Schatten norm in (19) is the Hilbert–Schmidt norm and the c-minimization closes in elementary form, as long as the positive part is inactive. With $\hat{L}$ the centered observable, write

$$s_2^2 \;:=\; \tfrac{1}{2}\operatorname{tr}(\sigma^{1/2}\hat{L}\,\sigma^{1/2}\hat{L}), \tag{25}$$

identified in Section 5 as the DRO variance $s_2^2(\hat{L})$. At this order $H_2(\rho\|\sigma) = \operatorname{tr}(\sigma^{-1/2}X\sigma^{-1/2}X)$, $X = \rho - \sigma$, is the geometric-mean member of the quantum χ^2 family of [Temme et al., 2010], and the corollary below is the Cauchy–Schwarz inequality for it (83), with equality as long as the worst-case state stays positive.

Corollary 3.2 (The Hilbert–Schmidt case $\alpha = 2$: exact square-root law)**.** *Let $\hat{L} \neq 0$ and $q_0 := 2s_2^2/\lambda_{\min}(\hat{L})^2$. For $0 < q \le q_0$ the infimum in (19) is attained at $c^* = \operatorname{tr}(L\sigma) - s_2\sqrt{2/q}$, and*

$$V_2(q) \;=\; \operatorname{tr}(L\sigma) + s_2\sqrt{2q}, \qquad \rho^* \;=\; \sigma^{1/2}\Big(I + \sqrt{\tfrac{q}{2}}\,\frac{\hat{L}}{s_2}\Big)\sigma^{1/2}. \tag{26}$$

For $q > q_0$, $V_2(q) < \operatorname{tr}(L\sigma) + s_2\sqrt{2q}$.

Proof. At $\alpha = 2$, $\beta = 1/4$, $\alpha' = 2$ and $\widetilde{q} = 1 + q$. For $c < \lambda_{\min}(L)$ the matrix $\widetilde{L}(c) = \sigma^{1/4}(L - cI)\sigma^{1/4}$ is positive, so with $u := \operatorname{tr}(L\sigma) - c > 0$,

$$\|\widetilde{L}(c)_+\|_2^2 = \operatorname{tr}\big[\sigma^{1/2}(\hat{L} + uI)\sigma^{1/2}(\hat{L} + uI)\big] = 2s_2^2 + u^2,$$

the cross term vanishing because $\operatorname{tr}(\sigma\hat{L}) = 0$. The objective of (19) becomes $\Phi(c) = \operatorname{tr}(L\sigma) - u + \sqrt{1+q}\sqrt{2s_2^2 + u^2}$, which is strictly convex in u with stationary point $u_*^2(1+q) = 2s_2^2 + u_*^2$, i.e. $u_* = s_2\sqrt{2/q}$, and value $\operatorname{tr}(L\sigma) - u_* + (1+q)u_* = \operatorname{tr}(L\sigma) + s_2\sqrt{2q}$. The condition $c^* = \operatorname{tr}(L\sigma) - u_* < \lambda_{\min}(L)$ reads $u_* > -\lambda_{\min}(\hat{L})$ (note $\lambda_{\min}(\hat{L}) < 0$ as $\operatorname{tr}(\sigma\hat{L}) = 0$, $\hat{L} \neq 0$), i.e. $q < q_0$. Since Φ is convex on $\mathbb{R}$ (Step 3 of the proof of Theorem 3.4, case (i)) and coincides with the function above on $(-\infty, \lambda_{\min}(L)]$, its stationary point c^* in that half-line is its global minimizer (for $q = q_0$, $c^* = \lambda_{\min}(L)$ is the endpoint, and Φ is nondecreasing to the right of a point where its left derivative vanishes). Then (21) — whose verification in Corollary 3.1 uses only that c^* is a minimizer with $\widetilde{L}_+^* \neq 0$, both exhibited here — with $\widetilde{L}_+^* = \widetilde{L}^*$, $\alpha' - 1 = 1$, gives $\rho^* = \sqrt{1+q}\,\sigma^{1/2}(\hat{L} + u_*I)\sigma^{1/2}/\sqrt{2s_2^2 + u_*^2} = \sigma^{1/2}(I + \hat{L}/u_*)\sigma^{1/2}$, which is (26). For $q > q_0$ the point $c_u := \operatorname{tr}(L\sigma) - u_*$ lies above $\lambda_{\min}(L)$, so $L - c_uI$ has a negative eigenvalue and $\|\widetilde{L}(c_u)_+\|_2 < \|\widetilde{L}(c_u)\|_2$ strictly; hence $V_2(q) \le \Phi(c_u) < \operatorname{tr}(L\sigma) - u_* + \sqrt{1+q}\sqrt{2s_2^2 + u_*^2} = \operatorname{tr}(L\sigma) + s_2\sqrt{2q}$. □ □

The worst-case state is a sandwiched linear tilt of σ, and q_0 is the budget at which it loses positivity — the quantum counterpart of the classical χ^2-ball DRO [Duchi and Namkoong, 2019, §2.1]. Section 5 establishes the square-root law for every α; at $\alpha = 2$ it is exact on an interval.

3.3.5 The fidelity endpoint $\alpha = 1/2$

At $\alpha = 1/2$, the parameters of Theorem 3.4 (case (ii)) specialize to $\gamma = 1/2$, $p = 1$, and $Q_{1/2}(\rho\|\sigma) = \operatorname{Fid}(\rho,\sigma)$, the **root fidelity** $\operatorname{Fid}(\rho,\sigma) := \operatorname{tr}|\sigma^{1/2}\rho^{1/2}|$, giving $H_{1/2} = 2(1 - \operatorname{Fid}(\rho,\sigma))$. Since $p = 1$ and $\sigma^{1/2}(cI - L)^{-1}\sigma^{1/2} \ge 0$, the Schatten 1-norm reduces to the trace, which gives the corollary below. The inequality $\le$ in it is Alberti's variational formula $\operatorname{Fid}(\rho,\sigma)^2 \le \operatorname{tr}(\rho X)\operatorname{tr}(\sigma X^{-1})$ [Alberti, 1983] at $X = cI - L$. Bounds on expectation values over a fidelity ball that use only the mean of the observable, or its mean and variance, are in [Weber et al., 2022, Thms. 1 and 2] (see also [Weber et al., 2021]); the corollary gives the value for a given observable.

Corollary 3.3 (Fidelity DRO)**.** *For $\sigma > 0$, $L = L^*$, and $0 \le q \le 2$ ($H_{1/2} = 2(1 - \mathrm{Fid}) < 2$ on $\mathcal{S}_n$),*

$$V_{1/2}(q) = \inf_{c > \lambda_{\max}(L)} \left\{ c - \frac{(1 - q/2)^2}{\mathrm{tr}\big[\sigma(cI - L)^{-1}\big]} \right\}.$$

Proof. Set $\alpha = 1/2$ in Theorem 3.4 (case (ii)). Cyclicity of the trace absorbs the sandwiching: $\|\sigma^{1/2}(cI - L)^{-1}\sigma^{1/2}\|_1 = \mathrm{tr}[\sigma(cI - L)^{-1}]$. The prefactor specializes to $(1 + (\alpha - 1)q)^{1/\alpha} = (1 - q/2)^2$. □ □

Remark 3.3 (Stieltjes structure at $\alpha = 1/2$). The denominator $f(c) := \mathrm{tr}[\sigma(cI - L)^{-1}]$ is the **operator Stieltjes transform** of L weighted by σ, reappearing in the proof of the Fuchs–van de Graaf inequality in Section 6.5; the collapse of the sandwiched norm to a σ-weighted trace is the spectral face of the coincidence $H_{1/2}^{\mathrm{meas}} = H_{1/2}$ of the measured and sandwiched divergences at $\alpha = \frac{1}{2}$ [Berta et al., 2017].

3.4 DRO for $\alpha = \infty$

As $\alpha \to \infty$ the sandwiched Rényi divergence converges to the **max-relative entropy** [Datta, 2009] $D_\infty(\rho\|\sigma) = \inf\{\lambda : \rho \le e^\lambda \sigma\} = \log Q_\infty(\rho\|\sigma)$, and we set

$$H_\infty(\rho\|\sigma) := Q_\infty(\rho\|\sigma) - 1, \qquad Q_\infty(\rho\|\sigma) := \|\sigma^{-1/2}\rho\,\sigma^{-1/2}\|_\infty = \inf\{\mu \ge 0 : \rho \le \mu\sigma\}.$$

Since Q_∞ is convex in ρ, the duality of Section 2 applies, and the constraint $H_\infty(\rho\|\sigma) \le q$ is the operator inequality $\rho \le (1 + q)\sigma$, making (11) a semidefinite program:

$$V_\infty(q) = \sup\{\mathrm{tr}(L\rho) : \rho \in \mathcal{S}_n,\ \rho \le (1 + q)\sigma\}. \tag{27}$$

The Fenchel strategy of Theorem 3.3 applies verbatim at $\alpha = \infty$, with $\beta = \frac{1}{2}$: $Q_\infty = \|Y\|_\infty$ is 1-homogeneous, so its conjugate is an indicator, of $\{\mathrm{tr}\, B_+ \le s\}$, and the s-minimization of (2) returns the trace of the positive part directly. The resulting formula (Theorem 3.5) is the dual of a quantum Neyman–Pearson problem — for $L \ge 0$, the dual form of the quantum relative Lorenz curve of Buscemi and Gour [Buscemi and Gour, 2017]: in the variable $T = \sigma^{-1/2}\rho\,\sigma^{-1/2}/(1 + q)$, (27) maximizes $(1 + q)\,\mathrm{tr}(\sigma^{1/2}L\sigma^{1/2}\,T)$ over tests $0 \le T \le I$ with $\mathrm{tr}(\sigma T) = 1/(1 + q)$.

3.4.1 DRO value

Theorem 3.5 (Quantum CVaR)**.** *For $\sigma > 0$, $L = L^*$, and $q \ge 0$,*

$$V_\infty(q) = \inf_{c \in \mathbb{R}} \{c + (1 + q)\ \mathrm{tr}\big[(\sigma^{1/2}(L - cI)\sigma^{1/2})_+\big]\}, \tag{28}$$

and for $q > 0$ the infimum is attained.

Proof. Steps 1–2 of the proof of Theorem 3.3 apply verbatim with Q_∞ in place of Q_α — Q_∞ is convex and continuous on the positive cone — and $\beta = \frac{1}{2}$: for $s > 0$,

$$\widehat{G}_\infty(s) := \sup_{\rho \in \mathcal{S}_n} \{\mathrm{tr}(L\rho) - sQ_\infty(\rho\|\sigma)\} = \inf_{c \in \mathbb{R}} \{c + \Psi_\infty(\sigma^{1/2}(L - cI)\sigma^{1/2})\},$$
$$\Psi_\infty(A) := \sup_{X \ge 0} \{\mathrm{tr}(AX) - s\|X\|_\infty\}.$$

By von Neumann's trace inequality, $\operatorname{tr}(AX) \le \|X\|_\infty \operatorname{tr} A_+$ for $X \ge 0$, with equality at $X = \lambda \mathbf{1}_{(0,\infty)}(A)$, $\lambda \ge 0$; so $\Psi_\infty(A) = \sup_{\lambda\ge 0} \lambda(\operatorname{tr} A_+ - s)$ is 0 if $\operatorname{tr} A_+ \le s$ and $+\infty$ otherwise, and $\widehat{G}_\infty(s) = \inf\{c : \operatorname{tr}[(\sigma^{1/2}(L - cI)\sigma^{1/2})_+] \le s\}$. Since $H_\infty = Q_\infty - 1$, the strong duality (2) reads

$$V_\infty(q) = \inf_{s\ge 0}\{(1+q)s + \widehat{G}_\infty(s)\} = \inf_{c\in\mathbb{R}} \inf_{s \ge \operatorname{tr}[(\sigma^{1/2}(L-cI)\sigma^{1/2})_+]} \{c + (1+q)s\},$$

which is (28). For $q > 0$ the infimum is attained: the objective $\Phi(c) := c + (1+q)\operatorname{tr}[(\sigma^{1/2}(L - cI)\sigma^{1/2})_+]$ is convex and continuous, $\Phi(c) \ge c \to +\infty$ as $c \to +\infty$, and for $c \le \lambda_{\min}(L)$ it equals $c + (1+q)(\operatorname{tr}(L\sigma) - c) \to +\infty$ as $c \to -\infty$. □ □

3.4.2 Optimal state and primal–dual identity

Corollary 3.4 (Worst-case state and primal–dual identity ($\alpha = \infty$))**.** *In the active regime $q \in (0, q_{\max})$, let c^* be a minimizer in (28), $\widetilde{L}^* := \sigma^{1/2}(L - c^*I)\sigma^{1/2}$, and $P_+ := \mathbf{1}_{(0,\infty)}(\widetilde{L}^*)$. Assume $\ker \widetilde{L}^* = \{0\}$. Then the supremum in $V_\infty(q)$ is attained at the unique*

$$\rho^* = (1+q)\,\sigma^{1/2} P_+ \sigma^{1/2}, \tag{29}$$

c^* satisfies $(1+q)\operatorname{tr}(\sigma P_+) = 1$, $(1+q)\sigma - \rho^* = (1+q)\sigma^{1/2}(I - P_+)\sigma^{1/2}$, and

$$V_\infty(q) = \operatorname{tr}(L\rho^*) = c^* + (1+q)\,\|\widetilde{L}^*_+\|_1. \tag{30}$$

Proof. In sandwiched coordinates $Y = \sigma^{-1/2}\rho\,\sigma^{-1/2}$, (27) reads $c^* + \max\{\operatorname{tr}(\widetilde{L}^*Y) : 0 \le Y \le (1+q)I,\ \operatorname{tr}(\sigma Y) = 1\}$. For feasible Y, $\operatorname{tr}(\widetilde{L}^*Y) \le \operatorname{tr}(\widetilde{L}^*_+ Y) \le (1+q)\operatorname{tr}\widetilde{L}^*_+$, with equality in the first step iff Y vanishes on the range of $\widetilde{L}^*_-$ and in the second iff $Y = (1+q)I$ on the range of P_+; and $c^* + (1+q)\operatorname{tr}\widetilde{L}^*_+ = V_\infty(q)$ by Theorem 3.5. When $\ker\widetilde{L}^* = \{0\}$ the two conditions force $Y^* = (1+q)P_+$, and stationarity of the objective at c^*, $0 = 1 - (1+q)\operatorname{tr}(\sigma P_+)$ (the objective is differentiable there, $\widetilde{L}^*$ having no zero eigenvalue), is the normalization $\operatorname{tr}(\sigma Y^*) = 1$; so Y^* is feasible and optimal, and unique. Undoing the sandwich gives (29); (30) is (28) at c^*. □ □

Remark 3.4 (Neyman–Pearson structure, and the degenerate case). The worst-case state (29) is a threshold test, with P_+ the Holevo–Helstrom optimal test [Holevo, 1973, Helstrom, 1976] — the projection onto the positive part of the noncommutative likelihood ratio — and c^* the Neyman–Pearson level. It is $\widetilde{L}^*$, not the spectral cut $\mathbf{1}_{\{L>c^*\}}$, that appears: the adversary tests the *sandwiched* observable, not L itself.

The hypothesis $\ker\widetilde{L}^* = \{0\}$ in Corollary 3.4 fails exactly when c^* is an eigenvalue of L: these are the exceptional budgets, in general whole intervals of q, on which c^* is constant and V_∞ is affine in q. The two equality conditions still pin Y^* off the kernel $\Pi_0 := \mathbf{1}_{\{0\}}(\widetilde{L}^*)$, but leave it free there subject only to the normalization $\operatorname{tr}(\sigma Y^*) = 1$; when c^* is a simple eigenvalue of L, Π_0 has rank one and the normalization fixes $Y_0 = y\,\Pi_0$, so the worst-case state is still unique, while at a multiple eigenvalue ($\operatorname{rank}\Pi_0 \ge 2$) it is non-unique,

$$\rho^* = (1+q)\,\sigma^{1/2}P_+\sigma^{1/2} + \sigma^{1/2}Y_0\,\sigma^{1/2}, \qquad 0 \le Y_0 \le (1+q)\Pi_0,$$

while $V_\infty(q)$ and (30) are unaffected, since $\widetilde{L}^*$ vanishes on Π_0. This is the quantum face of the randomization at the Value-at-Risk quantile in the classical CVaR [Rockafellar and Uryasev, 2000]: when the spectral measure carries an atom at the threshold, the optimal tail allocation may split it, and only the split, within a multiple eigenspace, is undetermined — never the value.

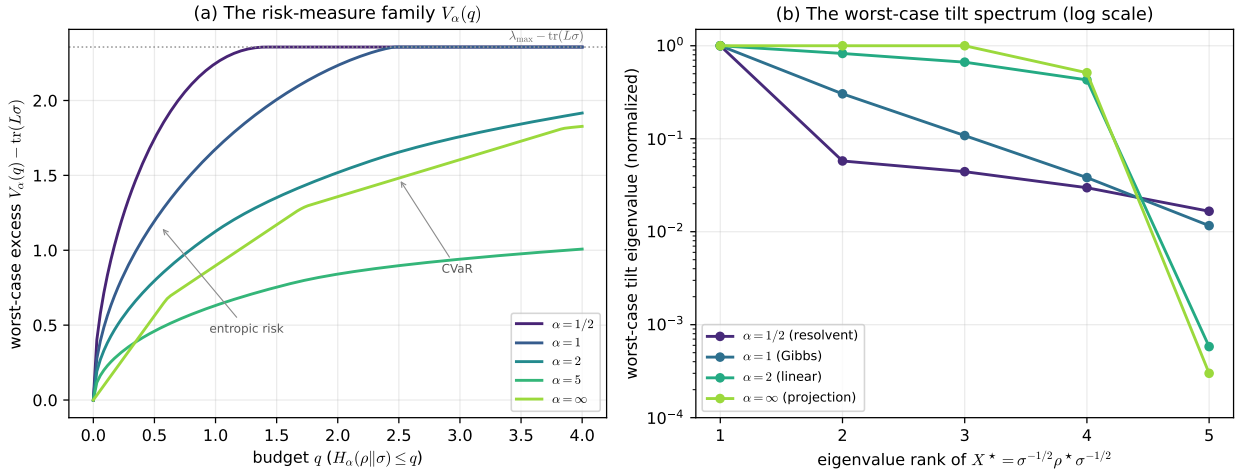


Figure 3: The quantum risk-measure family for a noncommuting pair on $\mathbb{C}^5$ ($\sigma = \mathrm{diag}(0.40, 0.25, 0.15, 0.12, 0.08)$, L tridiagonal with diagonal $(0, 0.8, 1.6, 2.4, 3.2)$ and coupling 0.35). *(a)* Worst-case excess $V_\alpha(q) - \mathrm{tr}(L\sigma)$ against the budget q for $\alpha \in \{1/2, 1, 2, 5, \infty\}$, from the entropic risk measure ($\alpha = 1$) to the quantum CVaR ($\alpha = \infty$), each saturating at $\lambda_{\max} - \mathrm{tr}(L\sigma)$ (beyond the plotted range for $\alpha \ge 2$); the affine segments of the $\alpha = \infty$ curve are the exceptional-budget intervals of Remark 3.4. *(b)* Spectrum of the worst-case tilt $X^* = \sigma^{-1/2}\rho^*\sigma^{-1/2}$ at $q = 1$ (sorted, normalized, log scale): full support for $\alpha \le 1$ (resolvent weighting at $\alpha = \frac{1}{2}$, Gibbs tilt at $\alpha = 1$), truncation $\propto (\widetilde{L}_+^*)^{\alpha'-1}$ for $\alpha > 1$, down to the Neyman–Pearson projection at $\alpha = \infty$, where this budget is exceptional and the fourth eigenvalue shows the fractional split on $\ker \widetilde{L}^*$.

3.5 Reading the dual as a quantum risk measure

Both faces of the family — the value functions and the optimizer — are shown in Figure 3.

When $[\sigma, L] = 0$, (19) reduces to the classical α-divergence DRO of the Born law $\sum_i p_i \, \delta_{l_i}$ (Lemma 4.1); Section 4 shows that the quantum DRO is a classical DRO for every pair, of a measure ω_α that encodes the noncommutativity.

4 The quantum spectral measures

The dual formulas of Section 3 express the DRO value through Schatten norms of sandwiched operators. The main result of this section is that this apparently quantum object is *exactly* a classical DRO: for each α there is a probability measure ω_α on the spectral interval $[\lambda_{\min}(L), \lambda_{\max}(L)]$, the **quantum spectral measure**, with $V_\alpha(q) = V_\alpha^{\mathrm{cl}}(q; \omega_\alpha)$ for every budget; we call this reduction **classicalization**. The measure comes from one transform identity per regime: stop-loss for $\alpha > 1$ from Heinävaara's tracial joint spectral measure [Heinävaara, 2025], Stieltjes for $\alpha < 1$ from the companion note [Rey-Bellet, 2026], Laplace at $\alpha = 1$ from Stahl's theorem.

4.1 The classicalization theorem

Fix $\sigma > 0$ and L Hermitian on $\mathbb{C}^n$ with eigenvalues $l_1 \le \cdots \le l_n$ and an orthonormal eigenbasis $|1\rangle, \ldots, |n\rangle$. With $\beta = (\alpha - 1)/2\alpha$, $\alpha' = \alpha/(\alpha - 1)$ and $p = -\alpha'$ as in (9) ($2\beta\alpha' = 1$; here β is used on both ranges, $\beta < 0$ for $\alpha < 1$), the **sandwiched pair** is

$$A = \sigma^\beta L \, \sigma^\beta, \qquad B = \sigma^{2\beta} \succ 0, \tag{31}$$

and we write $N_i := \langle i|\sigma^{-2\beta}|i\rangle > 0$. The dual formulas of Section 3 are traces of functions of the pencil combination $\sigma^\beta(L-cI)\sigma^\beta = A-cB$: a positive-part power $\mathrm{tr}(A-cB)_+^{\alpha'}$ for $\alpha > 1$, a resolvent power $\mathrm{tr}(cB-A)^{-p}$ for $\alpha < 1$. *Standing convention.* In the pencil analysis of Section 4.2 and in the structure statements of this section L has *simple* spectrum, $l_1 < \cdots < l_n$; the existence part of Theorem 4.1 needs no such assumption (Remark 4.1).

The classical DRO value. For a probability measure ν on a compact interval $J = [\ell_{\min}, \ell_{\max}] \subset \mathbb{R}$ and a budget $q \geq 0$,

$$V_\alpha^{\mathrm{cl}}(q;\,\nu) := \sup\Big\{\int \ell\, dP : P \in \mathcal{P}(J),\ H_\alpha^{\mathrm{cl}}(P\|\nu) \leq q\Big\}, \tag{32}$$

where H_α^{cl} is the classical Hellinger divergence of order α, the commutative case of Definition 3.1: $H_\alpha^{\mathrm{cl}}(P\|\nu) = \frac{1}{\alpha-1}\big(\int (dP_{\mathrm{ac}}/d\nu)^\alpha\, d\nu - 1\big)$ with P_{ac} the ν-absolutely continuous part of P and $H_\alpha^{\mathrm{cl}}(P\|\nu) = +\infty$ for $\alpha > 1$ unless $P \ll \nu$; at $\alpha = 1$ the relative entropy $D(P\|\nu) = \int \log(dP/d\nu)\, dP$; at $\alpha = \infty$, $H_\infty^{\mathrm{cl}}(P\|\nu) = \|dP/d\nu\|_{L^\infty(\nu)} - 1$ ($+\infty$ unless $P \ll \nu$).

Lemma 4.1 (Classical duals)**.** *Let $\nu \in \mathcal{P}(J)$ and $q \geq 0$. Then*

$$V_\alpha^{\mathrm{cl}}(q;\,\nu) = \begin{cases} \inf\limits_{c\in\mathbb{R}}\Big\{c + (1+(\alpha-1)q)^{1/\alpha}\,\big\|(\ell-c)_+\big\|_{L^{\alpha'}(\nu)}\Big\}, & \alpha \in (1,\infty), \\ \inf\limits_{c\in\mathbb{R}}\Big\{c + (1+q)\int(\ell-c)_+\, d\nu\Big\}, & \alpha = \infty, \\ \inf\limits_{t>0}\ \dfrac{1}{t}\Big\{q + \log\int e^{t\ell}\, d\nu\Big\}, & \alpha = 1, \\ \inf\limits_{c>\ell_{\max}}\Big\{c - (1+(\alpha-1)q)_+^{1/\alpha}\,\big\|(c-\ell)^{-1}\big\|_{L^p(\nu)}^{-1}\Big\}, & \alpha \in [\tfrac12, 1). \end{cases} \tag{33}$$

The proof is in Appendix A.

Theorem 4.1 (Classicalization)**.** *Let $\sigma > 0$ and let L be Hermitian on $\mathbb{C}^n$.*

(i) For every $\alpha \in [\frac12, \infty]$ there is a probability measure ω_α on $[\lambda_{\min}(L), \lambda_{\max}(L)]$ with mean $\mathrm{tr}(L\sigma)$ such that

$$\boxed{V_\alpha(q) = V_\alpha^{\mathrm{cl}}(q;\,\omega_\alpha), \qquad q \geq 0.} \tag{34}$$

It is the unique measure on that interval satisfying the transform identity of its regime — (44) for $\alpha \in (1,\infty]$, (47) for $\alpha \in [\frac12, 1)$, (48) for $\alpha = 1$ — and if $[L,\sigma] = 0$ it is the Born measure ν_L for every α.

(ii) If L has simple spectrum, $\omega_\alpha = \sum_i \omega_\alpha(\{l_i\})\,\delta_{l_i} + \omega_\alpha^{\mathrm{ac}}$ with atoms $\omega_\alpha(\{l_i\}) = N_i^{-\alpha'}$ for $\alpha \neq 1$ and $e^{\langle i|\log\sigma|i\rangle}$ for $\alpha = 1$, and an absolutely continuous part on (l_1, l_n), given in Proposition 4.1 ($\alpha \neq 1$) and (49) ($\alpha = 1$) as an integral of Heinävaara's density, which vanishes if and only if $[L,\sigma] = 0$ or $\alpha = \frac12$.

Related work. The existence of a positive measure with the transform identity is not ours in any regime. It is Heinävaara's theorem for $\alpha \in (1,\infty)$ [Heinävaara, 2025], the companion note for $\alpha \in (\frac12, 1)$ [Rey-Bellet, 2026], which is built on Heinävaara's measure, and Stahl's theorem at $\alpha = 1$ [Stahl, 2013]; at $\alpha = \infty$ it needs only the convexity of $c \mapsto \mathrm{tr}(A-cB)_+$, and at $\alpha = \frac12$ only trace cyclicity. At $\alpha = \infty$ the measure ω_∞ is the layer-cake measure of [Gour, 2026, Eqs. (15)–(16); Liu et al., 2025, Eq. (17), Lemma 2.2, Prop. 6.1], taken here for the pair $(\sigma^{1/2}L\sigma^{1/2}, \sigma)$, and for $L \geq 0$ the value V_∞ is, up to normalization, the quantum relative Lorenz curve of [Buscemi and Gour, 2017]. Concurrently and independently, Schmitt [Schmitt, 2026] obtains the same measure as the push-forward of Heinävaara's measure under $(a,b) \mapsto a/b$, with the stop-loss identity [Schmitt, 2026, Prop. 6.2] and the dichotomy between the commuting and the non-commuting case [Schmitt, 2026,

Thm. 6.4]. What the theorem adds is the identification with the DRO value at every budget and for every finite α, with the normalization (mass one, mean $\mathrm{tr}(L\sigma)$) and the atoms.

Remark 4.1 (Degenerate spectra). Simple spectrum enters only through the explicit atoms and the ray structure of (ii); existence and the transform identities in (i) hold for arbitrary multiplicities, since Heinävaara's identity, the Stieltjes representation of Theorem 4.3 and Stahl's theorem need no non-degeneracy. The analogue of (ii) for repeated eigenvalues follows by the same arguments; for $\alpha < 1$ it is carried out in [Rey-Bellet, 2026, Cor. A.1].

The fidelity endpoint. At $\alpha = \frac{1}{2}$ ($\beta = -\frac{1}{2}$, $p = 1$) trace cyclicity collapses the sandwich, $\mathrm{tr}[\sigma^{1/2}(cI - L)^{-1}\sigma^{1/2}] = \mathrm{tr}[\sigma(cI - L)^{-1}] = \int (c - \ell)^{-1}\, d\nu_L$, so

$$\omega_{1/2} = \nu_L = \sum_i \langle i|\sigma|i\rangle\, \delta_{l_i} \tag{35}$$

for every L, and the DRO dual Corollary 3.3 is the classical $\alpha = \frac{1}{2}$ DRO of the Born measure. It is the only order at which ω_α is the Born measure for every L (Theorem 5.3), the spectral trace of the coincidence $H^{\mathrm{meas}}_{1/2} = H_{1/2}$ (Remark 3.3).

4.2 Tools: the pencil and Heinävaara's measure

A *matrix pencil* (A, B) is the family $\{\ell B - A : \ell \in \mathbb{R}\}$; its *roots* are the zeros of $\det(\ell B - A)$, the eigenvalues of $B^{-1}A$ when B is invertible, and the pencil is *non-degenerate* if some real combination of A, B is invertible.

Lemma 4.2 (Pencil of the sandwiched pair)**.** *Let L have simple spectrum $l_1 < \cdots < l_n$ with unit eigenvectors $|1\rangle, \ldots, |n\rangle$, and let $\beta \in \mathbb{R}$. The pencil (A, B) of (31) is non-degenerate with roots $l_1, \ldots, l_n$. At l_i its unit eigenvector is*

$$|v_i\rangle = \frac{\sigma^{-\beta}|i\rangle}{N_i^{1/2}}, \qquad N_i = \langle i|\sigma^{-2\beta}|i\rangle,$$

with $\langle v_i|A|v_i\rangle = l_i/N_i$ and $\langle v_i|B|v_i\rangle = 1/N_i$. Moreover, for $b > 0$ and $\ell = a/b \notin \mathrm{spec}(L)$,

$$(bA - aB)^{-1} = b^{-1}\, \sigma^{-\beta}(L - \ell I)^{-1}\sigma^{-\beta}. \tag{36}$$

Proof. $\ell B - A = \sigma^\beta(\ell I - L)\sigma^\beta$ with σ^β invertible, so the roots are the l_i, B is invertible, and $\ker(l_i B - A) = \mathrm{span}\{\sigma^{-\beta}|i\rangle\}$; normalizing gives $|v_i\rangle$ and the two expectations, and inverting the three factors of $bA - aB = b\,\sigma^\beta(L - \ell I)\sigma^\beta$ gives (36). □ □

Theorem 4.2 (Tracial joint spectral measure [Heinävaara, 2025])**.** *Let A, B be Hermitian on $\mathbb{C}^n$.*

(i) There is a positive measure $\mu_{A,B}$ on $\mathbb{R}^2$ with $\mu_{A,B}(\{0\}) = 0$ such that, for every measurable $f : \mathbb{R} \to \mathbb{R}$ with $f(0) = 0$ and $\int_{-M}^{M} |f(t)/t|\, dt < \infty$ for all $M > 0$, and every $(x, y) \in \mathbb{R}^2$,

$$\mathrm{tr}\; H(f)(xA + yB) = \int_{\mathbb{R}^2} f(ax + by)\, d\mu_{A,B}(a, b), \qquad H(f)(x) := \int_0^1 \frac{1 - t}{t}\, f(xt)\, dt. \tag{37}$$

(ii) Let $\mu_{A,B} = \mu_c + \mu_s$ be the Lebesgue decomposition. If the pencil (A, B) is non-degenerate, μ_c has density

$$\mathfrak{h}_{A,B}(a, b) = \frac{1}{2\pi} \sum_{i=1}^{n} |\mathrm{Im}\, \lambda_i(C_1(a, b))|, \qquad C_1(a, b) = \Big(I - \tfrac{aA + bB}{a^2 + b^2}\Big)(bA - aB)^{-1}, \tag{38}$$

and if moreover A is invertible and $A^{-1}B$ has n distinct eigenvalues, then for $\varphi \in C_c^\infty(\mathbb{R}^2\setminus\{0\})$

$$\mu_s(\varphi) = \sum_{v\in E(A,B)} \int_0^1 \frac{1-t}{t}\,\varphi(t\langle v|A|v\rangle,\, t\langle v|B|v\rangle)\,dt, \tag{39}$$

$E(A,B)$ containing one unit eigenvector of the pencil per real root.

(iii) If $xA+yB \succeq 0$ for some real (x,y), then $\operatorname{supp}\mu_{A,B} \subseteq \{xa+yb \geq 0\}$; $\mu_c \equiv 0$ if and only if $[A,B]=0$; and $\operatorname{supp}\mu_c$ is contained in the joint numerical range $W(A,B) := \{(\langle v|A|v\rangle, \langle v|B|v\rangle) : \|v\|=1\}$.

Item (i) is [Heinävaara, 2025, Thm. 1.3] (uniqueness away from the origin is its Prop. 3.9) and (ii) is [Heinävaara, 2025, Thm. 1.6]; in (iii) the half-plane statement is [Heinävaara, 2025, Lemma 1.5], stated there for $(x,y)=(1,0)$ and valid in every direction by the same proof, and the dichotomy and the inclusion $\operatorname{supp}\mu_c \subseteq W(A,B)$ are [Heinävaara, 2024, Prop. 4]. The specialization we use is the one-sided power $f(t)=t_+^s$, $s>0$, which Heinävaara uses for integer s in the proof of [Heinävaara, 2025, Cor. 1.2] and which is stated for every $s>0$ in [Schmitt, 2026, Prop. 2.2]; it is admissible since $\int_{-M}^{M}|f(t)/t|\,dt = M^s/s$, with $H(f)(x) = x_+^s\int_0^1(1-t)t^{s-1}dt = x_+^s/(s(s+1))$:

$$\operatorname{tr}(xA+yB)_+^s = s(s+1)\int (ax+by)_+^s\,d\mu_{A,B}(a,b), \qquad s>0. \tag{40}$$

Lemma 4.3 (Support for the sandwiched pair)**.** *For the pair (31), any $\beta\in\mathbb{R}$ and arbitrary multiplicities of L: $b>0$ and $\ell := a/b \in [l_1,l_n]$ for $\mu_{A,B}$-a.e. (a,b).*

Proof. The matrices $B=\sigma^{2\beta}$, $A-l_1B = \sigma^\beta(L-l_1I)\sigma^\beta$ and $l_nB-A=\sigma^\beta(l_nI-L)\sigma^\beta$ are positive semidefinite, so by the half-plane statement of Theorem 4.2(iii) the measure $\mu_{A,B}$ is carried by $\{b\geq 0,\ l_1b\leq a\leq l_nb\}$. On this set $b=0$ forces $a=0$, and the origin carries no mass. □ □

Lemma 4.4 (Transform uniqueness)**.** *A finite measure ν on $[l_1,l_n]$ is determined by any one of: its stop-loss transform $c\mapsto\int(\ell-c)_+^{\alpha'}d\nu$ on $\mathbb{R}$ ($\alpha'\geq 1$); its Stieltjes transform $c\mapsto\int(c-\ell)^{-p}d\nu$ on (l_n,∞) ($p>0$); its Laplace transform $t\mapsto\int e^{t\ell}d\nu$ on $[0,\infty)$.*

The proof is in Appendix A.

For $\alpha<1$ the classicalization rests on a companion result — a Stieltjes representation of $\operatorname{tr}[(A+tB)^{-p}]$ built from Heinävaara's measure — which we import; it is Theorem 1.1 there, stated in the note's convention, in which the positive-definite matrix comes first; the pair (31) enters below with the roles exchanged.

Theorem 4.3 (Stieltjes-p representation [Rey-Bellet, 2026])**.** *Let $A\succ 0$ and $B\succeq 0$ be Hermitian on $\mathbb{C}^n$, and let $p\geq 1$. With $\lambda_1<\cdots<\lambda_r$ ($r\leq n$) the* distinct *eigenvalues of $A^{-1/2}BA^{-1/2}$ and $P_1,\dots,P_r$ its spectral projections, there is a unique positive finite measure ρ_p on $[\lambda_1,\lambda_r]$ with*

$$\operatorname{tr}[(A+tB)^{-p}] = \int (1+t\ell)^{-p}\,d\rho_p(\ell), \qquad t\geq 0. \tag{41}$$

Its atoms are $\rho_p(\{\lambda_s\}) = \operatorname{tr}[(P_sA^{-1}P_s)^p]$, $s=1,\dots,r$, and its continuous part has density

$$\frac{d\rho_p^{\mathrm{ac}}}{d\ell}(\ell) = p(p-1)\int_0^\infty a^{1-p}\,\mathfrak{h}_{A,B}(a,\,\ell a)\,da, \tag{42}$$

with $\mathfrak{h}_{A,B}$ the density (38); the continuous part vanishes iff $p=1$ or $[A,B]=0$.

The representation is obtained in [Rey-Bellet, 2026, §4.2] from (37) by inverting Heinävaara's transform on x^{-p} with a cutoff at the origin, whose boundary terms cancel against the singular rays of $\mu_{A,B}$; the weight a^{1-p} is integrable because $a\geq\lambda_{\min}(A)$ on $\operatorname{supp}\mu_c$, a by-product of the same argument.

4.3 The three transform identities

Each regime is one identity expressing the trace functional of the quantum dual as a classical transform of a measure that the identity itself defines.

4.3.1 Stop-loss identity, $\alpha \in (1, \infty]$

The identity of this regime is Heinävaara's, namely his (40) read in the spectral variable $\ell = a/b$; what the lemma adds is the normalization and the atoms for the sandwiched pair. At $\alpha = \infty$ and for a pair of states, the same push-forward and the same identity are [Schmitt, 2026, Prop. 6.2], obtained independently.

Lemma 4.5 (Stop-loss identity)**.** *Let $\alpha \in (1, \infty]$, so $\alpha' \in [1, \infty)$ ($\alpha' = 1$, $\beta = \frac{1}{2}$ at $\alpha = \infty$), and define ω_α by*

$$\int \varphi(\ell)\, d\omega_\alpha(\ell) := \alpha'(\alpha' + 1) \int \varphi(a/b)\, b^{\alpha'}\, d\mu_{A,B}(a, b). \tag{43}$$

Then ω_α is a probability measure on $[l_1, l_n]$ and, for every $c \in \mathbb{R}$,

$$\operatorname{tr}\big[\sigma^\beta (L - cI)\sigma^\beta\big]_+^{\alpha'} = \int (\ell - c)_+^{\alpha'}\, d\omega_\alpha(\ell). \tag{44}$$

If L has simple spectrum, $\omega_\alpha = \sum_i N_i^{-\alpha'} \delta_{l_i} + \omega_\alpha^{\mathrm{ac}}$, where $\omega_\alpha^{\mathrm{ac}}$, the image of $b^{\alpha'}\mu_c$, is absolutely continuous and vanishes iff $[L, \sigma] = 0$.

Proof. By Lemma 4.3, $\ell = a/b$ is defined $\mu_{A,B}$-a.e. with values in $[l_1, l_n]$, so (43) defines a positive measure there. Identity (40) with $s = \alpha'$ at $(x, y) = (1, -c)$ gives $\operatorname{tr}[\sigma^\beta(L - cI)\sigma^\beta]_+^{\alpha'} = \alpha'(\alpha' + 1) \int (a - cb)_+^{\alpha'}\, d\mu_{A,B} = \alpha'(\alpha' + 1) \int b^{\alpha'} (a/b - c)_+^{\alpha'}\, d\mu_{A,B}$, which is (44); at $(x, y) = (0, 1)$ it gives $\omega_\alpha(\mathbb{R}) = \operatorname{tr} B^{\alpha'} = \operatorname{tr} \sigma^{2\beta\alpha'} = \operatorname{tr} \sigma = 1$. For simple spectrum (and $0 \notin \operatorname{spec} L$, as (39) requires; otherwise apply what follows to $L + \kappa I$ and translate by κ, which is legitimate by Lemma 4.4 since the transform of $L + \kappa I$ is that of L shifted in c), (39) and Lemma 4.2 describe μ_s as the radial measures $(1 - t)t^{-1}dt$ on the segments $\{t(l_i/N_i, 1/N_i) : 0 < t \le 1\}$, on which $a/b = l_i$ and $b^{\alpha'} = (t/N_i)^{\alpha'}$, so

$$\omega_\alpha(\{l_i\}) = \alpha'(\alpha' + 1)\, N_i^{-\alpha'} \int_0^1 (1 - t)\, t^{\alpha' - 1}\, dt = N_i^{-\alpha'}. \tag{45}$$

The image of $b^{\alpha'}\mu_c$ under $(a, b) \mapsto a/b$ is absolutely continuous because $(\ell, b) \mapsto (\ell b, b)$ is a diffeomorphism of $\mathbb{R} \times (0, \infty)$ with Jacobian b (its density is computed in Proposition 4.1); it vanishes iff $\mu_c \equiv 0$ iff $[A, B] = 0$ (Theorem 4.2(iii)) iff $[L, \sigma] = 0$, since $[A, B] = \sigma^\beta [L, \sigma^{2\beta}]\sigma^\beta$ and $\sigma > 0$. □ □

The CVaR endpoint $\alpha = \infty$. At $\alpha = \infty$ ($\beta = \frac{1}{2}$, $\alpha' = 1$) the pair is $(A, B) = (\sigma^{1/2} L \sigma^{1/2}, \sigma)$ and Lemma 4.5 reads

$$\operatorname{tr}\big[(\sigma^{1/2}(L - cI)\sigma^{1/2})_+\big] = \int_c^{l_n} (\ell - c)\, d\omega_\infty,$$

the order-one stop-loss transform of ω_∞. The left side is convex in c, so at this endpoint the existence of ω_∞ needs nothing from Theorem 4.2; ω_∞ is the layer-cake measure of [Gour, 2026, Eq. (15); Liu et al., 2025, Eq. (17); Schmitt, 2026, Prop. 6.2] for this pair (see the related work after Theorem 4.1). Inserted into the max-divergence dual (28), it makes the $\alpha = \infty$ case of Theorem 4.1 the Rockafellar–Uryasev conditional value-at-risk [Rockafellar and Uryasev, 2000] at tail level $u = (1 + q)^{-1}$:

$$V_\infty(q) = V_\infty^{\mathrm{cl}}(q; \omega_\infty) = \mathrm{CVaR}_u(\omega_\infty) = \inf_{c \in \mathbb{R}} \left\{ c + (1 + q) \int_c^{l_n} (\ell - c)\, d\omega_\infty \right\}. \tag{46}$$

The quantum CVaR is dominated by its Born-measure shadow, $V_\infty(q) \le \mathrm{CVaR}_u(\nu_L)$, with equality for all q iff $[L,\sigma]=0$: this is the $\alpha=\infty$ case of the stop-loss domination Proposition 6.1.

4.3.2 Stieltjes identity, $\alpha \in [\frac{1}{2}, 1)$

Lemma 4.6 (Stieltjes identity [Rey-Bellet, 2026])**.** *Let $\alpha \in [\frac{1}{2},1)$, $p=\alpha/(1-\alpha)$, fix $c_0 > \lambda_{\max}(L)$, and let ρ_p be the measure of Theorem 4.3 for the pair*

$$A_0 := \sigma^\beta(c_0 I - L)\sigma^\beta \succ 0, \qquad B_0 := \sigma^{2\beta} \succ 0.$$

With $m(\ell') := c_0 - 1/\ell'$, define $\omega_\alpha := m_*(\ell'^{-p}\, d\rho_p)$. Then ω_α is a probability measure on $[l_1, l_n]$, independent of c_0, and for every $c > \lambda_{\max}(L)$

$$\mathrm{tr}\big[\sigma^\beta(cI-L)\sigma^\beta\big]^{-p} = \int (c-\ell)^{-p}\, d\omega_\alpha(\ell). \tag{47}$$

Its atoms sit at the eigenvalues of L, with weight $N_i^{-\alpha'}$ at a simple eigenvalue l_i; its continuous part, the image of $\ell'^{-p}\rho_p^{\mathrm{ac}}$, vanishes iff $p=1$ or $[L,\sigma]=0$.

Proof. This is the pencil form of the Stieltjes-p representation, [Rey-Bellet, 2026, Cor. A.1], for $P=\sigma^{2\beta}$ and $Q=\sigma^\beta L\sigma^\beta$. Indeed $cP-Q=\sigma^\beta(cI-L)\sigma^\beta$, $P^{-1/2}QP^{-1/2}=L$ and $(c_0P-Q,\,P)=(A_0,B_0)$, so the measure of the corollary is ω_α. It is carried by $[l_1,l_n]$, it is positive since $p\ge 1$, and the corollary gives (47) for every $c>\lambda_{\max}(L)$; by Lemma 4.4 it is the unique measure with this property, hence independent of c_0. Its mass is $\mathrm{tr}\,P^{-p}=\mathrm{tr}\,\sigma^{-2\beta p}=\mathrm{tr}\,\sigma=1$ (as $-2\beta p=1$). Its atom at a simple eigenvalue l_i is $\langle i|P^{-1}|i\rangle^p = N_i^p = N_i^{-\alpha'}$. Its continuous part is the image of $\ell'^{-p}\rho_p^{\mathrm{ac}}$ and vanishes iff $p=1$ or $[P,Q]=\sigma^\beta[\sigma^{2\beta},L]\sigma^\beta=0$, that is, iff $p=1$ or $[L,\sigma]=0$. □ □

At $\alpha=\frac{1}{2}$ the lemma returns (35): $p=1$ kills the continuous part, and the weights $N_i^{-\alpha'}=\langle i|\sigma|i\rangle$ are the Born weights.

4.3.3 Laplace identity, $\alpha=1$: the Stahl measure

At $\alpha=1$ the sandwiching exponent is $\beta=0$; the pair degenerates and the state-dependence sits in the Gibbs cumulant $F_1(t)=\log\mathrm{tr}\;e^{\log\sigma+tL}$ of Theorem 3.2.

Lemma 4.7 (Laplace identity (Stahl; Heinävaara))**.** *There is a unique probability measure ω_1 on $[l_1,l_n]$ with*

$$\mathrm{tr}\;e^{\log\sigma+tL} = \int e^{t\ell}\, d\omega_1(\ell), \qquad t\in\mathbb{R}, \qquad \omega_1 = \sum_{i=1}^n e^{\langle i|\log\sigma|i\rangle}\,\delta_{l_i} + \mathfrak{h}_1(\ell)\,d\ell \quad \textit{(simple spectrum)}, \tag{48}$$

where

$$\mathfrak{h}_1(\ell) = \frac{1}{2\pi}\int_{\mathbb{R}} e^a \sum_{j=1}^n \big|\mathrm{Im}\,\lambda_j\big((aI-\log\sigma)(L-\ell I)^{-1}\big)\big|\,da, \qquad \ell\in(l_1,l_n), \tag{49}$$

$\mathfrak{h}_1=0$ off $[l_1,l_n]$, and $\mathfrak{h}_1\equiv 0$ iff $[L,\sigma]=0$. The mean of ω_1 is $\mathrm{tr}(L\sigma)$.

Proof. The atoms and the density are those of [Heinävaara, 2024, §1.4, Thm. 24], and the derivation is Heinävaara's [Heinävaara, 2024, proof of Thm. 24]: it is his identity (37) for the pair $(\log\sigma,\ \lambda_{\max}(L)I-L)$ with the test function $f(u)=u^{-2}e^{-1/u}\mathbf{1}_{u>0}$, for which $H(f)(x)=e^{-1/x}\mathbf{1}_{x>0}$: a reciprocal spectral variable produces the exponential, and the shift $\ell=\lambda_{\max}(L)-b$ is absorbed. The support lies in $[l_1,l_n]$ because $\mu_{A,B}$ is carried by the joint numerical range $W(A,B)$ and the

segments from the origin to its points (Theorem 4.2(ii)–(iii), for simple spectrum; in general the support statement is part of Stahl's theorem [Stahl, 2013]), whose b-projection is $[0, l_n - l_1]$. Mass and mean are the transform and its derivative at $t = 0$, uniqueness is Lemma 4.4, and the dichotomy is that of Theorem 4.2(iii). □ □

The existence of a positive representing measure is Stahl's theorem [Stahl, 2013], the former BMV conjecture [Bessis et al., 1975], whose proof (and Eremenko's [Eremenko, 2015, Clivaz, 2016]) yields the density implicitly, as a contour integral whose positivity is not manifest; in Heinävaara's formula (49) it is read off the absolute values. By Jensen's inequality each atom $e^{\langle i|\log\sigma|i\rangle}$ is dominated by the Born weight $\langle i|\sigma|i\rangle$, the deficit being carried by $\mathfrak{h}_1$.

4.4 Proof of the classicalization theorem

Step 1 (dual coincidence). For $\alpha \in (1,\infty)$ insert (44) into Theorem 3.4(i): $\|(\sigma^\beta(L-cI)\sigma^\beta)_+\|_{\alpha'} = \big(\int(\ell-c)_+^{\alpha'} d\omega_\alpha\big)^{1/\alpha'} = \|(\ell-c)_+\|_{L^{\alpha'}(\omega_\alpha)}$. At $\alpha=\infty$ insert it ($\alpha'=1$) into Theorem 3.5. For $\alpha \in [\frac12, 1)$ insert (47) into Theorem 3.4(ii), where $\|\sigma^\gamma(cI-L)^{-1}\sigma^\gamma\|_p^{-1} = \big(\mathrm{tr}[\sigma^\beta(cI-L)\sigma^\beta]^{-p}\big)^{-1/p} = \|(c-\ell)^{-1}\|^{-1}_{L^p(\omega_\alpha)}$. At $\alpha=1$ insert (48) into Theorem 3.2. In each case the quantum dual becomes, verbatim, the corresponding row of Lemma 4.1 at $\nu=\omega_\alpha$; both hold for every $q\ge 0$, which is (34).

Step 2 (mean). $V_\alpha(0) = \mathrm{tr}(L\sigma)$ (Proposition 2.1) while $V_\alpha^{\mathrm{cl}}(0;\nu) = \int \ell\, d\nu$, since at $q=0$ the classical ball is $\{\nu\}$.

Step 3 (commuting case). If $[L,\sigma]=0$, in a joint eigenbasis $\mathrm{tr}[\sigma^\beta(L-cI)\sigma^\beta]_+^{\alpha'} = \sum_i \langle i|\sigma|i\rangle (l_i-c)_+^{\alpha'}$, $\mathrm{tr}[\sigma^\beta(cI-L)\sigma^\beta]^{-p} = \sum_i \langle i|\sigma|i\rangle (c-l_i)^{-p}$ and $\mathrm{tr}\, e^{\log\sigma + tL} = \sum_i \langle i|\sigma|i\rangle e^{tl_i}$ (using $2\beta\alpha' = -2\beta p = 1$): the transforms of ν_L. By Lemma 4.4, $\omega_\alpha = \nu_L$.

Step 4 (uniqueness). This is Lemma 4.4 applied to the identity of each regime.

Step 5 (structure). Under simple spectrum, Lemma 4.5, Lemma 4.6 and Lemma 4.7 give the atoms of (ii) and exhibit the remainder as absolutely continuous, vanishing iff $[L,\sigma]=0$ or $p=1$, i.e. $\alpha=\frac12$. □

4.5 Structure of the spectral measure

Proposition 4.1 (The spectral measure, explicitly)**.** *For $\sigma>0$, $L=L^*$ with simple spectrum $l_1<\cdots<l_n$, and $\alpha\in[\frac12,\infty]\setminus\{1\}$ ($\alpha'=1$, $\beta=\frac12$ at $\alpha=\infty$), the spectral measure of Theorem 4.1 has atoms $\omega_\alpha(\{l_i\}) = N_i^{-\alpha'}$ and continuous density, for $\ell\in(l_1,l_n)$,*

Case (i), *$\alpha>1$:*

$$\frac{d\omega_\alpha^{\mathrm{ac}}}{d\ell}(\ell) = \alpha'(\alpha'+1)\int_0^\infty b^{\alpha'+1}\,\mathfrak{h}_{A,B}(\ell b,\, b)\, db, \tag{50}$$

Case (ii), *$\alpha\in[\frac12,1)$:*

$$\frac{d\omega_\alpha^{\mathrm{ac}}}{d\ell}(\ell) = p(p-1)\,\ell'^{\,2-p}\int_0^\infty a^{1-p}\,\mathfrak{h}_{A_0,B_0}(a,\,\ell' a)\, da, \qquad \ell' := \frac{1}{c_0-\ell}, \tag{51}$$

for any fixed $c_0>\lambda_{\max}(L)$, with $\mathfrak{h}_{A,B}$ (resp. $\mathfrak{h}_{A_0,B_0}$) the continuous Heinävaara density (38) of the pair (A,B) of (31) (resp. (A_0,B_0) of Lemma 4.6).

Proof. The atoms are (45). In case (i), restrict (43) to μ_c and substitute $a=\ell b$ at fixed $b>0$ (so $da = b\,d\ell$): $\int\varphi\, d\omega_\alpha^{\mathrm{ac}} = \int\varphi(\ell)\big[\alpha'(\alpha'+1)\int_0^\infty b^{\alpha'+1}\mathfrak{h}_{A,B}(\ell b,b)\,db\big]\,d\ell$, which is (50). In case (ii), push the density (42) of ρ_p forward through $\ell = c_0 - 1/\ell'$ (so $d\ell'/d\ell = \ell'^2$): $d\omega_\alpha^{\mathrm{ac}}/d\ell(\ell) = \ell'^{-p}\,(d\rho_p^{\mathrm{ac}}/d\ell')(\ell')\cdot\ell'^2$, which is (51). □ □

For $n = 2$ and $\alpha < 1$, closed forms follow from [Rey-Bellet, 2026, §5]; Figure 4 shows a non-commuting $n = 4$ example.

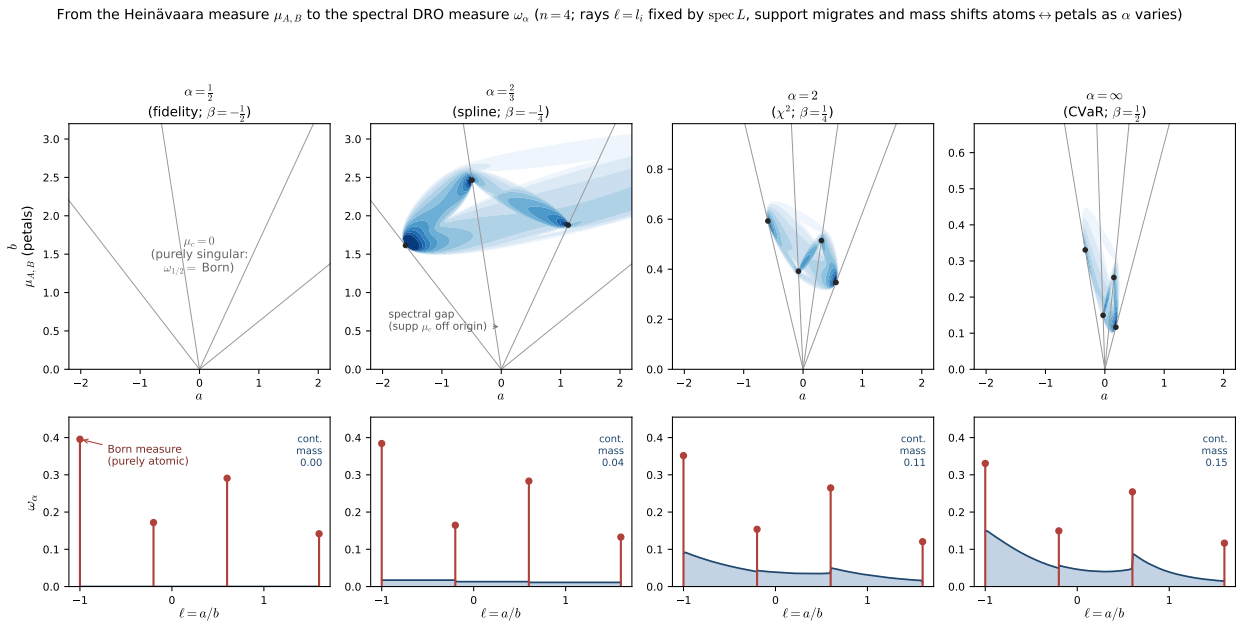


Figure 4: From the tracial measure to the spectral measure ($n = 4$; $\sigma = \mathrm{diag}(0.45, 0.30, 0.15, 0.10)$, L with eigenvalues $\{-1, -0.2, 0.6, 1.6\}$ rotated so that $[\sigma, L] \neq 0$). *Top:* the continuous part of $\mu_{A,B}$ in the (a, b) half-plane; the rays $\ell = l_i$ are fixed by $\operatorname{spec} L$, while the singular points $(l_i/N_i, 1/N_i)$ and the continuous density move with α. *Bottom:* ω_α on $[l_1, l_n]$. As α grows from $\frac{1}{2}$ to ∞, mass migrates from the atoms into the continuous part; at $\alpha = \frac{1}{2}$ the measure is the purely atomic Born measure.

5 Linearization and quantum Fisher metrics

A distributionally robust objective is often used through its *linearization*: at a small budget the worst-case value is the nominal value plus a variance term, the variance-regularized surrogate that makes DRO tractable in machine learning [Duchi and Namkoong, 2019, Lam, 2016, Gotoh et al., 2018], and the linearized goal-oriented divergence bounds of uncertainty quantification [Dupuis et al., 2016]. For the quantum sandwiched DRO the square-root law of Proposition 2.3 gives

$$V_\alpha(q) = \mathrm{tr}(L\sigma) + s_\alpha(\hat{L})\sqrt{2q} + o(\sqrt{q}), \qquad q \to 0^+, \tag{52}$$

with $\hat{L}$ the centered observable; the coefficient $s_\alpha^2(\hat{L})$, the **DRO variance**, is the quantum variance-regularizer, and computing it is the business of this section.

Theorem 5.1 (DRO variance)**.** *Let $\sigma > 0$, L Hermitian and not a multiple of the identity, and $\alpha \in [1/2, \infty)$. The DRO value linearizes as in (52), and the **DRO variance** has the two forms*

$$s_\alpha^2(\hat{L}) \;=\; \langle \hat{L}, \mathcal{H}_\alpha^{-1}(\hat{L}) \rangle \;=\; \tfrac{1}{\alpha}\,\mathrm{Var}_{\omega_\alpha}(\ell), \tag{53}$$

where $\mathcal{H}_\alpha$ is the Hessian of $H_\alpha(\cdot\|\sigma)$ at σ and ω_α the spectral measure of Theorem 4.1. At matched curvature the variance $\mathrm{Var}_{\omega_\alpha}(\ell) = \alpha\, s_\alpha^2(\hat{L})$ is non-increasing in α: constant, equal to $\mathrm{Var}_\sigma(L)$, if $[\sigma, L] = 0$, and strictly decreasing otherwise.

The theorem is proved by combining Theorem 5.2 (the linearization), Corollary 5.2 and Corollary 5.3 (the two forms of the variance) and Theorem 5.3 (the ordering), established in that order below.

5.1 The DRO variance bridge

For a divergence $\mathcal{D}$ with $\mathcal{D}(\cdot\|\sigma)$ twice continuously differentiable on a relative neighborhood of σ, the second variation defines a symmetric bilinear form on $T_\sigma\mathcal{S}_n := \{A = A^* \in M_n(\mathbb{C}) : \operatorname{tr} A = 0\}$, whose self-adjoint operator is the **Hessian superoperator** $\mathcal{H}$:

$$\langle A, \mathcal{H}(A)\rangle \;:=\; \frac{d^2}{d\epsilon^2}\,\mathcal{D}(\sigma + \epsilon A\|\sigma)\Big|_{\epsilon=0}, \qquad A \in T_\sigma\mathcal{S}_n. \tag{54}$$

We extend $\mathcal{H}$ to all Hermitian matrices by $\mathcal{H}(I) = 0$; equivalently $\mathcal{H}^{-1}$ is the Moore–Penrose pseudoinverse, and $\mathcal{H}^{-1}(A) = \mathcal{H}^{-1}(A|_T)$ where $A|_T := A - (\operatorname{tr} A/n)I$.

This curvature is what the DRO linearization sees: Theorem 5.2 below is the classical fact that the Hessian of a Legendre conjugate is the inverse Hessian [Crouzeix, 1977], on the tangent space of $\{\operatorname{tr}\rho = 1\}$, combined with the square-root law of Proposition 2.3; its classical counterpart is the variance expansion of [Lam, 2016, Thm. 3.1; Duchi and Namkoong, 2019; Duchi et al., 2021; Gotoh et al., 2018]. Recall $\operatorname{tr}(\hat{L}\sigma) = 0$; trace pairings $\operatorname{tr}(\hat{L}\,X)$ for $X \in T_\sigma\mathcal{S}_n$ depend only on $\hat{L}|_T$ since $\operatorname{tr}(I \cdot X) = 0$.

Theorem 5.2 (Hessian bridge)**.** *Let $\mathcal{D}$ satisfy (D1)–(D3), let $\mathcal{D}(\cdot\|\sigma)$ be C^2 on a relative neighborhood of σ with Hessian $\mathcal{H}$ positive definite on $T_\sigma\mathcal{S}_n$, and let L not be a multiple of the identity. Then $F''(0) = \operatorname{tr}(\hat{L}\,\mathcal{H}^{-1}(\hat{L})) > 0$, and*

$$V_{\mathcal{D}}(q;\hat{L}) \;=\; \sqrt{2q\,\operatorname{tr}(\hat{L}\,\mathcal{H}^{-1}(\hat{L}))} \;+\; o(\sqrt{q}), \qquad q \to 0^+. \tag{55}$$

Proof. Write

$$\varphi(t,\rho) := t\operatorname{tr}(\hat{L}\rho) - \mathcal{D}(\rho\|\sigma), \qquad F(t) := \sup_{\rho\in\mathcal{S}_n}\varphi(t,\rho),$$

so that F is the abstract CGF of (4) for the observable $\hat{L}$. Since $\mathcal{H} \succ 0$ and $\mathcal{D}(\cdot\|\sigma)$ is C^2, there is a relatively open convex neighborhood U of σ in $\mathcal{S}_n$, contained in the relative interior (as $\sigma \succ 0$), on which $\mathcal{D}(\cdot\|\sigma)$ is strictly convex. Maximizers of $\varphi(t,\cdot)$ exist by compactness and lower semicontinuity, and any maximizer ρ satisfies $\mathcal{D}(\rho\|\sigma) \le t\operatorname{tr}(\hat{L}(\rho-\sigma)) \le 2|t|\,\|\hat{L}\|$ (compare with $\rho = \sigma$, using $\operatorname{tr}(\hat{L}\sigma) = 0$ and $\|\rho-\sigma\|_1 \le 2$); by (D2), lower semicontinuity and compactness, all maximizers therefore lie in U for $|t| < \delta$, where $\varphi(t,\cdot)$ is strictly concave, so the maximizer $\rho^*(t)$ is unique and interior for $|t| < \delta$. We locate $\rho^*(t)$ by the implicit function theorem and then differentiate F by the envelope theorem, on $(-\delta,\delta)$.

Step 1 — first-order optimality and the multiplier. Write $g(\rho) := \nabla_1\mathcal{D}(\rho\|\sigma)$ for the differential of $\mathcal{D}(\cdot\|\sigma)$ in its first argument, a Hermitian-matrix valued map that is C^1 near σ by the C^2 hypothesis and satisfies the tangent projection $g(\sigma)|_T = 0$ since σ minimizes $\mathcal{D}(\cdot\|\sigma)$ subject to $\operatorname{tr}\rho = 1$ (the ambient gradient may be a scalar multiple of I). The only constraint active near the interior maximizer is $\operatorname{tr}\rho = 1$, with Lagrangian $\varphi(t,\rho) - \zeta(\operatorname{tr}\rho - 1)$; stationarity in ρ reads

$$t\,\hat{L} \;=\; g(\rho^*(t)) + \zeta(t)\,I. \tag{56}$$

Project onto the tangent space $T_\sigma\mathcal{S}_n = \{X = X^* : \operatorname{tr} X = 0\}$ by $X \mapsto X|_T := X - (\operatorname{tr} X/n)I$. The multiplier term drops out, $(\zeta I)|_T = 0$, leaving

$$G(\rho^*(t)) \;=\; t\,\hat{L}\big|_T, \qquad G(\rho) := g(\rho)\big|_T. \tag{57}$$

Strict convexity on U makes σ the unique zero of the tangent first-order condition $G(\rho) = 0$ in U, so at $t = 0$ (57) forces $\rho^*(0) = \sigma$.

Step 2 — implicit function theorem and $\dot\rho^*(0)$. Define $\Phi(t, \rho) := t\,\hat L|_T - G(\rho)$ on a relative neighborhood of $(0, \sigma)$, valued in $T_\sigma\mathcal{S}_n$. It is C^1 with $\Phi(0, \sigma) = 0$, and its partial derivative in ρ there is $\partial_\rho\Phi|_{(0,\sigma)} = -\,DG|_\sigma$. Now $DG|_\sigma$ is the Fréchet derivative of G at σ, i.e. for $A \in T_\sigma\mathcal{S}_n$,

$$DG|_\sigma(A) = \big(\partial_\rho g|_\sigma(A)\big)\big|_T = \mathcal{H}(A),$$

the second differential of $\mathcal{D}(\cdot\|\sigma)$ at σ of (54) (the projection $(\cdot)|_T$ is redundant because $\operatorname{tr}\mathcal{H}(A) = \langle I, \mathcal{H}(A)\rangle = \langle \mathcal{H}(I), A\rangle = 0$ by self-adjointness and $\mathcal{H}(I) = 0$). So $\partial_\rho\Phi|_{(0,\sigma)} = -\mathcal{H}$, invertible on $T_\sigma\mathcal{S}_n$ by hypothesis, and the implicit function theorem yields a unique C^1 curve $t \mapsto \tilde\rho(t)$ with $\tilde\rho(0) = \sigma$ solving $\Phi(t, \tilde\rho(t)) = 0$; it is the maximizer, $\tilde\rho(t) = \rho^*(t)$ for small $|t|$, because G is injective on the convex set U (strict convexity gives $\langle G(\rho_1) - G(\rho_2), \rho_1 - \rho_2\rangle > 0$ for $\rho_1 \neq \rho_2$ in U), so (57) has at most one solution in U, and both $\rho^*(t)$ and $\tilde\rho(t)$ lie in U and solve it. Differentiating $G(\rho^*(t)) = t\,\hat L|_T$ at $t = 0$ gives $\mathcal{H}(\dot\rho^*(0)) = \hat L|_T$, whence

$$\dot\rho^*(0) \;=\; \mathcal{H}^{-1}(\hat L|_T) \;=\; \mathcal{H}^{-1}(\hat L) \;\in\; T_\sigma\mathcal{S}_n \tag{58}$$

by the extension convention $\mathcal{H}^{-1}(\hat L) = \mathcal{H}^{-1}(\hat L|_T)$.

Step 3 — envelope theorem. The feasible set $\mathcal{S}_n$ is compact and independent of t, $\partial_t\varphi(t, \rho) = \operatorname{tr}(\hat L\rho)$ is jointly continuous, and the maximizer $\rho^*(t)$ is unique for $|t| < \delta$. Danskin's theorem [Bonnans and Shapiro, 2000, Thm. 4.13] therefore gives, with no differentiability of ρ^* required,

$$F'(t) \;=\; \partial_t\varphi(t, \rho^*(t)) \;=\; \operatorname{tr}(\hat L\,\rho^*(t)) \qquad \text{for } |t| < \delta.$$

In particular $F'(0) = \operatorname{tr}(\hat L\sigma) = 0$ because $\hat L$ is σ-centered, and since ρ^* is C^1 near 0 by Step 2, (58) yields

$$F''(0) \;=\; \operatorname{tr}(\hat L\,\dot\rho^*(0)) \;=\; \operatorname{tr}\big(\hat L\,\mathcal{H}^{-1}(\hat L)\big) \;=\; \langle \hat L|_T, \mathcal{H}^{-1}(\hat L|_T)\rangle \;>\; 0,$$

the I-component of $\hat L$ not contributing since $\mathcal{H}^{-1}(\hat L) \in T_\sigma\mathcal{S}_n$, and positivity holding because $\hat L|_T \neq 0$ when L is not a multiple of I. Feeding $F(0) = \varphi(0, \sigma) = 0$, $F'(0) = 0$ and $F''(0)$ into the square-root law of Proposition 2.3 gives (55). □ □

5.2 The sandwiched Hessian

We now recompute this Hessian for $\mathcal{D} = H_\alpha$: it is the metric of Takahashi and Fujiwara [Takahashi and Fujiwara, 2017, Lemma 4], obtained by the same divided-difference computation in [Wilde, 2025, Cor. 21]. The short computation fixes the DRO normalization (H_α rather than D_α) and produces the mean functions $m_\alpha = 1/h_\alpha$ used throughout.

Proposition 5.1 (Sandwiched Fisher kernel [Takahashi and Fujiwara, 2017, Wilde, 2025])**.** *Let* $\sigma = \sum_i p_i\,|i\rangle\langle i| > 0$ *and* $\alpha \in [1/2, \infty) \setminus \{1\}$. *For* $A \in T_\sigma\mathcal{S}_n$,

$$\frac{d^2}{d\epsilon^2}\,H_\alpha(\sigma + \epsilon A\,\|\,\sigma)\Big|_{\epsilon=0} \;=\; \sum_{i,j} h_\alpha(p_i, p_j)\,|A_{ij}|^2, \tag{59}$$

where, for $a, b > 0$ *with* $a \neq b$,

$$h_\alpha(a, b) \;=\; \frac{\alpha}{\alpha - 1}\cdot\frac{a^{(\alpha-1)/\alpha} - b^{(\alpha-1)/\alpha}}{a^{1/\alpha} - b^{1/\alpha}}\cdot(ab)^{(1-\alpha)/\alpha}, \tag{60}$$

with $h_\alpha(a, a) = \alpha/a$. *The formula extends continuously at* $\alpha = 1$ *to* $h_1(a, b) = (\log a - \log b)/(a - b)$, *agreeing with the Hessian of the Umegaki relative entropy* H_1.

Proof. Set $\gamma := (1-\alpha)/(2\alpha) = -\beta$ as in (9), $X_0 := \sigma^{1/\alpha}$ (eigenvalues $\pi_i := p_i^{1/\alpha}$), and $Y := \sigma^\gamma A\, \sigma^\gamma$. The identity $2\gamma + 1 = 1/\alpha$ gives $\sigma^\gamma\, \sigma\, \sigma^\gamma = \sigma^{1/\alpha} = X_0$, hence

$$\sigma^\gamma(\sigma + \epsilon A)\sigma^\gamma \;=\; X_0 + \epsilon\, Y, \qquad Q_\alpha(\sigma + \epsilon A\|\sigma) \;=\; \operatorname{tr}\big[(X_0 + \epsilon Y)^\alpha\big].$$

Since $H_\alpha = (Q_\alpha - 1)/(\alpha - 1)$ and $Q_\alpha(\sigma\|\sigma) = 1$, it suffices to expand the right side to second order in ϵ. The first variation is $\alpha \operatorname{tr}[X_0^{\alpha-1} Y] = \alpha \operatorname{tr}[\sigma^{(\alpha-1)/\alpha + 2\gamma} A] = \alpha \operatorname{tr} A = 0$. For the second, differentiate $\alpha \operatorname{tr}[(X_0 + \epsilon Y)^{\alpha-1} Y]$ once more via the Daleckiĭ–Krein formula [Bhatia, 1997, Thm. V.3.3] for $f'(x) = \alpha x^{\alpha-1}$ (smooth near the spectrum of $X_0 > 0$):

$$\frac{d^2}{d\epsilon^2} \operatorname{tr}\big[(X_0 + \epsilon Y)^\alpha\big]\Big|_0 \;=\; \sum_{i,j} f^{[1,1]}(\pi_i, \pi_j)\, |Y_{ij}|^2, \qquad f^{[1,1]}(a,b) := \frac{f'(a) - f'(b)}{a - b}, \quad f^{[1,1]}(a,a) = f''(a).$$

With $\pi_i^{\alpha-1} = p_i^{(\alpha-1)/\alpha}$ and, in the eigenbasis of σ, $|Y_{ij}|^2 = (p_i p_j)^{2\gamma} |A_{ij}|^2 = (p_i p_j)^{(1-\alpha)/\alpha} |A_{ij}|^2$, dividing by $\alpha - 1$ gives

$$\frac{d^2 H_\alpha}{d\epsilon^2}\bigg|_0 \;=\; \frac{\alpha}{\alpha - 1} \sum_{i,j} \frac{p_i^{(\alpha-1)/\alpha} - p_j^{(\alpha-1)/\alpha}}{p_i^{1/\alpha} - p_j^{1/\alpha}} (p_i p_j)^{(1-\alpha)/\alpha}\, |A_{ij}|^2 \;=\; \sum_{i,j} h_\alpha(p_i, p_j)\, |A_{ij}|^2,$$

the diagonal $h_\alpha(a,a) = \alpha/a$ coming from $f''(\pi) = \alpha(\alpha-1)\pi^{\alpha-2}$ at $\pi = a^{1/\alpha}$. At $\alpha = 1$, L'Hôpital in (60) gives $h_1(a,b) = (\log a - \log b)/(a-b)$, and the same computation for $\operatorname{tr} \rho \log \rho$ (Daleckiĭ–Krein for log; the term $-\operatorname{tr} \rho \log \sigma$ is linear) gives the Bogoliubov–Kubo–Mori Hessian $\sum_{ij} h_1(p_i, p_j) |A_{ij}|^2$ [Petz and Toth, 1993, Hiai and Petz, 2014, Ex. 7.18]. □ □

In detail, h_α/α is the Takahashi–Fujiwara metric [Takahashi and Fujiwara, 2017], to which the α-z information matrices of Wilde [Wilde, 2025] specialize at $z = \alpha$; the difference of normalizations ($\log Q_\alpha$ against $Q_\alpha - 1$) vanishes to second order at $\rho = \sigma$. The reciprocal kernel is a classical object: multiplying numerator and denominator of (60) by $(ab)^{(\alpha-1)/\alpha}$ shows

$$\frac{\alpha}{h_\alpha(a,b)} \;=\; M_{1/\alpha}(a,b), \qquad M_r(a,b) := \frac{r-1}{r} \cdot \frac{a^r - b^r}{a^{r-1} - b^{r-1}}, \tag{61}$$

the *power-difference mean* of Hiai and Kosaki [Hiai and Kosaki, 1999, §2] (limits at $r \in \{0, 1\}$ and on the diagonal understood by continuity).

Write $m_\alpha := 1/h_\alpha$ for the reciprocal kernel, the **DRO mean function** of the sandwiched Rényi family.

Corollary 5.1 (Inverse of the sandwiched Hessian)**.** *For $\alpha \in [1/2, \infty)$ the Hessian $\mathcal{H}_\alpha$ of $H_\alpha(\cdot\|\sigma)$ at σ is positive definite on $T_\sigma \mathcal{S}_n$, and for every Hermitian B with $\operatorname{tr}(\sigma B) = 0$, in the eigenbasis of σ,*

$$\mathcal{H}_\alpha^{-1}(B)_{ij} \;=\; m_\alpha(p_i, p_j)\, B_{ij}. \tag{62}$$

Proof. By (59), $\langle A, \mathcal{H}_\alpha A\rangle = \sum_{ij} h_\alpha(p_i, p_j) |A_{ij}|^2$ on $T_\sigma \mathcal{S}_n$, and $h_\alpha > 0$ on $(0,\infty)^2$ for $\alpha \geq 1/2$ (for $\alpha < 1$ both the prefactor and the divided difference in (60) are negative); this is positivity. With K_α the Schur multiplier by $h_\alpha(p_i, p_j)$ and $P_T(X) = X|_T$, $\mathcal{H}_\alpha = P_T K_\alpha P_T$ on $T_\sigma \mathcal{S}_n$ (the multiplier itself does not preserve tracelessness). Put $X := m_\alpha \circ B$ (entrywise product): since $m_\alpha(a,a) = a/\alpha$, $\operatorname{tr} X = \frac{1}{\alpha} \operatorname{tr}(\sigma B) = 0$, so $X \in T_\sigma \mathcal{S}_n$ and $P_T K_\alpha(X) = B|_T$, i.e. $X = \mathcal{H}_\alpha^{-1}(B|_T) = \mathcal{H}_\alpha^{-1}(B)$ under the extension convention of Section 5.1 — which is why $\hat{L}$ is σ-centered. □ □

Remark 5.1 (Monotonicity of the Hessian metric). For each $\alpha \in [1/2, \infty)$ the rescaled kernel h_α/α is a Petz monotone metric on $\mathcal{S}_n$, with representing function $\mathfrak{f}_\alpha(t) = M_{1/\alpha}(t, 1)$ [Petz, 1996, Hiai and Petz, 2014, Thm. 5.7] in the power-difference family of (61); it is monotone *if and only if* $\alpha \in (-\infty, -1] \cup [1/2, \infty)$ [Takahashi and Fujiwara, 2017, Hiai and Kosaki, 1999, Prop. 4.2], so the DPI threshold $\alpha = 1/2$ of the divergence is exactly the monotonicity threshold of its Hessian.

Corollary 5.2 (DRO variance for H_α)**.** *For $\alpha \in [1/2, \infty)$ the DRO variance s_α^2 of Proposition 2.3 for $\mathcal{D} = H_\alpha$ is*

$$s_\alpha^2(\hat{L}) \;:=\; F''(0) \;=\; \sum_{i,j} m_\alpha(p_i, p_j)\, |\hat{L}_{ij}|^2, \qquad m_\alpha(a, a) = a/\alpha. \tag{63}$$

In the commutative case $[\sigma, L] = 0$, $s_\alpha^2(\hat{L}) = (1/\alpha)\operatorname{Var}_\sigma(L)$.

Proof. The hypotheses of Theorem 5.2 hold for $\mathcal{D} = H_\alpha$: $Q_\alpha(\rho\|\sigma) = \operatorname{tr}(\sigma^{-\beta}\rho\sigma^{-\beta})^\alpha$ is C^∞ near σ, where the sandwiched argument is positive definite (likewise $\operatorname{tr}\rho\log\rho$ at $\alpha = 1$, with kernel the continuous extension h_1), and positivity of $\mathcal{H}_\alpha$ is Corollary 5.1. Since $\operatorname{tr}(\sigma\hat{L}) = 0$, (62) gives $\mathcal{H}_\alpha^{-1}(\hat{L})_{ij} = m_\alpha(p_i, p_j)\hat{L}_{ij}$, and the pairing $\operatorname{tr}(\hat{L}\,\mathcal{H}_\alpha^{-1}(\hat{L}))$ in the eigenbasis of σ is (63); the commutative case reads off the diagonal $m_\alpha(p_i, p_i) = p_i/\alpha$. □ □

Remark 5.2 (Quantum Fisher information and Cramér–Rao). $s_\alpha^2(\hat{L}) = \langle \hat{L}, \mathcal{H}_\alpha^{-1}(\hat{L})\rangle$ is the squared *co-metric* norm of the differential of $\rho \mapsto \operatorname{tr}(L\rho)$ in the α-information metric: for any C^1 curve ρ_t through $\rho_0 = \sigma$, Cauchy–Schwarz gives

$$\left|\tfrac{d}{dt}\operatorname{tr}(L\rho_t)\big|_{t=0}\right|^2 \;\le\; \langle \dot\rho_0, \mathcal{H}_\alpha(\dot\rho_0)\rangle\; s_\alpha^2(\hat{L}),$$

with equality iff $\dot\rho_0 \propto \mathcal{H}_\alpha^{-1}(\hat{L})$, the worst-case direction (58); the square-root law (55) is this Cramér–Rao inequality integrated over the α-information ball. At $\alpha = 1/2$ it is the SLD quantum Cramér–Rao bound [Helstrom, 1976, Braunstein and Caves, 1994], at $\alpha = 1$ the bound for the Kubo–Mori (Bogoliubov) metric [Petz and Toth, 1993]; for a general monotone metric and its covariance the inequality is in [Petz, 2002].

Corollary 5.3 (Variance of the spectral measure)**.** *For $\alpha \in [1/2, \infty)$ let ω_α be the spectral measure of Theorem 4.1. Then*

$$\mathbb{E}_{\omega_\alpha}[\ell] = \operatorname{tr}(L\sigma), \qquad \operatorname{Var}_{\omega_\alpha}(\ell) \;=\; \alpha\, s_\alpha^2(\hat{L}) \;=\; \alpha \sum_{i,j} m_\alpha(p_i, p_j)\, |\hat{L}_{ij}|^2. \tag{64}$$

Proof. Both sides of the variance identity are the coefficient of $\sqrt{2q}$ in $V_\alpha(q) = V_\alpha^{\mathrm{cl}}(q; \omega_\alpha)$ (34); on the quantum side $V_\alpha(q) = \operatorname{tr}(L\sigma) + s_\alpha(\hat{L})\sqrt{2q} + o(\sqrt{q})$ by Theorem 5.2 and Corollary 5.2, while on the classical side H_α^{cl} is the f-divergence with $f(t) = (t^\alpha - 1)/(\alpha - 1)$, $f''(1) = \alpha$, for which $V_\alpha^{\mathrm{cl}}(q; \nu) = \mathbb{E}_\nu[\ell] + \sqrt{2q\operatorname{Var}_\nu(\ell)/\alpha} + o(\sqrt{q})$ [Lam, 2016, Thm. 3.1; Duchi and Namkoong, 2019; Duchi et al., 2021]. The mean is the identity of Theorem 4.1. □ □

5.3 Examples

The function m_α at the three distinguished values $\alpha \in \{1/2, 1, 2\}$ recovers three classical means of $a, b > 0$, identified at the level of rescaled kernels h_α/α with normalized Petz monotone metrics on $\mathcal{S}_n$.

$\alpha = 1$ **(Kubo–Mori).** m_1 is the logarithmic mean, $s_1^2(\hat{L}) = \int_0^1 \operatorname{tr}(\hat{L}\sigma^s\hat{L}\sigma^{1-s})\,ds$ is the Kubo–Mori–Bogoliubov variance, and h_1 generates the Kubo–Mori–Bogoliubov metric [Petz, 1996, Hasegawa, 1993, Lesniewski and Ruskai, 1999].

$\alpha = 1/2$ **(Bures).** $h_{1/2}(a,b) = 1/(a+b)$, so $s^2_{1/2}(\hat{L}) = 2\,\mathrm{Var}_\sigma(L)$, twice the Born variance: $H_{1/2}$ is the squared Bures distance [Bures, 1969, Uhlmann, 1976, Fuchs and van de Graaf, 1999], its Hessian twice the Bures metric, and the rescaled kernel $2/(a+b)$ is the SLD/Bures metric [Helstrom, 1976, Braunstein and Caves, 1994].

$\alpha = 2$ **(geometric mean).** At $\alpha = 2$ both $1/\alpha$ and $(\alpha-1)/\alpha$ equal $1/2$, the divided-difference factor in (60) is identically 1, and

$$h_2(a,b) \;=\; \frac{2}{\sqrt{ab}}, \qquad m_2(a,b) \;=\; \tfrac{1}{2}\sqrt{ab}, \qquad m_2(a,a) = a/2.$$

Hence $s^2_2(\hat{L}) = \frac{1}{2}\,\mathrm{tr}[\sigma^{1/2}\hat{L}\,\sigma^{1/2}\hat{L}]$, the quantity s^2_2 of (25), so the square-root law at $\alpha = 2$ is the small-q face of the exact formula (26); the same trace is the second moment of ω_2 (Corollary 5.3), whose deficit against the Born variance is the Wigner–Yanase skew information [Wigner and Yanase, 1963],

$$\mathrm{Var}_\sigma(L) - \mathrm{Var}_{\omega_2}(\ell) = \mathrm{tr}(\sigma L^2) - \mathrm{tr}(\sigma^{1/2} L \sigma^{1/2} L) = -\tfrac{1}{2}\,\mathrm{tr}[\sigma^{1/2}, L]^2 =: I_{\mathrm{WY}}(\sigma, L), \tag{65}$$

non-negative and zero iff $[L,\sigma] = 0$. The rescaled kernel $h_2/\alpha = 1/\sqrt{ab}$ has representing function $\mathfrak{f}_2(t) = \sqrt{t}$, the **geometric-mean Petz metric** (the Petz metric associated with the operator-monotone function $f(t) = \sqrt{t}$) [Petz, 1996, Hiai and Kosaki, 1999]. This metric should not be confused with the standard Wigner–Yanase monotone metric; the Wigner–Yanase quantity that appears here is instead the variance deficit in (65).

$\alpha \to \infty$**.** Since $m_\alpha(a,b) \to 0$ pointwise, $s^2_\alpha(\hat{L}) \to 0$ and the square-root law degenerates: $H_\infty(\rho\|\sigma) = \|\sigma^{-1/2}\rho\sigma^{-1/2}\|_\infty - 1$ is a cone at $\rho = \sigma$, $H_\infty(\sigma + \epsilon A\|\sigma) = \epsilon\,\lambda_{\max}(\sigma^{-1/2} A \sigma^{-1/2})$, not a paraboloid, and $V_\infty(q)$ grows linearly in q. The spectral measure does not degenerate — $\mathrm{Var}_{\omega_\alpha}(\ell) \to \mathrm{Var}_{\omega_\infty}(\ell) > 0$ (Theorem 5.3) — only the budget-normalized coefficient $\mathrm{Var}_{\omega_\alpha}(\ell)/\alpha$ vanishes, and the exact CVaR formula (46) takes over.

Remark 5.3 (Higher-order terms). The expansion can be continued. By classicalization, every further term is that of the classical small-radius expansion of the f-divergence DRO of ω_α — [He and Lam, 2026] for general f, [Lam, 2016] for the relative entropy, [Gotoh et al., 2018] in penalty form — evaluated on the moments of ω_α: the coefficient of q is a skewness correction, proportional to $(2-\alpha)$ and to the third central moment of ω_α, which vanishes at $\alpha = 2$ in agreement with Corollary 3.2; the same coefficient can be reached intrinsically from the third derivative of $H_\alpha(\cdot\|\sigma)$ at σ.

5.4 Variance ordering

Comparing worst-case fluctuations across α requires a normalization: H_α has classical curvature $f''_\alpha(1) = \alpha$, so a budget-q ball has a different local size at each α, and comparing s^2_α at equal q conflates this scaling with the effect of interest. The clean comparison rescales to unit curvature — the Petz normalization H_α/α, under which every member of the family restricts to the *same* classical Fisher information on commuting directions [Takahashi and Fujiwara, 2017] — where the DRO variance is the spectral variance $\mathrm{Var}_{\omega_\alpha}(\ell) = \alpha\, s^2_\alpha(\hat{L})$ of Corollary 5.3, equal to $\mathrm{Var}_\sigma(L)$ for every α in the commuting case; any ordering across α is then a pure noncommutativity effect. The normalized mean function is the power-difference mean $\alpha\, m_\alpha = M_{1/\alpha}$ of (61) — the ladder M_2 (arithmetic), M_1 (logarithmic), $M_{1/2}$ (geometric), $M_0 = ab\log(a/b)/(a-b)$ of Table 1 read in the variable $1/\alpha$ — for which Hiai–Kosaki prove the pointwise order $M_r \le M_{r'}$, $r < r'$ [Hiai and Kosaki, 1999, Thm. 2.1]. The ordering on $[1/2, \infty)$ is reproved in Lemma 5.1 by a log-convexity argument in the hyperbolic parametrization of [Hiai and Kosaki, 1999, proof of Thm. 2.1], because it also gives strictness off the diagonal and the equality cases, which we did not find stated.

Lemma 5.1 (Hyperbolic representation of m_α)**.** *For $a, b > 0$ put $u := \frac{1}{2}\log(a/b)$ and*

$$\Psi(x) \;:=\; \frac{\sinh(xu)}{x} \;=\; \frac{1}{2}\int_{-u}^{u} e^{xt}\,dt, \qquad \Psi(0) := u.$$

Then for all $\alpha \in [1/2, \infty)$, with $\xi := 1/\alpha \in (0, 2]$,

$$\alpha\, m_\alpha(a, b) \;=\; \sqrt{ab}\,\frac{\Psi(\xi)}{\Psi(1-\xi)}, \tag{66}$$

the value at $a = b$ understood by continuity. Consequently, for $a \neq b$,

$$\frac{ab\,\log(a/b)}{a-b} \;<\; \alpha\, m_\alpha(a, b) \;\leq\; \frac{a+b}{2}, \tag{67}$$

with equality on the right iff $\alpha = 1/2$, the left value attained as $\alpha \to \infty$; and $\alpha \mapsto \alpha\, m_\alpha(a, b)$ is strictly decreasing on $[1/2, \infty)$.

The proof is in Appendix A.

Since Ψ is even, (66) gives $\alpha' m_{\alpha'}(a, b) = (\alpha\, m_\alpha(a^{-1}, b^{-1}))^{-1}$ for the Hölder conjugate α' — the duality $M_{1-r} = M_r^{(-)}$ of [Hiai and Kosaki, 1999, p. 907] — which pairs Bures ($\alpha = 1/2$) with the RLD kernel ($\alpha' = -1$), Kubo–Mori with the CVaR kernel ($\alpha' = \infty$), and the geometric mean with itself; hence the $\alpha \to \infty$ row of Table 1.

Table 1: DRO mean functions $m_\alpha = 1/h_\alpha$ for the sandwiched Rényi family, and the rescaled kernels h_α/α identifying the corresponding normalized Petz monotone metric.

α	$m_\alpha(a, b)$	Rescaled kernel h_α/α
$1/2$	$a + b$	SLD/Bures: $2/(a + b)$
1	$(a - b)/(\log a - \log b)$	Kubo–Mori: $(\log a - \log b)/(a - b)$
2	$\sqrt{ab}/2$	Geometric: $1/\sqrt{ab}$
$\to \infty$	$\alpha m_\alpha \downarrow ab\log(a/b)/(a - b)$	harmonic-log: $m_1(a^{-1}, b^{-1})$

Theorem 5.3 (Variance ordering)**.** *Let $\hat{L} \neq 0$. The map $\alpha \mapsto \mathrm{Var}_{\omega_\alpha}(\ell)$ is non-increasing on $[1/2, \infty]$, constant if $[\sigma, L] = 0$ and strictly decreasing otherwise, with*

$$\mathrm{Var}_\sigma(L) \;=\; \mathrm{Var}_{\omega_{1/2}}(\ell) \;\geq\; \mathrm{Var}_{\omega_\alpha}(\ell) \;\geq\; \mathrm{Var}_{\omega_\infty}(\ell) \;=\; \sum_{i,j} \frac{p_i p_j\,\log(p_i/p_j)}{p_i - p_j}\,|\hat{L}_{ij}|^2 \;>\; 0 \tag{68}$$

(kernel value p_i at $p_i = p_j$).

Here $\omega_{1/2} = \nu_L$ is the Born measure (35), so the left end is $\mathrm{Var}_{\nu_L}(\ell) = \mathrm{tr}(\sigma\hat{L}^2)$, and ω_∞ is the CVaR measure of Section 4.3.1; in the commuting case all variances equal $\mathrm{Var}_\sigma(L)$. That a variance built from a monotone metric lies below the Born variance, with a deficit of skew-information type, is known [Petz, 2002, Gibilisco et al., 2009, Hansen, 2008, Luo, 2003], and the monotonicity in α is [Hiai and Kosaki, 1999, Thm. 2.1] for the kernels and [Wilde, 2025, Thm. 30] for the sandwiched information matrices; the theorem adds strictness with the equality cases, and the fact that these are the variances of the laws ω_α, with ω_∞ the CVaR measure at the right end.

Proof. By Corollary 5.3, $\mathrm{Var}_{\omega_\alpha}(\ell) = \sum_{ij} \alpha\, m_\alpha(p_i, p_j)\,|\hat{L}_{ij}|^2$. By Lemma 5.1 each kernel entry $\alpha\, m_\alpha(p_i, p_j)$ is strictly decreasing in α when $p_i \neq p_j$ and constant ($= p_i$) when $p_i = p_j$; hence the

sum is strictly decreasing iff $\hat{L}_{ij} \neq 0$ for some pair with $p_i \neq p_j$, i.e. iff $[\sigma, L] \neq 0$, and in the commuting case a joint eigenbasis gives $\mathrm{Var}_{\omega_\alpha} = \mathrm{Var}_\sigma(L)$ for all α (with $\omega_\alpha = \sum_i p_i \delta_{l_i}$ independent of α).

Left endpoint. At $\alpha = 1/2$ the upper bound of (67) is an equality, $\alpha\, m_\alpha(a,b) = (a+b)/2$, so the sum is $\sum_{ij} \frac{p_i+p_j}{2}|\hat{L}_{ij}|^2 = \mathrm{tr}(\sigma \hat{L}^2) = \mathrm{Var}_\sigma(\hat{L})$, agreeing with the direct evaluation $\mathrm{Var}_{\nu_L}(\ell) = \mathrm{tr}(\sigma \hat{L}^2)$ against $\omega_{1/2} = \nu_L$.

Right endpoint. By Lemma 5.1, $\alpha\, m_\alpha(a,b) \downarrow ab\log(a/b)/(a-b)$ as $\alpha \to \infty$ (value a at $a = b$), so the finite sum $\sum_{ij} \alpha\, m_\alpha(p_i, p_j)|\hat{L}_{ij}|^2$ converges to the right side of (68), which is strictly positive for $\hat{L} \neq 0$ since every kernel entry is. This limit is $\mathrm{Var}_{\omega_\infty}(\ell)$ because $\omega_\alpha \to \omega_\infty$ weakly as $\alpha \to \infty$: by (44) and (46), $\int (\ell - c)_+^{\alpha'} d\omega_\alpha = \mathrm{tr}[\sigma^\beta (L - cI)\sigma^\beta]_+^{\alpha'} \to \mathrm{tr}[\sigma^{1/2}(L - cI)\sigma^{1/2}]_+ = \int (\ell - c)_+\, d\omega_\infty$ for every c (continuity of eigenvalues as $\beta \to \frac{1}{2}$, $\alpha' \to 1$), while $x^{\alpha'} \to x$ uniformly on $[0, l_n - l_1]$ and the functions 1, $(\cdot - c)_+$ span a dense subspace of $C([l_1, l_n])$, which carries every ω_α; the second moment is continuous under weak convergence. □ □

6 Concentration and distinguishability: the quantum deflation

The quantum DRO value lies *below* its measured shadow. Let $\nu_L = \sum_i \langle i|\sigma|i\rangle\, \delta_{l_i}$ be the Born measure of L in the state σ (35) and $V_\alpha^{\mathrm{meas}}(q) := V_\alpha^{\mathrm{cl}}(q; \nu_L)$ its classical DRO, the **measured DRO**: then $V_\alpha(q) \leq V_\alpha^{\mathrm{meas}}(q)$ for every budget, with equality for all q only if $[\sigma, L] = 0$ or $\alpha = \frac{1}{2}$. The deficit is noncommutativity, and this section shows that it is seen at three levels: pointwise in the dual variable, where, as in [Berta et al., 2017], Golden–Thompson and Araki–Lieb–Thirring give the comparison with the measured problem (Proposition 6.1); in the variance, where $\mathrm{Var}_{\omega_\alpha}(\ell) < \mathrm{Var}_\sigma(L)$ strictly for $\alpha > \frac{1}{2}$ when $[\sigma, L] \neq 0$ (Theorem 5.3); and, non-asymptotically, in every concentration bound that reads the second moment — the spectral measure ω_α of Theorem 4.1 is a *deflated* Born measure, and a Bernstein bound built on it is never larger than the same bound built on ν_L, and strictly smaller below saturation when $[\sigma, L] \neq 0$ and $\alpha \in (\frac{1}{2}, 2]$ (Proposition 6.3). The classical layer is standard — the φ-divergence DRO duals [Ben-Tal et al., 2013, Breuer and Csiszár, 2016], the concentration inequalities of Hoeffding, Bernstein and Bennett [Hoeffding, 1963, Bennett, 1962, Boucheron et al., 2013], and the Markov–Krein theory of Chebyshev systems [Karlin and Studden, 1966, Kreĭn and Nudel′man, 1977] and the stochastic orders [Shaked and Shanthikumar, 2007], whose proofs are collected in Appendix A — and the quantum input is the transform identities of Section 4 together with the deflated variance. Throughout, $\mu := \mathrm{tr}(L\sigma)$ is the common mean of ω_α and ν_L, $\hat{L} = L - \mu I$, $a := l_n - \mu$ and $b := \mu - l_1$ (both positive when $L \notin \mathbb{R}I$), and every upper bound below is two-sided once applied to $-L$ as well (the spectral measure of $-L$ is the reflection of ω_α; (8)).

One monotonicity is used throughout. Write, for a probability measure ν on $[l_1, l_n]$, $\Pi(c;\nu) := \int (\ell - c)_+^{\alpha'} d\nu$ ($\alpha > 1$), $M(t;\nu) := \int e^{t\ell} d\nu$ ($\alpha = 1$) and $S_p(c;\nu) := \int (c - \ell)^{-p} d\nu$ ($\alpha < 1$).

Lemma 6.1 (Transform monotonicity)**.** *If the transform of ν is pointwise dominated by that of ν' — $\Pi(c;\nu) \leq \Pi(c;\nu')$ for all $c \in \mathbb{R}$ ($1 < \alpha \leq \infty$, with $\alpha' = 1$ at $\alpha = \infty$), $M(t;\nu) \leq M(t;\nu')$ for all $t \geq 0$ ($\alpha = 1$), or $S_p(c;\nu) \leq S_p(c;\nu')$ for all $c > l_n$ ($\alpha < 1$) — then $V_\alpha^{\mathrm{cl}}(q;\nu) \leq V_\alpha^{\mathrm{cl}}(q;\nu')$ for every $q \geq 0$.*

Proof. By Lemma 4.1 the value is an infimum over the dual variable of an objective that increases with the transform at that point: $c + \kappa\, \Pi(c;\nu)^{1/\alpha'}$ with $\kappa = (1 + (\alpha - 1)q)^{1/\alpha}$ (and $c + (1+q)\Pi(c;\nu)$ at $\alpha = \infty$), $(q + \log M(t;\nu))/t$, and $c - \kappa\, S_p(c;\nu)^{-1/p}$. Pointwise domination of the objectives passes to the infima. □ □

6.1 Born domination

Proposition 6.1 (Born domination)**.** *For every $\alpha \in [\frac{1}{2}, \infty]$ the trace functional of the quantum dual (Theorem 3.4, Theorem 3.2, Theorem 3.5) is pointwise dominated by the corresponding transform of the Born measure ν_L:*

(i) for $\alpha \in (1, \infty)$ and every $c \in \mathbb{R}$,

$$\operatorname{tr}\big[\sigma^{\beta}(L - cI)\sigma^{\beta}\big]_{+}^{\alpha'} \;\le\; \operatorname{tr}\big[\sigma\,((L - cI)_{+})^{\alpha'}\big]; \tag{69}$$

(ii) at $\alpha = 1$, for every $t \ge 0$, $\operatorname{tr}\, e^{\log\sigma + tL} \le \operatorname{tr}\big[\sigma\, e^{tL}\big]$;

(iii) for $\alpha \in [\frac{1}{2}, 1)$ and every $c > l_n$, $\operatorname{tr}\big[\sigma^{\beta}(cI - L)\sigma^{\beta}\big]^{-p} \le \operatorname{tr}\big[\sigma\,(cI - L)^{-p}\big]$;

(iv) at $\alpha = \infty$, for every $c \in \mathbb{R}$, $\operatorname{tr}\big[\sigma^{1/2}(L - cI)\sigma^{1/2}\big]_{+} \le \operatorname{tr}\big[\sigma\,(L - cI)_{+}\big]$, i.e.

$$\omega_{\infty} \;\preceq_{\mathrm{cx}}\; \nu_L \tag{70}$$

in the convex order.

Consequently

$$V_{\alpha}(q) \;\le\; V_{\alpha}^{\mathrm{meas}}(q) \qquad \text{for all } q \ge 0, \tag{71}$$

with equality for all $q \ge 0$ if and only if $\alpha = \frac{1}{2}$ or $[\sigma, L] = 0$.

Proof. Write $X = L - cI$. (i) $\sigma^{\beta} X \sigma^{\beta} \le \sigma^{\beta} X_{+} \sigma^{\beta}$, so $\operatorname{tr}[(\sigma^{\beta} X \sigma^{\beta})_{+}^{\alpha'}] \le \operatorname{tr}[(\sigma^{\beta} X_{+} \sigma^{\beta})^{\alpha'}] \le \operatorname{tr}[\sigma^{2\beta\alpha'} X_{+}^{\alpha'}]$, by Weyl monotonicity and the Araki–Lieb–Thirring inequality [Lieb and Thirring, 1976, Araki, 1990] ($\operatorname{tr}(B^{1/2} A B^{1/2})^{r} \le \operatorname{tr}(B^{r/2} A^{r} B^{r/2})$ for $A, B \ge 0$, $r \ge 1$) with $r = \alpha'$; and $2\beta\alpha' = 1$. (ii) is the Golden–Thompson inequality [Golden, 1965, Thompson, 1965]. (iii) is Araki–Lieb–Thirring with $A = (cI - L)^{-1}$, $B = \sigma^{-2\beta}$, $r = p$, since $-2\beta p = 1$. (iv) is the Weyl step of (i) with $\alpha' = 1$; its two sides are the order-one stop-loss transforms of ω_{∞} and ν_L, two laws with equal means, so it is the convex order. The dual objectives of Theorem 3.4, Theorem 3.2 and Theorem 3.5 are increasing functions of the left sides, and those of Lemma 4.1 for ν_L the same functions of the right sides, so (71) follows by taking infima. If $\alpha = \frac{1}{2}$ or $[\sigma, L] = 0$ then $\omega_{\alpha} = \nu_L$ (Theorem 4.1) and equality holds. Conversely, equality in (71) for all small q forces equal square-root coefficients (Corollary 5.3), i.e. $\operatorname{Var}_{\omega_{\alpha}}(\ell) = \operatorname{Var}_{\nu_L}(\ell)$, and at $\alpha = \infty$ equality of the two CVaRs at every level forces $\omega_{\infty} = \nu_L$; either contradicts Theorem 5.3 when $\alpha > \frac{1}{2}$ and $[\sigma, L] \ne 0$. □ □

The proposition is the dual-variable form of the comparison of the measured and the sandwiched divergences by Berta, Fawzi and Tomamichel [Berta et al., 2017, Prop. 5 and Thm. 6], through the same two trace inequalities, Golden–Thompson at $\alpha = 1$ and Araki–Lieb–Thirring otherwise, with equality only for commuting pairs; (71) is the data-processing inequality [Frank and Lieb, 2013, Beigi, 2013] for the measurement of L, and at $\alpha = 1$ item (ii) with (71) is the robustness bound of Yamamoto and Bouten [Yamamoto and Bouten, 2009] and of Vladimirov, Petersen and James [Vladimirov et al., 2018, eq. (53)]. What we add is the convex-order endpoint (iv) and the equality case for the value.

6.2 Moment certificates

Fix α and write $\omega = \omega_{\alpha}$. Let $\mathcal{M}_k$ be the probability measures on $[l_1, l_n]$ whose first k moments agree with those of ω; since $\omega \in \mathcal{M}_k$,

$$\inf_{\nu \in \mathcal{M}_k} V_{\alpha}^{\mathrm{cl}}(q;\, \nu) \;\le\; V_{\alpha}(q) \;\le\; \sup_{\nu \in \mathcal{M}_k} V_{\alpha}^{\mathrm{cl}}(q;\, \nu). \tag{72}$$

What makes (72) sharp and computable is the theory of Chebyshev systems [Karlin and Studden, 1966, Kreĭn and Nudel′man, 1977]. A function f is $(k{+}1)$**-convex** if its divided differences of order $k+1$ are nonnegative ($f^{(k+1)} \geq 0$ for $f \in C^{k+1}$). For an interior moment point, $\mathcal{M}_k$ contains exactly two representing measures of minimal index (interior atoms counting 1, endpoint atoms $\frac{1}{2}$), the **lower and upper principal representations** $\nu_-^{(k)}$ and $\nu_+^{(k)}$, distinguished by $l_n \in \operatorname{supp} \nu_+^{(k)}$; they are the Gauss, Gauss–Radau and Gauss–Lobatto quadratures of the moment data. The first two levels are explicit: $\nu_-^{(1)} = \delta_\mu$ and $\nu_+^{(1)}$ is the two-point measure on $\{l_1, l_n\}$ with mean μ, the *Hoeffding measure*; at $k=2$, with $v := \operatorname{Var}_\omega(\ell)$,

$$\nu_+^{(2)} = \left(1 - \tfrac{v}{v+a^2}\right) \delta_{x_+} + \tfrac{v}{v+a^2}\, \delta_{l_n}, \qquad x_+ = \mu - \frac{v}{a}, \tag{73}$$

which we call the *Bernstein measure*; it is the extremal law of Bennett's lemma [Bennett, 1962, Hoeffding, 1963], and its moment generating function is the Bennett bound of [Gourgoulias et al., 2020]. Symmetrically $\nu_-^{(2)}$ is carried by $\{l_1, x_-\}$, $x_- = \mu + v/b$. The Markov–Krein theorem [Karlin and Studden, 1966, Ch. III–IV] states that for every $(k{+}1)$-convex f and $\nu \in \mathcal{M}_k$,

$$\int f\, d\nu_-^{(k)} \;\leq\; \int f\, d\nu \;\leq\; \int f\, d\nu_+^{(k)}, \tag{74}$$

both bounds attained. Whether the dual integrands of Lemma 4.1 — $(\ell - c)_+^{\alpha'}$, $e^{t\ell}$, $(c-\ell)^{-p}$ — lie in this cone decides where the certificate is sharp:

Proposition 6.2 (Convexity-cone stratification)**.** *Fix an integer $k \geq 0$ and work on $[l_1, l_n]$. (i) For every $p > 0$ and $c > l_n$, $\ell \mapsto (c-\ell)^{-p}$ is $(k{+}1)$-convex, and so is $\ell \mapsto e^{t\ell}$ for every $t \geq 0$. (ii) For $\alpha' > 0$, $\ell \mapsto (\ell - c)_+^{\alpha'}$ is $(k{+}1)$-convex for every $c \in \mathbb{R}$ if and only if $k \leq \alpha'$; for $k > \alpha'$ it fails at every interior c. Consequently, whenever $k \leq \alpha'$ — and, by (i), for every k when $\alpha \leq 1$ —*

$$V_\alpha^{\mathrm{cl}}(q;\, \nu_-^{(k)}) \;\leq\; V_\alpha(q) \;\leq\; V_\alpha^{\mathrm{cl}}(q;\, \nu_+^{(k)}), \tag{75}$$

and both bounds are attained on $\mathcal{M}_k$.

Parts (i) and (ii) are classical [Karlin and Studden, 1966, Denuit et al., 1998]; the proof is in Appendix A. The certificate follows from (i)–(ii) because (74) then brackets the transform of ω uniformly in the dual variable, and Lemma 6.1 turns the bracket into (75).

Remark 6.1 (The ladder for $\alpha \leq 1$: Gaussian quadrature and Padé). For $\alpha \leq 1$ the ladder is unlimited, and it is a classical object under another name. The extremal measures $\nu_\pm^{(k)}$ are the principal (Gauss, Radau, Lobatto) quadrature measures of the truncated Hausdorff moment problem, their nodes and weights governed by the orthogonal polynomials of ω_α [Golub and Meurant, 2010]; their Cauchy transforms — nodes the poles, weights the residues — are the Padé approximants to the Stieltjes transform of ω_α [Baker and Graves-Morris, 1996]; the same principal measures give the sharp brackets (75) through (74).

At $\alpha = 1$ this is the mechanism of Bessis–Moussa–Villani [Bessis et al., 1975], who conjectured the positivity of the representing measure of $\operatorname{tr} e^{\log\sigma + tL}$ precisely to license monotone two-sided Padé bounds on the partition function, with Gibbs–Bogoliubov as the $k=1$ rung; Stahl's theorem [Stahl, 2013] licenses them, and classicalization makes them certificates for the DRO value.

The bounds are monotone in k — $\mathcal{M}_{k+1} \subseteq \mathcal{M}_k$, so (74) at level k orders the transforms of $\nu_\pm^{(k+1)}$ and $\nu_\pm^{(k)}$, hence the values by Lemma 6.1 — and the upper certificates converge to $V_\alpha(q)$: by Hausdorff determinacy $\nu_\pm^{(k)} \to \omega_\alpha$ weakly, so the objective at $\nu_+^{(k)}$ converges to that at ω_α for each fixed dual variable, which bounds $\limsup_k V_\alpha^{\mathrm{cl}}(q;\nu_+^{(k)})$ by $V_\alpha(q)$ from above. The moments are computable: at $\alpha = 1$, $m_k = \frac{d^k}{dt^k} \operatorname{tr} e^{\log\sigma + tL}\big|_{t=0}$, and for $\alpha < 1$ they are the coefficients of $\operatorname{tr}[\sigma^\beta(cI - L)\sigma^\beta]^{-p}$ in powers of $1/c$ (Lemma 4.4). Figure 5 shows the ladder closing on a qutrit.

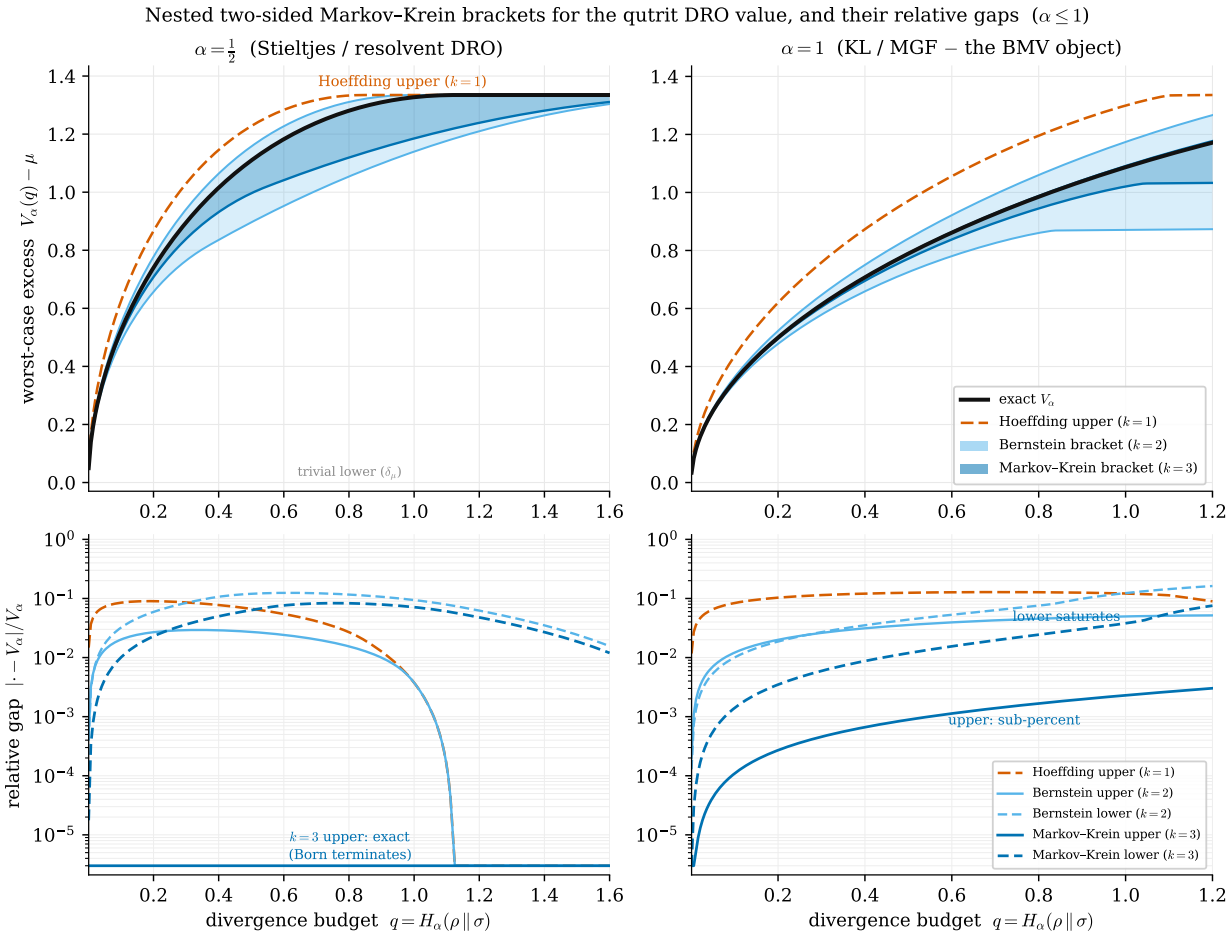


Figure 5: Nested Markov–Krein certificates for a qutrit ($L = \mathrm{diag}(0,1,2)$, σ tilted so that $[\sigma, L] \neq 0$) at $\alpha = \frac{1}{2}$ and $\alpha = 1$: the exact excess $V_\alpha(q) - \mu$ inside the Hoeffding cap ($k = 1$), the Bernstein bracket ($k = 2$) and the $k = 3$ bracket, with relative gaps below. At $\alpha = \frac{1}{2}$ three moments pin the three-atom Born measure and the $k = 3$ upper gap vanishes; at $\alpha = 1$ the density of ω_1 keeps every bound two-sided, the $k = 3$ upper gap staying below one percent.

6.3 The spectral Bernstein bound and the deflation

The **spectral variance**

$$v \;:=\; \mathrm{Var}_{\omega_\alpha}(\ell) \;=\; \alpha\, s_\alpha^2(\hat{L}) \;=\; \alpha \sum_{i,j} m_\alpha(p_i,p_j)\,|\hat{L}_{ij}|^2 \tag{76}$$

is computable in closed form for every α by Corollary 5.3 (at $\alpha=\infty$ by (68)), and provides the $k=2$ level of the hierarchy.[1] Let $\mathcal{M}(\mu,v)$ be the probability measures on $[l_1,l_n]$ with mean μ and variance v; note $v \le ab$ (the Bhatia–Davis inequality [Bhatia and Davis, 2000]).

Theorem 6.1 (Spectral Bernstein bound)**.** *Let $\alpha \in [\frac{1}{2},\infty]$, $q \ge 0$, and ρ with $H_\alpha(\rho\|\sigma) \le q$.*

(i) For $\alpha \in [\frac{1}{2},2]$, $V_\alpha^{\mathrm{cl}}(q;\nu_-^{(2)}) \le V_\alpha(q) \le V_\alpha^{\mathrm{cl}}(q;\nu_+^{(2)})$, both bounds attained on $\mathcal{M}(\mu,v)$; in particular

$$\mathrm{tr}\big(L(\rho-\sigma)\big) \;\le\; \mathcal{B}_\alpha(q;v) := V_\alpha^{\mathrm{cl}}(q;\,\nu_+^{(2)}) - \mu. \tag{77}$$

(ii) For $\alpha=2$, with $\mathrm{Var}_{\omega_2}(\ell)$ as in (65),

$$\boxed{\mathrm{tr}\big(L(\rho-\sigma)\big) \;\le\; \min\Big(\sqrt{q\;\mathrm{Var}_{\omega_2}(\ell)}\,,\; a\Big).} \tag{78}$$

(iii) For $\alpha=\infty$, with $u=(1+q)^{-1}$,

$$\mathrm{tr}\big(L(\rho-\sigma)\big) \;\le\; \sup_{\nu\in\mathcal{M}(\mu,v)} \mathrm{CVaR}_u(\nu) - \mu \;=\; \min\big(\,q\,b,\;\sqrt{q\,v},\;a\,\big). \tag{79}$$

Proof. (i) For $\alpha \in (1,2]$ one has $\alpha' \ge 2$, and for $\alpha \le 1$ the cone is unlimited (Proposition 6.2), so (75) holds at $k=2$ with $\mathcal{M}_2 = \mathcal{M}(\mu,v)$. (ii) $\nu_+^{(2)}$ is a two-point law; the classical χ^2-DRO of a law ν with mean μ and variance v equals $\mu+\sqrt{qv}$ as long as the linear tilt $1+\sqrt{q/v}\,(\ell-\mu)$ is nonnegative on the support [Duchi and Namkoong, 2019, §2.1] — for $\nu_+^{(2)}$, until $\sqrt{qv}=a$ at the atom x_+ — and a beyond; the variance identity is (65). (iii) is Hürlimann's theorem [Hürlimann, 2002, Thm. 3.1]. □ □

In (iii), Hürlimann's formula for the maximal CVaR at given range, mean and variance, the three terms are attained on $q \le v/b^2$, on $v/b^2 \le q \le a^2/v$ and on $q \ge a^2/v$, each by a law with at most three atoms. At $\alpha=1$ the upper certificate of (i) is the two-point bound of Bennett and Hoeffding [Bennett, 1962, Hoeffding, 1963], in the form of [Gourgoulias et al., 2020], where its sharpness over the moment class is also proved. The quantum content of the theorem is its input v: ω_α and ν_L share (μ,l_1,l_n), so the deflation of v is the whole difference between the spectral bound and its measured counterpart.

Proposition 6.3 (Quantum deflation of the Bernstein bound)**.** *Let $\alpha \in [\frac{1}{2},2]$ and $q \ge 0$. The certificate $v \mapsto \mathcal{B}_\alpha(q;v)$ of (77) is nondecreasing on $(0,ab]$, and, for $q>0$, strictly increasing below saturation: $\mathcal{B}_\alpha(q;v) < \mathcal{B}_\alpha(q;v')$ whenever $v<v'$ and $\mathcal{B}_\alpha(q;v') < a$. Consequently, for every ρ with $H_\alpha(\rho\|\sigma) \le q$,*

$$\mathrm{tr}\big(L(\rho-\sigma)\big) \;\le\; \mathcal{B}_\alpha\big(q;\mathrm{Var}_{\omega_\alpha}(\ell)\big) \;\le\; \mathcal{B}_\alpha\big(q;\mathrm{Var}_\sigma(L)\big), \tag{80}$$

and the second inequality is strict whenever $[\sigma,L]\neq 0$, $\alpha>\frac{1}{2}$, $q>0$, and the measured bound is below saturation.

[1] The names *Hoeffding* and *Bernstein* refer, as classically [Hoeffding, 1963, Bennett, 1962], to the moment data consumed — support, resp. support and variance — and not to the quantum Hoeffding bound of hypothesis testing [Hayashi, 2007, Audenaert et al., 2008] or the matrix Bernstein inequalities [Ahlswede and Winter, 2002, Tropp, 2015]. We use *Bernstein* for the whole mean–variance–support tier; classically, the sharp two-point bound of that tier is Bennett's and its explicit form $\sqrt{2vq}+aq/3$ is Bernstein's [Bennett, 1962, Boucheron et al., 2013].

At $\alpha = \infty$ the chain (80) holds for (79), which is nondecreasing in v; its second inequality is strict exactly when $\sqrt{q\,\mathrm{Var}_{\omega_\infty}(\ell)} < \min(qb,\, a)$, i.e. when the $\sqrt{qv}$ term is the active one at the deflated variance, and at small budgets, where the support term qb is active for both measures, it is an equality.

Proof. Let $v < v'$ and write $\nu_+ := \nu_+^{(2)}(v)$, $\nu'_+ := \nu_+^{(2)}(v')$. Since $x_+(v) = \mu - v/a > \mu - v'/a = x_+(v')$, the law ν_+ is supported in $[x_+(v'), l_n]$ and has mean μ, while ν'_+ is the two-point law on the endpoints of that interval with mean μ; hence $\int f\,d\nu_+ \le \int f\,d\nu'_+$ for every convex f (Jensen along the chord), i.e. $\nu_+ \preceq_{\mathrm{cx}} \nu'_+$. The dual integrands are convex, so Lemma 6.1 gives $\mathcal{B}_\alpha(q; v) \le \mathcal{B}_\alpha(q; v')$. For strictness, let c' be a minimizer of the ν'_+-objective, which exists in $(-\infty, l_n)$ (resp. (l_n, ∞) for $\alpha < 1$, $t' \in (0, \infty)$ for $\alpha = 1$) when $q > 0$ and $\mathcal{B}_\alpha(q; v') < a$, the objective being continuous and coercive with value $\ge a$ at $c \ge l_n$; there the integrand — $(\ell - c')_+^{\alpha'}$ with $c' < l_n$, $e^{t'\ell}$, $(c' - \ell)^{-p}$ — is convex and not affine on $[x_+(v'), l_n]$, and ν_+ charges the interior point $x_+(v)$, so the chord inequality is strict and $\mathcal{B}_\alpha(q; v) \le \mathrm{objective}_{\nu_+}(c') < \mathrm{objective}_{\nu'_+}(c') = \mathcal{B}_\alpha(q; v')$. The chain (80) is (77) followed by monotonicity with $v = \mathrm{Var}_{\omega_\alpha}(\ell) \le \mathrm{Var}_{\nu_L}(\ell) = \mathrm{Var}_\sigma(L)$, strict for $\alpha > \frac{1}{2}$ and $[\sigma, L] \ne 0$ by Theorem 5.3; and $\nu_L \in \mathcal{M}(\mu, \mathrm{Var}_\sigma(L))$ makes the right side the measured certificate. □ □

Since $\alpha \mapsto \mathrm{Var}_{\omega_\alpha}(\ell)$ is strictly decreasing for $[\sigma, L] \ne 0$ (Theorem 5.3), the deflation grows with α: from none at the fidelity endpoint, where $\omega_{1/2} = \nu_L$, to the Wigner–Yanase skew information at $\alpha = 2$ — (78) reads $\sqrt{q(\mathrm{Var}_\sigma(L) - I_{\mathrm{WY}}(\sigma, L))}$ against the measured $\sqrt{q\,\mathrm{Var}_\sigma(L)}$ (65) — and to the harmonic-logarithmic variance (68) at $\alpha = \infty$. Figure 6 shows it on a qutrit, across α and across budgets.

The certificates of Theorem 6.1 are sharp but quadrature-shaped. Explicit formulas exist at three *anchors* — $\alpha = \frac{1}{2}$, where the two-point certificate is solved by the Bures angle, $\alpha = 1$, where Bernstein's inequality applies, and $\alpha = 2$, where the value is a square root — and the monotonicity of H_α^{cl} in α (Remark 3.1) transports each anchor to the orders above it, at the price of a constant. Write, for $w \in [0, 1]$ and $\mathrm{Fid} := 1 - q/2$, $\Delta_q^*(w) := \sin(2\arcsin\sqrt{w} + \arccos \mathrm{Fid})\,\sqrt{1 - \mathrm{Fid}^2}$ for $w \le \mathrm{Fid}^2$ and $\Delta_q^*(w) := 1 - w$ otherwise, the fidelity profile of Theorem 6.3.

Proposition 6.4 (Explicit Bernstein bounds)**.** *Let $q \ge 0$, ρ with $H_\alpha(\rho\|\sigma) \le q$, and $v = \mathrm{Var}_{\omega_\alpha}(\ell)$.*

(i) For $\alpha \in [\frac{1}{2}, 1)$ and $q \le 2$,

$$\mathrm{tr}(L(\rho - \sigma)) \;\le\; \frac{a^2 + v}{a}\,\Delta_q^*\Big(\frac{v}{v + a^2}\Big). \tag{81}$$

(ii) For $\alpha \in [1, 2)$,

$$\mathrm{tr}(L(\rho - \sigma)) \;\le\; \min\Big(\sqrt{2vq} + \frac{aq}{3}\,,\; a\Big). \tag{82}$$

(iii) For $\alpha \in [2, \infty)$, $\mathrm{tr}(L(\rho - \sigma)) \le \min\big(\sqrt{vq},\; \sqrt{v(\kappa_\alpha(q)^2 - 1)},\; a\big)$, $\kappa_\alpha(q) := (1 + (\alpha - 1)q)^{1/\alpha}$.

Proof. For (i)–(ii), $\mathrm{tr}(L(\rho - \sigma)) \le \mathcal{B}_\alpha(q; v) = V_\alpha^{\mathrm{cl}}(q; \nu_+^{(2)}) - \mu$ by Theorem 6.1(i), with $\nu_+^{(2)} = \nu_+^{(2)}(v)$ the two-point law (73); this is where the deflated variance enters. For an anchor $\alpha_0 \le \alpha$ within the range where the $k = 2$ certificate is valid, monotonicity gives $V_\alpha^{\mathrm{cl}}(q; \nu_+^{(2)}) \le V_{\alpha_0}^{\mathrm{cl}}(q; \nu_+^{(2)})$. Thus (i) $\alpha_0 = \frac{1}{2}$: $\nu_+^{(2)}$ is the Bernoulli(w) law on $\{x_+, l_n\}$, $w = v/(v + a^2)$, of mean μ, and by the computation in the proof of Theorem 6.3 its fidelity DRO excess is $(l_n - x_+)\,\Delta_q^*(w)$ with $l_n - x_+ = (a^2 + v)/a$; at $\alpha = \frac{1}{2}$ the inclusion is an equality. (ii) $\alpha_0 = 1$: a law with mean μ, variance v and support in $(-\infty, l_n]$ obeys Bernstein's majorant $\log \int e^{t\hat{\ell}}\,d\nu \le vt^2/(2(1 - ta/3))$, $0 < t < 3/a$, and optimizing the $\alpha = 1$ dual gives $\sqrt{2vq} + aq/3$ (Appendix A). (iii) For $\alpha \ge 2$ we do **not** invoke the $k = 2$ Markov–Krein

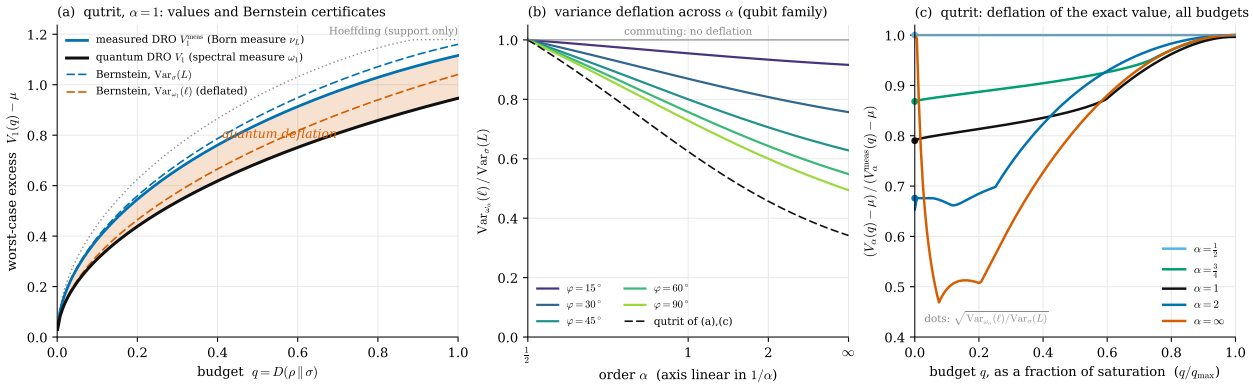


Figure 6: The quantum deflation. *(a)* A qutrit ($L = \mathrm{diag}(0, 1, 2)$, σ with eigenvalues $(0.90, 0.07, 0.03)$ rotated so that $[\sigma, L] \neq 0$; Born weights $(0.47, 0.24, 0.29)$) at $\alpha = 1$: worst-case excess against the budget $q = D(\rho\|\sigma)$ for the measured DRO (Born measure ν_L) and the quantum DRO (spectral measure ω_1), with the two $k = 2$ Bernstein certificates of Theorem 6.1 built on $\mathrm{Var}_\sigma(L)$ and on the deflated $\mathrm{Var}_{\omega_1}(\ell) = 0.62\ \mathrm{Var}_\sigma(L)$, and the Hoeffding bound above them. The shaded band is the deflation; the deflated certificate lies below the exact *measured* value. *(b)* The ratio $\mathrm{Var}_{\omega_\alpha}(\ell)/\,\mathrm{Var}_\sigma(L)$ against α (axis linear in $1/\alpha$) for the qubit $\sigma = \mathrm{diag}(0.9, 0.1)$ and $L_\varphi = \cos\varphi\, Z + \sin\varphi\, X$, φ from 15° to 90°, and (dashed) for the qutrit of (a): equal to 1 at $\alpha = \frac{1}{2}$ and for commuting pairs, strictly decreasing in α otherwise (Theorem 5.3), down to the harmonic-logarithmic variance at $\alpha = \infty$. *(c)* For the qutrit of (a), the ratio of the exact quantum to the exact measured excess against the budget, as a fraction of the saturation budget, at $\alpha = \frac{1}{2}, \frac{3}{4}, 1, 2, \infty$; the dots at $q = 0$ are $\sqrt{\mathrm{Var}_{\omega_\alpha}(\ell)/\,\mathrm{Var}_\sigma(L)}$ (Corollary 5.3). No deflation at $\alpha = \frac{1}{2}$; at $\alpha = \infty$ the small-budget regime is linear for both measures, so the ratio starts at 1 and the deflation is strongest at intermediate budgets.

certificate, whose sharp validity stops at $\alpha = 2$. Instead classicalization gives a probability P with $H_\alpha^{\mathrm{cl}}(P\|\omega_\alpha) \le q$. Since $H_2^{\mathrm{cl}} \le H_\alpha^{\mathrm{cl}}$, Cauchy–Schwarz (Hammersley–Chapman–Robbins) directly yields $\int \hat{\ell}\, dP \le \sqrt{vq}$. The second estimate follows from the exact α-dual at the shifted level $\tilde{c} = c - \mu$ and Lyapunov's inequality $\|(\hat{\ell} - \tilde{c})_+\|_{\alpha'} \le \|\hat{\ell} - \tilde{c}\|_2 = \sqrt{v + \tilde{c}^2}$; minimizing $\tilde{c} + \kappa_\alpha(q)\sqrt{v + \tilde{c}^2}$ over $\tilde{c}$ gives $\sqrt{v(\kappa_\alpha(q)^2 - 1)}$ (Appendix A). Finally a is the support bound. □ □

The anchors are $\alpha = \frac{1}{2}$ (fidelity) in (i), $\alpha = 1$ (Bernstein) in (ii) and $\alpha = 2$ (Cauchy–Schwarz) in (iii), whose second term is the Lyapunov bound. In (i) equality holds at $\alpha = \frac{1}{2}$ for the certificate; in (ii) at $\alpha = 1$, $q = D(\rho\|\sigma)$ and v is the Kubo–Mori variance. At $\alpha = \frac{1}{2}$, where $v = \mathrm{Var}_\sigma(L)$, (i) is, after a reflection of the observable, the mean–variance bound of Weber et al. [Weber et al., 2022, Thm. 2], with the same validity condition, here the $k = 2$ certificate, which for $\alpha \in (\frac{1}{2}, 1)$ holds with the smaller variance $\mathrm{Var}_{\omega_\alpha}(\ell)$; for a classical pair (ii) is Bernstein's inequality [Boucheron et al., 2013]. As $q \to 0$ the three bounds behave as $2\sqrt{qv}$, $\sqrt{2qv}$ and $\sqrt{qv}$ against the sharp $\sqrt{2qv/\alpha}$ of Corollary 5.3: loss factors $\sqrt{2\alpha}$, $\sqrt{\alpha}$ and $\sqrt{\alpha/2}$, each equal to 1 at its anchor. The Lyapunov term in (iii), of order $\sqrt{2qv(\alpha-1)/\alpha}$, is worse than $\sqrt{qv}$ at small q but sublinear in q, hence better at large budgets; all bounds are monotone in v and inherit the deflation.

Remark 6.2 (The $\alpha = 2$ anchor is Hammersley–Chapman–Robbins). The Cauchy–Schwarz step of (iii) is the classical Hammersley–Chapman–Robbins inequality $(\mathbb{E}_P g - \mathbb{E}_Q g)^2 \le \mathrm{Var}_Q(g)\, \chi^2(P\|Q)$ [Hammersley, 1950, Chapman and Robbins, 1951], and its quantum form at $\alpha = 2$,

$$\left|\mathrm{tr}(L(\rho - \sigma))\right|^2 \;\le\; \mathrm{Var}_{\omega_2}(\ell)\; H_2(\rho\|\sigma), \tag{83}$$

follows directly by splitting $\mathrm{tr}(\hat{L}X) = \mathrm{tr}[(\sigma^{1/4}\hat{L}\sigma^{1/4})(\sigma^{-1/4}X\sigma^{-1/4})]$, $X = \rho - \sigma$: here $H_2 = \mathrm{tr}(\sigma^{-1/2}X\sigma^{-1/2}X)$ is the geometric-mean member of the quantum χ^2 family of [Temme et al., 2010], and the variance factor is the spectral variance (65), deflated by the skew information. It is exact: equality holds at $X \propto \sigma^{1/2}\hat{L}\sigma^{1/2}$ as long as ρ stays positive, i.e. for $H_2 \le q_0$, where (83) is the DRO value itself (Corollary 3.2); and for $\rho = \sigma + \varepsilon A$, $H_2 = \frac{1}{2}\varepsilon^2\langle A, \mathcal{H}_2 A\rangle$ exactly, so (83) is the quantum Cramér–Rao bound of Remark 5.2.

6.4 The Hoeffding bound

The $k = 1$ certificate uses only (μ, l_1, l_n), which ω_α shares with ν_L; it therefore needs neither the spectral measure nor the classicalization, only Born domination. Write $\Delta = a + b$ and $\theta = b/\Delta$, the weight of $\nu_+^{(1)}$ at l_n.

Theorem 6.2 (Spectral Hoeffding bound)**.** *For every $\alpha \in [\frac{1}{2}, \infty]$, $q \ge 0$, and every ρ with $H_\alpha(\rho\|\sigma) \le q$:*

$$-\Delta \cdot \eta_\alpha(q,\, 1-\theta) \;\le\; \mathrm{tr}\big(L(\rho - \sigma)\big) \;\le\; \Delta \cdot \eta_\alpha(q,\, \theta), \tag{84}$$

where $\eta_\alpha(q, \theta)$ is the ***Hoeffding profile****, the excess DRO value of a Bernoulli(θ) on $\{0, 1\}$,*

$$\eta_\alpha(q, \theta) := V_\alpha^{\mathrm{cl}}(q;\, \theta\delta_1 + (1-\theta)\delta_0) - \theta. \tag{85}$$

Proof. By (71), $V_\alpha(q) \le V_\alpha^{\mathrm{cl}}(q; \nu_L)$. The dual integrands are convex ($k = 1 \le \alpha'$ always, and Proposition 6.2(i)), and ν_L has mean μ on $[l_1, l_n]$, so Jensen along the chord bounds each transform of ν_L by that of the Hoeffding measure $\nu_+^{(1)}$; Lemma 6.1 gives $V_\alpha^{\mathrm{cl}}(q; \nu_L) \le V_\alpha^{\mathrm{cl}}(q; \nu_+^{(1)}) = \mu + \Delta\, \eta_\alpha(q, \theta)$, the last step by the affine rescaling of $[l_1, l_n]$ onto $\{0, 1\}$. The lower bound is the upper bound for $-L$, whose Hoeffding parameter is $1 - \theta$. □ □

The classical statement is Hoeffding's [Hoeffding, 1963], at $\alpha = 1$ in the form of [Gourgoulias et al., 2020]; the quantum step is data processing. At $\alpha = \frac{1}{2}$ the bound, with the closed form (88) of its

profile, is the robustness interval of Weber et al. [Weber et al., 2022, Thm. 1]; at $\alpha = 1$ it follows from the measured bound discussed after Proposition 6.1; what the theorem adds is the statement for all α at once. The bound is sharp given only the data (μ, l_1, l_n). For $\alpha \in (\frac{1}{2}, \infty)$ it is attained in the quantum model at every budget precisely when $[\sigma, L] = 0$ and L has at most two distinct eigenvalues: only then does ω_α collapse onto ν_L (Theorem 4.1(i)), discrete on $\{l_1, l_n\}$, hence equal to $\nu_+^{(1)}$; at $\alpha = \frac{1}{2}$, where $\omega_{1/2} = \nu_L$, it is attained for every L with at most two distinct eigenvalues, commuting with σ or not. For $\alpha \in (\frac{1}{2}, \infty)$ and a noncommuting L with two eigenvalues, ω_α carries a density (Theorem 4.1(ii)) and the bound is strictly loose at every budget below saturation, and for L with three or more eigenvalues it is generically only approached (Figure 5). At $\alpha = \infty$ the integrand $(\ell - c)_+$ is affine above c and the bound is attained at small budgets for every L with simple spectrum, commuting or not: for $q \le w_1/(1 - w_1)$, $w_1 := \omega_\infty(\{l_1\}) = 1/\langle 1|\sigma^{-1}|1\rangle$, the bottom $(1 - u)$-mass of ω_∞ sits at l_1 and $V_\infty(q) - \mu = qb = \Delta\, \eta_\infty(q, \theta)$. At $\alpha = \infty$ the profile is piecewise linear, $\eta_\infty(q, \theta) = \min(\theta q,\, 1 - \theta)$, the CVaR of a Bernoulli(θ).

6.5 Trace distance and fidelity

Applied to the Helstrom observable, the bounds of this section become inequalities between a divergence and the trace distance — Pinsker-type inequalities [Pinsker, 1964, Csiszár, 1967, Hiai et al., 1981, Fuchs and van de Graaf, 1999]. Throughout, $\sigma > 0$, $T(\rho, \sigma) := \frac{1}{2}\|\rho - \sigma\|_1 = \max_{0 \le L \le I} \operatorname{tr}(L(\rho - \sigma))$, attained at the Helstrom projector $L_0 := \mathbb{1}_{\{\rho > \sigma\}}$ [Helstrom, 1976], of some rank $k \in \{1, \dots, n-1\}$ when $T > 0$; $\theta_0 := \operatorname{tr}(\sigma L_0) \in (0, 1)$; and $L_H := 2L_0 - I$, so that $L_H^2 = I$, $\operatorname{tr}(L_H(\rho - \sigma)) = \|\rho - \sigma\|_1$, $\operatorname{tr}(\sigma L_H) = 2\theta_0 - 1$ and $\operatorname{Var}_\sigma(L_H) = 4\theta_0(1 - \theta_0)$.

Corollary 6.1 (Pinsker-type inequalities at $\alpha = 1$)**.** *Let $r := D(\rho\|\sigma)$ and $s_1^2 := s_1^2(\hat{L}_H) = \operatorname{Var}_{\omega_1}(\ell)$, the Kubo–Mori variance of L_H (Corollary 5.3).*

(i) $T(\rho, \sigma) \le \eta_1(r, \theta_0) \le \sup_{\theta \in (0,1)} \eta_1(r, \theta) \le \sqrt{r/2}$.

(ii)

$$\|\rho - \sigma\|_1 \;\le\; \sqrt{2 s_1^2\, r} + \tfrac{2}{3}(1 - \theta_0)\, r, \qquad s_1^2 \;\le\; \operatorname{Var}_\sigma(L_H) = 4\theta_0(1 - \theta_0) \;\le\; 1, \tag{86}$$

with $s_1^2 < 4\theta_0(1 - \theta_0)$ whenever $[\sigma, L_H] \neq 0$.

Proof. (i) Theorem 6.2 at $\alpha = 1$ for $L = L_0$, which has $l_1 = 0$, $l_n = 1$, $\Delta = 1$ and $\theta = \theta_0$, gives $T = \operatorname{tr}(L_0(\rho - \sigma)) \le \eta_1(r, \theta_0)$. For a binary pair the mean shift is the total variation, so $\eta_1(r, \theta)$ is the largest total variation of a binary pair (P, Q) with $Q = \mathrm{Ber}(\theta)$ and $D(P\|Q) \le r$; its supremum over θ is the optimal classical curve, and $\eta_1(r, \theta) \le \sqrt{r/2}$ is Pinsker's inequality for binary pairs. (ii) Proposition 6.4(ii) for $L = L_H$, with $q = r$, $\operatorname{tr}(L_H(\rho - \sigma)) = \|\rho - \sigma\|_1$ and $a = \lambda_{\max}(\hat{L}_H) = 1 - (2\theta_0 - 1) = 2(1 - \theta_0)$. The variance comparison is Theorem 5.3 at $\alpha = 1$, strict iff $[\sigma, L_H] \neq 0$, and $\operatorname{Var}_\sigma(L_H) = 1 - (2\theta_0 - 1)^2$. The regime of improvement: $\sqrt{2s_1^2 r} + 2(1 - \theta_0)r/3 < \sqrt{2r}$ iff $\sqrt{r} < 3\sqrt{2}\,(1 - s_1)/(2(1 - \theta_0))$. □ □

Part (ii) improves on Pinsker's $\|\rho - \sigma\|_1 \le \sqrt{2r}$ exactly when $r < 9(1 - s_1)^2/(2(1 - \theta_0)^2)$, in particular for all small $r > 0$ as soon as $s_1^2 < 1$. In (i) the middle term is the optimal classical trade-off between relative entropy and total variation [Fedotov et al., 2003] and the last inequality is Pinsker's; all of (i) is known: the reduction to a binary pair is the argument of Hiai, Ohya and Tsukada [Hiai et al., 1981], and the sharp bound is [Audenaert and Eisert, 2005, Thm. 4]. Replacing s_1^2 by $4\theta_0(1 - \theta_0)$ gives the *measured* Bernstein–Pinsker bound, which is classical [Boucheron et al., 2013]; the strict inequality $s_1^2 < 4\theta_0(1 - \theta_0)$ is the quantum layer, Proposition 6.3 at $\alpha = 1$ for the Helstrom observable, and the only place in this subsection where the spectral measure, hence Stahl's theorem,

is used. The $\alpha = 2$ member needs no inversion: (83) with $L = L_H$ reads $\|\rho-\sigma\|_1^2 \le \mathrm{Var}_{\omega_2}(\ell)\, H_2(\rho\|\sigma)$ with $\mathrm{Var}_{\omega_2}(\ell) = 4\theta_0(1-\theta_0) - I_{\mathrm{WY}}(\sigma, L_H)$.

Remark 6.3 (No quantum penalty). Let $\Phi_\alpha(q) := \sup_{\theta\in(0,1)} \eta_\alpha(q,\theta)$, the largest total variation of a binary pair with $H_\alpha^{\mathrm{cl}} \le q$ [Harremoës and Vajda, 2011, Sason and Verdú, 2016]. The argument of Corollary 6.1(i), which is that of [Hiai et al., 1981] and uses only monotonicity under measurements, runs verbatim at every $\alpha \in [\frac12, \infty]$: $T(\rho,\sigma) \le \eta_\alpha(q,\theta_0) \le \Phi_\alpha(q)$ whenever $H_\alpha(\rho\|\sigma) \le q$, while the commuting binary pairs $\sigma = \mathrm{diag}(1-\theta,\theta)$, $\rho = \mathrm{diag}(1-\theta-x, \theta+x)$ have $H_\alpha = H_\alpha^{\mathrm{cl}}$ and $T = x$, and realize $\eta_\alpha(q,\theta)$ (a continuous objective over a compact feasible set). Hence

$$\sup\{T(\rho,\sigma) : n \ge 1,\ \sigma > 0,\ H_\alpha(\rho\|\sigma) \le q\} \ = \ \Phi_\alpha(q)\,: \tag{87}$$

the optimal trace-distance versus H_α inequality is the classical one — at $\alpha = 1$ the theorem of Hiai–Ohya–Tsukada [Hiai et al., 1981, Audenaert and Eisert, 2005, Thm. 4], at $\alpha = \frac12$ the sharpness of Fuchs–van de Graaf (Theorem 6.3). This does not contradict the deflation: deflation is an observable-by-observable statement, and the extremal pairs of the universal curve are commuting, so there is nothing to deflate.

The fidelity endpoint. At $\alpha = \frac12$ the spectral measure is the Born measure and $V_{1/2}(q) = V_{1/2}^{\mathrm{meas}}(q)$ (35); for the Helstrom projector $\nu_{L_0} = \mathrm{Ber}(\theta_0)$ and the fidelity dual Corollary 3.3 is a two-point computation, which is Fuchs–van de Graaf. The argument is the original proof of Fuchs and van de Graaf [Fuchs and van de Graaf, 1999] in dual form: Alberti's formula $\mathrm{Fid}(\rho,\sigma)^2 \le \mathrm{tr}(\rho X)\,\mathrm{tr}(\sigma X^{-1})$ [Alberti, 1983] for $X = cI - L_0$ is the data-processing inequality for the fidelity followed by a classical Cauchy–Schwarz step.

Theorem 6.3 (Fuchs–van de Graaf via DRO)**.** *Let* $\mathrm{Fid} := \mathrm{Fid}(\rho,\sigma)$ *(Section 3.3.5),* $\psi := \arccos \mathrm{Fid} \in [0, \frac{\pi}{2})$ *and, for* $\theta \in [0,1]$, $\phi_\theta := \arcsin\sqrt{\theta}$. *Then*

$$T(\rho,\sigma) \ \le\ \Delta^*(\theta_0) \ \le\ \sqrt{1-\mathrm{Fid}^2}, \qquad \Delta^*(\theta) := \begin{cases} \sin(2\phi_\theta + \psi)\,\sin\psi, & \theta \le \mathrm{Fid}^2, \\ 1-\theta, & \theta \ge \mathrm{Fid}^2, \end{cases} \tag{88}$$

The second inequality is an equality iff $\theta_0 = \theta^* := \frac12(1-\sqrt{1-\mathrm{Fid}^2})$, *and for* $\theta_0 \le \mathrm{Fid}^2$

$$T(\rho,\sigma) \ \le\ \sqrt{1-\mathrm{Fid}^2}\ \cos(2(\phi_{\theta_0} - \phi_{\theta^*})). \tag{89}$$

Proof. With $q = H_{1/2}(\rho\|\sigma) = 2(1-\mathrm{Fid})$ and $L = L_0$, Corollary 3.3 reads $V_{1/2}(q) = \inf_{c>1}\{c - \mathrm{Fid}^2/[(1-\theta_0)/c + \theta_0/(c-1)]\}$, which by Lemma 4.1 is the classical fidelity DRO of $\mathrm{Ber}(\theta_0)$; so $T = \mathrm{tr}(L_0(\rho-\sigma)) \le V_{1/2}(q) - \theta_0 = \sup\{x : \mathrm{Fid}_{\mathrm{cl}}(\mathrm{Ber}(\theta_0+x), \mathrm{Ber}(\theta_0)) \ge \mathrm{Fid}\}$. The classical root fidelity of binary laws is an inner product of unit vectors, $\sqrt{(1-\theta)(1-\theta-x)} + \sqrt{\theta(\theta+x)} = \cos(\phi_{\theta+x} - \phi_\theta)$, so the constraint is $\phi_{\theta_0+x} \le \phi_{\theta_0} + \psi$: $x \le \sin^2(\phi_{\theta_0}+\psi) - \theta_0$ if $\phi_{\theta_0} + \psi \le \frac{\pi}{2}$, i.e. $\theta_0 \le \mathrm{Fid}^2$, and $x \le 1-\theta_0$ otherwise. The identity $\sin^2(\phi+\psi) - \sin^2\phi = \sin(2\phi+\psi)\sin\psi$ gives the first form of Δ^*, and expanding $\sin(\phi_\theta+\psi) = \sqrt{\theta}\,\mathrm{Fid} + \sqrt{1-\theta}\sqrt{1-\mathrm{Fid}^2}$ the second. On $\theta \le \mathrm{Fid}^2$ the argument $2\phi_\theta + \psi$ ranges over $[\psi, \pi-\psi]$, so Δ^* has the unique maximizer $2\phi_\theta + \psi = \frac{\pi}{2}$, i.e. $\theta = \sin^2(\frac{\pi}{4} - \frac{\psi}{2}) = \frac12(1-\sin\psi) = \theta^*$, with value $\sin\psi = \sqrt{1-\mathrm{Fid}^2}$; on $\theta \ge \mathrm{Fid}^2$, $1-\theta \le 1-\mathrm{Fid}^2 = \sin^2\psi < \sin\psi$. Finally $2\phi_{\theta^*} + \psi = \frac{\pi}{2}$ turns $\sin(2\phi_{\theta_0}+\psi)\sin\psi$ into $\sqrt{1-\mathrm{Fid}^2}\cos(2(\phi_{\theta_0} - \phi_{\theta^*}))$, which is (89). □ □

Here $\Delta^*(\theta) = \eta_{1/2}(2(1-\mathrm{Fid}), \theta)$ is the Hoeffding profile at $\alpha = \frac12$, which for $\theta \le \mathrm{Fid}^2$ reads $\Delta^*(\theta) = (1-\mathrm{Fid}^2)(1-2\theta) + 2\,\mathrm{Fid}\sqrt{1-\mathrm{Fid}^2}\sqrt{\theta(1-\theta)}$. The first inequality in (88) is the bound of Weber et al. [Weber et al., 2022, Thm. 1] for the Helstrom observable $2L_0 - I$; its angle form (89),

and the fact that its maximum over θ_0 is the Fuchs–van de Graaf bound, are what the theorem adds. The outer inequality is the upper half of Fuchs–van de Graaf [Fuchs and van de Graaf, 1999]; the lower half $1 - \mathrm{Fid} \le T$ is not a DRO statement. Nothing in this argument deflates: at $\alpha = \frac{1}{2}$ the $k = 2$ certificate of Theorem 6.1 uses the Born variance, and for the Helstrom projector, a two-point Born measure of maximal variance $\theta_0(1-\theta_0)$, it sharpens nothing. The lever that remains is the spectrum of σ, which constrains θ_0.

Proposition 6.5 (Spectrum-dependent Fuchs–van de Graaf)**.** *Let $p_1 \ge \cdots \ge p_n$ be the eigenvalues of σ and $\mathcal{K}_k(\sigma) := [\sum_{i>n-k} p_i,\ \sum_{i\le k} p_i]$ the k-th Ky Fan interval. Then*

$$T(\rho,\sigma) \;\le\; \max_{1\le k\le n-1}\ \max_{\theta\in\mathcal{K}_k(\sigma)} \Delta^*(\theta) \;\le\; \sqrt{1-\mathrm{Fid}^2}, \tag{90}$$

and the second inequality is strict iff $\theta^* \notin \bigcup_k \mathcal{K}_k(\sigma)$.

Proof. $\theta_0 = \mathrm{tr}(\sigma L_0)$ with L_0 a rank-k projector, and by Ky Fan's maximum principle [Bhatia, 1997, Ch. III] $\mathrm{tr}(\sigma P)$ ranges exactly over $\mathcal{K}_k(\sigma)$ as P ranges over rank-k projectors: the endpoints at the spectral projectors onto the k smallest, resp. largest, eigenvalues, the interior by continuity on the connected Grassmannian. Hence $T \le \Delta^*(\theta_0) \le \max_{\mathcal{K}_k(\sigma)} \Delta^*$ by Theorem 6.3, and the maximum over k removes the unknown rank. Since θ^* is the unique maximizer of Δ^* on $[0,1]$, the constrained maximum is $< \sqrt{1-\mathrm{Fid}^2}$ iff no Ky Fan interval contains θ^*. □ □

The refinement is classical in content: it depends on the pair only through $\mathrm{spec}\,\sigma$ and the Born weight θ_0, and noncommutativity works against it. For commuting pairs L_0 is diagonal and θ_0 is a sum of k eigenvalues of σ, so (90) holds with the finite set of k-fold partial sums in place of $\mathcal{K}_k(\sigma)$; the interval is the relaxation forced by the connectedness of the Grassmannian.

7 Discussion

The quantum DRO problem for the sandwiched Rényi family has turned out to be a classical problem in disguise, and the disguise is informative. The dual (Theorem 3.4) is the optimized certainty equivalent with a sandwiched Schatten norm in place of an $L^{\alpha'}$ norm; the classicalization (Theorem 4.1) trades that norm for the law ω_α of a single real random variable, whose transform of the right kind — stop-loss, Laplace or Stieltjes — is a sandwiched trace; the linearization (Theorem 5.1) identifies the variance of ω_α with a generalized variance in the sense of Petz, dual to a quantum Fisher metric; and the comparison with the Born measure (Proposition 6.1, Proposition 6.3) says what noncommutativity does to the worst case: it deflates it. Equality at all budgets characterizes the commuting case (apart from $\alpha = \frac{1}{2}$), while strictness at a given budget depends on whether saturation or another variance-insensitive branch is active. Locally the deflation is quantified by a variance that decreases from the Born variance to the harmonic-logarithmic one as α runs from $\frac{1}{2}$ to ∞. Pinsker and Fuchs–van de Graaf are the Helstrom-observable instances of the undeflated tier, which is why they carry no quantum penalty; the deflated tier gives the Bernstein-type refinement of Pinsker.

Several questions are left open. The sandwiched family is one of several quantum Rényi families, and the DRO problem is equally natural for Petz's quasi-entropies and for the α-z divergences [Petz, 1986, Audenaert and Datta, 2015], which carry the operational meaning in the error-exponent regime for $\alpha < 1$ [Hayashi, 2007, Audenaert et al., 2008]; whether their duals close and classicalize, and with which spectral measures, we do not know. The measured Rényi divergences are a further case, of a different flavour: their Donsker–Varadhan-type variational formulas [Berta et al., 2017]

give the DRO dual directly as an optimization over positive operators rather than over a scalar, and since the measured divergence is dominated by the sandwiched one by data processing and dominates the classical divergence of any fixed measurement, the corresponding robust value lies between V_α and the measured shadow V_α^{meas} of Section 6 — where in that window it falls, and whether it too classicalizes, is open. A further family is that of the quantum f-divergences defined through integral representations over hockey-stick divergences [Hirche and Tomamichel, 2024, Liu et al., 2025], whose DRO problem we have not studied. The size of the deflation is understood for a fixed pair, through $\mathrm{Var}_{\omega_\alpha}(\ell)/\mathrm{Var}_\sigma(L)$ (Theorem 5.3, Figure 6), but not as a function of the dimension: for a generic pair in high dimension, or for local observables of a many-body Gibbs state, we expect the ratio to be small, and quantifying this would turn the deflation from a comparison into a resource. In the Padé bracket of Remark 6.1 only the upper certificates are shown to converge; the lower principal representations also converge weakly to ω_α, and we have not pursued the convergence of the lower certificates. Everything here is finite-dimensional, as is Heinävaara's tracial measure; an infinite-dimensional classicalization, for normal states of a von Neumann algebra, would need a replacement for it. Finally, the transport-type ambiguity sets of classical DRO — Wasserstein balls — have quantum counterparts whose duals are of a different nature, and we intend to take them up separately.

Acknowledgments

This work was supported by the National Science Foundation under grants DMS-2307115 and DMS-2606221. I thank the Département de Mathématiques of the École Polytechnique Fédérale de Lausanne for its hospitality while this work was completed. The author used Claude (Anthropic) extensively in the preparation of this paper: for drafting and restructuring the text, for cross-checking proofs and references, for numerical verification of the inequalities and equality cases stated here, and for the figures and their generating code. All mathematical claims were verified by the author, who takes full responsibility for them.

Appendix A. Classical moment tools: proofs

This appendix collects the proofs that are classical in content: the classical duals (Lemma 4.1), transform uniqueness (Lemma 4.4), the hyperbolic representation of the mean functions (Lemma 5.1), the convexity-cone stratification (Proposition 6.2) and the explicit Lyapunov and Bernstein bounds (Proposition 6.4). Notation is that of the sections in which the statements appear.

Proof of Lemma 4.1. Let $q_{\max}^{\mathrm{cl}}(\nu) := H_\alpha^{\mathrm{cl}}(\delta_{\ell_{\max}}\|\nu)$. For $q < q_{\max}^{\mathrm{cl}}(\nu)$ this is Proposition 2.1 for the commutative model $(\mathcal{P}(J), \int \ell\, dP)$, i.e. the classical divergence-DRO duality for a scalar random variable: the Cressie–Read representation for $\alpha \neq 1$ ([Duchi and Namkoong, 2021, Lemma 1] for $\alpha \in (1,\infty)$, [Ben-Tal et al., 2013, Pichler and Schlotter, 2020] for $\alpha \in [\frac{1}{2}, 1)$, and [Ben-Tal and Teboulle, 2007, Ben-Tal et al., 2013, Shapiro, 2017] for general φ-divergences), the relative-entropy DRO at $\alpha = 1$ [Petersen et al., 2000, Chowdhary and Dupuis, 2013, Breuer and Csiszár, 2016, Glasserman and Xu, 2014], and the Rockafellar–Uryasev representation of the conditional value-at-risk at $\alpha = \infty$ [Rockafellar and Uryasev, 2000]. For $q \geq q_{\max}^{\mathrm{cl}}(\nu)$ both sides equal $\ell_{\max}$: the value because $\delta_{\ell_{\max}}$ is feasible, and each infimum because, with $w := \nu(\{\ell_{\max}\})$, bounding the transform from below by the contribution of the atom shows that the objective is $\geq \ell_{\max}$ for every dual variable precisely when $q \geq q_{\max}^{\mathrm{cl}}(\nu)$, while it tends to $\ell_{\max}$ as $c \to \ell_{\max}$ (resp. $t \to \infty$). For $\alpha < 1$ and $q \geq 1/(1-\alpha)$ the ball is all of $\mathcal{P}(J)$ and the positive part in the prefactor makes the formula read $\ell_{\max}$. □

Proof of Lemma 4.4. Each transform returns every moment of ν, which determine a measure on a compact interval (polynomials are dense in $C([l_1, l_n])$): for $c < -\max_i |l_i|$ one has $\int (\ell - c)_+^{\alpha'} d\nu = (-c)^{\alpha'} \sum_k \binom{\alpha'}{k} (-1/c)^k \int \ell^k d\nu$ with non-vanishing coefficients when α' is not an integer, while for integer α' differentiating α' times gives $(-1)^{\alpha'} \alpha'! \, \nu((c, \infty))$ for a.e. c; for $c > l_n$ large, $\int (c - \ell)^{-p} d\nu = c^{-p} \sum_k \binom{-p}{k} (-1/c)^k \int \ell^k d\nu$ with all coefficients non-zero; and the Laplace transform is entire with derivatives $\int \ell^k d\nu$ at $t = 0$. □

Proof of Lemma 5.1. Both sides of (66) are homogeneous of degree one in (a, b), so we may normalize $ab = 1$, i.e. $(a, b) = (e^u, e^{-u})$, and by symmetry take $u > 0$. From (60), with $\eta := 1 - \xi = (\alpha - 1)/\alpha$ and $(ab)^\eta = 1$,

$$\alpha \, m_\alpha(e^u, e^{-u}) \;=\; (\alpha - 1) \, \frac{e^{\xi u} - e^{-\xi u}}{e^{\eta u} - e^{-\eta u}} \;=\; \frac{\eta}{\xi} \cdot \frac{\sinh(\xi u)}{\sinh(\eta u)} \;=\; \frac{\Psi(\xi)}{\Psi(\eta)},$$

including $\alpha = 1$ ($\eta = 0$) by the convention $\Psi(0) = u$, which reproduces the logarithmic mean.

The integral formula exhibits Ψ as the two-sided Laplace transform of the measure $\frac{1}{2}\mathbf{1}_{[-u,u]} \, dt$: it is even, and strictly log-convex by Hölder's inequality (equality would force $e^{xt} \propto e^{yt}$ a.e. on $[-u, u]$, i.e. $x = y$). Hence $\Lambda := \log \Psi$ is even, Λ' is odd and strictly increasing, and $\Lambda' > 0$ on $(0, \infty)$. The function

$$\psi(\xi) := \Lambda(\xi) - \Lambda(1 - \xi), \qquad \psi'(\xi) = \Lambda'(\xi) + \Lambda'(1 - \xi),$$

is strictly increasing on $(0, 2)$: for $\xi \in (0, 1]$ both terms of ψ' are nonnegative and not both zero, while for $\xi \in (1, 2)$, $\psi'(\xi) = \Lambda'(\xi) - \Lambda'(\xi - 1) > 0$. Since $\alpha \, m_\alpha = \sqrt{ab} \, e^{\psi(\xi)}$ with $\xi = 1/\alpha$ decreasing in α, $\alpha \, m_\alpha$ is strictly decreasing in α, with supremum $\psi(2) = \Lambda(2) - \Lambda(1) = \log \cosh u$ (as $\Lambda(-1) = \Lambda(1)$), attained only at $\xi = 2$, where $\sqrt{ab} \cosh u = (a + b)/2$, and infimum $\psi(0^+) = \Lambda(0) - \Lambda(1) = \log(u/\sinh u)$, where $\sqrt{ab} \, u/\sinh u = ab \log(a/b)/(a - b)$. □

Proof of Proposition 6.2. (i) Both functions are absolutely monotone on $[l_1, l_n]$: for $c > l_n$, $\frac{d^j}{d\ell^j}(c - \ell)^{-p} = p(p + 1) \cdots (p + j - 1) \, (c - \ell)^{-p-j} > 0$ and $(e^{t\ell})^{(j)} = t^j e^{t\ell} \geq 0$ for every $j \geq 1$; all derivatives being nonnegative, they are $(k{+}1)$-convex for every k. (ii) Here $(k{+}1)$-convexity is the statement that the $(k{+}1)$-st distributional derivative is a nonnegative measure. Let $\chi_c^\lambda := (\ell - c)_+^\lambda / \Gamma(\lambda + 1)$, entire in λ, with $(\chi_c^\lambda)' = \chi_c^{\lambda - 1}$ and $\chi_c^{-1-j} = \delta_c^{(j)}$. Then $\frac{d^{k+1}}{d\ell^{k+1}} (\ell - c)_+^{\alpha'} = \Gamma(\alpha' + 1) \, \chi_c^{\alpha' - k - 1}$. For $k < \alpha'$ the right-hand side is the nonnegative locally integrable function $\Gamma(\alpha' + 1)\Gamma(\alpha' - k)^{-1} (\ell - c)_+^{\alpha' - k - 1}$, and at $k = \alpha' \in \mathbb{N}$ it is $\alpha'! \, \delta_c$. For $k > \alpha'$ it is $\alpha'! \, \delta_c^{(k - \alpha')}$ when $\alpha' \in \mathbb{N}$ and a finite-part distribution otherwise; neither has order zero, so neither is a measure. The obstruction sits at c, so it is seen on $[l_1, l_n]$ whenever c is interior (for $c < l_1$ and non-integer α' the sign of $\Gamma(\alpha' - k)$ decides, so $(k{+}1)$-convexity may fail there as well; for $c \geq l_n$ the function vanishes on $[l_1, l_n]$). The certificate (75) follows as stated after the proposition. □

Proof of Proposition 6.4, Lyapunov term of (iii) and Bernstein step of (ii). Write $\tilde{c} = c - \mu$ and $Y = \hat{\ell} - \tilde{c}$. Throughout, Lemma 4.1 expresses $V_\alpha(q) - \mu$ through the transform of ω_α at $\tilde{c}$, and by Lemma 6.1 any pointwise majorant of the transform majorizes the value. (iii) For $\alpha' \leq 2$, Lyapunov's inequality on the probability measure ω_α gives $\|Y_+\|_{\alpha'} \leq \|Y_+\|_2 \leq (\mathbb{E} Y^2)^{1/2} = (v + \tilde{c}^2)^{1/2}$, so, writing $\kappa := \kappa_\alpha(q)$, $V_\alpha(q) - \mu \leq \inf_{\tilde{c}} \{\tilde{c} + \kappa \sqrt{v + \tilde{c}^2}\}$. For $\kappa > 1$ the infimum is attained at $\tilde{c}^* = -\sqrt{v/(\kappa^2 - 1)}$ with value $\sqrt{v(\kappa^2 - 1)}$; intersect with the trivial bound $V_\alpha - \mu \leq a$. (ii) Under a law ν with mean μ, variance v and support in $(-\infty, l_n]$ — ω_1 or the certificate $\nu_+^{(2)}$ — the centered variable $\hat{\ell}$ has mean 0, variance v and $\hat{\ell} \leq a$, so Bennett's lemma [Bennett, 1962, Boucheron et al., 2013, §2.7] gives $\log \int e^{t\hat{\ell}} d\nu \leq (v/a^2)(e^{ta} - ta - 1) \leq vt^2/(2(1 - ta/3))$ for $0 < t < 3/a$, the second step, to Bernstein's majorant, being the series comparison $e^u - u - 1 = \sum_{k \geq 2} u^k/k! \leq \frac{u^2}{2} \sum_{k \geq 2} (u/3)^{k-2}$. The $\alpha = 1$ case of Lemma 4.1 reads $V_1^{\mathrm{cl}}(q; \nu) - \mu = \inf_{t > 0} \{(q + \log \int e^{t\hat{\ell}} d\nu)/t\}$, so the majorant gives

$V_1(q) - \mu \le \inf_{0<t<3/a}\{q/t + vt/(2(1-ta/3))\}$. With $u = ta/3 \in (0,1)$ the objective is $\frac{aq}{3u} + \frac{3v}{2a}\frac{u}{1-u}$, stationary where $\frac{1-u}{u} = \frac{3}{a}\sqrt{\frac{v}{2q}}$, and substituting this value of u gives $\frac{aq}{3} + \sqrt{\frac{qv}{2}} + \sqrt{\frac{qv}{2}} = \sqrt{2vq} + \frac{aq}{3}$, which is Bernstein's inequality [Boucheron et al., 2013]. □

References

Rudolf Ahlswede and A. Winter. Strong converse for identification via quantum channels. *IEEE Transactions on Information Theory*, 48(3):569–579, 2002. doi: 10.1109/18.985947.

Amir Ahmadi-Javid. Entropic value-at-risk: a new coherent risk measure. *Journal of Optimization Theory and Applications*, 155(3):1105–1123, 2012. doi: 10.1007/s10957-011-9968-2.

Peter M. Alberti. A note on the transition probability over C*-algebras. *Lett. Math. Phys.*, 7:25–32, 1983. doi: 10.1007/BF00398708.

Huzihiro Araki. On an inequality of Lieb and Thirring. *Letters in Mathematical Physics*, 19(2): 167–170, 1990. doi: 10.1007/bf01045887.

P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Math. Finance*, 9: 203–228, 1999. doi: 10.1111/1467-9965.00068.

Rami Atar, Kenny Chowdhary, and Paul Dupuis. Robust bounds on risk-sensitive functionals via Rényi divergence. *SIAM/ASA Journal on Uncertainty Quantification*, 3(1):18–33, 2015. doi: 10.1137/130939730.

K. M. R. Audenaert and N. Datta. α-z-relative Rényi entropies. *J. Math. Phys.*, 56:022202, 2015. doi: 10.1063/1.4906367.

K. M. R. Audenaert, Michael Nussbaum, Arleta Szkoła, and Frank Verstraete. Asymptotic error rates in quantum hypothesis testing. *Communications in Mathematical Physics*, 279(1):251–283, 2008. doi: 10.1007/s00220-008-0417-5.

Koenraad M. R. Audenaert and Jens Eisert. Continuity bounds on the quantum relative entropy. *J. Math. Phys.*, 46(10):102104, 2005. doi: 10.1063/1.2044667.

George A. Baker, Jr. and Peter Graves-Morris. *Padé Approximants*, volume 59 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2 edition, 1996. doi: 10.1017/cbo9780511530074.

Panagiotis Kl. Barkoutsos, Giacomo Nannicini, Anton Robert, Ivano Tavernelli, and Stefan Woerner. Improving variational quantum optimization using CVaR. *Quantum*, 4:256, 2020. doi: 10.22331/q-2020-04-20-256.

S. Beigi. Sandwiched Rényi divergence satisfies data processing inequality. *J. Math. Phys.*, 54: 122202, 2013. doi: 10.1063/1.4838855.

A. Ben-Tal and M. Teboulle. An old-new concept of convex risk measures: the optimized certainty equivalent. *Math. Finance*, 17:449–476, 2007. doi: 10.1111/j.1467-9965.2007.00311.x.

A. Ben-Tal, A. Ben-Israel, and M. Teboulle. Certainty equivalents and information measures: duality and extremal principles. *Journal of Mathematical Analysis and Applications*, 157(1):211–236, 1991. doi: 10.1016/0022-247x(91)90145-p.

A. Ben-Tal, Dick den Hertog, Anja De Waegenaere, Bertrand Melenberg, and Gijs Rennen. Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, 59 (2):341–357, 2013. doi: 10.1287/mnsc.1120.1641.

George Bennett. Probability inequalities for the sum of independent random variables. *Journal of the American Statistical Association*, 57(297):33–45, 1962. doi: 10.1080/01621459.1962.10482149.

M. Berta, O. Fawzi, and M. Tomamichel. On variational expressions for quantum relative entropies. *Lett. Math. Phys.*, 107:2239–2265, 2017. doi: 10.1007/s11005-017-0990-7.

Dimitris Bertsimas and Ioana Popescu. Optimal inequalities in probability theory: a convex optimization approach. *SIAM Journal on Optimization*, 15(3):780–804, 2005. doi: 10.1137/s10526 23401399903.

D. Bessis, P. Moussa, and M. Villani. Monotonic converging variational approximations to the functional integrals in quantum statistical mechanics. *J. Math. Phys.*, 16:2318–2325, 1975. doi: 10.1063/1.522463.

R. Bhatia. *Matrix Analysis*. Number 169 in Graduate Texts in Mathematics. Springer, 1997. doi: 10.1007/978-1-4612-0653-8.

Rajendra Bhatia and Chandler Davis. A Better Bound on the Variance. *Amer. Math. Monthly*, 107 (4):353–357, 2000. doi: 10.1080/00029890.2000.12005203.

Jeremiah Birrell, Paul Dupuis, Markos A. Katsoulakis, Luc Rey-Bellet, and Jie Wang. Distributional robustness and uncertainty quantification for rare events, 2019. arXiv:1911.09580.

Jeremiah Birrell, Paul Dupuis, Markos A. Katsoulakis, Luc Rey-Bellet, and Jie Wang. Variational representations and neural network estimation of Rényi divergences. *SIAM Journal on Mathematics of Data Science*, 3(4):1093–1116, 2021. doi: 10.1137/20M1368926.

Jose Blanchet and Karthyek Murthy. Quantifying distributional model risk via optimal transport. *Mathematics of Operations Research*, 44(2):565–600, 2019. doi: 10.1287/moor.2018.0936.

J. F. Bonnans and A. Shapiro. *Perturbation Analysis of Optimization Problems*. Springer Series in Operations Research. Springer, New York, 2000. doi: 10.1007/978-1-4612-1394-9.

Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, Oxford, 2013. doi: 10.1093/acprof:oso/9780199535255.001.0001.

S. L. Braunstein and C. M. Caves. Statistical distance and the geometry of quantum states. *Phys. Rev. Lett.*, 72:3439–3443, 1994. doi: 10.1103/physrevlett.72.3439.

Thomas Breuer and Imre Csiszár. Systematic stress tests with entropic plausibility constraints. *Journal of Banking & Finance*, 37(5):1552–1559, 2013. doi: 10.1016/j.jbankfin.2012.04.013.

Thomas Breuer and Imre Csiszár. Measuring distribution model risk. *Mathematical Finance*, 26(2): 395–411, 2016. doi: 10.1111/mafi.12050.

D. Bures. An extension of Kakutani's theorem on infinite product measures to the tensor product of semifinite w^*-algebras. *Trans. Amer. Math. Soc.*, 135:199–212, 1969. doi: 10.1090/s0002-9947-1969-0236719-2.

Francesco Buscemi and Gilad Gour. Quantum relative Lorenz curves. *Phys. Rev. A*, 95(1):012110, 2017. doi: 10.1103/PhysRevA.95.012110.

E. A. Carlen. Trace inequalities and quantum entropy: an introductory course. In R. Sims and D. Ueltschi, editors, *Entropy and the Quantum*, volume 529 of *Contemp. Math.*, pages 73–140. Amer. Math. Soc., 2010. doi: 10.1090/conm/529/10428.

Douglas G. Chapman and Herbert Robbins. Minimum variance estimation without regularity assumptions. *Ann. Math. Statist.*, 22:581–586, 1951. doi: 10.1214/aoms/1177729548.

Kenny Chowdhary and Paul Dupuis. Distinguishing and integrating aleatoric and epistemic variation in uncertainty quantification. *ESAIM: Mathematical Modelling and Numerical Analysis*, 47(3): 635–662, 2013. doi: 10.1051/m2an/2012038.

Matthias Christandl and Renato Renner. Reliable quantum state tomography. *Physical Review Letters*, 109(12):120403, 2012. doi: 10.1103/PhysRevLett.109.120403.

Fabien Clivaz. Stahl's theorem (aka BMV conjecture): Insights and intuition on its proof. In Marius Mantoiu, Georgi Raikov, and Rafael Tiedra de Aldecoa, editors, *Spectral Theory and Mathematical Physics*, pages 107–117, Cham, 2016. Springer International Publishing. ISBN 978-3-319-29992-1. doi: 10.1007/978-3-319-29992-1_6.

Noel Cressie and Timothy R. C. Read. Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society, Series B*, 46(3):440–464, 1984. doi: 10.1111/j.2517-6161.1984.tb01318.x.

J. P. Crouzeix. A relationship between the second derivatives of a convex function and of its conjugate. *Math. Programming*, 13(1):364–365, 1977. doi: 10.1007/BF01584350.

I. Csiszár. Information-type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungar.*, 2:299–318, 1967.

Nilanjana Datta. Min- and max-relative entropies and a new entanglement monotone. *IEEE Transactions on Information Theory*, 55(6):2816–2826, 2009. doi: 10.1109/TIT.2009.2018325.

Erick Delage and Yinyu Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):595–612, 2010. doi: 10.1287/opre.1090.0741.

Michel Denuit, Claude Lefèvre, and Moshe Shaked. The s-convex orders among real random variables, with applications. *Mathematical Inequalities & Applications*, 1(4):585–613, 1998. doi: 10.7153/mia-01-56.

M. J. Donald. On the relative entropy. *Comm. Math. Phys.*, 105:13–34, 1986. doi: 10.1007/bf01212339.

J. C. Duchi and H. Namkoong. Variance-based regularization with convex objectives. *J. Mach. Learn. Res.*, 20(68):1–55, 2019.

J. C. Duchi and H. Namkoong. Learning models with uniform performance via distributionally robust optimization. *Ann. Statist.*, 49(3):1378–1406, 2021. doi: 10.1214/20-AOS2004.

John C. Duchi, Peter W. Glynn, and Hongseok Namkoong. Statistics of Robust Optimization: A Generalized Empirical Likelihood Approach. *Math. Oper. Res.*, 46(3):946–969, 2021. doi: 10.1287/moor.2020.1085.

Paul Dupuis, Markos A. Katsoulakis, Yannis Pantazis, and Petr Plecháč. Path-space information bounds for uncertainty quantification and sensitivity analysis of stochastic dynamics. *SIAM/ASA J. Uncertain. Quantif.*, 4(1):80–111, 2016. doi: 10.1137/15M1025645.

Paul Dupuis, Markos A. Katsoulakis, Yannis Pantazis, and Luc Rey-Bellet. Sensitivity analysis for rare events based on Rényi divergence. *The Annals of Applied Probability*, 30(4):1507–1533, 2020. doi: 10.1214/19-AAP1468.

A. Eremenko. Herbert Stahl's proof of the BMV conjecture. *Sb. Math.*, 206:87–92, 2015. doi: 10.1070/sm2015v206n01abeh004447.

Philippe Faist and Renato Renner. Practical and reliable error bars in quantum tomography. *Physical Review Letters*, 117(1):010404, 2016. doi: 10.1103/PhysRevLett.117.010404.

A. A. Fedotov, P. Harremoës, and F. Topsøe. Refinements of Pinsker's inequality. *IEEE Trans. Inform. Theory*, 49(6):1491–1498, 2003. doi: 10.1109/TIT.2003.811927.

Steven T. Flammia and Yi-Kai Liu. Direct fidelity estimation from few Pauli measurements. *Physical Review Letters*, 106(23):230501, 2011. doi: 10.1103/PhysRevLett.106.230501.

R. L. Frank and E. H. Lieb. Monotonicity of a relative Rényi entropy. *J. Math. Phys.*, 54:122201, 2013. doi: 10.1063/1.4838835.

C. A. Fuchs and J. van de Graaf. Cryptographic distinguishability measures for quantum-mechanical states. *IEEE Trans. Inform. Theory*, 45(4):1216–1227, 1999. doi: 10.1109/18.761271.

Rui Gao and Anton J. Kleywegt. Distributionally robust stochastic optimization with Wasserstein distance. *Mathematics of Operations Research*, 48(2):603–655, 2023. doi: 10.1287/moor.2022.1275.

Paolo Gibilisco, Fumio Hiai, and Dénes Petz. Quantum Covariance, Quantum Fisher Information, and the Uncertainty Relations. *IEEE Trans. Inform. Theory*, 55(1):439–443, 2009. doi: 10.1109/TIT.2008.2008142.

Paul Glasserman and Xingbo Xu. Robust risk measurement and model risk. *Quantitative Finance*, 14(1):29–58, 2014. doi: 10.1080/14697688.2013.822989.

Sidney Golden. Lower bounds for the Helmholtz function. *Physical Review*, 137(4B):B1127–B1128, 1965. doi: 10.1103/physrev.137.b1127.

Gene H. Golub and Gérard Meurant. *Matrices, Moments and Quadrature with Applications*. Princeton Series in Applied Mathematics. Princeton University Press, Princeton, NJ, 2010. doi: 10.1515/9781400833887.

Jun-ya Gotoh, Michael Jong Kim, and Andrew E. B. Lim. Robust empirical optimization is almost the same as mean–variance optimization. *Operations Research Letters*, 46(4):448–452, 2018. doi: 10.1016/j.orl.2018.05.005.

Gilad Gour. Quantum Noncommutativity Uniquely Determines Relative Entropy, 2026. arXiv:2607.01712.

Konstantinos Gourgoulias, Markos A. Katsoulakis, Luc Rey-Bellet, and Jie Wang. How biased is your model? Concentration inequalities, information and model bias. *IEEE Transactions on Information Theory*, 66(5):3079–3097, 2020. doi: 10.1109/TIT.2020.2977067.

Ji Guan, Wang Fang, and Mingsheng Ying. Robustness verification of quantum classifiers. In *Computer Aided Verification (CAV 2021), Part I*, volume 12759 of *Lecture Notes in Computer Science*, pages 151–174. Springer, 2021. doi: 10.1007/978-3-030-81685-8_7.

Jeongwan Haah, Aram W. Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu. Sample-optimal tomography of quantum states. *IEEE Transactions on Information Theory*, 63(9):5628–5641, 2017. doi: 10.1109/TIT.2017.2719044.

J. M. Hammersley. On estimating restricted parameters. *J. Roy. Statist. Soc. Ser. B*, 12:192–240, 1950. doi: 10.1111/j.2517-6161.1950.tb00056.x.

Frank Hansen. Metric adjusted skew information. *Proc. Natl. Acad. Sci. U.S.A.*, 105(29):9909–9916, 2008. doi: 10.1073/pnas.0803323105.

Lars Peter Hansen and Thomas J. Sargent. *Robustness*. Princeton University Press, Princeton, NJ, 2008. doi: 10.1515/9781400829385.

Peter Harremoës and Igor Vajda. On pairs of f-divergences and their joint range. *IEEE Transactions on Information Theory*, 57(6):3230–3235, 2011. doi: 10.1109/tit.2011.2137353.

H. Hasegawa. α-divergence of the non-commutative information geometry. *Rep. Math. Phys.*, 33 (1–2):87–93, 1993. doi: 10.1016/0034-4877(93)90043-e.

Masahito Hayashi. Error exponent in asymmetric quantum hypothesis testing and its application to classical-quantum channel coding. *Physical Review A*, 76(6):062301, 2007. doi: 10.1103/physreva.76.062301.

Shengyi He and Henry Lam. Higher-order expansion and Bartlett correctability of distributionally robust optimization. *Mathematics of Operations Research*, 51(2):1538–1584, 2026. doi: 10.1287/moor.2023.0191.

Otte Heinävaara. *Tracial joint spectral measures*. PhD thesis, Princeton University, Princeton, NJ, 2024. URL https://www.its.caltech.edu/~oeh/thesis_final.pdf.

Otte Heinävaara. Tracial joint spectral measures. *Inventiones Mathematicae*, 239(2):505–526, 2025. doi: 10.1007/s00222-024-01309-6.

C. W. Helstrom. *Quantum Detection and Estimation Theory*. Academic Press, New York, 1976.

F. Hiai and Hideki Kosaki. Means for matrices and comparison of their norms. *Indiana University Mathematics Journal*, 48(3):899–936, 1999. doi: 10.1512/iumj.1999.48.1665.

F. Hiai and M. Mosonyi. Different quantum f-divergences and the reversibility of quantum operations. *Rev. Math. Phys.*, 29:1750023, 2017. doi: 10.1142/s0129055x17500234.

F. Hiai and D. Petz. *Introduction to Matrix Analysis and Applications*. Universitext. Springer, Cham, 2014. doi: 10.1007/978-3-319-04150-6.

F. Hiai, Masanori Ohya, and Makoto Tsukada. Sufficiency, KMS condition and relative entropy in von Neumann algebras. *Pacific Journal of Mathematics*, 96(1):99–109, 1981. doi: 10.2140/pjm.1981.96.99.

F. Hiai, M. Mosonyi, D. Petz, and C. Bény. Quantum f-divergences and error correction. *Rev. Math. Phys.*, 23:691–747, 2011. doi: 10.1142/s0129055x11004412.

Christoph Hirche and Marco Tomamichel. Quantum Rényi and f-divergences from integral representations. *Comm. Math. Phys.*, 405(9):208, 2024. doi: 10.1007/s00220-024-05087-3.

Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963. doi: 10.1080/01621459.1963.10500830.

A. S. Holevo. Statistical decision theory for quantum systems. *J. Multivariate Anal.*, 3(4):337–394, 1973. doi: 10.1016/0047-259X(73)90028-6.

Hsin-Yuan Huang, Richard Kueng, and John Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16(10):1050–1057, 2020. doi: 10.1038/s41567-020-0932-7.

Werner Hürlimann. Analytical Bounds for two Value-at-Risk Functionals. *ASTIN Bull.*, 32(2): 235–265, 2002. doi: 10.2143/AST.32.2.1028.

Vojkan Jakšić, Yoshiko Ogata, Yan Pautrat, and Claude-Alain Pillet. Entropic fluctuations in quantum statistical mechanics. An introduction. In *Quantum Theory from Small to Large Scales (Les Houches Summer School, Volume 95, August 2010)*. Oxford University Press, Oxford, 2012. doi: 10.1093/acprof:oso/9780199652495.003.0004.

Samuel Karlin and William J. Studden. *Tchebycheff Systems: With Applications in Analysis and Statistics*. Interscience Publishers, John Wiley & Sons, New York, 1966.

Markos A. Katsoulakis, Luc Rey-Bellet, and Jie Wang. Scalable information inequalities for uncertainty quantification. *Journal of Computational Physics*, 336:513–545, 2017. doi: 10.1016/j.jcp.2017.02.020.

Mark G. Kreĭn and Adolf A. Nudel′man. *The Markov Moment Problem and Extremal Problems*, volume 50 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI, 1977. doi: 10.1090/mmono/050.

Daniel Kuhn, Soroosh Shafiee, and Wolfram Wiesemann. Distributionally robust optimization. *Acta Numerica*, 34:579–804, 2025. doi: 10.1017/S0962492924000084.

Henry Lam. Robust sensitivity analysis for stochastic systems. *Mathematics of Operations Research*, 41(4):1248–1275, 2016. doi: 10.1287/moor.2015.0776.

A. Lesniewski and M. B. Ruskai. Monotone Riemannian metrics and relative entropy on noncommutative probability spaces. *J. Math. Phys.*, 40:5702–5724, 1999. doi: 10.1063/1.533053.

Ke Li and Yongsheng Yao. Operational Interpretation of the Sandwiched Rényi Divergence of Order 1/2 to 1 as Strong Converse Exponents. *Communications in Mathematical Physics*, 405(2):22, 2024. doi: 10.1007/s00220-024-04934-8.

E. H. Lieb and Walter E. Thirring. Inequalities for the moments of the eigenvalues of the Schrödinger Hamiltonian and their relation to Sobolev inequalities. In *Studies in Mathematical Physics: Essays in Honor of Valentine Bargmann*, pages 269–303. Princeton University Press, Princeton, NJ, 1976. doi: 10.1515/9781400868940-014.

Göran Lindblad. Completely positive maps and entropy inequalities. *Communications in Mathematical Physics*, 40(2):147–151, 1975. doi: 10.1007/BF01609396.

Po-Chieh Liu, Christoph Hirche, and Hao-Chung Cheng. Layer Cake Representations for Quantum Divergences, 2025. arXiv:2507.07065.

Sirui Lu, Lu-Ming Duan, and Dong-Ling Deng. Quantum adversarial machine learning. *Physical Review Research*, 2(3):033212, 2020. doi: 10.1103/PhysRevResearch.2.033212.

Shunlong Luo. Wigner–Yanase skew information and uncertainty relations. *Phys. Rev. Lett.*, 91 (18):180403, 2003. doi: 10.1103/PhysRevLett.91.180403.

Peyman Mohajerin Esfahani and Daniel Kuhn. Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations. *Mathematical Programming*, 171(1–2):115–166, 2018. doi: 10.1007/s10107-017-1172-1.

M. Mosonyi and T. Ogawa. Quantum hypothesis testing and the operational interpretation of the quantum Rényi relative entropies. *Comm. Math. Phys.*, 334(3):1617–1648, 2015. doi: 10.1007/s00220-014-2248-x.

M. Müller-Lennert, F. Dupuis, O. Szehr, S. Fehr, and M. Tomamichel. On quantum Rényi entropies: a new generalization and some properties. *J. Math. Phys.*, 54:122203, 2013. doi: 10.1063/1.4838856.

Michael Nussbaum and Arleta Szkoła. The Chernoff lower bound for symmetric quantum hypothesis testing. *Ann. Statist.*, 37(2):1040–1057, 2009. doi: 10.1214/08-AOS593.

Ryan O'Donnell and John Wright. Efficient quantum tomography. In *Proceedings of the 48th Annual ACM Symposium on Theory of Computing (STOC 2016)*, pages 899–912, 2016. doi: 10.1145/2897518.2897544.

Masanori Ohya and Dénes Petz. *Quantum Entropy and Its Use*. Texts and Monographs in Physics. Springer, Berlin, 2 edition, 2004. doi: 10.1007/978-3-642-57997-4. First edition 1993.

Ian R. Petersen, Matthew R. James, and Paul Dupuis. Minimax optimal control of stochastic uncertain systems with relative entropy constraints. *IEEE Transactions on Automatic Control*, 45(3):398–412, 2000. doi: 10.1109/9.847720.

D. Petz. Quasi-entropies for finite quantum systems. *Rep. Math. Phys.*, 23:57–65, 1986. doi: 10.1016/0034-4877(86)90067-4.

D. Petz. Monotone metrics on matrix spaces. *Linear Algebra Appl.*, 244:81–96, 1996. doi: 10.1016/0024-3795(94)00211-8.

D. Petz and G. Toth. The Bogoliubov inner product in quantum statistics. *Lett. Math. Phys.*, 27:205–216, 1993. doi: 10.1007/bf00739578.

Dénes Petz. A variational expression for the relative entropy. *Comm. Math. Phys.*, 114(2):345–349, 1988. doi: 10.1007/BF01225040.

Dénes Petz. Covariance and Fisher information in quantum mechanics. *J. Phys. A*, 35(4):929–939, 2002. doi: 10.1088/0305-4470/35/4/305.

Georg Pflug and David Wozabal. Ambiguity in portfolio selection. *Quantitative Finance*, 7(4):435–442, 2007. doi: 10.1080/14697680701455410.

Alois Pichler and Ruben Schlotter. Entropy based risk measures. *European Journal of Operational Research*, 285(1):223–236, 2020. doi: 10.1016/j.ejor.2019.01.016.

M. S. Pinsker. *Information and Information Stability of Random Variables and Processes*. Holden-Day, San Francisco, 1964.

Yixian Qiu, Josep Lumbreras, Xiufan Li, and Patrick Rebentrost. Quantum Tilted Loss in Variational Optimization: Theory and Applications, 2026. arXiv:2605.02850.

Hamed Rahimian and Sanjay Mehrotra. Frameworks and results in distributionally robust optimization. *Open J. Math. Optim.*, 3:Art. 4, 85 pp., 2022. doi: 10.5802/ojmo.15.

Alfréd Rényi. On measures of entropy and information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1*, pages 547–561, Berkeley, CA, 1961. University of California Press.

Luc Rey-Bellet. A Stieltjes representation of order p for $\mathrm{tr}[(A+tB)^{-p}]$: Heinävaara's measure, B-splines, and Stahl's theorem, 2026. arXiv:2611.03208.

R. T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *J. Risk*, 2:21–42, 2000. doi: 10.21314/jor.2000.038.

R. Tyrrell Rockafellar. *Convex Analysis*. Number 28 in Princeton Mathematical Series. Princeton University Press, Princeton, NJ, 1970. doi: 10.1515/9781400873173.

Domingos S. P. Salazar. Universal thermodynamic uncertainty relation for quantum f divergences. *Physical Review E*, 113:064150, 2026. doi: 10.1103/n2bt-jcm6.

Igal Sason and Sergio Verdú. f-divergence inequalities. *IEEE Transactions on Information Theory*, 62(11):5973–6006, 2016. doi: 10.1109/tit.2016.2603151.

Herbert Scarf. A min-max solution of an inventory problem. In K. J. Arrow, S. Karlin, and H. Scarf, editors, *Studies in the Mathematical Theory of Inventory and Production*, pages 201–209. Stanford University Press, Stanford, CA, 1958.

Lukas Schmitt. Relative entropy representations via tracial joint spectral measures, 2026. arXiv:2609.12547.

Moshe Shaked and J. George Shanthikumar. *Stochastic Orders*. Springer Series in Statistics. Springer, New York, 2007. doi: 10.1007/978-0-387-34675-5.

Alexander Shapiro. Distributionally robust stochastic programming. *SIAM Journal on Optimization*, 27(4):2258–2275, 2017. doi: 10.1137/16M1058297.

H. R. Stahl. Proof of the BMV conjecture. *Acta Math.*, 211:255–290, 2013. doi: 10.1007/s11511-013-0104-z.

Kaito Takahashi and Akio Fujiwara. Information geometry of sandwiched Rényi α-divergence. *Journal of Physics A: Mathematical and Theoretical*, 50(16):165301, 2017. doi: 10.1088/1751-8121/aa6326.

Kristan Temme, Michael J. Kastoryano, Mary Beth Ruskai, Michael M. Wolf, and Frank Verstraete. The χ^2-divergence and mixing times of quantum Markov processes. *J. Math. Phys.*, 51:122201, 2010. doi: 10.1063/1.3511335.

Colin J. Thompson. Inequality with applications in statistical mechanics. *Journal of Mathematical Physics*, 6(11):1812–1813, 1965. doi: 10.1063/1.1704727.

M. Tomamichel. *Quantum Information Processing with Finite Resources*, volume 5 of *SpringerBriefs Math. Phys.* Springer, 2016. doi: 10.1007/978-3-319-21891-5.

Joel A. Tropp. An introduction to matrix concentration inequalities. *Foundations and Trends in Machine Learning*, 8(1–2):1–230, 2015. doi: 10.1561/2200000048.

A. Uhlmann. The "transition probability" in the state space of a $*$-algebra. *Rep. Math. Phys.*, 9: 273–279, 1976. doi: 10.1016/0034-4877(76)90060-4.

A. Uhlmann. Relative entropy and the Wigner-Yanase-Dyson-Lieb concavity in an interpolation theory. *Comm. Math. Phys.*, 54(1):21–32, 1977. doi: 10.1007/BF01609834.

Tim van Erven and Peter Harremoës. Rényi divergence and Kullback–Leibler divergence. *IEEE Transactions on Information Theory*, 60(7):3797–3820, 2014. doi: 10.1109/TIT.2014.2320500.

Igor G. Vladimirov, Ian R. Petersen, and Matthew R. James. Risk-sensitive performance criteria and robustness of quantum systems with a relative entropy description of state uncertainty, 2018. arXiv:1802.00250.

Maurice Weber, Nana Liu, Bo Li, Ce Zhang, and Zhikuan Zhao. Optimal provable robustness of quantum classification via quantum hypothesis testing. *npj Quantum Information*, 7:76, 2021. doi: 10.1038/s41534-021-00410-5.

Maurice Weber, Abhinav Anand, Alba Cervera-Lierta, Jakob S. Kottmann, Thi Ha Kyaw, Bo Li, Alán Aspuru-Guzik, Ce Zhang, and Zhikuan Zhao. Toward reliability in the NISQ era: Robust interval guarantee for quantum measurements on approximate states. *Phys. Rev. Research*, 4(3): 033217, 2022. doi: 10.1103/PhysRevResearch.4.033217.

Wolfram Wiesemann, Daniel Kuhn, and Melvyn Sim. Distributionally robust convex optimization. *Operations Research*, 62(6):1358–1376, 2014. doi: 10.1287/opre.2014.1314.

E. P. Wigner and M. M. Yanase. Information contents of distributions. *Proc. Natl. Acad. Sci. USA*, 49:910–918, 1963. doi: 10.1073/pnas.49.6.910.

M. M. Wilde. Quantum Fisher information matrices from Rényi relative entropies, 2025. arXiv:2510.02218.

M. M. Wilde, A. Winter, and D. Yang. Strong converse for the classical capacity of entanglement-breaking and Hadamard channels via a sandwiched Rényi relative entropy. *Comm. Math. Phys.*, 331:593–622, 2014. doi: 10.1007/s00220-014-2122-x.

Naoki Yamamoto and Luc Bouten. Quantum Risk-Sensitive Estimation and Robustness. *IEEE Trans. Automat. Control*, 54(1):92–107, 2009. doi: 10.1109/TAC.2008.2009571.

Meiyi Zhu and Osvaldo Simeone. Risk-averse Decision Making via Quantum Measurement Design, 2026. arXiv:2609.10482.

Luc Rey-Bellet
Department of Mathematics and Statistics, University of Massachusetts Amherst,
710 N. Pleasant St., Amherst, MA 01003, USA
E-mail address: luc@cns.umass.edu