A generalization of Costa's Entropy Power Inequality

Luca Tamanini

Abstract—Aim of this short note is to study Shannon's entropy power along entropic interpolations, thus generalizing Costa's concavity theorem. We shall provide two proofs of independent interest: the former by Γ-calculus, hence applicable to more abstract frameworks; the latter with an explicit remainder term, reminiscent of [28], allowing us to characterize the case of equality.

Index Terms—Entropy, information theory, entropy power, Schrödinger problem, heat equation

I. INTRODUCTION AND STATEMENT OF THE RESULT

GIVEN a random variable X with density u in $\mathbb{R}^n$, Shannon's entropy and entropy power are respectively defined as

$$\mathcal{H}(X) = \mathcal{H}(u) := -\int_{\mathbb{R}^n} u \log u \, d\mathcal{L}^n,$$
$$\mathcal{N}(X) = \mathcal{N}(u) := \exp\left(\frac{2}{n}\mathcal{H}(u)\right), \quad \text{(I.1)}$$

where $\mathcal{L}^n$ denotes the n-dimensional Lebesgue measure. If Y is a second random variable, then the Entropy Power Inequality (EPI) states that $\mathcal{N}(X + Y) \geq \mathcal{N}(X) + \mathcal{N}(Y)$, see [24], [25]. As a refinement, in the case $Y = \sqrt{t}Z$ with Z a standard normal Costa [8] proved that Shannon's entropy power is concave along $X + \sqrt{t}Z$. In the language of semigroups [4], this means that $\mathcal{N}$ is concave along the heat flow, namely: if u is a non-negative probability density on $\mathbb{R}^n$ and P_t is the semigroup associated to the heat equation, i.e. $\mathsf{P}_t u = u_t$ if and only if $\frac{\partial}{\partial t} u_t = \Delta u_t$ and $u|_{t=0} = u$, then

$$\frac{d^2}{dt^2} \mathcal{N}(\mathsf{P}_t u) \leq 0, \qquad \forall t > 0. \quad \text{(I.2)}$$

In fact, inequality is strict for all $t > 0$ unless u is an isotropic Gaussian distribution (in which case, equality holds for all $t > 0$). The proof, originally quite involved, was eventually simplified in [9], [10], [28]. This result plays an important role in information theory, as it strengthens the EPI (see [23] for an exhaustive list of references), and it is also useful in connection to some functional inequalities, e.g. the dimensional logarithmic Sobolev inequality, as pointed out in [1, Chapter 10].

The proof of (I.2) relies on De Bruijn's identity $\frac{d}{dt}\mathcal{H}(\mathsf{P}_t u) = \mathcal{I}(\mathsf{P}_t u)$ relating Shannon's entropy and Fisher information, the latter being defined as

$$\mathcal{I}(u) := \int_{\mathbb{R}^n} \frac{|\nabla u|^2}{u} \, d\mathcal{L}^n = \int_{\mathbb{R}^n} |\nabla \log u|^2 u \, d\mathcal{L}^n.$$

An interpolation problem strictly related with the heat semigroup is the so-called Schrödinger system, which reads as follows: given two probabilty measures $\mu = u\mathcal{L}^n$, $\nu = v\mathcal{L}^n$ and a parameter $T > 0$, find two non-negative Borel functions f^T, g^T (also called "decomposition") such that

$$u = f^T \mathsf{P}_T g^T, \qquad v = g^T \mathsf{P}_T f^T. \quad \text{(I.3)}$$

If this system is solvable, then $\rho_t^T := \mathsf{P}_t f^T \mathsf{P}_{T-t} g^T$ is a probability density which interpolates between u at time $t = 0$ and v at time $t = T$ and we will eventually refer either to it or to $\mu_t^T := \rho_t^T \mathcal{L}^n$ as "T-entropic interpolation". It is worth mentioning that the heat flow is a particular entropic interpolation: indeed, if $\mu = u\mathcal{L}^n$ and $\nu = \mathsf{P}_T u\mathcal{L}^n$ (in the sequel, with a slight abuse of notation we will write $\nu = \mathsf{P}_T\mu$ for sake of brevity), then (I.3) is trivially solved by $f^T = u$ and $g^T = 1$.

From a physical point of view (see [16] for a detailed discussion and [6] for a more recent insight), μ and ν can be thought of as probability distributions of a cloud of independent Brownian particles observed at two different times and the entropic interpolation is the most-likely evolution between them. Indeed, in a quite surprising way [18], if one maximizes the relative entropy

$$\mathsf{Q} \mapsto \mathcal{H}(\mathsf{Q} \,|\, \mathsf{R}) := -\int_\Omega \log\left(\frac{d\mathsf{Q}}{d\mathsf{R}}\right)(\omega) d\mathsf{Q}(\omega)$$

among all probability measures Q on $\Omega := C([0, T], \mathbb{R}^n)$ under the marginal constraints $(X_0)_\sharp \mathsf{Q} = \mu$ and $(X_T)_\sharp \mathsf{Q} = \nu$ (where $X_t : \Omega \to \mathbb{R}^n$, $X_t(\omega) := \omega_t$, is the canonical process and R is the law of the reversible Brownian motion on $\mathbb{R}^n$, i.e. the Brownian motion with $\mathcal{L}^n$ as initial distribution), then the marginal flow of the (unique) maximizer is exactly μ_t^T. This explains why the interpolation μ_t^T is called "entropic". Moreover, up to reparametrization T can be interpreted either as a diffusion parameter or (as in the present paper) as the time interval between the two observations. In the former case, the link between Schrödinger problem and optimal transport appears naturally by large deviations theory, whereas in the latter the convergence of the T-entropic interpolation towards the heat flow as $T \to \infty$ is rather easy to guess. With this

physical interpretation in mind, we recognize the following quantity

$$\mathscr{E}_T(\mu,\nu) := \frac{1}{2}\int_{\mathbb{R}^n} |v_t^T|^2 \,\mathrm{d}\mu_t^T - \frac{1}{2}\mathfrak{I}(\mu_t^T)$$

as the total energy of the system and by the conservation of energy principle it is not surprising that $\mathscr{E}_T(\mu,\nu)$ does not depend on t, although the right-hand side might a priori do (cf. [15, Lemma 3.2] for a rigorous proof). In the definition above

$$v_t^T := \nabla \log \mathsf{P}_{T-t}g^T - \nabla \log \mathsf{P}_t f^T$$

is the velocity field driving the T-entropic interpolation, since ρ_t^T and v_t^T are linked together by the continuity equation, namely $\partial_t \rho_t^T + \mathrm{div}(v_t^T \rho_t^T) = 0$, as proved for instance in [12, Proposition 4.3] in a very general framework.

In this paper we show that if we look at the entropy power $\mathcal{N}$ defined in (I.1) along the entropic interpolation $(\rho_t^T)_{t\in[0,T]}$ rather than along the heat flow on $[0,T]$, then a generalization of Costa's EPI (I.2) involving $\mathcal{N}$, $\mathfrak{I}$ and $\mathscr{E}_T$ can be deduced, at least if $(\rho_t^T)_{t\in[0,T]}$ interpolates between two suitable measures. For a more precise statement, let us first define $\mathcal{S}'$ as the space of L^∞ functions with bounded support and the Schwartz-like space

$$\mathcal{S}'' := \left\{ \begin{array}{l} f \in L^\infty \cap C^\infty(\mathbb{R}^n): \\ \|x^\alpha D^\beta f\|_\infty < \infty, \forall \alpha \in \mathbb{N}^n, \beta \in \mathbb{N}^n \setminus \{0_n\} \\ |\log f(x)| \le C(1+|x|^2) \text{ for some } C > 0 \end{array} \right\},$$

where 0_n is the null n-tuple; relying on these two spaces, let us introduce the class Υ_T of *regular* constraints for (I.3) as follows

$$\Upsilon_T := \left\{ \begin{array}{l} (\mu,\nu): \mu,\nu \in \mathscr{P}(\mathbb{R}^n) \text{ and} \\ \exists f^T, g^T \in \mathcal{S}' \cup \mathcal{S}'' \text{ solving (I.3)} \end{array} \right\}.$$

By [12, Proposition 2.1] we know that all couples (μ,ν) of absolutely continuous measures with bounded densities and supports belong to Υ_T. Furthermore, also $(u\mathcal{L}^n, \mathsf{P}_T u\mathcal{L}^n) \in \Upsilon_T$ for any $u \in \mathcal{S}' \cup \mathcal{S}''$, since in the associated decomposition $g^T \equiv 1$ and constant functions belong to $\mathcal{S}''$: this is the reason behind the choice of excluding $\beta = 0_n$ (but still asking for $f \in L^\infty$) in the definition of $\mathcal{S}''$. As a consequence of this simple fact, any statement valid for all $(\mu,\nu) \in \Upsilon_T$ implies as a byproduct a particular statement for the heat flow: this is the case of Theorem I.1 below, where Costa's EPI (I.2) appears as a particular case of its "entropic" version (I.4).

After this preamble, we can finally state our main results.

Theorem I.1. *Let $(\mu,\nu) \in \Upsilon_T$ for some $T > 0$ and denote by $(\rho_t^T)_{t\in[0,T]}$ the T-entropic interpolation between μ and ν. Then $t \mapsto \mathcal{N}(\rho_t^T)$ belongs to $C([0,T]) \cap C^2((0,T))$ and it holds*

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathcal{N}(\rho_t^T) \le \frac{4}{n^2}\mathcal{N}(\rho_t^T)\mathfrak{I}(\rho_t^T)\mathscr{E}_T(\mu,\nu). \qquad \text{(I.4)}$$

If $\nu = \mathsf{P}_T\mu$, then (I.4) reduces to Costa's inequality (I.2).

As a consequence of the proof provided in Section III, we can also characterize the equality case in (I.4). Roughly speaking, it always reduces to the equality case in (I.2), so

that in particular inequality in (I.4) is always strict along non-trivial entropic interpolations.

Theorem I.2. *With the same assumptions and notations as in Theorem I.1, there exists $t \in (0,T)$ where (I.4) holds with equality if and only if either $\mu = u\mathcal{L}^n$, $\nu = \mathsf{P}_T u\mathcal{L}^n$ or $\nu = u\mathcal{L}^n$, $\mu = \mathsf{P}_T u\mathcal{L}^n$ for some isotropic Gaussian distribution u.*

As pointed out by Villani [28], the concavity of Shannon's entropy power along the heat flow can also be deduced by relying on the so-called Γ-calculus, introduced by Bakry and Émery [5] in the study of hypercontractive diffusions, although this approach does not allow to obtain a precise error term. This means that Costa's result holds not only in the Euclidean setting but also on Riemannian manifolds with non-negative Ricci curvature and even more generally (and with suitable modifications) on Riemannian manifolds with Ricci curvature bounded from below by some constant $K \in \mathbb{R}$. In this case (I.2) becomes

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathcal{N}(\mathsf{P}_t u) \le -\frac{4K}{n}\mathcal{N}(\mathsf{P}_t u)\mathfrak{I}(\mathsf{P}_t u), \qquad \forall t > 0, \quad \text{(I.5)}$$

as recently proved in [19]. In a completely analogous fashion, if we move from the Euclidean to the Riemannian framework, Theorem I.1 reads as follows.

Theorem I.3. *Let (M,g) be an m-dimensional smooth, connected and complete Riemannian manifold without boundary, $V \in C^2(M)$ and $\mathfrak{m} = e^{-V}\mathrm{vol}$, where vol is the volume measure. Assume that for some $K \in \mathbb{R}$ and $n \ge m$ the Bakry-Émery Ricci tensor $\mathrm{Ric}_{V,n}$ satisfies the lower bound*

$$\mathrm{Ric}_{V,n} := \mathrm{Ric}_g + \mathrm{Hess}(V) - \frac{\nabla V \otimes \nabla V}{n-m} \ge Kg. \quad \text{(I.6)}$$

Let $\mu,\nu \ll \mathfrak{m}$ be probability measures with bounded densities and supports and denote by $(\rho_t^T)_{t\in[0,T]}$ the T-entropic interpolation between them. Then $t \mapsto \mathcal{N}(\rho_t^T)$ belongs to $C([0,T]) \cap C^2((0,T))$ and it holds

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathcal{N}(\rho_t^T) \le \frac{4}{n^2}\mathcal{N}(\rho_t^T)\mathfrak{I}(\rho_t^T)\mathscr{E}_T(\mu,\nu)$$
$$- \frac{2K}{n}\mathcal{N}(\rho_t^T)\left(\int_M |v_t^T|^2\rho_t^T\,\mathrm{d}\mathfrak{m} + \mathfrak{I}(\rho_t^T)\right). \quad \text{(I.7)}$$

If M is compact and $\nu = \mathsf{P}_T\mu$, then (I.7) reduces to (I.5).

Of course, this change of framework needs some remarks. Both in (I.5) and (I.7) it is understood that the reference measure in the definition of $\mathcal{H}$, $\mathcal{N}$, and $\mathfrak{I}$ is no longer $\mathcal{L}^n$ but $\mathfrak{m}$. As concerns the semigroup P_t, it denotes the diffusion semigroup associated with the Witten Laplacian $\mathsf{L} = \Delta_g - \nabla V \cdot \nabla$, where Δ_g is the Laplace-Beltrami operator: this means that $\mathsf{P}_t u$ solves $\partial_t \mathsf{P}_t u = \mathsf{L}\mathsf{P}_t u$. It is such a semigroup that has to be considered in (I.3), in the definition of the entropic interpolation (ρ_t^T) as well as in (I.5). The reader who is not familiar with the language of Markov semigroups is addressed e.g. to [4].

As regards the link between (I.5) and (I.7), it is still formally true that the former is a particular case of the

latter, as we shall discuss in Section II-A, but a rigorous proof is technical without compactness assumption. Already (I.5) requires more effort than (I.2). The reason preventing us from saying that, in full generality, (I.7) reduces to (I.5) when $\nu = \mathsf{P}_T\mu$ is the fact that (I.7) will be proven under a boundedness assumption on the supports of μ, ν, whereas the support of $\mathsf{P}_T\mu$ is the whole manifold: thus $\nu = \mathsf{P}_T\mu$ is never satisfied, unless M is compact.

In the rest of the paper we shall give two different proofs of Theorem I.1. In Section II we present a first abstract argument based on Γ-calculus, which proves Theorems I.1 and I.3 at the same time. In Section III we provide a second (algebraic) proof of Theorem I.1, whence Theorem I.2 immediately follows.

II. PROOF BY Γ-CALCULUS

Otto and Γ-calculus are powerful tools: the former allows to obtain in a rather easy way heuristic explanations for technical statements on the Wasserstein space; the latter is an abstract formalism based on semigroup theory, which fits well to diffusions in both the Euclidean and Riemannian setting. For this reason in Section II-A we first provide a heuristics for Theorems I.1 and I.3 to hold, while Section II-B is devoted to the real proof by Γ-calculus.

A. Heuristics

After Otto's seminal work [22], it is well established that a formal Riemannian structure is associated with the Wasserstein space $(\mathscr{P}_2(M), W_2)$. This means that we can treat $\mathcal{H}$ as a smooth function defined on a manifold (or, in an even simpler way, on $\mathbb{R}^n$) and $(\mu_t^T)_{t \in [0,T]}$ as a smooth trajectory on it. Furthermore, it is also well known that several PDEs on M can be lifted to gradient flow equations on $\mathscr{P}_2(M)$ w.r.t. the Wasserstein metric of suitable functionals: this is the case of the heat flow, which reads as the gradient flow of $-\mathcal{H}$, namely $\dot{\mu}_t = \nabla\mathcal{H}(\mu_t)$ where $\mu_t := \mathsf{P}_t\mu$, $\mu \in \mathscr{P}_2(M)$. Since in $(\mathscr{P}_2(M), W_2)$ it is more common to work with measures rather than the corresponding densities w.r.t. $\mathfrak{m}$, with a slight abuse we keep the same notations introduced before, e.g. $\mathcal{H}(\mu)$ denotes $\mathcal{H}(u)$ provided $\mu = u\mathfrak{m}$; analogously for $\mathcal{N}$ and $\mathcal{I}$.

As (I.7) is a statement on the second derivative of $t \mapsto \mathcal{N}(t) := \mathcal{N}(\mu_t^T)$, let us differentiate it twice. The first derivative reads as

$$\mathcal{N}'(t) = \mathcal{N}(t) \cdot \frac{2}{n}\langle\nabla\mathcal{H}(\mu_t^T), \dot{\mu}_t^T\rangle,$$

so that the second derivative is given by

$$\frac{\mathcal{N}''(t)}{\mathcal{N}(t)} = \frac{4}{n^2}\big(\langle\nabla\mathcal{H}(\mu_t^T), \dot{\mu}_t^T\rangle\big)^2 + \frac{2}{n}\mathrm{Hess}(\mathcal{H})(\dot{\mu}_t^T, \dot{\mu}_t^T)$$
$$+ \frac{2}{n}\langle\nabla\mathcal{H}(\mu_t^T), \ddot{\mu}_t^T\rangle.$$

A look at the dynamical aspects of entropic interpolations is required in order to move forward; more precisely, the "acceleration" of $t \mapsto \mu_t^T$ (or, in more geometric terms,

the covariant derivative of $t \mapsto \dot{\mu}_t^T$ along $t \mapsto \mu_t^T$) has to be determined. The desired information is provided by the following Newton's law (see [6])

$$\ddot{\mu}_t^T = \frac{1}{2}\nabla|\nabla\mathcal{H}(\mu_t^T)|^2, \qquad (\text{II.1})$$

so that the previous identity becomes

$$\frac{\mathcal{N}''(t)}{\mathcal{N}(t)} = \Big(\frac{4}{n^2}\big(\langle\nabla\mathcal{H}(\mu_t^T), \dot{\mu}_t^T\rangle\big)^2 + \frac{2}{n}\mathrm{Hess}(\mathcal{H})(\dot{\mu}_t^T, \dot{\mu}_t^T)$$
$$+ \frac{2}{n}\mathrm{Hess}(\mathcal{H})(\nabla\mathcal{H}(\mu_t^T), \nabla\mathcal{H}(\mu_t^T)).$$

Now we rely on the geometric structure of $(\mathscr{P}_2(M), W_2)$ and, more specifically, on the role played by the curvature-dimension condition (I.6) in connection with the Boltzmann entropy $-\mathcal{H}$. After the seminal works [26], [27], [11] it is well known that (I.6) is equivalent to the (K, n)-convexity of $-\mathcal{H}$. Let us recall that a functional $\mathcal{F} : \mathscr{P}_2(M) \to \mathbb{R}$ is said to be (K, n)-convex provided

$$\mathrm{Hess}(\mathcal{F}) \geq K\mathrm{Id} + \frac{1}{n}\nabla\mathcal{F} \otimes \nabla\mathcal{F}. \qquad (\text{II.2})$$

This information turns into an upper bound on $\mathrm{Hess}(\mathcal{H})$, whence

$$\frac{\mathcal{N}''(t)}{\mathcal{N}(t)} \leq \frac{2}{n^2}\big(\langle\nabla\mathcal{H}(\mu_t^T), \dot{\mu}_t^T\rangle\big)^2 - \frac{2}{n^2}|\nabla\mathcal{H}(\mu_t^T)|^4$$
$$- \frac{2K}{n}\big(|\dot{\mu}_t^T|^2 + |\nabla\mathcal{H}(\mu_t^T)|^2\big).$$

It we further note that $|\langle\nabla\mathcal{H}(\mu_t^T), \dot{\mu}_t^T\rangle| \leq |\nabla\mathcal{H}(\mu_t^T)||\dot{\mu}_t^T|$ by Cauchy-Schwarz inequality and use De Bruijn's identity, which reads as $|\nabla\mathcal{H}|^2 = \mathcal{I}$ in Otto's formalism (namely the Fisher information is nothing but the squared norm of the gradient of Shannon's entropy), we obtain

$$\frac{\mathcal{N}''(t)}{\mathcal{N}(t)} \leq \frac{2}{n^2}\mathcal{I}(\mu_t^T)\big(|\dot{\mu}_t^T|^2 - \mathcal{I}(\mu_t^T)\big)^2 - \frac{2K}{n}\big(|\dot{\mu}_t^T|^2 + \mathcal{I}(\mu_t^T)\big).$$

It only remains to remark that the speed of a curve (μ_t) solving the continuity equation with drift v_t can be expressed as $|\dot{\mu}_t| = \|v_t\|_{L^2(\mu_t)}$, so that we recognize $\int_M |v_t^T|^2 \, \mathrm{d}\mu_t^T$ in (I.7) as $|\dot{\mu}_t^T|^2$ and, as a consequence, $|\dot{\mu}_t^T|^2 - \mathcal{I}(\mu_t^T)$ as $2\mathscr{E}_T(\mu, \nu)$.

The fact that (I.5) is a particular case of (I.7) is not surprising, since, as already mentioned, the heat flow is a particular case of entropic interpolation. If $\mu = u\mathfrak{m}$ and $\nu = \mathsf{P}_T u\mathfrak{m}$, then the Schrödinger system (I.3) is solved by $f^T = u$ and $g^T = 1$, whence

$$|\dot{\mu}_t^T|^2 = \mathcal{I}(\mu_t^T) = \mathcal{I}(\mathsf{P}_t u) \qquad \text{and} \qquad \mathscr{E}_T(\mu, \nu) = 0.$$

Plugging these identities into (I.7) yields (I.5).

Remark II.1. The heuristics described in this section actually applies to a wider class of variational problems, known as *generalized Schrödinger problems* and introduced in [14]. Given $\mathcal{F} : \mathscr{P}_2(M) \to \mathbb{R} \cup \{+\infty\}$ and $\mu, \nu \in \mathscr{P}_2(M)$, they read as the following action minimizing problem

$$\inf_{(\nu_t)_{t \in [0,T]}} \int_0^T \Big(\frac{1}{2}|\dot{\nu}_t|^2 + \frac{1}{2}|\nabla\mathcal{F}|^2(\nu_t)\Big)\mathrm{d}t,$$

where the infimum runs over all paths $(\nu_t)_{t\in[0,T]}$ joining μ to ν, and they indeed generalize the dynamic formulation of the entropic cost à la Benamou-Brenier (see [13]). A rigorous investigation of these problems has been started in [21] (the setting is even more general, since $\mathscr{P}_2(M)$ is replaced by an abstract metric space), but the possibility to turn the heuristic arguments of this section into rigorous ones for this class of problems seems, at present, out of reach: the main difficulty is the lack of information on the regularity of the minimizers, more precisely the lack of a Newton's law, which is instead available for the Schrödinger problem, as we are about to see in the next section. ∎

B. Proof of the result

Let us first discuss Theorem I.1. In order to turn the heuristic approach presented above into a precise one, rigorous counterparts of (II.1) and (II.2) are required. As concerns the former, we shall rely on the following formulas for the first and second derivatives of the entropy along entropic interpolations, computed for the first time in [17]:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{H}(\rho_t^T) = -\int_{\mathbb{R}^n}\langle\nabla\rho_t^T,\nabla\vartheta_t^T\rangle\,\mathrm{d}\mathcal{L}^n, \tag{II.3a}$$

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathcal{H}(\rho_t^T) = -\int_{\mathbb{R}^n}\Big(\Gamma_2(\vartheta_t^T)+\Gamma_2(\log\rho_t^T)\Big)\rho_t^T\,\mathrm{d}\mathcal{L}^n, \tag{II.3b}$$

where $\vartheta_t^T := \log\mathsf{P}_{T-t}g^T - \log\mathsf{P}_t f^T$ and Γ_2 is the iterated carré du champ operator defined as

$$\Gamma_2(\phi) := \frac{1}{2}\Delta|\nabla\phi|^2 - \langle\nabla\phi,\nabla\Delta\phi\rangle, \qquad \forall\phi\in C_c^\infty(\mathbb{R}^n).$$

As $\mathsf{P}_t\phi$ is the convolution of ϕ with the n-dimensional Gaussian density having 0_n mean and $2t\mathrm{Id}_n$ as covariance matrix, P_t maps $\mathcal{S}'\cup\mathcal{S}''$ into $\mathcal{S}''$ for all $t>0$; actually, a stronger statement holds: for any $\phi\in\mathcal{S}'\cup\mathcal{S}''$, $\mathsf{P}_t\phi$ is, locally in $t\in(0,\infty)$, uniformly bounded by an integrable function and the same is true for $|\partial_t\mathsf{P}_t\phi|$ and $|\partial_t^2\mathsf{P}_t\phi|$. Therefore the same arguments that justify the (twice) differentiability of the entropy along the heat flow in Costa's EPI, namely of $t\mapsto -\int_{\mathbb{R}^n}\Phi(\mathsf{P}_t\phi)\,\mathrm{d}\mathcal{L}^n$ for $\phi\in\mathcal{S}''$ with $\Phi(z):=z\log z$, allow to deduce that

$$\alpha(s,t) := -\int_{\mathbb{R}^n}\Phi(\mathsf{P}_s f^T)\mathsf{P}_{T-t}g^T\,\mathrm{d}\mathcal{L}^n$$

$$\beta(s,t) := -\int_{\mathbb{R}^n}\Phi(\mathsf{P}_{T-s}g^T)\mathsf{P}_t f^T\,\mathrm{d}\mathcal{L}^n$$

are C^2 on $(0,T)\times(0,T)$ and continuous up to the boundary; since $\mathcal{H}(\rho_t^T) = \alpha(t,t)+\beta(t,t)$, as a byproduct $t\mapsto\mathcal{H}(\rho_t^T)$ belongs to $C([0,T])\cap C^2((0,T))$. The validity of (II.3a), (II.3b) for all $t\in(0,T)$ is then a matter of computations (see the already cited [17]).

On the other hand, it is not difficult to verify that in $\mathbb{R}^n$ it holds

$$\Gamma_2(\phi) = |\mathrm{Hess}(\phi)|_{\mathsf{HS}}^2 \geq \frac{1}{n}(\Delta\phi)^2, \qquad \forall\phi\in C_c^\infty(\mathbb{R}^n) \tag{II.4}$$

and this replaces (II.2) with $K=0$. With this premise, the heuristic argument of the previous section becomes fully rigorous in the following way.

Proof of Theorem I.1. As $t\mapsto\mathcal{H}(\rho_t^T)$ is $C([0,T])\cap C^2((0,T))$, so is $t\mapsto\mathcal{N}(\rho_t^T)$. For sake of brevity set $\mathcal{N}(t):=\mathcal{N}(\rho_t^T)$ and write its first derivative as

$$\mathcal{N}'(t) = \mathcal{N}(t)\cdot\frac{2}{n}\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{H}(\rho_t^T),$$

so that by (II.3a), (II.3b) and (II.4) $\mathcal{N}''$ can be estimated as follows

$$\frac{\mathcal{N}''(t)}{\mathcal{N}(t)} = \frac{4}{n^2}\Big(\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{H}(\rho_t^T)\Big)^2 + \frac{2}{n}\frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathcal{H}(\rho_t^T)$$

$$= \frac{4}{n^2}\Big(\int_{\mathbb{R}^n}\langle\nabla\rho_t^T,\nabla\vartheta_t^T\rangle\,\mathrm{d}\mathcal{L}^n\Big)^2$$

$$\quad -\frac{2}{n}\int_{\mathbb{R}^n}\Big(\Gamma_2(\vartheta_t^T)+\Gamma_2(\log\rho_t^T)\Big)\rho_t^T\,\mathrm{d}\mathcal{L}^n$$

$$\leq \frac{4}{n^2}\Big(\int_{\mathbb{R}^n}\langle\nabla\rho_t^T,\nabla\vartheta_t^T\rangle\,\mathrm{d}\mathcal{L}^n\Big)^2$$

$$\quad -\frac{2}{n^2}\int_{\mathbb{R}^n}\Big((\Delta\vartheta_t^T)^2+(\Delta\log\rho_t^T)^2\Big)\rho_t^T\,\mathrm{d}\mathcal{L}^n.$$

Then by Jensen's inequality and integration by parts

$$\int_{\mathbb{R}^n}\Big((\Delta\vartheta_t^T)^2+(\Delta\log\rho_t^T)^2\Big)\rho_t^T\,\mathrm{d}\mathcal{L}^n$$

$$\geq \Big(\int_{\mathbb{R}^n}\Delta\vartheta_t^T\,\rho_t^T\,\mathrm{d}\mathcal{L}^n\Big)^2 + \Big(\int_{\mathbb{R}^n}\Delta\log\rho_t^T\,\rho_t^T\,\mathrm{d}\mathcal{L}^n\Big)^2$$

$$= \Big(-\int_{\mathbb{R}^n}\langle\nabla\rho_t^T,\nabla\vartheta_t^T\rangle\,\mathrm{d}\mathcal{L}^n\Big)^2$$

$$\quad + \Big(-\int_{\mathbb{R}^n}|\nabla\log\rho_t^T|^2\,\rho_t^T\,\mathrm{d}\mathcal{L}^n\Big)^2$$

$$= \Big(\int_{\mathbb{R}^n}\langle\nabla\log\rho_t^T,\nabla\vartheta_t^T\rangle\rho_t^T\,\mathrm{d}\mathcal{L}^n\Big)^2 + \mathcal{I}(\rho_t^T)^2$$

and plugging this inequality into the previous one yields

$$\frac{\mathcal{N}''(t)}{\mathcal{N}(t)} \leq \frac{2}{n^2}\Big(\int_{\mathbb{R}^n}\langle\nabla\log\rho_t^T,\nabla\vartheta_t^T\rangle\rho_t^T\,\mathrm{d}\mathcal{L}^n\Big)^2 - \frac{2}{n^2}\mathcal{I}(\rho_t^T)^2.$$

By Cauchy-Schwarz inequality the first summand on the right-hand side can be controlled as

$$\Big(\int_{\mathbb{R}^n}\langle\nabla\log\rho_t^T,\nabla\vartheta_t^T\rangle\,\rho_t^T\mathrm{d}\mathcal{L}^n\Big)^2$$
$$\leq \int_{\mathbb{R}^n}|\nabla\log\rho_t^T|^2\rho_t^T\mathrm{d}\mathcal{L}^n\int_{\mathbb{R}^n}|\nabla\vartheta_t^T|^2\rho_t^T\mathrm{d}\mathcal{L}^n \tag{II.5}$$

and this implies

$$\frac{\mathcal{N}''(t)}{\mathcal{N}(t)} \leq \frac{2}{n^2}\mathcal{I}(\rho_t^T)\Big(\underbrace{\int_{\mathbb{R}^n}|\nabla\vartheta_t^T|^2\rho_t^T\mathrm{d}\mathcal{L}^n - \mathcal{I}(\rho_t^T)}_{=2\mathscr{E}_T(\mu,\nu)}\Big),$$

whence (I.4). Finally, if $\nu=\mathsf{P}_T\mu$ (and say $\mu=u\mathcal{L}^n$), then as already said the associated Schrödinger system (I.3) is solved by $f^T=u$ and $g^T=1$. By the very definition of the total energy $\mathscr{E}_T$, this implies $\mathscr{E}_T(\mu,\nu)=0$ and plugging this information into (I.4) yields (I.2). □

Let us now discuss Theorem I.3. The fact that $t \mapsto \mathcal{H}(\rho_t^T)$ belongs to $C([0,T]) \cap C^2((0,T))$ and (II.3a), (II.3b) hold true for all $t \in (0,1)$ (with $\mathbb{R}^n$, $\mathcal{L}^n$ replaced by M, $\mathfrak{m}$ respectively) is justified by [12, Proposition 4.8]. On the other hand, the curvature-dimension assumption (I.6) is equivalent to the generalized Bochner inequality

$$\Gamma_2(\phi) \geq K|\nabla\phi|^2 + \frac{1}{n}(\mathsf{L}\phi)^2, \qquad \forall \phi \in C_c^\infty(M), \quad \text{(II.6)}$$

which replaces (II.2); of course, in the definition of Γ_2 on M, L substitutes Δ. After this digression, the proof of Theorem I.3 is a minor modification of the previous one.

Proof of Theorem I.3. Computing $\mathcal{N}''$ as in the proof of Theorem I.1 and using (II.6) instead of (II.4) we get

$$\frac{\mathcal{N}''(t)}{\mathcal{N}(t)} \leq \frac{4}{n^2}\left(\int_M \langle \nabla\rho_t^T, \nabla\vartheta_t^T\rangle \, \mathrm{d}\mathfrak{m}\right)^2$$
$$- \frac{2}{n^2}\int_M \left((\mathsf{L}\vartheta_t^T)^2 + (\mathsf{L}\log\rho_t^T)^2\right)\rho_t^T \, \mathrm{d}\mathfrak{m}$$
$$- \frac{2K}{n}\int_M \left(|\nabla\vartheta_t^T|^2 + |\nabla\log\rho_t^T|^2\right)\rho_t^T \, \mathrm{d}\mathfrak{m}.$$

Now it suffices to follow the same argument as in the proof of Theorem I.1 (replacing Δ, $\mathcal{L}^n$ with L, $\mathfrak{m}$ respectively, the same integration by parts formula is valid) and keep track of the additional term until the end to obtain (I.7).

If M is compact and $\nu = \mathsf{P}_T\mu$ with $\mu = u\mathfrak{m}$, then $\mathscr{E}_T(\mu, \nu) = 0$ as in Theorem I.1 and moreover $\rho_t^T = \mathsf{P}_t u$, $|\nabla\vartheta_t^T| = |\nabla\log\mathsf{P}_t u|$. □

Remark II.2. A further way to see that (I.7) implies (I.5) relies on the long-time behavior of T-entropic interpolations and the energy $\mathscr{E}_T(\mu, \nu)$, investigated in [7]. From an intuitive point of view, the more the time parameter T grows, the less the final condition $v = g^T\mathsf{P}_T f^T$ in (I.3) is influent, so that in the limit (I.3) is in fact a decoupled system and the T-entropic interpolation is nothing but the heat flow starting at μ.

Under the further assumption that $K \geq 0$ and $\mathfrak{m}$ is a probability (for $K > 0$ this is always true thanks to [26, Theorem 4.26]), by [7, Theorem 1.2] we know that $\mathscr{E}_T(\mu, \nu) \to 0$ as $T \to \infty$. Moreover, if $\mu = u\mathfrak{m}$ and $\nu = v\mathfrak{m}$, by [7, Lemma 3.6] we also know that

$$\lim_{T\to\infty} f^T = u, \qquad \lim_{T\to\infty} g^T = v$$

and

$$\lim_{T\to\infty} \mathsf{P}_T f^T = \lim_{T\to\infty} \mathsf{P}_T g^T = 1,$$

where all limits are in $L^p(\mathfrak{m})$ for any $p \in [1, \infty)$, so that $\rho_t^T \to \mathsf{P}_t u$ in $L^p(\mathfrak{m})$ as $T \to \infty$. Therefore it is intuitively clear that (I.5) can be recovered as the long-time limit of (I.7), since as $T \to \infty$ we expect the first term on the right-hand side of (I.7) to vanish, whereas

$$\lim_{T\to\infty} \int_M |v_t^T|^2 \rho_t^T \, \mathrm{d}\mathfrak{m} + \mathcal{I}(\rho_t^T) = 2\mathcal{I}(\mathsf{P}_t u).$$

To turn this sketch of proof into a rigorous demonstration, one should only pass through an integrated version of (I.7) and argue by dominated convergence. ■

Remark II.3. For the reader who is familiar with the Lott-Sturm-Villani [20], [26] and Ambrosio-Gigli-Savaré [2], [3] theories of synthetic Ricci lower bounds in metric measure spaces, Theorem I.3 actually holds true on any $\mathrm{RCD}(K,n)$ space. Indeed, the C^2-regularity of $t \mapsto \mathcal{H}(\rho_t^T)$ has been proved in [12, Proposition 4.8] in the RCD setting and (II.6) is known to be equivalent to the $\mathrm{RCD}(K,n)$ condition by [11]. ■

III. Proof of Theorem I.1 with deficit

In this section we shall give a direct proof of Theorem I.1 with an (almost) exact error term, in the same spirit of [28]. This means that we shall put aside (II.4) and argue by explicit computations: the disadvantage is the validity of the approach only in the Euclidean setting, but as an advantage we are able to characterize the case of equality in (I.4), thus proving Theorem I.2.

Inspired by [28], let us first observe that, thanks to the computations carried out in the proof of Theorem I.1 (and in particular thanks to the formula for $\mathcal{N}''$), (I.4) is equivalent to

$$\int_{\mathbb{R}^n} \left(\Gamma_2(\vartheta_t^T) + \Gamma_2(\log\rho_t^T)\right)\rho_t^T \, \mathrm{d}\mathcal{L}^n$$
$$- \frac{2}{n}\left(\int_{\mathbb{R}^n} \langle\nabla\rho_t^T, \nabla\vartheta_t^T\rangle \, \mathrm{d}\mathcal{L}^n\right)^2 + \frac{2}{n}\mathcal{I}(\rho_t^T)\mathscr{E}_T(\mu, \nu) \geq 0 \tag{III.1}$$

and define

$$A_1(\lambda) := \sum_{i,j=1}^n \int_{\mathbb{R}^n} \left(\partial_{ij}\vartheta_t^T + \lambda\delta_{ij}\right)^2 \rho_t^T \, \mathrm{d}\mathcal{L}^n,$$
$$A_2(\eta) := \sum_{i,j=1}^n \int_{\mathbb{R}^n} \left(\partial_{ij}\log\rho_t^T + \eta\delta_{ij}\right)^2 \rho_t^T \, \mathrm{d}\mathcal{L}^n.$$

By expanding A_1 as a binomial in λ, using integration by parts and the fact that ρ_t^T is a probability density, we get

$$A_1(\lambda) = \sum_{i,j=1}^n \int_{\mathbb{R}^n} \left(\partial_{ij}\vartheta_t^T\right)^2 \rho_t^T \, \mathrm{d}\mathcal{L}^n$$
$$- 2\sum_{i=1}^n \lambda \int_{\mathbb{R}^n} \partial_i\rho_t^T \partial_i\vartheta_t^T \, \mathrm{d}\mathcal{L}^n + \lambda^2 n$$
$$= \int_{\mathbb{R}^n} \Gamma_2(\vartheta_t^T)\rho_t^T \, \mathrm{d}\mathcal{L}^n$$
$$- 2\lambda \int_{\mathbb{R}^n} \langle\nabla\rho_t^T, \nabla\vartheta_t^T\rangle \, \mathrm{d}\mathcal{L}^n + \lambda^2 n,$$

the second identity being motivated by Bochner's identity in $\mathbb{R}^n$, namely

$$\sum_{i,j=1}^n (\partial_{ij}\phi)^2 = |\mathrm{Hess}(\phi)|_{\mathsf{HS}}^2 = \frac{1}{2}\Delta|\nabla\phi|^2 - \langle\nabla\phi, \nabla\Delta\phi\rangle = \Gamma_2(\phi).$$

In particular, if we choose

$$\lambda^* = \frac{1}{n}\int_{\mathbb{R}^n} \langle\nabla\rho_t^T, \nabla\vartheta_t^T\rangle \, \mathrm{d}\mathcal{L}^n,$$

then

$$A_1(\lambda^*) = \int_{\mathbb{R}^n} \Gamma_2(\vartheta_t^T)\rho_t^T \, d\mathcal{L}^n - \frac{1}{n}\Big(\int_{\mathbb{R}^n} \langle \nabla \rho_t^T, \nabla \vartheta_t^T \rangle \, d\mathcal{L}^n\Big)^2.$$

Arguing in the same way for A_2, we obtain

$$A_2(\eta) = \int_{\mathbb{R}^n} \Gamma_2(\log \rho_t^T)\rho_t^T \, d\mathcal{L}^n - 2\eta \mathfrak{I}(\rho_t^T) + \eta^2 n$$

and if we choose

$$\eta^* = \frac{1}{n}\mathfrak{I}(\rho_t^T),$$

then

$$A_2(\eta^*) = \int_{\mathbb{R}^n} \Gamma_2(\log \rho_t^T)\rho_t^T \, d\mathcal{L}^n - \frac{1}{n}\mathfrak{I}(\rho_t^T)^2.$$

Hence

$$A_1(\lambda^*) + A_2(\eta^*)$$
$$= \int_{\mathbb{R}^n} \Gamma_2(\vartheta_t^T)\rho_t^T \, d\mathcal{L}^n + \int_{\mathbb{R}^n} \Gamma_2(\log \rho_t^T)\rho_t^T \, d\mathcal{L}^n$$
$$- \frac{2}{n}\Big(\int_{\mathbb{R}^n} \langle \nabla \rho_t^T, \nabla \vartheta_t^T \rangle \, d\mathcal{L}^n\Big)^2$$
$$+ \frac{1}{n}\Big(\int_{\mathbb{R}^n} \langle \nabla \rho_t^T, \nabla \vartheta_t^T \rangle \, d\mathcal{L}^n\Big)^2 - \frac{1}{n}\mathfrak{I}(\rho_t^T)^2$$

and by (II.5) and the very definition of $\mathscr{E}_T(\mu, \nu)$

$$\frac{1}{n}\Big(\int_{\mathbb{R}^n} \langle \nabla \rho_t^T, \nabla \vartheta_t^T \rangle \, d\mathcal{L}^n\Big)^2 - \frac{1}{n}\mathfrak{I}(\rho_t^T)^2 \le \frac{2}{n}\mathfrak{I}(\rho_t^T)\mathscr{E}_T(\mu, \nu).$$
(III.2)

Plugging this inequality into the previous identity exactly yields (III.1), since trivially $A_1(\lambda^*) + A_2(\eta^*) \ge 0$. Note that if equality occurs in (III.1) for some t, then in particular equality must hold in (III.2) and this is true if and only if (II.5) is actually an identity. As Cauchy-Schwarz inequality is an equality only for parallel vectors, this means that

$$\text{either} \quad \nabla \log \rho_t^T = \nabla \vartheta_t^T, \qquad \text{or} \quad \nabla \log \rho_t^T = -\nabla \vartheta_t^T.$$

By definition of ρ_t^T and ϑ_t^T, this means that

$$\text{either} \quad \nabla \log \mathsf{P}_t f^T = 0, \qquad \text{or} \quad \nabla \log \mathsf{P}_{T-t} g^T = 0,$$

namely either f^T or g^T is constant and this implies that the entropic interpolation $(\mu_t^T)_{t \in [0,T]}$ is in fact a forward or backward heat flow. Therefore the case of equality in the entropic EPI (I.4) reduces to equality in Costa's EPI (I.2) and the latter is already well understood (see e.g. [8, Theorem 3]).

ACKNOWLEDGEMENT

REFERENCES

[1] Cécile Ané, Sébastien Blachère, Djalil Chafaï, Pierre Fougères, Ivan Gentil, Florent Malrieu, Cyril Roberto, and Grégory Scheffer. *Sur les inégalités de Sobolev logarithmiques*. Société mathématique de France Paris (10), 2000.
[2] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. Metric measure spaces with Riemannian Ricci curvature bounded from below. *Duke Mathematical Journal*, 163(7):1405–1490, 2014.
[3] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. Bakry-Émery curvature-dimension condition and Riemannian Ricci curvature bounds. *The Annals of Probability*, 43(1):339–404, 2015.
[4] Dominique Bakry, Ivan Gentil, and Michel Ledoux. em Analysis and geometry of Markov diffusion operators. Springer, 2014.
[5] Dominique Bakry and Michel Émery. Diffusions hypercontractives. *Séminaire de probabilités, XIX, 1983/84*, Lecture Notes in Math. 1123:177–206, 1985.
[6] Giovanni Conforti. A second order equation for Schrödinger bridges with applications to the hot gas experiment and entropic transportation cost. *Probability Theory and Related Fields*, 174(1-2):1–47, 2019.
[7] Giovanni Conforti and Luca Tamanini. A formula for the time derivative of the entropic cost and applications. *Journal of Functional Analysis*, 280(11):108964, 2021.
[8] Max Costa. A new entropy power inequality. *IEEE Transactions on Information Theory*, 31(6):751–760, 1985.
[9] Amir Dembo. Simple proof of the concavity of the entropy power with respect to added Gaussian noise. *IEEE Transactions on Information Theory*, 35(4):887–888, 1989.
[10] Amir Dembo, Thomas M. Cover and Joy A. Thomas. Information theoretic inequalities. *IEEE Transactions on Information theory*, 37(6):1501–1518.
[11] Matthias Erbar, Kazumasa Kuwada, and Karl-Theodor Sturm. On the equivalence of the entropic curvature-dimension condition and Bochner's inequality on metric measure spaces. *Inventiones mathematicae*, 201(3), 2014.
[12] Nicola Gigli and Luca Tamanini. Second order differentiation formula on $RCD^*(K,N)$ spaces. *Journal of the European Mathematical Society*, 23:1727–1795, 2021.
[13] Nicola Gigli and Luca Tamanini. Benamou-Brenier and duality formulas for the entropic cost on $RCD^*(K,N)$ spaces. *Probability Theory and Related Fields*, 176(1-2):1–34, 2020.
[14] Ivan Gentil, Christian Léonard, and Luigia Ripani. Dynamical aspects of generalized Schrödinger problem via Otto calculus - A heuristic point of view. *Revista Matemática Iberoamericana*, 36(4):1071–1112, 2018.
[15] Ivan Gentil, Christian Léonard, Luigia Ripani, and Luca Tamanini. An entropic interpolation proof of the HWI inequality. *Stochastic Processes and their Applications*, 130(2):907–923, 2020.
[16] Christian Léonard. From the Schrödinger problem to the Monge-Kantorovich problem. *Journal of Functional Analysis*, 262(4):1879–1920, 2012.
[17] Christian Léonard. On the convexity of the entropy along entropic interpolations. *Preprint, arXiv:1310.1274v*, 2013.
[18] Christian Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport *Discrete and Continuous Dynamical Systems. Series A*, 34(4):1533–1574, 2014.
[19] Songzi Li and Xiang-Dong Li. On the Shannon entropy power on Riemannian manifolds and Ricci flow. *Preprint, arXiv:2001.00410*, 2020.
[20] John Lott and Cédric Villani. Ricci curvature for metric-measure spaces via optimal transport. em Annals of Mathematics. Second Series, 169(3):903–991, 2009.
[21] Léonard Monsaingeon, Luca Tamanini, and Dmitry Vorotnikov. The dynamical Schrödinger problem in abstract metric spaces. *Preprint, arXiv:2012.12005*, 2020.
[22] Felix Otto. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations*, 26(1-2):101–174, 2001.
[23] Olivier Rioul. Information theoretic proofs of entropy power inequalities. *IEEE Transactions on Information Theory*, 57(1):33–55, 2010.
[24] Claude E. Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948.
[25] Aart J. Stam. Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Information and Control*, 2(2):101–112, 1959.
[26] Karl-Theodor Sturm. On the geometry of metric measure spaces. I. *Acta Mathematica*, 196(1):65–131, 2006.

[27] Karl-Theodor Sturm. On the geometry of metric measure spaces. II. *Acta Mathematica*, 196(1):133–177, 2006.

[28] Cédric Villani. A short proof of the "concavity of entropy power". *IEEE Transactions on Information Theory*, 46(4):1695–1696, 2006.

Luca Tamanini was born in Tione di Trento, Italy. He received the B.Sc. and M.Sc. degrees in Mathematics from the University of Trento in 2012 and 2014 respectively, and a joint Ph.D. degree from Université Paris Nanterre and SISSA - International School for Advanced Studies in 2017.

From 2017 to 2019 he was a postdoc within the Stochastic Analysis group at the Institut for Applied Mathematics of the University of Bonn. He was then, from 2019 to 2021, a postdoc in the mixed research group MOKAPLAN (Université Paris Dauphine & INRIA Paris). He is currently working as postdoc at the Bocconi Institute for Data Science and Analytics in Milan.