

Efficient computation of the asymptotics of extensive-rank HCIZ integrals

Antoine Maillard[★], Jean-Christophe Mourrat[◦] 

September 9, 2026

Abstract

We study the high-dimensional asymptotics of Harish-Chandra–Itzykson–Zuber (HCIZ) integrals in the extensive-rank regime. The limit of these integrals is governed by a one-dimensional boundary-value hydrodynamical problem originally derived by Matytsin (1994) and rigorously proved by Guionnet and Zeitouni (2002). Despite its wide-ranging applications, explicit solutions to this problem are known only in a few specific cases. In this work, we introduce an efficient numerical scheme based on a particle discretization and prove its convergence to the continuous boundary-value problem for generic boundary densities. We validate our approach against known analytical solutions and apply it to generic densities, uncovering interesting dynamical phenomena. The high-dimensional limit of HCIZ integrals appears in various contexts, from the large deviations of random matrix spectra to the limiting free energy of disordered systems, high-dimensional statistics, and machine learning. As such, our contribution opens the way towards the numerical exploration of a wide range of high-dimensional models that were previously intractable.

Contents

1	Introduction and main results	2
1.1	Asymptotics of the HCIZ integral	2
1.2	Main results	7
1.3	Discussion and consequences	11
1.4	Structure of the paper	13
2	Efficient minimization algorithm	14
2.1	Parametrization by the spacings	15
2.2	Time discretization	16
2.3	Preconditioning	16
2.4	Global scaling and the θ parameter	18
2.5	Initialization, and final algorithm	19
2.6	Practical implementation and final remarks	20
3	Proof of Theorem 1.2	20
3.1	Small-regularization limit	21
3.2	Upper bound	23
3.3	Lower bound	25
3.4	Remark on an alternative formulation	31

[★] Inria Paris & DI ENS, PSL University, Paris, France. antoine.maillard@inria.fr

[◦] Department of Mathematics, ENS Lyon and CNRS, Lyon, France. jean-christophe.mourrat@ens-lyon.fr.

1 Introduction and main results

1.1 Asymptotics of the HCIZ integral

1.1.1 Notation and definition

Notation – Throughout the paper, we adopt the following notation. For $E \subseteq \mathbb{R}$, we denote by $\mathcal{P}(E)$ the set of probability measures on E . We also let $\mathcal{P}_{\text{abs.}}(\mathbb{R})$ denote the set of probability measures that are absolutely continuous with respect to the Lebesgue measure and $\mathcal{P}_c(\mathbb{R})$ the set of compactly supported probability measures. Moreover, for $\mu \in \mathcal{P}(\mathbb{R})$, we let $\Sigma(\mu) := \int \log|x - y| d\mu(x) d\mu(y)$. For a given $\alpha \geq 0$, we denote by $\alpha \# \mu$ the push-forward measure under multiplication by α , i.e. $\int f(x)(\alpha \# \mu)(dx) = \int f(\alpha x) \mu(dx)$ for test functions f . For $\beta \in \{1, 2\}$, we set $\mathbb{K} = \mathbb{R}$ for $\beta = 1$ and $\mathbb{K} = \mathbb{C}$ for $\beta = 2$. For $A \in \mathbb{K}^{d \times d}$, we denote by $A^\dagger$ the Hermitian conjugate of A , and set $\text{tr}(A) := (1/d)\text{Tr}(A)$. We use $\mathcal{H}_d$ to denote the set of symmetric ($\beta = 1$) or Hermitian ($\beta = 2$) $d \times d$ matrices. For $A \in \mathcal{H}_d$, we denote by $\hat{\mu}_A := (1/d) \sum_{i=1}^d \delta_{\lambda_i}$ its empirical eigenvalue distribution, where $(\lambda_i)_{i=1}^d$ are the (real) eigenvalues of A . $\mathbb{S}_\beta^{d-1}$ denotes the unit sphere in $\mathbb{K}^d$. Finally, $\mathcal{U}_\beta(d)$ denotes the group of orthogonal ($\beta = 1$) or unitary ($\beta = 2$) matrices, and $d\rho_\beta$ is the Haar measure on this compact group.

Definition – We consider the “Harish-Chandra–Itzykson–Zuber” (HCIZ) integral [1, 2], defined as follows. For $A, B \in \mathcal{H}_d$, let

$$I_d^{(\beta)}(A, B) := \frac{2}{\beta d^2} \log \int_{\mathcal{U}_\beta(d)} d\rho_\beta(O) \exp \left\{ \frac{\beta d}{2} \text{Tr}[O A O^\dagger B] \right\}. \quad (1)$$

We also define, for $\theta \geq 0$,

$$I_d^{(\beta)}(\theta, A, B) := I_d^{(\beta)}(\sqrt{\theta} A, \sqrt{\theta} B) = \frac{2}{\beta d^2} \log \int_{\mathcal{U}_\beta(d)} d\rho_\beta(O) \exp \left\{ \frac{\beta \theta d}{2} \text{Tr}[O A O^\dagger B] \right\}. \quad (2)$$

While one could fix $\theta = 1$ without loss of generality, it will sometimes be useful in our discussion to be able to vary this scaling parameter.

1.1.2 Related literature

The “spherical” Harish-Chandra–Itzykson–Zuber (HCIZ) integral of eq. (1) is a ubiquitous mathematical object with wide-ranging applications across multiple disciplines. It was initially introduced in the study of Lie algebras [1], and has found applications in fields as diverse as harmonic analysis, representation theory [3], quantum gravity and quantum field theory [2], and integrable systems [4]. We refer to McSwiggen [3] for a review of these applications, and discuss other motivations in more detail below.

High-dimensional asymptotics – In this work, we consider the high-dimensional regime, i.e. the behavior of $I_d^{(\beta)}$ in eq. (1) when the dimension d gets very large. Notably, an exact finite-dimensional determinantal formula exists in the unitary case $\beta = 2$ [1, 2, 5]:

$$I_d^{(2)}(A_d, B_d) = \frac{1}{d^2} \log \det \left[\left(e^{da_i b_j} \right)_{1 \leq i, j \leq d} \right] - \frac{1}{d^2} \sum_{c \in \{a, b\}} \sum_{i < j} \log(c_j - c_i) + K_d, \quad (3)$$

where $K_d \in \mathbb{R}$ is a constant independent of (A_d, B_d) , and $a_1 \leq \dots \leq a_d$, $b_1 \leq \dots \leq b_d$ are the eigenvalues of A_d and B_d . However, evaluating the asymptotic limit of eq. (3) as $d \rightarrow \infty$ is

highly non-trivial, and current results on the limit of $I_d^{(\beta)}$ as $d \rightarrow \infty$ do not rely on this explicit formula. There are many motivations for considering the HCIZ integral in the high-dimensional regime: without aiming to be exhaustive, we detail three such motivations below.

I: Large deviations in random matrix theory – A central motivation for evaluating high-dimensional HCIZ integrals stems from the theory of large deviations for the spectra of random matrices [6]. Notably, a recent line of work has directly related the large deviations of the extreme eigenvalues of many random matrix ensembles to the asymptotics of the HCIZ integral when one of the two matrices A_d, B_d has finite rank [7–15]. The more generic case of A_d, B_d with growing rank as $d \rightarrow \infty$, on the other hand, is related to the rate function for the large deviations of the entire spectral density of the Dyson Brownian motion [16–19]. It also appears in the large deviations of the spectra of correlated random matrix models [20], and in the joint distribution of the eigenvalues of the sample covariance matrix of Gaussian data with a generic covariance matrix [21, 22].

II: Disordered systems and spin glasses – In statistical mechanics, the high-dimensional asymptotics of the HCIZ integral frequently appear in the limiting free energies of disordered systems and mean-field spin glasses under rotationally-invariant couplings [23–25]. Recently, it was also shown to be an important ingredient in the free energy of constraint satisfaction problems over symmetric matrices, such as ellipsoid fitting [26] or an average-case variant of the matrix discrepancy problem [20]. The partial differential equations governing the asymptotics of the HCIZ integral have also been shown to arise in the macroscopic fluctuation theory of the Dyson gas [27]. Finally, it also connects to the theory of stochastic processes, e.g. via the statistics of non-intersecting Brownian bridges; see [28] and the references therein.

III: High-dimensional statistics and machine learning – In recent years, the HCIZ integral has also emerged in several studies in theoretical data science and high-dimensional statistics to characterize the information-theoretic limits and computational phase transitions of statistical estimation problems. Concretely, the low-rank integral (where one of the two matrices A, B in eq. (1) has a finite rank as $d \rightarrow \infty$) plays an important role to derive such quantities in low-rank matrix estimation, compressed sensing, and generalized linear models [9, 25, 29, 30]. Conversely, the extensive-rank HCIZ integral (when A and B both have rank $\Theta(d)$) appears in high-rank matrix denoising and factorization under rotationally invariant setups [31–35]. Most recently, this extensive-rank regime has been actively deployed to derive exact asymptotics for the learning dynamics and generalization capabilities of models of extensive-width neural networks and simple models of attention [36–42].

Previous results – The limit of $I_d^{(\beta)}$ in the low-rank regime has been extensively studied and characterized [23, 43, 44]. Conversely, the extensive-rank regime is governed by a one-dimensional boundary-value hydrodynamical problem, first derived by Matytsin [45] and later proved by Guionnet and Zeitouni [16, 17], see also [4, 18, 46] for extensions and discussions. These results form the starting point of our study, and are stated below in detail. Recent work has also expanded these extensive-rank asymptotic evaluations to a generalization of the HCIZ integral to rectangular matrices [47].

1.1.3 Matytsin’s solution

We now state the asymptotics of I_d , first obtained by Matytsin [45] and proved by Guionnet and Zeitouni [16, 17]. First we need to define admissible trajectories between two probability measures.

Definition 1.1 (Admissible trajectories). For any $\mu, \nu \in \mathcal{P}(\mathbb{R})$, we denote by $\text{Adm}(\mu, \nu)$ the set of all pairs (ρ, v) of measurable functions from $(0, 1) \times \mathbb{R}$ to $\mathbb{R}$ such that the following conditions hold.

1. For each $t \in (0, 1)$, the function $\rho(t, \cdot)$ is the density of a probability measure which is absolutely continuous with respect to the Lebesgue measure.
2. We have the integrability condition

$$\int_0^1 \int_{\mathbb{R}} |v(t, x)| \rho(t, x) dx dt < +\infty. \quad (4)$$

3. The continuity equation

$$\partial_t \rho + \partial_x(\rho v) = 0 \quad (5)$$

is satisfied in the weak sense.

4. We have

$$\lim_{t \rightarrow 0} \int \rho(t, x) dx = \mu \quad \text{and} \quad \lim_{t \rightarrow 1} \int \rho(t, x) dx = \nu,$$

in the sense of weak convergence of probability measures.

Notice that the condition in eq. (4) is required in order to make sense of the continuity equation (5). The conditions of eqs. (4) and (5) imply the continuity of the path $t \mapsto \int \rho(t, x) dx$ for the topology of weak convergence of probability measures, see e.g. [48, Lemma 8.1.2]. We are now ready to state the asymptotics of $I_d^{(\beta)}$.

Theorem 1.1 ([16, 17, 45, 49]). *Let $\mu, \nu \in \mathcal{P}(\mathbb{R})$, and for each $d \geq 1$, let $A_d, B_d \in \mathcal{H}_d$, with empirical eigenvalue distributions $\hat{\mu}_{A_d}, \hat{\mu}_{B_d}$. Assume that*

- (i) $\hat{\mu}_{A_d} \rightarrow \mu$ and $\hat{\mu}_{B_d} \rightarrow \nu$ as $d \rightarrow \infty$ (in the sense of weak convergence);
- (ii) There exists $K > 0$ such that $\text{supp}(\hat{\mu}_{A_d}) \subseteq [-K, K]$ and $\text{supp}(\hat{\mu}_{B_d}) \subseteq [-K, K]$ for all $d \geq 1$;

Then $I_{\text{HCIZ}}(\mu, \nu) := \lim_{d \rightarrow \infty} I_d^{(\beta)}(A_d, B_d)$ exists, is independent of $\beta \in \{1, 2\}$, and is continuous in both μ and ν with respect to the weak convergence on $\mathcal{P}([-K, K])$.

Moreover, if we assume further that $\min\{\Sigma(\mu), \Sigma(\nu)\} > -\infty$, then the following statements hold.

- (1) $I_{\text{HCIZ}}(\mu, \nu)$ is given by

$$I_{\text{HCIZ}}(\mu, \nu) = -\frac{3}{4} + \frac{1}{2} \left[\int \mu(dx) x^2 + \int \nu(dx) x^2 \right] - \frac{1}{2} [\Sigma(\mu) + \Sigma(\nu)] - J(\mu, \nu), \quad (6)$$

where

$$J(\mu, \nu) := \frac{1}{2} \inf_{(\rho, v) \in \text{Adm}(\mu, \nu)} \int_0^1 dt \int dx \rho(t, x) \left[v(t, x)^2 + \frac{\pi^2}{3} \rho(t, x)^2 \right]. \quad (7)$$

- (2) The infimum in eq. (7) is attained by a unique path (ρ, v) , which satisfies (in the weak sense) the equation

$$\partial_t(\rho v) + \partial_x \left[\rho v^2 - \frac{\pi^2}{3} \rho^3 \right] = 0. \quad (8)$$

Further, (ρ, v) are smooth in the interior of $\Omega := \{(t, x) \in (0, 1) \times \mathbb{R} : \rho(t, x) > 0\}$, which is bounded. In particular, eqs. (5) and (8) hold everywhere in the interior of Ω .

We notice that having both measures uniformly compactly supported is not strictly necessary for some of these results. For instance, one can define $I_{\text{HCIZ}}(\mu, \nu)$ whenever μ is compactly supported and ν has bounded first moment: we refer to [50, Section 2] for more details.

Scaling parameter θ – Under the same assumptions as in Theorem 1.1, we define

$$I_{\text{HCIZ}}(\theta, \mu, \nu) := \lim_{d \rightarrow \infty} I_d^{(\beta)}(\theta, A_d, B_d) \quad (9)$$

where $I_d^{(\beta)}$ is defined in eq. (2), and we have

$$I_{\text{HCIZ}}(\theta, \mu, \nu) = -\frac{3}{4} - \frac{1}{2} \log \theta + \frac{\theta}{2} \left[\int \mu(dx) x^2 + \int \nu(dx) x^2 \right] - \frac{1}{2} [\Sigma(\mu) + \Sigma(\nu)] - J(\theta, \mu, \nu), \quad (10)$$

where $J(\theta, \mu, \nu) := J(\sqrt{\theta} \# \mu, \sqrt{\theta} \# \nu)$ as given by eq. (7). We will keep the shorthand notation $I_{\text{HCIZ}}(\mu, \nu) = I_{\text{HCIZ}}(\theta = 1, \mu, \nu)$.

We now make a series of additional remarks on Theorem 1.1.

- (I) Interestingly, eqs. (5) and (8) can be interpreted as the hydrodynamical description of a fluid under an *attractive* force, i.e. a negative pressure field $P = -(\pi^2/3)\rho^3$ [18]. Another key interpretation of these dynamics arises from random matrix theory, see e.g. [6]. To describe this interpretation, let us consider $\beta = 1$ for simplicity. One can show that $\rho(t, x) dx$ is the limiting spectral distribution of the constrained Dyson Brownian motion

$$M_t := A_d + \sqrt{t}W, \quad (11)$$

where $(W_{ij})_{i \leq j} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, (1 + \delta_{ij})/d)$, under the constraint that $M_{t=1}$ has asymptotic eigenvalue density equal to ν . This interpretation is also key to some derivations of Theorem 1.1, which use this relation of the HCIZ integral to the large deviations of the Dyson Brownian motion [16–18].

- (II) Notice that the limit in eq. (6) is independent of $\beta \in \{1, 2\}$. This is a generic property known as “Zuber’s 1/2 rule” [18, 51].

- (III) It was shown in [44, Theorem 6] that (under some technical assumptions)

$$\lim_{a \rightarrow 0} \left[\frac{1}{a} I_{\text{HCIZ}}((1-a)\delta_0 + a\mu, \nu) \right] = \int_{\mathbb{R}} \tilde{I}_\nu(t) \mu(dt), \quad (12)$$

where (with $\hat{\mu}_B \rightarrow \nu$ as $d \rightarrow \infty$)

$$\tilde{I}_\nu(t) := \lim_{d \rightarrow \infty} \frac{2}{\beta d} \log \int_{\mathcal{U}_\beta(d)} d\rho_\beta(O) \exp \left\{ \frac{\beta dt}{2} (O^\dagger B_d O)_{11} \right\}$$

is the “rank-one” HCIZ integral. While eq. (12) shows a smooth transition between the asymptotic formulas for the “low-rank” and “extensive-rank” regimes of the HCIZ integral, the proof of Collins and Śniady [44] shows it using bounds on the finite-dimensional integral. As pointed out in [6], it is an interesting question to derive eq. (12) directly from Theorem 1.1 instead, and to characterize the behavior of the trajectories when one of the two densities has vanishingly small rank.

The need for a numerical scheme – Somewhat disappointingly, explicit analytical solutions to the hydrodynamical problem of Theorem 1.1 are very scarce. To our knowledge, the following cases are the only ones amenable to an exact analytical treatment.

(C1) For any μ, ν satisfying the hypotheses of Theorem 1.1, one can expand $I_{\text{HCIZ}}(\theta, \mu, \nu)$ in powers of θ as $\theta \downarrow 0$ (a *high-temperature* expansion) [52]. Conversely, there are formulas in the physics literature [18] for the first orders of the expansion of $I_{\text{HCIZ}}(\theta, \mu, \nu)$ as $\theta \rightarrow \infty$ (a *low-temperature* expansion).

(C2) If $\nu = \mu \boxplus \sigma_{\text{s.c.}}$, where

$$\sigma_{\text{s.c.}}(dx) := \frac{\sqrt{4-x^2}}{2\pi} \mathbb{1}_{\{|x| \leq 2\}} dx$$

is the semicircle law and $\boxplus$ is the additive free convolution [53], then the Dyson Brownian motion in eq. (11) is unconstrained, as ν corresponds then to the *typical* eigenvalue density of $M_{t=1}$. Writing $\rho_t := \rho(t, x) dx$ for the probability measure at time t , we have (see e.g. [31, 54]) that $\rho_t = \mu \boxplus \sigma_{\text{s.c.}, \sqrt{t}}$ for any $t \in [0, 1]$, where $\sigma_{\text{s.c.}, r}$ is the semicircle law with variance r^2 . Further, this allows us to express $I_{\text{HCIZ}}(\mu, \mu \boxplus \sigma_{\text{s.c.}})$ in terms of a simple Gaussian integral and obtain the formula

$$I_{\text{HCIZ}}(\mu, \mu \boxplus \sigma_{\text{s.c.}}) = -\frac{1}{4} - \Sigma(\mu \boxplus \sigma_{\text{s.c.}}) + \int \mu(dx) x^2. \quad (13)$$

The measure ρ_t can be computed analytically from the knowledge of μ using its $\mathcal{R}$ -transform: we refer to [19, 53] for more details. Moreover, the velocity field can be expressed as:

$$v(t, x) = -\lim_{\varepsilon \downarrow 0} \text{Re}[g_{\rho_t}(x + i\varepsilon)], \quad (14)$$

where we define the Stieltjes transform of any probability measure $\mu \in \mathcal{P}(\mathbb{R})$ as

$$g_\mu(z) := \int \frac{1}{x-z} d\mu(x), \quad (15)$$

with $z \in \mathbb{C}_+ := \{w \in \mathbb{C} : \text{Im}(w) > 0\}$.

(C3) In [18], it is shown that when μ and ν are both centered¹ semicircular laws, the measure ρ_t remains semicircular for any $t \in [0, 1]$, and $I_{\text{HCIZ}}(\mu, \nu)$ admits a simple expression. Without loss of generality, we can consider $I_{\text{HCIZ}}(\theta, \mu, \nu)$ for $\mu = \nu = \sigma_{\text{s.c.}}$, which yields:

$$I_{\text{HCIZ}}(\theta, \sigma_{\text{s.c.}}, \sigma_{\text{s.c.}}) = \frac{1}{2} \left[\sqrt{1+4\theta^2} - 1 - \log \left(\frac{1 + \sqrt{1+4\theta^2}}{2} \right) \right]. \quad (16)$$

(C4) Finally, one can leverage the finite-dimensional determinantal formula of eq. (3) if μ or ν is the uniform distribution over a real interval. Indeed, in this case the computation of the determinant in eq. (3) reduces to computing

$$\det \left[\left(e^{j\rho_k} \right)_{1 \leq j, k \leq d} \right] = \det \left[\left((e^{\rho_k})^j \right)_{1 \leq j, k \leq d} \right] = \prod_{j < k} |e^{\rho_j} - e^{\rho_k}|,$$

for some coefficients $(\rho_j)_{j=1}^d$. One can then exploit this Vandermonde determinant to extract high-dimensional asymptotics for the HCIZ integral, see e.g. [28] for an example in the physics literature.

While these specific solvable scenarios serve as the backbone for some of the statistical applications mentioned previously, computing $I_{\text{HCIZ}}(\mu, \nu)$ for generic distributions μ, ν has largely remained mathematically intractable. In this work, we overcome this limitation by designing a convergent numerical discretization scheme capable of evaluating the extensive-rank HCIZ integral for generic boundary densities. We benchmark our algorithm against several of these known

¹Centering can be assumed without loss of generality, since $I_d^{(\beta)}(A, B) = I_d^{(\beta)}[A - \text{tr}(A)\text{I}_d, B - \text{tr}(B)\text{I}_d] + \text{tr}(A)\text{tr}(B)$ for any $A, B \in \mathcal{H}_d$.

analytical cases, and open the door to the numerical exploration of a wide range of previously intractable high-dimensional models. In the context of high-dimensional statistics, we anticipate that this will enable the exploration of much more general settings beyond Bayes-optimal estimation or extreme-temperature limits of free energies, and we plan to explore some of these applications in subsequent work.

1.2 Main results

As is clear from eq. (6), in order to compute $I_{\text{HCIZ}}(\mu, \nu)$ for arbitrary (μ, ν) , it is enough to compute the functional $J(\mu, \nu)$ defined in eq. (7). Recall that we have

$$J(\mu, \nu) := \inf_{(\rho, v) \in \text{Adm}(\mu, \nu)} G(\rho, v), \quad (17)$$

with

$$G(\rho, v) := \frac{1}{2} \int_0^1 dt \int_{\mathbb{R}} dx \left(v(t, x)^2 + \frac{\pi^2}{3} \rho(t, x)^2 \right) \rho(t, x). \quad (18)$$

Our main theoretical result is the validity of a discretization scheme to compute $J(\mu, \nu)$ in eq. (17) up to arbitrary accuracy. We first define the scheme, before stating our theorem.

1.2.1 The discretization scheme

We fix $\mu_0, \mu_1 \in \mathcal{P}(\mathbb{R})$ with $\Sigma(\mu_0)$ and $\Sigma(\mu_1)$ finite. We design a particle-based approximation scheme to compute $J(\mu_0, \mu_1)$, keeping time as a continuous variable for now.

For every integer $N \geq 1$, we denote by $\text{Adm}_N(\mu_0, \mu_1)$ the set of all trajectories $\{(x_i, v_i)\}_{i=1}^N$, with $x_i, v_i \in \mathcal{C}([0, 1], \mathbb{R})$, that satisfy the following constraints.

(H1) $(x_i(0))_{1 \leq i \leq N}$ and $(x_i(1))_{1 \leq i \leq N}$ are fixed such that for every $i \in \{1, \dots, N\}$:

$$\int_{-\infty}^{x_i(0)} d\mu_0(x) = \frac{i}{N+1} \quad \text{and} \quad \int_{-\infty}^{x_i(1)} d\mu_1(x) = \frac{i}{N+1}. \quad (19)$$

We may at times find it convenient to also use the convention $x_0(t) = -\infty$ and $x_{N+1}(t) = +\infty$ for every $t \in [0, 1]$.

(H2) For every $i \in \{1, \dots, N\}$ and $t \in [0, 1]$, we have $x_i(t) \leq x_{i+1}(t)$.

(H3) For every $i \in \{1, \dots, N\}$ and $t \in [0, 1]$:

$$x_i(t) = x_i(0) + \int_0^t v_i(s) ds.$$

Notice that μ_0 and μ_1 have no atoms since $\Sigma(\mu_0)$ and $\Sigma(\mu_1)$ are finite, which ensures that $(x_i(0))_i, (x_i(1))_i$ in eq. (19) are well-defined and strictly increasing. Within the set $\text{Adm}_N(\mu_0, \mu_1)$, the trajectories $x_1, \dots, x_N$ are affine functions of the trajectories $v_1, \dots, v_N$. By (H3) we must have:

$$\int_0^1 v_i(s) ds = x_i(1) - x_i(0) \quad \text{for all } i \in [N]. \quad (20)$$

Moreover, (H2) imposes that for every $i \in \{1, \dots, N-1\}$ and $t \in [0, 1]$:

$$\int_0^t (v_{i+1}(s) - v_i(s)) ds \geq x_i(0) - x_{i+1}(0). \quad (21)$$

Conversely, any $(v_1, \dots, v_N)$ satisfying eqs. (20) and (21) uniquely defines an element of $\text{Adm}_N(\mu_0, \mu_1)$. Finally, for every $\{(x_i, v_i)\}_{i=1}^N \in \text{Adm}_N(\mu_0, \mu_1)$, and any $\varepsilon > 0$, we define

$$G_{N, \varepsilon}(\{(x_i, v_i)\}) := \frac{1}{2N} \sum_{i=1}^N \int_0^1 \left(v_i(t)^2 + \frac{\pi^2}{3(N+1)^2} \left(x_{i+1}(t) - x_i(t) + \frac{\varepsilon}{N+1} \right)^{-2} \right) dt, \quad (22)$$

and we set

$$J_{N,\varepsilon}(\mu_0, \mu_1) := \inf_{\text{Adm}_N(\mu_0, \mu_1)} G_{N,\varepsilon}. \quad (23)$$

1.2.2 Validity of the discretization scheme

The following theorem shows that the discretization scheme introduced above asymptotically approaches the value of the HCIZ integral.

Theorem 1.2 (Validity of the numerical method). *Let $\mu_0, \mu_1 \in \mathcal{P}_c(\mathbb{R}) \cap \mathcal{P}_{\text{abs.}}(\mathbb{R})$ be such that $\mu_0(dx) = \rho_0(x)dx$, $\mu_1(dx) = \rho_1(x)dx$ with $\rho_0, \rho_1 \in L^3(\mathbb{R})$. We have*

$$\lim_{\varepsilon \rightarrow 0} \liminf_{N \rightarrow \infty} J_{N,\varepsilon}(\mu_0, \mu_1) = \lim_{\varepsilon \rightarrow 0} \limsup_{N \rightarrow \infty} J_{N,\varepsilon}(\mu_0, \mu_1) = J(\mu_0, \mu_1).$$

Notice that the actual numerical scheme that we implement in Section 2 also involves a discretization of the time variable into small intervals. It is however straightforward to verify that for each fixed N and ε , this fully discretized optimization problem converges to $J_{N,\varepsilon}(\mu, \nu)$ as the time-step gaps are sent to zero, and we focus on proving Theorem 1.2.

1.2.3 First applications

In Section 2, we present an efficient numerical scheme, which involves an additional time discretization step, and is based on the Newton update combined with a preconditioned conjugate gradient method. We also discuss potential alternatives, as well as limitations of our scheme, in Section 2. Here, we present some applications, focusing first on the known “benchmark” cases mentioned above.

1. Free convolution (C2) – We consider item (C2), i.e. $\nu = \mu \boxplus \sigma_{\text{s.c.}}$. In Fig. 1, we show the

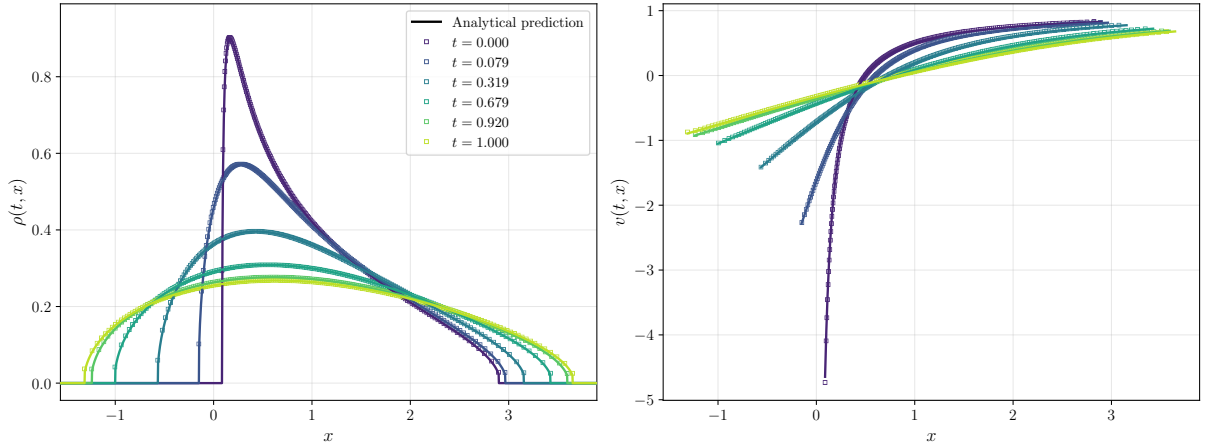


Figure 1: The density $\rho(t, x)$ (left) and the velocity field $v(t, x)$ (right) when $\mu = \mu_{\text{MP},\kappa}$ is the Marchenko-Pastur density with ratio $\kappa = 2$ (see eq. (24)), and $\nu = \mu \boxplus \sigma_{\text{s.c.}}$. We show the values of the density and velocity at various times $t \in [0, 1]$. Solid lines are the analytical predictions detailed in (C2), and squares represent a sample of 200 particles. For this plot, we used $N = 2^{12}$ total particles, and $T = 2^{10}$ time steps, see Section 2 for more details.

density and velocity evolution in the example where $\mu = \mu_{\text{MP},\kappa}$ is the Marchenko-Pastur density with ratio $\kappa \geq 1$ [55]:

$$\frac{d\mu_{\text{MP},\kappa}}{dx} = \frac{\kappa}{2\pi} \frac{\sqrt{(\lambda_+ - x)(x - \lambda_-)}}{x} \mathbb{1}\{\lambda_- \leq x \leq \lambda_+\}, \quad \lambda_{\pm} := (1 \pm \kappa^{-1/2})^2, \quad (24)$$

and we compare our numerical solver to the free convolution $\rho_t = \mu \boxplus \sigma_{\text{s.c.}, \sqrt{t}}$ and the velocity field v given by eq. (14). We find excellent agreement with the predictions.

2. Semicircular laws (C3) – In Fig. 2, we benchmark our method in the case where $\mu_0 = \mu_1 = \sigma_{\text{s.c.}}$, while varying the scaling parameter θ . We compare the predictions to the analytical

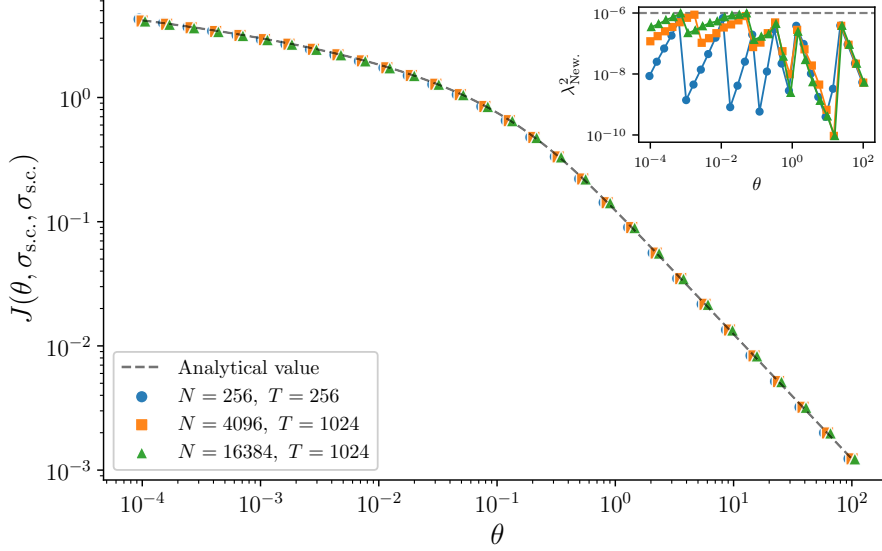


Figure 2: The value of $J(\theta, \sigma_{\text{s.c.}}, \sigma_{\text{s.c.}})$, as computed by Algorithm 1, against the asymptotic value of eq. (25). This comparison is performed for a wide range of values of θ , and different values of (N, T) . The inset shows the Newton decrement at the solution (see Section 2).

result of [18], cf. eq. (16). Equivalently, by eq. (10):

$$J(\theta, \sigma_{\text{s.c.}}, \sigma_{\text{s.c.}}) = \theta - \frac{1}{2} \log \theta - \frac{\sqrt{1 + 4\theta^2}}{2} + \frac{1}{2} \log \left(\frac{1 + \sqrt{1 + 4\theta^2}}{2} \right). \quad (25)$$

We find excellent agreement, validating here as well our method, for both large and small values of the scaling parameter θ , and already at relatively small values of N, T .

3. Singular boundary measures – For $\mu = \mu_{\text{MP},1}$ as in (24) and $\nu = \mu \boxplus \sigma_{\text{s.c.}, 1/\theta}$, we compute $I_{\text{HCIZ}}(\theta, \mu, \nu)$ for a range of values of θ and compare to the analytical predictions. This is Fig. 3. Note that here the density of μ is not in L^3 since $d\mu(x)/dx \sim x^{-1/2}$ as $x \downarrow 0$, but the formula in (13), suitably generalized to arbitrary θ , is still valid. We can thus check that our solver also performs very well in this case, suggesting that the L^3 assumption in Theorem 1.2 might not be necessary.

4. Application to generic densities – Crucially, our method applies to generic boundary measures (μ, ν) for which the value of $I_{\text{HCIZ}}(\mu, \nu)$ is not known. As a first example, we show in Fig. 4 the trajectories found by our numerical solver when $\nu = \sigma_{\text{s.c.}, 1/2}$ is a semicircle with variance $1/2$, and μ is the *symmetrized* Marchenko-Pastur density for $\kappa \geq 1$:

$$\frac{d\mu}{dx} = \frac{\kappa}{4\pi} \frac{\sqrt{(\lambda_+ - |x|)(|x| - \lambda_-)}}{|x|} \mathbb{1}_{\{\lambda_- \leq |x| \leq \lambda_+\}}, \quad \lambda_{\pm} := (1 \pm \kappa^{-1/2})^2, \quad (26)$$

Notice that the initial measure μ has disconnected support components, which merge during the evolution at around $t_{\text{merge}} \simeq 0.08$. We also witness the behavior of the velocity field v around the merging time numerically; since the components are merging, the velocity field cannot be Lipschitz continuous, and we witness its rather singular behavior numerically. In this setting, no analytical solution is known to the variational problem of eq. (7), and our work is the first to

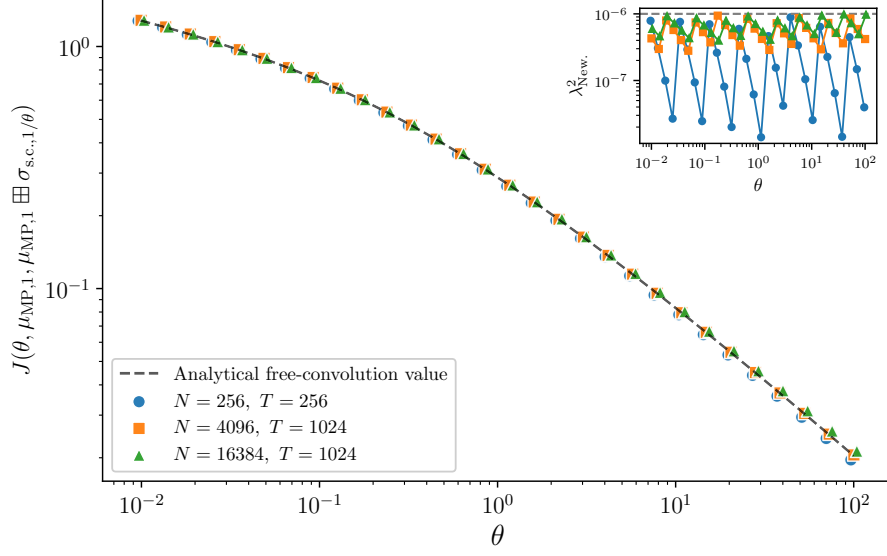


Figure 3: Similar quantities as in Fig. 2, but for $\mu = \mu_{\text{MP},1}$ and $\nu = \mu_{\text{MP},1} \boxplus \sigma_{\text{s.c.},1/\theta}$.

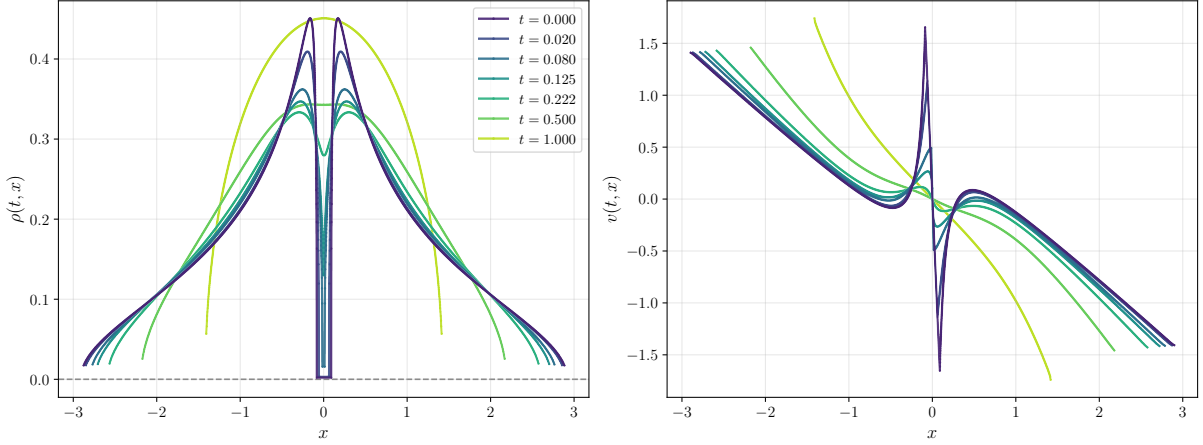


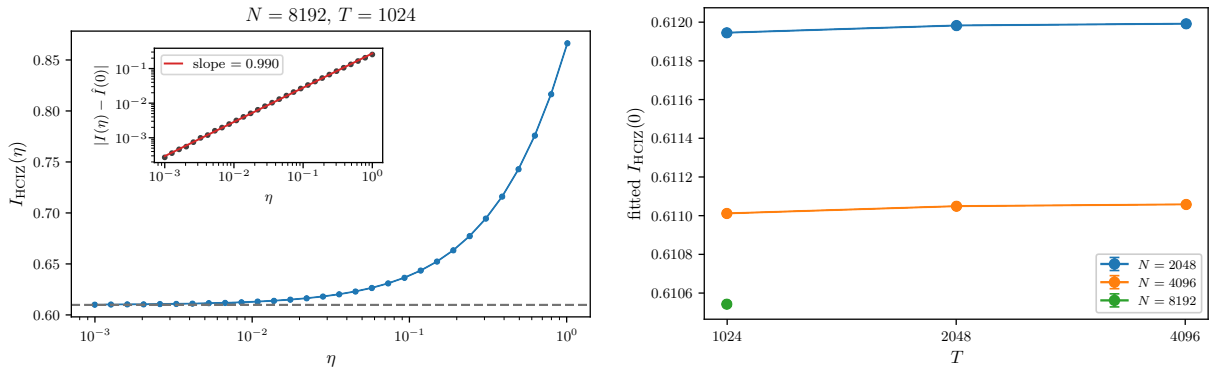
Figure 4: The density $\rho(t, x)$ (left) and the velocity field $v(t, x)$ (right) when μ is the *symmetrized* Marchenko-Pastur density with ratio $\kappa = 2$ (see eq. (26)), and $\nu = \sigma_{\text{s.c.},0.5}$, a semicircle law with variance $1/2$. We show the values of the density and velocity at various times $t \in [0, 1]$. Squares represent a sample of the particles, and the solid lines are linear interpolations of the density and velocity evaluated from the numerical discretization. For this plot, we used $N = 2^{12}$ total particles, and $T = 2^{10}$ time steps, see Section 2 for more details.

provide a simple numerical scheme to evaluate the asymptotic value of the HCIZ integral and the corresponding trajectories.

5. Singular matrices and regularization – Let $\mu, \nu \in \mathcal{P}_c(\mathbb{R})$, and assume that either μ or ν is singular: for concreteness we assume $\Sigma(\mu) = -\infty$ while $\Sigma(\nu) > -\infty$. While there is no known asymptotic formula for $I_{\text{HCIZ}}(\mu, \nu)$ in this case, one can use the continuity of $I_{\text{HCIZ}}(\mu, \nu)$ given in Theorem 1.1 to write

$$I_{\text{HCIZ}}(\mu, \nu) = \lim_{\eta \rightarrow 0} I_{\text{HCIZ}}(\mu_\eta, \nu),$$

where μ_η is any regularization, that is compactly supported (uniformly in η), and such that $\mu_\eta \rightarrow \mu$ as $\eta \downarrow 0$ in the sense of weak convergence. Applying our numerical scheme to the regularized measure allows to access the value of $I_{\text{HCIZ}}(\mu, \nu)$ even for strongly singular measures: we illustrate this in Fig. 5, where we consider $\mu = \mu_{\text{MP}, \kappa=0.5}$, which has an atom of mass $1/2$ in $x = 0$.



(a) The value of $I_{\text{HCIZ}}(\eta)$ as a function of η for a large value of N, T . We estimate the value at $\eta = 0$ from a power-law fit $I_{\text{HCIZ}}(\eta) \sim \hat{I}(0) + A\eta^\alpha$, and the inset suggests a linear decay ($\alpha = 1$). (b) The value of $I_{\text{HCIZ}}(0)$, as fitted from $I_{\text{HCIZ}}(\eta)$ for small η (see the left figure), for various values of N and T . Notice the small vertical range of order 10^{-3} .

Figure 5: Let $\mu = \mu_{\text{MP}, 1/2}$ and $\nu = \sigma_{\text{s.c.}}$. We let $\mu_\eta := \mu \boxplus \sigma_{\text{s.c.}, \eta}$ for $\eta > 0$ and compute $I_{\text{HCIZ}}(\eta) := I_{\text{HCIZ}}(\mu_\eta, \nu)$.

1.3 Discussion and consequences

We discuss here some properties of our discretization scheme and consequences of our analysis. A more detailed discussion of the implementation detail, as well as an analysis of convergence properties and computation time, is postponed to Section 2.

Mathematical remarks – We make here a few remarks stemming from these first examples.

- Perhaps the most transparent derivation of Theorem 1.1 is through the large deviations of the Dyson Brownian motion [16–18]. At finite N , the corresponding particle action contains the nonlocal inverse-square interaction

$$\Psi_{N,i}(\lambda) := \frac{1}{N^2} \sum_{j \neq i} \frac{1}{(\lambda_i - \lambda_j)^2}, \quad (27)$$

where $\lambda_1 < \dots < \lambda_N$ are the eigenvalues evolving under the Dyson Brownian motion (see e.g. [53]). Heuristically, assuming that $(1/N) \sum_{i=1}^N \delta_{\lambda_i} \rightarrow \rho(x) dx$, for an index i such that $\rho(\lambda_i) > 0$, the local quantile spacing satisfies, for k small compared with N ,

$$\lambda_{i+k} - \lambda_i \simeq \frac{k}{N\rho(\lambda_i)}. \quad (28)$$

Substitution into eq. (27) gives the approximation

$$\Psi_{N,i}(\boldsymbol{\lambda}) \simeq \rho(\lambda_i)^2 \sum_{k \in \mathbb{Z} \setminus \{0\}} \frac{1}{k^2} = \frac{\pi^2}{3} \rho(\lambda_i)^2. \quad (29)$$

Indeed, particles at macroscopic distance contribute only $O(N^{-1})$ to eq. (27), whereas the singular near-neighbor terms contribute at order one. Averaging over the particles then yields

$$\frac{1}{N} \sum_{i=1}^N \Psi_{N,i}(\boldsymbol{\lambda}) \simeq \frac{\pi^2}{3} \int \rho(x)^3 dx, \quad (30)$$

which explains the term proportional to ρ^3 in eq. (7). This (very) heuristic argument is at the heart of the physics derivation of Theorem 1.1, see [18, 45]. Our discretization in eq. (22) does not retain or truncate the full interaction in eq. (27). Instead, we directly discretize its continuum limit and use nearest-neighbor spacings to approximate the density while, preserving the coefficient $\pi^2/3$ in eq. (29). This makes our numerical scheme completely local in space, and simplifies its implementation and analysis.

- The assumption in Theorem 1.2 that $\rho(0, \cdot), \rho(1, \cdot) \in L^3$ can probably be weakened, as seen in the example of Fig. 3. In any case, singular boundary densities can be appropriately regularized, as seen in the example of Fig. 5. A more quantitative understanding of the tradeoffs involved in regularizing the density would be interesting to explore.
- Fig. 4 illustrates a topological change in the support of the density: the two connected components of $\text{supp } \rho(0, \cdot)$ merge during the evolution. The numerical results do not determine whether the velocity field $v(t, \cdot)$ remains continuous at the merging time. More generally, investigating the regularity properties of the solutions, in particular near such critical times, is an interesting mathematical question, and a likely necessary ingredient for a sharp quantitative understanding of convergence rates.

Generalization: free boundary – There are several natural extensions and generalizations of our approach. Among them is the numerical evaluation of variational principles in which one of the boundary measures entering the HCIZ functional is itself optimized. More precisely, given a confining potential F and a fixed measure $\nu \in \mathcal{P}(\mathbb{R})$, one may consider

$$L(\theta, F, \nu) := \sup_{\mu \in \mathcal{P}(\mathbb{R})} \left\{ \frac{1}{2} \Sigma(\mu) - \frac{1}{2} \int_{\mathbb{R}} F(x) \mu(dx) + \frac{1}{2} I_{\text{HCIZ}}(\theta, \mu, \nu) \right\}, \quad (31)$$

under suitable assumptions ensuring that the functional is well defined and that the supremum is finite. Variational problems of this form arise in several contexts discussed in Section 1.1.2, including spectral large-deviation problems for random matrices, the computation of free energies in disordered systems, and problems in high-dimensional statistics and machine learning. At the level of our particle discretization, the optimization over μ can be incorporated by treating the initial particle positions $x_i(0)$ as additional optimization variables, while keeping the terminal configuration associated with ν fixed. The discrete objective must then be augmented by consistent discretizations of the boundary terms $\Sigma(\mu)$ and $\int F d\mu$. This leads to a free-endpoint version of the algorithm introduced in Section 2. In an extended version of this work, we plan to study this extension systematically, including its numerical stability and convergence properties, and to analyze some of its first consequences for the applications mentioned above.

Alternative approaches.— We conclude by discussing alternative approaches to the numerical computation of the high-dimensional asymptotics of the HCIZ integral. A natural possibility is to solve directly the Euler equation in eq. (8), coupled with the continuity equation in eq. (5). Defining

$$f(t, x) := v(t, x) + i\pi\rho(t, x),$$

these two equations are equivalent to the complex inviscid Burgers' equation

$$\partial_t f + f \partial_x f = 0. \quad (32)$$

The corresponding real system for (ρ, v) has a rather complex structure: it is elliptic wherever $\rho > 0$ and hyperbolic on $\{\rho = 0\}$, see [4]. Consequently, the initial-value problem (i.e. evaluating $(\rho(1, \cdot), v(1, \cdot))$ from $(\rho(0, \cdot), v(0, \cdot))$) is generally ill-posed in standard function spaces, although some solutions may be constructed under analyticity assumptions. For the same reason, a direct shooting method for the boundary-value problem is expected to be numerically unstable.

Another idea is to extend eq. (32), and look for a function $f(t, z)$ that is holomorphic on a suitable domain and satisfies

$$\partial_t f + f \partial_z f = 0.$$

The method of complex characteristics then yields an implicit representation of the solution and recovers several of the explicit solutions mentioned above, as discussed in [4, 18, 45]. For general boundary densities, however, the function $f(t, x)$ cannot be analytically extended to the whole complex plane, which precludes the use of the method of characteristics, see [4].

In [4], an interesting relation to integrable systems was unveiled. The author showed that the most probable trajectory under the constrained Dyson Brownian Motion (the “instanton” in the language of [18]) can be exactly mapped to the attractive Calogero-Moser system. Given initial positions and velocities $(x_i(0), v_i(0))$, this yields a formula of the type $x(t) = \text{eig}(M(t))$, for a real non-symmetric matrix $M(t)$ that depends only on the initial conditions. However, in general $M(t)$ ceases to have real eigenvalues for a large enough $t \in [0, 1]$, except for a fine-tuned choice of the initial velocities. This difficulty was noticed in [4], and prevented the author from leveraging this idea for a numerical solution to the boundary-value problem.

Perhaps the most promising alternative to our particle discretization is to solve directly the convex optimization problem in eq. (7) using Eulerian techniques developed for dynamic optimal transport. If the term proportional to ρ^3 is removed in eq. (7), the resulting minimization problem is precisely the dynamic Benamou–Brenier formulation of quadratic optimal transport [56]. Benamou and Brenier discretized this problem on a space–time grid and solved the resulting convex problem using an augmented-Lagrangian method. These methods can be adapted to handle additional terms similar to our ρ^3 , see e.g. [57]. We refer to [58, Chapter 7] for more details on the fluid-dynamics formulation of optimal transport, and numerical schemes tailored to this formulation. Our particle formulation, on the other hand, does not require a fixed spatial grid and exhibits strong numerical performance even for non-smooth densities and velocity fields (cf. the examples considered above), while this might cause difficulties in a grid-discretization approach. For instance, for each fixed $t \in (0, 1)$, the density $\rho(t, \cdot)$ vanishes rather abruptly, typically with a square-root singularity, and this may require specific attention in a grid-discretization approach. An implementation of this alternative method for the asymptotics of the HCIZ integral, and a quantitative comparison to our algorithm, is an interesting future direction.

1.4 Structure of the paper

In Section 2, we present in detail the minimization algorithm used to compute the discretized version of the problem whose validity is established in Theorem 1.2. In Section 3, we present the proof of Theorem 1.2. We detail some additional analytical considerations on the HCIZ integral and Matytsin’s solution, as well as some technical aspects of the proof and derivation, in Appendix A. Finally, some additional numerical experiments on the convergence properties, and the runtime of our algorithm, are presented in Appendix B.

Acknowledgements and use of AI – We thank Alice Guionnet, Govind Menon, Gabriel Peyré and Adrien Taylor for interesting discussions at various stages of this work. We used modern

AI tools, primarily GPT-5.6 Sol as well as earlier versions of GPT and Claude Opus 5, to help designing and improving the numerical code accompanying this work. The manuscript itself was written entirely by the authors, and AI tools were used only to proofread the manuscript. The authors take full responsibility for the content and for any remaining errors.

2 Efficient minimization algorithm

We detail here the numerical procedure used to solve the finite-dimensional variational problem obtained from Theorem 1.2.

Time discretization – We start by defining a simple time discretization scheme. For any $T \geq 1$ and $\Delta t := (\Delta t_k)_{k=1}^T$ such that $\Delta t_k \geq 0$ and $\sum_{k=1}^T \Delta t_k = 1$, we define $\text{Adm}_{N,\Delta t}(\mu_0, \mu_1)$ to be the set of all trajectories $\{(x_{i,k}, v_{i,k})\}$ for $1 \leq i \leq N$ and $0 \leq k \leq T$, satisfying the following.

($\tilde{\text{H1}}$) $(x_{i,0}) = x_i(0)$ and $x_{i,T} = x_i(1)$ as given by eq. (19).

($\tilde{\text{H2}}$) For every $i \in \{1, \dots, N\}$ and $k \in \{0, \dots, T\}$, we have $x_{i,k} \leq x_{i+1,k}$.

($\tilde{\text{H3}}$) For every $i \in \{1, \dots, N\}$ and $k \in \{0, \dots, T\}$:

$$x_{i,k} = x_{i,0} + \sum_{l=1}^k \Delta t_l v_{i,l}.$$

We also let for $k \in \{0, \dots, T\}$:

$$w_k := \begin{cases} \frac{\Delta t_k}{2} & \text{if } k \in \{0, T\}, \\ \frac{\Delta t_k + \Delta t_{k+1}}{2} & \text{if } 1 \leq k \leq T-1. \end{cases} \quad (33)$$

Then, for any $\{(x_{i,k}, v_{i,k})\} \in \text{Adm}_{N,\Delta t}(\mu_0, \mu_1)$, we let

$$\begin{aligned} G_{N,\Delta t,\varepsilon}(\{(x_{i,k}, v_{i,k})\}) &:= \frac{1}{2N} \sum_{k=1}^T \sum_{i=1}^N \Delta t_k v_{i,k}^2 \\ &+ \frac{\pi^2}{6N(N+1)^2} \sum_{k=0}^T \sum_{i=1}^N w_k \left(x_{i+1,k} - x_{i,k} + \frac{\varepsilon}{N+1} \right)^{-2}. \end{aligned} \quad (34)$$

Notice that we use a trapezoidal discretization scheme for the second term in eq. (34). As we detail in Appendix A.2, the objective is to have a global time discretization error in $G_{N,\Delta t,\varepsilon}(\{(x_{i,k}, v_{i,k})\})$ of size $\mathcal{O}(\|\Delta t\|_\infty^2) = \mathcal{O}(1/T^2)$, which is consistent with what we numerically observe (experiments on this point are given in Appendix B.1). We set

$$J_{N,\Delta t,\varepsilon}(\mu_0, \mu_1) := \inf_{\text{Adm}_{N,\Delta t}(\mu_0, \mu_1)} G_{N,\Delta t,\varepsilon}. \quad (35)$$

As mentioned above, it is straightforward to check (by discretizing a near-optimal continuous trajectory on one hand, and building an affine interpolation of a discrete trajectory on the other hand, similarly to the proof shown in Section 3) that

$$\lim_{\substack{T \rightarrow \infty \\ \|\Delta t\|_\infty \rightarrow 0}} J_{N,\Delta t,\varepsilon}(\mu_0, \mu_1) = J_{N,\varepsilon}(\mu_0, \mu_1).$$

Equation (35) defines a *convex* minimization problem over a set of linear constraints (equalities and inequalities). We now detail the specifics of the minimization procedure and algorithm.

2.1 Parametrization by the spacings

We first change variables, to express the objective solely as a function of two variables: the (rescaled) particle spacings

$$\delta_{i,k} := (N+1)(x_{i+1,k} - x_{i,k}), \quad (1 \leq i \leq N-1, 0 \leq k \leq T), \quad (36)$$

and the position of the first particle $(x_{1,k})_{k=0}^T$. This allows us to rewrite the objective in eq. (34) as (notice that we removed the term $i = N$ from the first sum, since $x_{N+1,k} = +\infty$ by convention):

$$\begin{aligned} G_{N,\Delta t,\varepsilon}(\delta, x_1) &= \frac{\pi^2}{6N} \sum_{k=0}^T \sum_{i=1}^{N-1} \frac{w_k}{(\delta_{i,k} + \varepsilon)^2} + \frac{1}{2N} \sum_{k=1}^T \frac{1}{\Delta t_k} \sum_{i=1}^N \left[x_{1,k} - x_{1,k-1} + \frac{1}{N+1} \sum_{j=1}^{i-1} (\delta_{j,k} - \delta_{j,k-1}) \right]^2. \end{aligned} \quad (37)$$

One should minimize eq. (37) over $(x_{1,k})_{k=0}^{T-1}$ and $(\delta_{i,k})_{\substack{1 \leq i \leq N-1, \\ 1 \leq k \leq T-1}}$, under the sole constraint that $\delta_{i,k} \geq 0$, and recalling that the boundary values are fixed as:

$$\begin{cases} x_{1,0} &= x_1(0), \\ x_{1,T} &= x_1(1), \\ \delta_{i,0} &= (N+1)(x_{i+1}(0) - x_i(0)), \\ \delta_{i,T} &= (N+1)(x_{i+1}(1) - x_i(1)), \end{cases} \quad (38)$$

where $(x_i(0), x_i(1))$ are given by eq. (19). Parameterizing by the spacings has thus simplified the linear constraints to $\delta_{i,k} \geq 0$, which imposes the non-crossing of the particle trajectories. Further, the minimization over $(x_{1,k})_{k=0}^{T-1}$ is a simple quadratic optimization problem which we can solve analytically. This yields (using the same notation $G_{N,\Delta t,\varepsilon}$ slightly abusively):

$$\begin{aligned} G_{N,\Delta t,\varepsilon}(\delta) &= \frac{1}{2} [\bar{x}(0) - \bar{x}(1)]^2 + \frac{\pi^2}{6N} \sum_{k=0}^T \sum_{i=1}^{N-1} \frac{w_k}{(\delta_{i,k} + \varepsilon)^2} + \frac{1}{2(N+1)^2} \sum_{k=1}^T \Delta t_k (\mathrm{d}\delta_k^\top) L_{N-1}^{-1} (\mathrm{d}\delta_k), \end{aligned} \quad (39)$$

where we defined the following quantities:

$$\begin{cases} \bar{x}(t) &:= \frac{1}{N} \sum_{i=1}^N x_i(t), \quad (t \in \{0, 1\}) \\ (\mathrm{d}\delta)_{i,k} &:= \frac{\delta_{i,k} - \delta_{i,k-1}}{\Delta t_k}, \quad (1 \leq i \leq N-1 \text{ and } 1 \leq k \leq T) \\ L_{N-1} &:= N \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \cdots & 0 \\ 0 & -1 & 2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 2 \end{pmatrix} \in \mathbb{R}^{(N-1) \times (N-1)}. \end{cases} \quad (40)$$

We detail the derivation of eq. (39) for completeness in Appendix A.3. The first term in eq. (39) is a deterministic function of the boundary densities, and is not part of the optimization objective.

On the constraint $\delta_{i,k} \geq 0$ – Notice that $G_{N,\Delta t,\varepsilon}(\delta)$ is strongly convex on its domain $\{\delta_{i,k} \geq 0\}$ and therefore has a unique minimum. Rather than imposing the constraint $\delta_{i,k} \geq 0$, we extend the domain of definition to all $\delta_{i,k} \in \mathbb{R}$ and replace $(\delta_{i,k} + \varepsilon)^{-2}$ by $\sigma_\varepsilon(\delta_{i,k})$, with

$$\sigma_\varepsilon(x) := \begin{cases} (x + \varepsilon)^{-2} & \text{if } x \geq 0, \\ 6\varepsilon^{-2} - 8\varepsilon^{-3}(x + \varepsilon) + 3\varepsilon^{-4}(x + \varepsilon)^2 & \text{if } x < 0. \end{cases} \quad (41)$$

It is trivial to check that σ_ε is of class $\mathcal{C}^2$ and convex over $\mathbb{R}$. In practice, we choose $\varepsilon > 0$ very small (10^{-7} in the simulations presented), and we check that, at the output of the algorithm, we have $\delta_{i,k} \geq 0$, which ensures that the minimum we reach is indeed the global minimum of $G_{N,\Delta t,\varepsilon}$. This will allow us to use methods designed for *unconstrained* convex minimization problems. To sum up, we extend the definition of $G_{N,\Delta t,\varepsilon}$ to arguments that are not necessarily in the nonnegative quadrant, and minimize

$$\begin{aligned} G_{N,\Delta t,\varepsilon}(\delta) &= \frac{1}{2} [\bar{x}(0) - \bar{x}(1)]^2 + \frac{\pi^2}{6N} \sum_{k=0}^T \sum_{i=1}^{N-1} w_k \sigma_\varepsilon(\delta_{i,k}) + \frac{1}{2(N+1)^2} \sum_{k=1}^T \Delta t_k (\mathrm{d}\delta_k^\top) L_{N-1}^{-1}(\mathrm{d}\delta_k), \end{aligned} \quad (42)$$

over all $\delta_{i,k} \in \mathbb{R}$.

2.2 Time discretization

We now detail the choice of Δt used in the algorithm. To motivate it, we use a free probability interpretation of the solution (ρ, ν) in Theorem 1.1; we refer to [53] for an introduction to free probability in the context of random matrices. As explained in [49, p. 535], under the setting of Theorem 1.1, there exist self-adjoint operators X_0, X_1, S in a common W^* -probability space such that X_0 has law μ , X_1 has law ν , S is a semi-circular variable that is free from (X_0, X_1) , and for every $t \in [0, 1]$, the measure $\rho(t, x) \mathrm{d}x$ is the law of the free Brownian bridge

$$X_t := (1-t)X_0 + tX_1 + \sqrt{t(1-t)}S. \quad (43)$$

In particular, for small times $t \ll 1$, X_t is, to first order, close to the additive free convolution $X_0 + \sqrt{t}S$ (and similarly for times $t \rightarrow 1$). For a given value $T \geq 1$ (assumed even for simplicity), this motivates a non-linear time step schedule $(\Delta t_k)_{k=1}^T$, defined by:

$$t_k := \begin{cases} \frac{2k^2}{T^2} & , \quad (k \in \{0, \dots, T/2\}), \\ \frac{4k}{T} - 1 - \frac{2k^2}{T^2} & , \quad (k \in \{T/2 + 1, \dots, T\}). \end{cases} \quad (44)$$

$$\Delta t_k = t_k - t_{k-1} = \begin{cases} \frac{4}{T^2} \left(k - \frac{1}{2}\right) & , \quad (k \in \{1, \dots, T/2\}), \\ \Delta t_{T-k+1} & , \quad (k \in \{T/2 + 1, \dots, T\}). \end{cases} \quad (45)$$

2.3 Preconditioning

2.3.1 An ill-conditioned minimization problem

Notice that the Hessian of $G_{N,\Delta t,\varepsilon}$ is naturally ill-conditioned. We obtain from eq. (42) (we also state the gradient for completeness):

$$\begin{cases} \nabla_{i,k} G_{N,\Delta t,\varepsilon} = \frac{\pi^2}{6N} w_k \sigma'_\varepsilon(\delta_{i,k}) + \frac{1}{(N+1)^2} \left([L_{N-1}^{-1} \otimes L_{T-1}(\Delta t)] \delta \right)_{i,k}, \end{cases} \quad (46a)$$

$$\begin{cases} \nabla^2 G_{N,\Delta t,\varepsilon} = \frac{\pi^2}{6N} \text{Diag} [w_k \sigma''_\varepsilon(\delta_{i,k})] + \frac{1}{(N+1)^2} L_{N-1}^{-1} \otimes L_{T-1}(\Delta t), \end{cases} \quad (46b)$$

where we define a Laplacian matrix for $p \geq 1$ and any $a = (a_1, \dots, a_{p+1}) \in \mathbb{R}^{p+1}$ such that $a_i \geq 0$ and $\sum_{i=1}^{p+1} a_i = 1$:

$$L_p(a) := \begin{pmatrix} a_1^{-1} + a_2^{-1} & -a_2^{-1} & 0 & \cdots & 0 \\ -a_2^{-1} & a_2^{-1} + a_3^{-1} & -a_3^{-1} & \cdots & 0 \\ 0 & -a_3^{-1} & a_3^{-1} + a_4^{-1} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_p^{-1} + a_{p+1}^{-1} \end{pmatrix} \in \mathbb{R}^{p \times p}, \quad (47)$$

This is consistent with eq. (40), as we also set by convention

$$L_p := L_p \left(\left(\frac{1}{p+1} \right)_{i=1}^{p+1} \right).$$

From eq. (46b), we can see from an informal argument that the Hessian is generically ill-conditioned. It is well known that the eigenvalues of L_{p-1} satisfy

$$\lambda_k(L_{p-1}) = 4p \sin^2 \left(\frac{k\pi}{2p} \right), \quad (1 \leq k \leq p-1) \quad (48)$$

Assuming $\Delta t_k = 1/T$ for simplicity, this informal argument decomposes eq. (46b) into two terms:

- A diagonal and positive matrix arising from the ρ^3 term in eq. (18), with eigenvalues of order $\Theta(N^{-1}T^{-1})$ assuming the spacings are of order $\delta_{ik} = \Theta(1)$.
- A second term arising from the ρv^2 “optimal transport” objective: it is positive definite, with smallest eigenvalue of order $\Theta(N^{-3}T^{-1})$, and largest eigenvalue scaling as $\Theta(N^{-1}T)$.

From this heuristic argument, we expect the condition number of the Hessian to be $\Theta(T^2)$.

Alternative parametrization – Notice that if we were to consider the simple optimal transport problem, we would drop the first term in eq. (46b), which would yield a very large Hessian condition number $\Theta(N^2T^2)$. This apparent ill-conditioning would however only be due to a bad choice of parametrization: the optimal transport objective is proportional to $\int \rho v^2$, and its discretization is therefore quadratic in the velocities $v_{i,k}$: if we were to use these velocities as parameters, the ill-conditioning of the Hessian would disappear¹. However, this does not extend to our setting: using the velocities as parameters, the Hessian of the objective becomes

$$\nabla_v^2 G_{N,\Delta t,\varepsilon} = \frac{\pi^2}{6N} A_{N,T}^\top \text{Diag} \left[(w_k \sigma_\varepsilon''(\delta_{i,k}))_{\substack{1 \leq i \leq N-1 \\ 1 \leq k \leq T-1}} \right] A_{N,T} + \frac{1}{N} \text{Diag} \left[(\Delta t_k)_{k=1}^T \right] \otimes I_N, \quad (49)$$

where $A_{N,T} := (N+1)(\mathcal{I}_T \otimes D_N)$, with $\mathcal{I}_T$ the discrete integration operator in time (restricted to the linear subspace of velocities which are compatible with the boundary conditions), and D_N the first-difference operator in the particle index. Informally, $A_{N,T}$ maps velocities to the corresponding spacings. Considering constant time steps for simplicity, it is easy to see that the smallest and largest singular values of $A_{N,T}$ scale as

$$\sigma_{\min}(A_{N,T}) = \Theta(T^{-1}), \quad \sigma_{\max}(A_{N,T}) = \Theta(N). \quad (50)$$

Assuming that the spacings remain of order $\delta_{i,k} = \Theta(1)$ and bounded away from zero, eq. (49) thus yields a total condition number $\kappa(\nabla_v^2 G_{N,\Delta t,\varepsilon}) = \Theta(N^2)$. In the regimes considered numerically, where $N \gtrsim T$, the resulting condition number is no smaller than the $\Theta(T^2)$ condition number obtained in spacing variables. Since parametrizing with the velocities also requires a careful handling of the non-crossing constraints, this is further motivation for considering a parametrization of the objective by the spacings.

¹Notice that some bad conditioning could still arise during optimization from the linear constraints arising from the non-crossing conditions, but we do not discuss this here.

2.3.2 Preconditioning procedure

In order to overcome this ill-conditioning issue, we use a preconditioning procedure that approximates the inverse Hessian of eq. (46b). We build this preconditioner by replacing the first, diagonal, term of eq. (46b) by a block-diagonal average of its values, i.e. we define:

$$\begin{cases} H(\delta) &:= \mathbf{I}_{N-1} \otimes \text{Diag}[(\Gamma_k(\delta_k))_{k=1}^{T-1}] + \frac{1}{(N+1)^2} L_{N-1}^{-1} \otimes L_{T-1}(\Delta t), \\ \Gamma_k(\delta_k) &:= \frac{\pi^2 \omega_k}{6N(N-1)} \sum_{i=1}^{N-1} \sigma''_{\varepsilon}(\delta_{i,k}). \end{cases} \quad (51)$$

Notice that we expect $\Gamma_k(\delta_k) = \Theta(N^{-1}T^{-1})$. In the preconditioning procedure, we need to compute $H(\delta)^{-1}z$ in a computationally efficient manner, for $z \in \mathbb{R}^{N-1} \otimes \mathbb{R}^{T-1}$. Let us denote the eigenbasis decomposition

$$L_{N-1} := 4 \sum_{i=1}^{N-1} \sin^2\left(\frac{i\pi}{2N}\right) v_i^{(N)} \left[v_i^{(N)}\right]^\top.$$

Importantly, we have

$$(v_i^{(N)})_j := \sqrt{\frac{2}{N}} \sin\left(\frac{ij\pi}{N}\right), \quad (1 \leq i, j \leq N-1).$$

For any $u \in \mathbb{R}^N$, this allows us to compute the vector

$$\text{DST}(u) := (v_i^{(N)} \cdot u)_{i=1}^{N-1} = \left(\frac{2}{\sqrt{N}} \sum_{j=1}^{N-1} \sin\left(\frac{ij\pi}{N}\right) u_j \right)_{i=1}^{N-1}$$

in $\mathcal{O}(N \log N)$ operations using classical fast Fourier transform algorithms.

Moreover, notice that the matrix of eq. (51) restricted to a given sine spatial mode is tri-diagonal and positive definite. Therefore, when computing $w := H(\delta)^{-1}z$ for $z = (z_1, \dots, z_{N-1}) \in \mathbb{R}^{N-1} \otimes \mathbb{R}^{T-1}$, we need to solve $(N-1)$ independent tri-diagonal inverse problems, and each of these can be solved in $\mathcal{O}(T)$ operations using a cyclic reduction algorithm [59]. Combined with the discrete sine transform algorithm for the space modes, this allows us to solve the whole linear system in $\mathcal{O}(NT \log N)$ operations.

2.4 Global scaling and the θ parameter

While the global scaling parameter $\theta > 0$ in $I_{\text{HCIZ}}(\theta, \mu, \nu)$ can be formally incorporated in a rescaling of (μ, ν) , it can be useful to keep it explicitly, in particular to avoid numerical issues in the limits $\theta \ll 1$ and $\theta \gg 1$. As detailed in Appendix A.1, one can take this rescaling into account by modifying the variational principle of eqs. (17) and (18) as:

$$\begin{cases} J(\theta, \mu, \nu) &:= \inf_{(\rho, v) \in \text{Adm}_\theta(\mu, \nu)} G(\theta, \rho, v), \\ G(\theta, \rho, v) &:= \frac{1}{2} \int_0^{\theta^{-1}} du \int_{\mathbb{R}} dx \left(v(u, x)^2 + \frac{\pi^2}{3} \rho(u, x)^2 \right) \rho(u, x). \end{cases} \quad (52)$$

The definition of admissible trajectories in Definition 1.1 remains strictly identical, except that the time interval is now $u \in (0, \theta^{-1})$ (changing the name of the variable). The minimization algorithm can also be adapted to handle this rescaling: we detail how its different ingredients are affected in Appendix A.4.

2.5 Initialization, and final algorithm

Initialization – As mentioned above, we consider a minimization problem which generalizes the simple one-dimensional optimal transport. This motivates initializing the algorithm at the solution to this problem, which is simply given by a linear interpolation of the quantiles of μ and ν , i.e. we let

$$\delta_{i,k}^{(0)} := (N+1) [(1-t_k)(x_{i+1}(0) - x_i(0)) + t_k(x_{i+1}(1) - x_i(1))], \quad (53)$$

and we recall that $(x_i(0))_{i=1}^N$ and $(x_i(1))_{i=1}^N$ are the quantiles of $\mu_0 = \mu$ and $\mu_1 = \nu$, given in eq. (19).

Preconditioned conjugate gradient – In Appendix A.5, we recall the classical preconditioned conjugate gradient (PCG) algorithm. Given a positive-definite symmetric matrix A , a vector b , and a positive-definite symmetric matrix H (the *preconditioner*), it returns $\text{PCG}(H; A, b)$ which is the unique solution to the system $Ax = b$. In our simulations, we used a relative precision target $\varepsilon_{\text{PCG}} = 10^{-3}$ for the PCG algorithm.

Final algorithm – We are now ready to state our minimization algorithm. We use a simple Newton method, using the PCG algorithm to efficiently find a solution to the Newton system. We state it in Algorithm 1. Notice furthermore that the simple form of the Hessian in eq. (46b) allows us to perform the associated matrix-vector multiplications in time $\mathcal{O}(NT)$ rather than the naive $\mathcal{O}(N^2T^2)$. All in all, the complexity of each Newton iteration is dictated to first order by multiplying the number of iterations of the preconditioned conjugate gradient algorithm, and the complexity of inverting a linear system associated to the preconditioner $H(\delta)$, which is $\mathcal{O}(NT \log N)$ as detailed in Section 2.3.

Algorithm 1: Newton method with preconditioned conjugate gradient.

Result: The value $J_{N,\Delta t,\varepsilon}(\mu_0, \mu_1) = \inf_{\text{Adm}_{N,\Delta t}(\mu_0, \mu_1)} G_{N,\Delta t,\varepsilon}$, and the corresponding spacings $\delta_{i,k}$.

Input: The quantiles $\{x_i(0), x_i(1)\}_{i=1}^N$ of eq. (19), the time spacings $\{\Delta t_k\}_{k=1}^T$ of eq. (45), a small $\varepsilon > 0$;

Initialize $t = 0$ and $\delta_{i,k} = \delta_{i,k}^{(0)}$ according to eq. (53);

while not converged do

Compute the Newton direction

$$p^{(t)} = -\text{PCG}\left(H(\delta^{(t)}); \nabla^2 G_{N,\Delta t,\varepsilon}(\delta^{(t)}), \nabla G_{N,\Delta t,\varepsilon}(\delta^{(t)})\right);$$

- $\nabla G_{N,\Delta t,\varepsilon}$ and $\nabla^2 G_{N,\Delta t,\varepsilon}$ are given in eq. (46).
- $H(\delta^{(t)})$ is given in eq. (51), along with a fast FFT-based procedure for solving the associated linear systems.

Find the step size $\alpha > 0$ with backtracking line search [60];

Set $\delta^{(t+1)} = \delta^{(t)} + \alpha p^{(t)}$;

$t \leftarrow t + 1$;

end

Stopping criterion – We use as a stopping criterion the value of the Newton decrement $\lambda_{\text{New.}}^2 := [(\nabla G_{N,\Delta t,\varepsilon})^\top (\nabla^2 G_{N,\Delta t,\varepsilon})^{-1} \nabla G_{N,\Delta t,\varepsilon}]^{-1/2}$, which is exactly the decrease in the quadratic approximation obtained by a full Newton step [60]. We stop whenever $\lambda_{\text{New.}}^2 < \varepsilon_{\text{tol.}}$, for a given $\varepsilon_{\text{tol.}} > 0$ (we take $\varepsilon_{\text{tol.}} = 10^{-6}$ in the results presented in this paper).

2.6 Practical implementation and final remarks

Implementation – An implementation of Algorithm 1 is provided in a public [Github repository](#) [61]: it uses the Pytorch library [62], and all simulations shown were made on a single Nvidia GPU (either V100, RTX6000 or RTX8000), with running times for solving each instance ranging from a few seconds to a few hours for the largest sizes and hardest problems shown. We used $\varepsilon = 10^{-7}$ generically in all the instances shown, and we checked that the resulting spacings $\delta_{i,k}$ are positive at convergence, so that none of the non-crossing constraints $\delta_{i,k} > 0$ is saturated.

Convergence properties and computation time – In Appendix B, we investigate the numerical convergence and total runtime of the algorithm for several pairs of boundary measures and a range of discretization sizes N and T . For fixed N , the objective values generally exhibit convergence at a rate $\mathcal{O}(T^{-2})$, consistent with the trapezoidal discretization used in eq. (34) for the ρ^3 term. For fixed T , we observe an error of order $\mathcal{O}(N^{-1})$, which can be further improved by Richardson extrapolation [63]. We note that this $\mathcal{O}(1/N)$ accuracy is also commensurate with the discretization of the boundary measures we use: in practical applications, the evaluation of the HCIZ integral is often done alongside that of various linear statistics of these distributions, whose evaluation from these quantiles generically induces another error $\mathcal{O}(1/N)$.

Alternative approaches to the minimization procedure – There are many local optimization algorithms for large-scale convex optimization problems. We detail here our implementation of the Newton–conjugate gradient algorithm, as we empirically found this approach to outperform first-order methods (even preconditioned), such as Adam [64], and approximate second-order methods, such as LBFGS, that are based on a low-rank approximation of the Hessian [65]. Alternatively to replacing the function $(x + \varepsilon)^{-2}$ for $x \geq 0$ by the convex continuation $\sigma_\varepsilon(x)$ defined for all $x \in \mathbb{R}$ and checking *a posteriori* that the minimizer found satisfies the positivity constraint, this constraint can be implemented e.g. by a penalty function or by methods such as the augmented Lagrangian; we refer to [66] for more details. We leave a study of a quantitative comparison to alternative procedures (including also different formulations of the problem, as discussed in Section 1.3) to future work.

3 Proof of Theorem 1.2

This section is devoted to the proof of Theorem 1.2. As a continuous limit of $G_{N,\varepsilon}$ in eq. (22), it will be convenient to introduce, for each $\varepsilon > 0$, the quantity

$$G_\varepsilon(\rho, v) := \frac{1}{2} \int_0^1 \int_{\mathbb{R}} \left(v(t, x)^2 + \frac{\pi^2}{3} \left(\frac{\rho(t, x)}{1 + \varepsilon \rho(t, x)} \right)^2 \right) \rho(t, x) \, dx \, dt, \quad (54)$$

as well as

$$J_\varepsilon(\mu_0, \mu_1) := \inf_{(\rho, v) \in \text{Adm}(\mu_0, \mu_1)} G_\varepsilon(\rho, v). \quad (55)$$

We then have $G(\rho, v) = G_{\varepsilon=0}(\rho, v)$ and $J(\mu_0, \mu_1) = J_{\varepsilon=0}(\mu_0, \mu_1)$. We first recall the following theorem from [49].

Theorem 3.1 (Properties of the minimizer of G [49]). *For every $\mu_0, \mu_1 \in \mathcal{P}_c(\mathbb{R})$ with $\Sigma(\mu_0)$ and $\Sigma(\mu_1)$ finite, the functional G admits a unique minimizer (ρ^*, v^*) over the set $\text{Adm}(\mu_0, \mu_1)$. Moreover, there exists a compact set $K \subset \mathbb{R}$ such that for all $t \in (0, 1)$, the function $\rho^*(t, \cdot)$ is supported on K .*

Remark 3.1. For $(\rho, v) \in \text{Adm}(\mu_0, \mu_1)$, the function v is an element of $L^1((0, 1) \times \mathbb{R}; \rho(t, x) \, dt \, dx)$, and it is in this sense that the uniqueness of the minimizer stated in Theorem 3.1 should be understood. In other words, any two minimizers (ρ_1, v_1) and (ρ_2, v_2) of G over $\text{Adm}(\mu_0, \mu_1)$ must be such that $(\rho_1, \rho_1 v_1) = (\rho_2, \rho_2 v_2)$ on a set of full Lebesgue measure.

Recall the definition of the discretization scheme in Section 1.2.1. The proof of Theorem 1.2 will be decomposed into a number of intermediate results. We first prove Proposition 3.2, which shows continuity with respect to the limit $\varepsilon \rightarrow 0$ in the definition of $J_\varepsilon(\mu_0, \mu_1)$. We then move to Propositions 3.3 and 3.4, proved respectively in Sections 3.2 and 3.3. These propositions give upper and lower bounds on $J(\mu_0, \mu_1)$ and, together with Proposition 3.2, directly imply Theorem 1.2. We finish with a remark in Section 3.4 on a conjectural alternative formulation of $J(\mu_0, \mu_1)$ that is more directly related to our discretization scheme.

3.1 Small-regularization limit

We start by showing that the quantity J_ε , defined in eq. (55), converges to J as ε tends to zero¹. First, we clearly have that $G_\varepsilon \leq G$, so that $J_\varepsilon \leq J$. In order to show the converse inequality, we consider approximate minimizers of J_ε , and show that, after passing to a subsequence, we can obtain a candidate minimizer for the functional J as a limit of these approximate minimizers. In essence, the argument is very similar to one that would guarantee the existence of minimizers for J , and is based on the convexity of the functional G seen as a function of $(\rho, \rho v)$.

We remark that the existence of a (unique) minimizer for G_ε for a fixed $\varepsilon > 0$ would be more subtle to prove, as the functional G_ε is less coercive than G ; in particular, approximate minimizers of G_ε might be such that, for $t \in (0, 1)$, the measure $\rho(t, x) dx$ converges to a probability measure that is not absolutely continuous with respect to the Lebesgue measure. This difficulty might be addressed e.g. by assuming that μ_0 and μ_1 are absolutely continuous with respect to the Lebesgue measure and using some arguments in the spirit of the theory of optimal transport, or by enlarging the space of minimizer candidates; but for our purposes we will simply avoid this issue throughout.

Proposition 3.2. *For every $\mu_0, \mu_1 \in \mathcal{P}(\mathbb{R})$, we have*

$$\lim_{\varepsilon \rightarrow 0} J_\varepsilon(\mu_0, \mu_1) = J(\mu_0, \mu_1). \quad (56)$$

Proof of Proposition 3.2. The existence of the limit on the left-hand side is guaranteed by monotone convergence. The inequality $\leq$ between the left and the right sides of eq. (56) is clear, so we turn to the converse bound. Without loss of generality, we can assume that the left side of eq. (56) is finite; we denote it by $\mathcal{I} \in [0, +\infty)$. For each $\varepsilon \in (0, 1]$, we let $(\rho_\varepsilon, v_\varepsilon) \in \text{Adm}(\mu_0, \mu_1)$ be such that

$$G_\varepsilon(\rho_\varepsilon, v_\varepsilon) \leq \varepsilon + J_\varepsilon(\mu_0, \mu_1), \quad (57)$$

and we write $m_\varepsilon := \rho_\varepsilon v_\varepsilon$. By the continuity equation and [48, Lemma 8.1.2], we have for every $\phi \in C_c^\infty(\mathbb{R})$ and $0 < s < t < 1$ that

$$\int \phi(x)(\rho_\varepsilon(t, x) - \rho_\varepsilon(s, x)) dx = \int_s^t \int_{\mathbb{R}} \partial_x \phi(x) m_\varepsilon(u, x) dx du,$$

and thus, by the Kantorovich duality (e.g. [67, Theorem 1.14]), we have

$$W_1(\rho_\varepsilon(s, x) dx, \rho_\varepsilon(t, x) dx) \leq \int_s^t \int_{\mathbb{R}} |m_\varepsilon(t, x)| dx dt,$$

where W_1 denotes the L^1 -Wasserstein distance. We also observe that

$$\begin{aligned} \int_s^t \int_{\mathbb{R}} |m_\varepsilon(t, x)| dx dt &\leq \left(\int_s^t \int_{\mathbb{R}} \frac{|m_\varepsilon(t, x)|^2}{\rho_\varepsilon(t, x)} dx dt \right)^{\frac{1}{2}} \left(\int_s^t \int_{\mathbb{R}} \rho_\varepsilon(t, x) dx dt \right)^{\frac{1}{2}} \\ &\leq \sqrt{(2\mathcal{I} + 2)|t - s|}. \end{aligned}$$

¹This could be phrased as the Γ -convergence of G_ε to G , but we will refrain from using this language here.

The functions $t \mapsto \rho_\varepsilon(t, x)dx$ are thus uniformly equicontinuous with respect to the W_1 distance on probability measures. Since $\lim_{t \rightarrow 0} \rho_\varepsilon(t, x)dx = \mu_0$ is fixed, after passing to a subsequence, we can assume that these functions converge to some continuous map $t \mapsto \mu(t, dx)$ by the Arzelà–Ascoli theorem. The display above also shows that $|m_\varepsilon(t, x)|dxdt$ is a non-negative measure with a total mass that is uniformly bounded over ε ; we can therefore extract a subsequence that converges vaguely to a non-negative measure $\mathbf{m}$ over $[0, 1] \times \mathbb{R}$. One can directly check that the pair $(\mu(t, dx)dt, \mathbf{m})$ satisfies the continuity equation in the weak sense, in the sense that $\partial_t \mu + \partial_x \mathbf{m} = 0$.

We now observe that $\mu(t, dx)dt$ is in fact absolutely continuous with respect to the Lebesgue measure. Indeed, we first observe that for all $M \in (0, +\infty)$ and $\varepsilon > 0$ sufficiently small (in terms of M),

$$\begin{aligned} \int_0^1 \int_{\mathbb{R}} \mathbb{1}_{\{\rho_\varepsilon(t, x) \geq M\}} \rho_\varepsilon(t, x) dx dt &\leq 4M^{-2} \int_0^1 \int_{\mathbb{R}} \left(\frac{\rho_\varepsilon(t, x)}{1 + \varepsilon \rho_\varepsilon(t, x)} \right)^2 \rho_\varepsilon(t, x) dx dt \\ &\leq 24\pi^{-2} M^{-2} (\mathcal{I} + 1). \end{aligned}$$

Hence, for every measurable set $A \subset [0, 1] \times \mathbb{R}$ and $\varepsilon > 0$ sufficiently small, we have

$$\int_0^1 \int_{\mathbb{R}} \mathbb{1}_A(t, x) \rho_\varepsilon(t, x) dx dt \leq M|A| + 24\pi^{-2} M^{-2} (\mathcal{I} + 1), \quad (58)$$

where $|A|$ denotes the Lebesgue measure of A . Recall that for every bounded Borel set A , one can find an open set U containing A such that $|U \setminus A|$ is as small as desired (see e.g. [68, Exercise A.4 and solution]). Using this, the estimate above, and the Portmanteau theorem, we obtain that $\mu(t, dx)dt$ is absolutely continuous with respect to the Lebesgue measure. We can therefore write $\mu(t, dx)dt = \rho(t, x)dxdt$ for some probability density $\rho \in L^1([0, 1] \times \mathbb{R})$. Let us fix $K \in [0, +\infty)$ and denote

$$\chi_K(x) := (3 - K^{-1}) \int_0^x (u \wedge K)^2 du. \quad (59)$$

The function χ_K is convex over $[0, +\infty)$, and for $\varepsilon > 0$ sufficiently small (in terms of K), we have $\chi_K(x) \leq x^3/(1 + \varepsilon x)^2$ for all $x \geq 0$. The weak convergence of ρ_ε to ρ and the convexity of χ_K ensure that

$$\int_0^1 \int_{\mathbb{R}} \chi_K(\rho(t, x)) dx dt \leq \liminf_{\varepsilon \rightarrow 0} \int_0^1 \int_{\mathbb{R}} \chi_K(\rho_\varepsilon(t, x)) dx dt,$$

and the comparison between χ_K and $x \mapsto x^3/(1 + \varepsilon x)^2$ for small $\varepsilon > 0$ yields that

$$\int_0^1 \int_{\mathbb{R}} \chi_K(\rho(t, x)) dx dt \leq \liminf_{\varepsilon \rightarrow 0} \int_0^1 \int_{\mathbb{R}} \left(\frac{\rho_\varepsilon(t, x)}{1 + \varepsilon \rho_\varepsilon(t, x)} \right)^2 \rho_\varepsilon(t, x) dx dt.$$

By the monotone convergence theorem, we can then send K to infinity on the left side to obtain that

$$\int_0^1 \int_{\mathbb{R}} \rho(t, x)^3 dx dt \leq \liminf_{\varepsilon \rightarrow 0} \int_0^1 \int_{\mathbb{R}} \left(\frac{\rho_\varepsilon(t, x)}{1 + \varepsilon \rho_\varepsilon(t, x)} \right)^2 \rho_\varepsilon(t, x) dx dt. \quad (60)$$

We next notice that, for every $x, y \in \mathbb{R}$,

$$\sup_{t \in \mathbb{R}} (2tx - t^2y) = \begin{cases} \frac{x^2}{y} & \text{if } y > 0, \\ 0 & \text{if } x = y = 0, \\ +\infty & \text{otherwise.} \end{cases} \quad (61)$$

Hence, for every $\phi \in C_c^\infty([0, 1]; \mathbb{R})$, we have

$$\int_0^1 \int_{\mathbb{R}} (2\phi m_\varepsilon - \phi^2 \rho_\varepsilon)(t, x) \, dx \, dt \leq \int_0^1 \int_{\mathbb{R}} \frac{|m_\varepsilon(t, x)|^2}{\rho_\varepsilon(t, x)} \, dx \, dt \leq 2\mathcal{I} + 2\varepsilon. \quad (62)$$

Passing to the limit, we obtain that

$$\int 2\phi(t, x) \, d\mathbf{m}(t, x) - \int_0^1 \int_{\mathbb{R}} \phi(t, x)^2 \rho(t, x) \, dx \, dt \leq 2\mathcal{I}. \quad (63)$$

Replacing ϕ by $\lambda\phi$ for arbitrary $\lambda \in \mathbb{R}$ and calculating the discriminant of the resulting quadratic function of λ , we obtain

$$\left(\int \phi(t, x) \, d\mathbf{m}(t, x) \right)^2 \leq 2\mathcal{I} \int_0^1 \int_{\mathbb{R}} \phi(t, x)^2 \rho(t, x) \, dx \, dt.$$

By the Riesz representation theorem, we deduce that there exists $v \in L^2([0, 1] \times \mathbb{R}; \rho(t, x) \, dx \, dt)$ such that $d\mathbf{m}(t, x) = v(t, x) \rho(t, x) \, dx \, dt$. Recalling the first inequality in (62), and using also (60), we thus obtain that for every $\phi \in C_c^\infty([0, 1]; \mathbb{R})$,

$$\frac{1}{2} \int_0^1 \int_{\mathbb{R}} (2\phi v \rho - \phi^2 \rho + \frac{\pi^2}{3} \rho^3)(t, x) \, dx \, dt \leq \mathcal{I}. \quad (64)$$

Optimizing over ϕ yields that

$$\frac{1}{2} \int_0^1 \int_{\mathbb{R}} \left(v(t, x)^2 + \frac{\pi^2}{3} \rho(t, x)^2 \right) \rho(t, x) \, dx \, dt \leq \mathcal{I},$$

and we recall that the pair (ρ, v) satisfies the continuity equation (5). We have thus upper-bounded the right-hand side of (56) by the left-hand side of this display, as desired. $\square$

3.2 Upper bound

Our next goal is to upper bound $J_\varepsilon(\mu_0, \mu_1)$ by $\liminf_{N \rightarrow \infty} J_{N, \varepsilon}(\mu_0, \mu_1)$. We will show that from each $\{(x_i, v_i)\}_{i=1}^N \in \text{Adm}_N(\mu_0, \mu_1)$, we can build a continuous pair (ρ_N, v_N) such that $G_\varepsilon(\rho_N, v_N) = G_{N, \varepsilon}(\{(x_i, v_i)\})$. A difficulty will arise from the fact that $\rho_N(0, x) \, dx$ is almost but not exactly equal to μ_0 . As discussed in the paragraph preceding Proposition 3.2, we prefer to avoid having to construct minimizers for G_ε , so instead of simply letting N tend to infinity to construct some $(\rho, v) \in \text{Adm}(\mu_0, \mu_1)$ with $G_\varepsilon(\rho, v) \leq J_{N, \varepsilon}(\mu_0, \mu_1)$, we also let $\varepsilon \rightarrow 0$ at the same time and obtain a candidate for G in place of G_ε . We thus do not really show that $J_\varepsilon \leq \liminf_N J_{N, \varepsilon}$, but only the following weaker statement, which is sufficient for our purposes.

Proposition 3.3. *For every $\mu_0, \mu_1 \in \mathcal{P}(\mathbb{R})$ with $\Sigma(\mu_0), \Sigma(\mu_1)$ finite, we have*

$$J(\mu_0, \mu_1) \leq \lim_{\varepsilon \rightarrow 0} \liminf_{N \rightarrow +\infty} J_{N, \varepsilon}(\mu_0, \mu_1). \quad (65)$$

Proof of Proposition 3.3. Let $\{(x_i, v_i)\} \in \text{Adm}_N(\mu_0, \mu_1)$. Without loss of generality, we assume that $x_i(t) < x_{i+1}(t)$ for every $i \in \{1, \dots, N-1\}$ and $t \in [0, 1]$ (indeed, if this were not the case, we could modify the v 's ever so slightly to enforce this constraint while changing $G_{N, \varepsilon}(x, v)$ infinitesimally). For every $t \in [0, 1]$ and $i \in \{1, \dots, N\}$, we set

$$f(t, x_i(t)) := \frac{i-1}{N-1} \quad \text{and} \quad v(t, x_i(t)) := v_i(t).$$

We extend $f(t, \cdot)$ and $v(t, \cdot)$ to continuous functions that are affine on the intervals $[x_i(t), x_{i+1}(t)]$ ($i \in \{1, \dots, N-1\}$), with $f(t, x) = 0$ for $x \leq x_1(t)$ and $f(t, x) = 1$ for $x \geq x_N(t)$; for definiteness,

we can also extend $v(t, \cdot)$ so that it is constant on $(-\infty, x_1(t)]$ and on $[x_N(t), +\infty)$, but this will not play any role. Explicitly, for every $i \in \{1, \dots, N-1\}$ and $y \in [0, x_{i+1}(t) - x_i(t)]$, we have

$$f(t, x_i(t) + y) = \frac{i-1}{N-1} + \frac{y}{(N-1)(x_{i+1}(t) - x_i(t))}.$$

Hence, for every $y \in (0, x_{i+1}(t) - x_i(t))$, we have

$$\partial_x f(t, x_i(t) + y) = \frac{1}{(N-1)(x_{i+1}(t) - x_i(t))} \quad (66)$$

and

$$(\partial_t f + v_i(t) \partial_x f)(t, x_i(t) + y) = -\frac{y(v_{i+1}(t) - v_i(t))}{(N-1)(x_{i+1}(t) - x_i(t))^2}.$$

Combining the two previous displays yields that

$$\partial_t f(t, x_i(t) + y) + \left(v_i(t) + \frac{y(v_{i+1}(t) - v_i(t))}{x_{i+1}(t) - x_i(t)} \right) \partial_x f(t, x_i(t) + y) = 0,$$

and the expression between the large parentheses is $v(t, x_i(t) + y)$. We have thus shown that the continuity equation

$$\partial_t f + v \partial_x f = 0 \quad (67)$$

is satisfied at every point of differentiability of f . Since the function f is Lipschitz continuous and vanishes on $\{(t, x) : x \leq x_1(t)\}$, we have for every $t \in [0, 1]$ and $x \in \mathbb{R}$ that

$$f(t, x) = f(0, x) + \int_0^t \partial_s f(s, x) \, ds = \int_{-\infty}^x \partial_x f(t, y) \, dy,$$

where the derivatives $\partial_t f$ and $\partial_x f$ are well-defined almost everywhere, and thus for every test function $\phi \in C_c^\infty((0, 1) \times \mathbb{R})$, we have

$$\int_0^1 \int_{\mathbb{R}} \partial_x \phi \partial_t f = - \int_0^1 \int_{\mathbb{R}} \partial_t \partial_x \phi f = \int_0^1 \int_{\mathbb{R}} \partial_t \phi \partial_x f.$$

Setting $\rho(t, x) := \partial_x f(t, x)$ (see eq. (66)), we deduce from (67) and the previous display that the pair (ρ, v) satisfies the continuity equation (5) in the weak sense. For every $i \in \{1, \dots, N-1\}$ and $t \in [0, 1]$, we use the fact that $\rho(t, y)$ is constant over $(x_i(t), x_{i+1}(t))$ and integrates to $1/(N-1)$ there, while $v(t, y)$ is affine on this interval, to obtain that

$$\int_{x_i(t)}^{x_{i+1}(t)} v(t, y)^2 \rho(t, y) \, dy \leq \frac{1}{2(N-1)} (v_i(t)^2 + v_{i+1}(t)^2).$$

Recalling the expression for ρ in (66), we also have that

$$\int_{x_i(t)}^{x_{i+1}(t)} \left(\frac{\rho(t, y)}{1 + \varepsilon \rho(t, y)} \right)^2 \rho(t, y) \, dy = \frac{1}{(N-1)^3} \left(x_{i+1}(t) - x_i(t) + \frac{\varepsilon}{N-1} \right)^{-2}.$$

Combining these estimates yields that

$$\begin{aligned} & \frac{1}{2} \int_0^1 \int_{\mathbb{R}} \left(v(t, y)^2 + \frac{\pi^2}{3} \left(\frac{\rho(t, y)}{1 + \varepsilon \rho(t, y)} \right)^2 \right) \rho(t, y) \, dy \, dt \\ & \leq \frac{1}{2(N-1)} \sum_{i=1}^{N-1} \int_0^1 \left(\frac{v_i^2(t) + v_{i+1}^2(t)}{2} + \frac{\pi^2}{3(N-1)^2} \left(x_{i+1}(t) - x_i(t) + \frac{\varepsilon}{N-1} \right)^{-2} \right) dt, \end{aligned}$$

and in particular,

$$G_\varepsilon(\rho, v) \leq \left(\frac{N+1}{N-1}\right)^3 G_{N,\varepsilon}(\{(x_i, v_i)\}). \quad (68)$$

At this stage we have described a procedure that, for each N, ε and $\{(x_i, v_i)\} \in \text{Adm}_N(\mu_0, \mu_1)$, constructs some (ρ, v) such that (68) holds. As mentioned just before the statement of Proposition 3.3, we now need to face the problem that $\rho(0, x)dx$ is not exactly equal to μ_0 . To do so, we pick subsequences $N_k \rightarrow +\infty$ and $\varepsilon_k \rightarrow 0$ as well as $\{(x_i^{(k)}, v_i^{(k)})\} \in \text{Adm}_{N_k}(\mu_0, \mu_1)$ in such a way that

$$\lim_{k \rightarrow +\infty} G_{N_k, \varepsilon_k}(\{(x_i^{(k)}, v_i^{(k)})\}) = \lim_{\varepsilon \rightarrow 0} \liminf_{N \rightarrow +\infty} J_{N, \varepsilon}(\mu_0, \mu_1),$$

which without loss of generality we can assume to be finite. We next construct $(\rho^{(k)}, v^{(k)})$ as above so that

$$G_{\varepsilon_k}(\rho^{(k)}, v^{(k)}) \leq \left(\frac{N_k+1}{N_k-1}\right)^3 G_{N_k, \varepsilon_k}(\{(x_i^{(k)}, v_i^{(k)})\}),$$

and in particular, possibly after extracting a further subsequence to guarantee the existence of the limit in k ,

$$\lim_{k \rightarrow +\infty} G_{\varepsilon_k}(\rho^{(k)}, v^{(k)}) \leq \lim_{\varepsilon \rightarrow 0} \liminf_{N \rightarrow +\infty} J_{N, \varepsilon}(\mu_0, \mu_1). \quad (69)$$

The key point now is that $\rho^{(k)}(0, x)dx$ converges to μ_0 as k tends to infinity. Indeed, writing

$$f_k(x) := \int_{-\infty}^x \rho^{(k)}(0, u)du \quad \text{and} \quad f(x) := \int_{-\infty}^x d\mu_0,$$

we recall that, for every $i \in \{1, \dots, N_k\}$,

$$f_k(x_i^{(k)}(0)) = \frac{i-1}{N_k-1} \quad \text{and} \quad f(x_i^{(k)}(0)) = \frac{i}{N_k+1},$$

This and the monotonicity of f_k and f yield that

$$\lim_{k \rightarrow \infty} \|f_k - f\|_{L^\infty(\mathbb{R})} = 0,$$

which implies the announced convergence in law of $\rho^{(k)}(0, x)dx$ to μ_0 . Notice the strong similarity between eqs. (69) and (57): in both cases, we have a sequence of candidates $(\rho_\varepsilon, v_\varepsilon)$ for G_ε (possibly along a subsequence of ε 's), with an upper bound on the limit of $G_\varepsilon(\rho_\varepsilon, v_\varepsilon)$, and we aim to construct a candidate (ρ, v) such that $G(\rho, v)$ is smaller than or equal to this limit. The only small difference between our present situation and the one encountered in the proof of Proposition 3.2 comes from the fact that now $\rho^{(k)}(0, x)dx$ is not equal to μ_0 , but only converges to it as k tends to infinity. This entails essentially no modification to the argument there, so we can conclude that there exists a candidate $(\rho, v) \in \text{Adm}(\mu_0, \mu_1)$ such that

$$G(\rho, v) \leq \lim_{\varepsilon \rightarrow 0} \liminf_{N \rightarrow +\infty} J_{N, \varepsilon}(\mu_0, \mu_1),$$

as desired. □

3.3 Lower bound

We now turn to the converse bound.

Proposition 3.4. *Let $\mu_0 = \rho(0, x)dx$, $\mu_1 = \rho(1, x)dx \in \mathcal{P}_c(\mathbb{R}) \cap \mathcal{P}_{\text{abs.}}(\mathbb{R})$ with $\rho(0, \cdot), \rho(1, \cdot) \in L^3(\mathbb{R})$. For every $\varepsilon > 0$, we have*

$$\limsup_{N \rightarrow \infty} J_{N, \varepsilon}(\mu_0, \mu_1) \leq J(\mu_0, \mu_1). \quad (70)$$

3.3.1 Proof strategy and preliminary results

Ideally, we would aim, for each $(\rho, v) \in \text{Adm}(\mu_0, \mu_1)$, to construct some paths $\{(x_i, v_i)\} \in \text{Adm}_N(\mu_0, \mu_1)$ such that $G_{N,\varepsilon}(\{(x_i, v_i)\})$ is close to $G_\varepsilon(\rho, v)$. At a heuristic level, what one needs to do is clear: for each $i \in \{1, \dots, N\}$, we simply fix $x_i(0)$ as in eq. (19), and let $\partial_t x_i(t) = v(t, x_i(t))$.

The main problem is that a control of the error in this argument would require some estimate on the Lipschitz regularity of v in the space variable. Denoting by (ρ^*, v^*) the minimizer for G , it is not in general the case that v^* is smooth (and the situation for G_ε can only be worse). Some regularity results on (ρ^*, v^*) can be found in [49], and it is shown in particular that ρ^* and v^* are smooth on the set $\{(t, x) \in (0, 1) \times \mathbb{R} : \rho^*(t, x) > 0\}$. However $\partial_x v^*$ may indeed blow up near the boundary of this set — this is clearly necessary if the number of connected components of $\{x : \rho^*(t, x) > 0\}$ changes with time, as was already discussed in the paragraph below (26). We will therefore need to employ a regularization argument, i.e. to start by showing that one can find some sufficiently smooth pair (ρ, v) such that $G(\rho, v)$ is close to optimal. This smoothing argument is facilitated by the fact that we know in advance from [49] that ρ^* is compactly supported (under the assumption that μ_0 and μ_1 are), so that we can first lift it away from zero on some sufficiently large compact set, and then regularize ρ and ρv via a convolution procedure. While one might a priori expect to be able to show that $\limsup_{N \rightarrow +\infty} J_{N,\varepsilon}(\mu_0, \mu_1) \leq J_\varepsilon(\mu_0, \mu_1)$ for each fixed ε , we again only obtain a result in which the limit $\varepsilon \rightarrow 0$ is involved, because doing so allows us to exploit this known property of ρ^* .

Before going into the smoothing argument, we first describe how to control the error between the continuous $G_\varepsilon(\rho, v)$ and suitable discrete approximations when the vector field v is sufficiently regular. We denote by Reg the space of bounded continuous functions $v = v(t, x) : \mathbb{R}_+ \times \mathbb{R} \rightarrow \mathbb{R}$ that are differentiable in x and with $\partial_x v$ bounded and uniformly continuous. Fixing $v \in \text{Reg}$, we denote by $(\phi_t(x))_{t \geq 0, x \in \mathbb{R}}$ the flow associated with the differential equation

$$\begin{cases} \partial_t \phi_t(x) = v(t, \phi_t(x)), \\ \phi_0(x) = x. \end{cases} \quad (71)$$

As is well-known, for each $t \geq 0$, the mapping ϕ_t is a C^1 -diffeomorphism; we denote its inverse by ϕ_t^{-1} . We have

$$\partial_x \phi_t(x) = \exp \left(\int_0^t \partial_x v(s, \phi_s(x)) \, ds \right), \quad (72)$$

and in particular, the first derivatives of ϕ_t and ϕ_t^{-1} are bounded over $[0, T] \times \mathbb{R}$ for every $T < +\infty$.

For such a regular velocity field v and a function ρ satisfying the continuity equation (5), we can represent $\rho(t, x) \, dx$ as the image of μ_0 through the mapping ϕ_t (see e.g. [48, Lemma 8.1.6]). That is, assuming that $\mu_0 = \rho(0, x) \, dx$, we have for every $f \in C_c^\infty(\mathbb{R})$ that

$$\int f(x) \rho(t, x) \, dx = \int f(\phi_t(x)) \rho(0, x) \, dx.$$

From this, we deduce that

$$\rho(t, x) = \partial_x \phi_t^{-1}(x) \rho(0, \phi_t^{-1}(x)). \quad (73)$$

These observations will be useful in relation with the following result.

Proposition 3.5. *Let $v \in \text{Reg}$ and ρ be such that $(\rho, v) \in \text{Adm}(\mu_0, \mu_1)$, with $\mu_0 = \rho(0, x) \, dx \in \mathcal{P}_c(\mathbb{R})$ and $\rho(0, \cdot)$ a continuous function. For every $t \in [0, 1]$, we define $(x_i(t))_{1 \leq i \leq N}$ such that for every $i \in \{1, \dots, N\}$,*

$$\int_{-\infty}^{x_i(t)} \rho(t, \cdot) = \frac{i}{N+1}.$$

For every $i \in \{1, \dots, N\}$, the function $t \mapsto x_i(t)$ is differentiable, with

$$\partial_t x_i(t) = v(t, x_i(t)), \quad (74)$$

and setting $v_i(t) := v(t, x_i(t))$, we have

$$|G_{N,\varepsilon}(\{(x_i, v_i)\}) - G_\varepsilon(\rho, v)| \leq \frac{5}{\sqrt{N}} \left(\|v\|_{L^\infty}^2 + \varepsilon^{-2} \right) + (1 + 4\varepsilon^{-1})\omega(2MN^{-\frac{1}{2}}),$$

where $M \in \mathbb{R}$ is any number such that

$$\bigcup_{t \in [0,1]} \text{supp } \rho(t, \cdot) \subseteq [-M, M], \quad (75)$$

and ω is any function such that for every $t \in [0, 1]$ and $x, y \in \mathbb{R}$,

$$|v(t, y) - v(t, x)| + |\rho(t, x) - \rho(t, y)| \leq \omega(|y - x|). \quad (76)$$

Remark 3.2. For M_0 such that $\text{supp } \rho(0, \cdot) \subseteq [-M_0, M_0]$, choosing

$$M := M_0 + \|\partial_x v\|_{L^\infty([0,1] \times \mathbb{R})}$$

ensures that (75) holds. Similarly, using (72) and (73), we can choose ω in terms of $\|\partial_x v\|_{L^\infty([0,1] \times \mathbb{R})}$ and a modulus of continuity of $\partial_x v$ and $\rho(0, \cdot)$ only to ensure that (76) holds, while making sure that $\lim_{r \rightarrow 0} \omega(r) = 0$.

Proof of Proposition 3.5. Writing

$$f(t, x) := \int_{-\infty}^x \rho(t, u) du,$$

we see from the continuity equation (5) that

$$\partial_t f + v \partial_x f = 0.$$

In particular, for each $x \in \mathbb{R}$, the function $t \mapsto f(t, \phi_t(x))$ is constant, where $(\phi_t(x))$ is as in (71). Using this observation with x replaced by $x_i(0)$ yields (74). By the intermediate-value theorem, for each $t \in [0, 1]$ and $i \in \{1, \dots, N-1\}$, there exists $y_i(t) \in [x_i(t), x_{i+1}(t)]$ such that

$$\frac{1}{N+1} = f(t, x_{i+1}(t)) - f(t, x_i(t)) = \rho(t, y_i(t))(x_{i+1}(t) - x_i(t)).$$

We can thus write

$$G_{N,\varepsilon}(\{(x_i, v_i)\}) = \frac{1}{2N} \sum_{i=1}^N \int_0^1 v(t, x_i(t))^2 dt + \frac{\pi^2}{6N} \sum_{i=1}^{N-1} \int_0^1 \left(\frac{\rho(t, y_i(t))}{1 + \varepsilon \rho(t, y_i(t))} \right)^2 dt.$$

For each $t \in [0, 1]$, we clearly have that $x_i(t) \in [-M, M]$, so there must be fewer than $\sqrt{N}$ indices $i \in \{1, \dots, N-1\}$ with $x_{i+1}(t) - x_i(t) \geq 2MN^{-\frac{1}{2}}$. Denoting by $\mathcal{I}(t)$ the complementary set of indices in $\{1, \dots, N-1\}$, we thus have that

$$\begin{aligned} & \left| G_{N,\varepsilon}(\{(x_i, v_i)\}) - \frac{1}{2N} \int_0^1 \sum_{i \in \mathcal{I}(t)} \left(v(t, x_i(t))^2 + \frac{\pi^2}{3} \left(\frac{\rho(t, y_i(t))}{1 + \varepsilon \rho(t, y_i(t))} \right)^2 \right) dt \right| \\ & \leq \frac{1}{2\sqrt{N}} \left(\|v\|_{L^\infty}^2 + \frac{\pi^2}{3} \varepsilon^{-2} \right). \end{aligned}$$

Recalling that $\int_{x_i(t)}^{x_{i+1}(t)} \rho(t, \cdot) = (N+1)^{-1}$, we can write for $i \in \mathcal{I}(t)$ that

$$\left| v(t, x_i(t)) - (N+1) \int_{x_i(t)}^{x_{i+1}(t)} v(t, x) \rho(t, x) dx \right| \leq \omega(2MN^{-\frac{1}{2}}),$$

and likewise,

$$\left| \left(\frac{\rho(t, y_i(t))}{1 + \varepsilon \rho(t, y_i(t))} \right)^2 - (N+1) \int_{x_i(t)}^{x_{i+1}(t)} \left(\frac{\rho(t, x)}{1 + \varepsilon \rho(t, x)} \right)^2 \rho(t, x) dx \right| \leq 2\varepsilon^{-1} \omega(2MN^{-\frac{1}{2}}).$$

Up to these error terms, we can therefore replace the quantity

$$\frac{1}{2N} \int_0^1 \sum_{i \in \mathcal{I}(t)} \left(v(t, x_i(t))^2 + \frac{\pi^2}{3} \left(\frac{\rho(t, y_i(t))}{1 + \varepsilon \rho(t, y_i(t))} \right)^2 \right) dt$$

by

$$\frac{N+1}{2N} \int_0^1 \sum_{i \in \mathcal{I}(t)} \int_{x_i(t)}^{x_{i+1}(t)} \left(v(t, x)^2 + \frac{\pi^2}{3} \left(\frac{\rho(t, x)}{1 + \varepsilon \rho(t, x)} \right)^2 \right) \rho(t, x) dx dt.$$

The difference between this quantity and $G_\varepsilon(\rho, v)$ is at most

$$\left(\frac{1}{2N} + \frac{1}{2\sqrt{N}} \right) \left(\|v\|_{L^\infty}^2 + \frac{\pi^2}{3} \varepsilon^{-2} \right),$$

so the proof is complete. $\square$

As discussed earlier, the proof of Proposition 3.4 requires a smoothing operation on (ρ, v) . The smoothing causes a discrepancy on the initial and final measures $\rho(0, x) dx$ and $\rho(1, x) dx$. To avoid this, we will squeeze the convolved objects onto the time interval $[\eta, 1 - \eta]$ for some small $\eta > 0$, and use the next lemma to bridge time 0 to time η and time $1 - \eta$ to time 1.

Lemma 3.6. *Let $\eta \in (0, 1]$, and let $\mu_0 = \rho(0, x) dx$, $\mu_\eta = \rho(\eta, x) dx \in \mathcal{P}_c(\mathbb{R})$ with $\rho(0, \cdot), \rho(\eta, \cdot) \in L^3(\mathbb{R})$. We let $(x_i(0))_{1 \leq i \leq N}$ and $(x_i(\eta))_{1 \leq i \leq N}$ be defined as in (19), except with time 1 replaced by time η . For each $i \in \{1, \dots, N\}$, we denote by $(x_i(t))_{t \in [0, \eta]}$ the linear interpolation from $x_i(0)$ to $x_i(\eta)$, and we set $v_i(t) := \eta^{-1}(x_i(\eta) - x_i(0))$. We have*

$$\begin{aligned} \limsup_{N \rightarrow +\infty} \frac{1}{2N} \sum_{i=1}^N \int_0^\eta \left(v_i(t)^2 + \frac{\pi^2}{3(N+1)^2} (x_{i+1}(t) - x_i(t))^{-2} \right) dt \\ \leq \frac{1}{2\eta} W_2(\mu_0, \mu_\eta)^2 + \frac{\pi^2 \eta}{12} \int (\rho(0, x)^3 + \rho(\eta, x)^3) dx, \end{aligned}$$

where $W_2(\mu_0, \mu_\eta)$ denotes the L^2 -Wasserstein distance between μ_0 and μ_η .

Proof of Lemma 3.6. Since v_i is constant on $[0, \eta]$, we can write

$$\frac{1}{2N} \sum_{i=1}^N \int_0^\eta v_i(t)^2 dt = \frac{\eta}{2N} \sum_{i=1}^N \left(\frac{x_i(\eta) - x_i(0)}{\eta} \right)^2 = \frac{1}{2\eta} \frac{1}{N} \sum_{i=1}^N (x_i(\eta) - x_i(0))^2. \quad (77)$$

By construction $x_i(0) = F_0^{-1}\left(\frac{i}{N+1}\right)$ and $x_i(\eta) = F_\eta^{-1}\left(\frac{i}{N+1}\right)$, where F_0^{-1} and F_η^{-1} denote the inverse cumulative distribution functions of μ_0 and μ_η respectively. Since μ_0 and μ_η have compact support, these functions are bounded and, since they are monotone, they are of bounded variation. As a result, the function

$$u \mapsto \left(F_\eta^{-1}(u) - F_0^{-1}(u) \right)^2$$

is also of bounded variation. It is therefore Riemann-integrable (see e.g. [69, Corollary 14.3]), and thus

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N (x_i(\eta) - x_i(0))^2 = \int_0^1 (F_\eta^{-1}(u) - F_0^{-1}(u))^2 du. \quad (78)$$

The right-hand side of this expression is $W_2(\mu_0, \mu_\eta)^2$ (see e.g. [67, Chapter 2]).

For the second term in the integral to be estimated, we first observe that, since the mapping $t \mapsto x_{i+1}(t) - x_i(t)$ is affine and $z \mapsto z^{-2}$ is convex on $(0, +\infty)$, we have by Jensen's inequality that

$$(x_{i+1}(t) - x_i(t))^{-2} \leq \left(1 - \frac{t}{\eta}\right) (x_{i+1}(0) - x_i(0))^{-2} + \frac{t}{\eta} (x_{i+1}(\eta) - x_i(\eta))^{-2}.$$

Integrating in time, we thus obtain that

$$\int_0^\eta (x_{i+1}(t) - x_i(t))^{-2} dt \leq \frac{\eta}{2} (x_{i+1}(0) - x_i(0))^{-2} + \frac{\eta}{2} (x_{i+1}(\eta) - x_i(\eta))^{-2}. \quad (79)$$

We now proceed to estimate $\sum_i (x_{i+1}(0) - x_i(0))^{-2}$; the second term on the right side of the previous display can be estimated in the same way. By Hölder's inequality, we have

$$\left(\frac{1}{N+1}\right)^3 = \left(\int_{x_i(0)}^{x_{i+1}(0)} \rho(0, x) dx\right)^3 \leq (x_{i+1}(0) - x_i(0))^2 \int_{x_i(0)}^{x_{i+1}(0)} \rho(0, x)^3 dx,$$

that is,

$$\frac{1}{(N+1)^2} (x_{i+1}(0) - x_i(0))^{-2} \leq (N+1) \int_{x_i(0)}^{x_{i+1}(0)} \rho(0, x)^3 dx.$$

Summing this inequality over i and dividing by N gives

$$\frac{1}{N(N+1)^2} \sum_{i=1}^N (x_{i+1}(0) - x_i(0))^{-2} \leq \frac{N+1}{N} \int_{\mathbb{R}} \rho(0, x)^3 dx.$$

Combining this with (79) yields that

$$\frac{1}{2N} \sum_{i=1}^N \int_0^\eta \frac{\pi^2}{3(N+1)^2} (x_{i+1}(t) - x_i(t))^{-2} dt \leq \frac{\pi^2 \eta}{12} \frac{N+1}{N} \int (\rho(0, x)^3 + \rho(\eta, x)^3) dx.$$

Taking the limsup as $N \rightarrow \infty$, and recalling eq. (78), we obtain the announced result. $\square$

3.3.2 Proof of Proposition 3.4

By Theorem 3.1, there exists $(\rho, v) \in \text{Adm}(\mu_0, \mu_1)$ with ρ compactly supported such that $G(\rho, v) = J(\mu_0, \mu_1)$. We let $M \geq 1$ be such that for all $t \in [0, 1]$, the support of $\rho(t, \cdot)$ is contained in $[-(M-1), M-1]$. We extend ρ and v to all times $t \in \mathbb{R}$ by setting $v(t, x) = 0$ for $t \notin [-1, 1]$ and

$$\rho(t, x) = \begin{cases} \rho(0, x) & \text{if } t < 0, \\ \rho(1, x) & \text{if } t > 1. \end{cases}$$

The extended (ρ, v) still satisfies the continuity equation (5) in the weak sense, as does the pair $(\mathbb{1}_{[-M, M]}(x)/(2M), 0)$. We write $m := \rho v$, fix $\delta_1 \in (0, \frac{1}{4}]$, and set

$$\rho_{\delta_1}(t, x) := (1 - \delta_1)\rho(t, x) + \frac{\delta_1}{2M} \mathbb{1}_{[-M, M]}(x) \quad \text{and} \quad m_{\delta_1} := (1 - \delta_1)m.$$

Using the convexity of the mapping in eq. (61), as well as that of $r \mapsto r^3$ over $[0, +\infty)$, we have

$$\frac{1}{2} \int_0^1 \int_{\mathbb{R}} \left(\frac{m_{\delta_1}(t, x)^2}{\rho_{\delta_1}(t, x)} + \frac{\pi^2}{3} \rho_{\delta_1}(t, x)^3 \right) dx dt \leq (1 - \delta_1)G(\rho, v) + \frac{\pi^2}{6} \frac{\delta_1}{(2M)^2}.$$

We give ourselves $\delta_2 \in (0, \frac{1}{4}]$ and a nonnegative even function $\varphi \in C_c^\infty(\mathbb{R})$ supported in $[-1, 1]$ with $\int \varphi = 1$, and set

$$\varphi_{\delta_2}(t, x) := \delta_2^{-2} \varphi(t/\delta_2) \varphi(x/\delta_2), \quad \rho_\delta := \rho_{\delta_1} \star \varphi_{\delta_2}, \quad \text{and} \quad m_\delta := m_{\delta_1} \star \varphi_{\delta_2},$$

where $\star$ denotes the convolution in the time and space variables, and where we write $\delta = (\delta_1, \delta_2)$. By linearity, the pair (ρ_δ, m_δ) still satisfies the continuity equation (in the sense that $\partial_t \rho_\delta + \partial_x m_\delta = 0$). Notice also that ρ_δ is constant on $(-\infty, -\delta_2]$, as well as on $[1 + \delta_2, +\infty)$. For convenience, we introduce the shorthand

$$\mu_{0,\delta} := \rho_\delta(-\delta_2, x) dx, \quad \mu_{1,\delta} := \rho_\delta(1 + \delta_2, x) dx \in \mathcal{P}_c(\mathbb{R}).$$

The function $\rho_\delta(-\delta_2, \cdot)$ is the convolution (in space only) of $\rho_{\delta_1}(0, \cdot)$ with $\delta_2^{-1} \varphi(\cdot/\delta_2)$; and similarly for $\rho_\delta(1 + \delta_2, \cdot)$. In particular, we have that

$$\lim_{\delta \rightarrow 0} W_2(\mu_0, \mu_{0,\delta}) = \lim_{\delta \rightarrow 0} W_2(\mu_1, \mu_{1,\delta}) = 0. \quad (80)$$

We give ourselves $\eta \in [\delta_2, \frac{1}{4}]$, and use convexity to write that

$$\begin{aligned} & \frac{1}{2} \int_{-\eta}^{1+\eta} \int_{\mathbb{R}} \left(\frac{m_\delta(t, x)^2}{\rho_\delta(t, x)} + \frac{\pi^2}{3} \rho_\delta(t, x)^3 \right) dx dt \\ & \leq \frac{1}{2} \int_{-\eta}^{1+\eta} \int_{\mathbb{R}} \left(\frac{m_{\delta_1}^2}{\rho_{\delta_1}} + \frac{\pi^2}{3} \rho_{\delta_1}^3 \right) \star \varphi_{\delta_2}(t, x) dx dt \\ & \leq \frac{1}{2} \int_{-\eta-\delta_2}^{1+\eta+\delta_2} \int_{\mathbb{R}} \left(\frac{m_{\delta_1}(t, x)^2}{\rho_{\delta_1}(t, x)} + \frac{\pi^2}{3} \rho_{\delta_1}(t, x)^3 \right) dx dt \\ & \leq (1 - \delta_1) G(\rho, v) + \frac{\pi^2}{6} \frac{\delta_1}{(2M)^2} (1 + 2\eta + 2\delta_2) + (\eta + \delta_2) \frac{\pi^2}{6} \int_{\mathbb{R}} (\rho(0, x)^3 + \rho(1, x)^3) dx. \end{aligned} \quad (81)$$

Since we imposed $\eta \geq \delta_2$, we have that

$$\rho_\delta(-\eta, x) dx = \mu_{0,\delta} \quad \text{and} \quad \rho_\delta(1 + \eta, x) dx = \mu_{1,\delta}. \quad (82)$$

The point of the parameter δ_1 is to ensure a uniform lower bound on the density, namely that

$$\text{for every } x \in [-M, M], \quad \rho_\delta(t, x) \geq \frac{\delta_1}{4M}. \quad (83)$$

We set

$$v_\delta(t, x) := \begin{cases} \frac{m_\delta(t, x)}{\rho_\delta(t, x)} & \text{if } \rho_\delta(t, x) > 0, \\ 0 & \text{otherwise.} \end{cases}$$

By construction, the pair (ρ_δ, v_δ) satisfies the continuity equation (5), with $v_\delta \in L^2([0, 1] \times \mathbb{R}; \rho_\delta(t, x) dx dt)$. Thanks to the smoothing introduced with the parameter δ_2 , the function v_δ is infinitely differentiable on the set $\{(t, x) : \rho_\delta(t, x) > 0\}$, and in particular on $[0, 1] \times [-M, M]$ by (83). Recalling that $\rho(t, \cdot)$ and $m(t, \cdot)$ are supported in $[-M + 1, M - 1]$, we also see that for every $|x| \geq M - \frac{1}{2}$, the quantity $\rho_\delta(t, x)$ does not depend on t and $v_\delta(t, x) = 0$. In particular, we have that $v_\delta \in \text{Reg}$. We are thus in position to appeal to Proposition 3.5 for the pair (ρ_δ, v_δ) . One problem though is that if we apply it as stated, this proposition will not give us trajectories that belong to $\text{Adm}(\mu_0, \mu_1)$, since the initial and final measures will not be quite right. To remedy this, we take the (ρ_δ, v_δ) trajectory on the time interval $[-\eta, 1 + \eta]$ and shrink it to $[\eta, 1 - \eta]$, and we then leverage Lemma 3.6 to bridge the gap from time 0 to time η and from time $1 - \eta$ to time 1. That is, we set

$$\begin{cases} \rho_{\delta,\eta}(t, x) & := \rho_\delta \left(\frac{1 + 2\eta}{1 - 2\eta} \left(t - \frac{1}{2} \right) + \frac{1}{2}, x \right), \\ v_{\delta,\eta}(t, x) & := \frac{1 + 2\eta}{1 - 2\eta} v_\delta \left(\frac{1 + 2\eta}{1 - 2\eta} \left(t - \frac{1}{2} \right) + \frac{1}{2}, x \right). \end{cases}$$

Notice that the pair $(\rho_{\delta,\eta}, v_{\delta,\eta})$ still satisfies the continuity equation. We have

$$\begin{aligned} \int_{\eta}^{1-\eta} \int_{\mathbb{R}} \left(v_{\delta,\eta}(t, x)^2 + \frac{\pi^2}{3} \rho_{\delta,\eta}(t, x)^2 \right) \rho_{\delta,\eta}(t, x) dx dt \\ \leq \frac{1+2\eta}{1-2\eta} \int_{-\eta}^{1+\eta} \int_{\mathbb{R}} \left(v_{\delta}(t, x)^2 + \frac{\pi^2}{3} \rho_{\delta}(t, x)^2 \right) \rho_{\delta}(t, x) dx dt, \end{aligned} \quad (84)$$

and the latter integral has been estimated in (81). Slightly adapting Proposition 3.5 so that the time interval $[0, 1]$ there is replaced by $[\eta, 1 - \eta]$, we obtain trajectories $\{(x_{\delta,\eta,i}(t), v_{\delta,\eta,i}(t))\}$ defined for $t \in [\eta, 1 - \eta]$, with $\partial_t x_{\delta,\eta,i} = v_{\delta,\eta,i}$, such that for all $i \in \{1, \dots, N\}$:

$$\begin{cases} \int_{-\infty}^{x_{\delta,\eta,i}(\eta)} d\mu_{0,\delta}(x) dx &= \frac{i}{N+1}, \\ \int_{-\infty}^{x_{\delta,\eta,i}(1-\eta)} d\mu_{1,\delta}(x) dx &= \frac{i}{N+1}, \end{cases}$$

and furthermore

$$\begin{aligned} \left| \frac{1}{2N} \sum_{i=1}^N \int_{\eta}^{1-\eta} \left(v_{\delta,\eta,i}(t)^2 + \frac{\pi^2}{3(N+1)^2} \left(x_{\delta,\eta,i+1}(t) - x_{\delta,\eta,i}(t) + \frac{\varepsilon}{N+1} \right)^{-2} \right) dt \right. \\ \left. - \frac{1}{2} \int_{\eta}^{1-\eta} \int_{\mathbb{R}} \left(v_{\delta,\eta}(t, x)^2 + \frac{\pi^2}{3} \left(\frac{\rho_{\delta,\eta}(t, x)}{1 + \varepsilon \rho_{\delta,\eta}(t, x)} \right)^2 \right) \rho_{\delta,\eta}(t, x) dx dt \right| \leq \text{err}(N, \delta, \eta), \end{aligned}$$

where $\text{err}(N, \delta, \eta)$ is an error term such that, for each fixed δ, η , we have $\lim_{N \rightarrow +\infty} \text{err}(N, \delta, \eta) = 0$. We next extend the trajectories $\{(x_{\delta,\eta,i}(t), v_{\delta,\eta,i}(t))\}$ to the full interval $t \in [0, 1]$ using Lemma 3.6¹. Recalling eq. (80), as well as eqs. (81) and (84), we send N to infinity, then δ to zero, and finally η to zero, to obtain that

$$\limsup_{N \rightarrow \infty} J_{N,\varepsilon}(\mu_0, \mu_1) \leq J(\mu_0, \mu_1),$$

as desired. □

3.4 Remark on an alternative formulation

We remark that, by taking a formal $N \rightarrow \infty$ limit, Theorem 1.2 suggests the following possible alternative formulation for $J(\mu_0, \mu_1)$. For each $\mu \in \mathcal{P}(\mathbb{R})$ and $\omega \in [0, 1]$, let

$$F_{\mu}^{-1}(\omega) := \inf \{s \geq 0 : \mu((-\infty, s]) \geq \omega\}$$

denote the inverse cumulative distribution function of μ . Let $\text{Adm}'(\mu_0, \mu_1)$ denote the space of all continuous functions $Z \in C([0, 1]^2; \mathbb{R})$ that satisfy the following properties.

1. We have $Z(0, \cdot) = F_{\mu_0}^{-1}$ and $Z(1, \cdot) = F_{\mu_1}^{-1}$.
2. For every $t \in [0, 1]$, the function $Z(t, \cdot)$ is non-decreasing.
3. The function Z is differentiable on $(0, 1)^2$.

¹Technically, this yields trajectories for which the velocities have jumps, while we required the velocities to be continuous in the definition of $\text{Adm}_N(\mu_0, \mu_1)$ (see Section 1.2.1), but it is easy to see that the set $\text{Adm}_N(\mu_0, \mu_1)$ can be extended to include such trajectories without causing a change in the value of $J_{N,\varepsilon}(\mu_0, \mu_1)$.

We speculate that for every $\mu_0, \mu_1 \in \mathcal{P}_c(\mathbb{R})$ with $\Sigma(\mu_0)$ and $\Sigma(\mu_1)$ finite, the quantity $J(\mu_0, \mu_1)$ is equal to

$$\inf_{Z \in \text{Adm}'(\mu_0, \mu_1)} \frac{1}{2} \int_0^1 \int_0^1 \left((\partial_t Z(t, \omega))^2 + \frac{\pi^2}{3(\partial_\omega Z(t, \omega))^2} \right) d\omega dt, \quad (85)$$

and also to

$$\lim_{\varepsilon \rightarrow 0} \inf_{Z \in \text{Adm}'(\mu_0, \mu_1)} \frac{1}{2} \int_0^1 \int_0^1 \left((\partial_t Z(t, \omega))^2 + \frac{\pi^2}{3(\varepsilon + \partial_\omega Z(t, \omega))^2} \right) d\omega dt. \quad (86)$$

Notice that, as in the discrete approximation, the set $\text{Adm}'(\mu_0, \mu_1)$ is convex and the functionals appearing inside the infima in eqs. (85) and in (86) are strictly convex in Z . Moreover, the quantity in eq. (86) is more directly related to the discretization we perform in Section 2.

References

- [1] Harish-Chandra. “Differential operators on a semisimple Lie algebra”. In: *American Journal of Mathematics* 79.1 (1957), pp. 87–120.
- [2] C. Itzykson and J.-B. Zuber. “The planar approximation. II”. In: *Journal of Mathematical Physics* 21.3 (1980), pp. 411–421.
- [3] C. McSwiggen. “The Harish-Chandra integral: An introduction with examples”. In: *L’Enseignement Mathématique* 67.3 (2021), pp. 229–299.
- [4] G. Menon. *The complex Burgers equation, the HCIZ integral and the Calogero–Moser system*. Unpublished manuscript. 2017. URL: <https://www.dam.brown.edu/people/menon/talks/cmsa.pdf> (visited on 05/22/2026).
- [5] T. Tao. *The Harish-Chandra–Itzykson–Zuber Integral Formula*. What’s new. Feb. 8, 2013. URL: <https://terrytao.wordpress.com/2013/02/08/the-harish-chandra-itzykson-zuber-integral-formula/> (visited on 05/22/2026).
- [6] A. Guionnet. “Rare events in random matrix theory”. In: *International Congress of Mathematicians 2022*. EMS Press, 2023, pp. 1008–1052.
- [7] A. Maillard. “Large deviations of extreme eigenvalues of generalized sample covariance matrices”. In: *Europhysics Letters* 133.2 (2021), p. 20005.
- [8] P. Mergny and M. Potters. “Right large deviation principle for the top eigenvalue of the sum or product of invariant random matrices”. In: *Journal of Statistical Mechanics: Theory and Experiment* 2022.6 (2022), p. 063301.
- [9] P. Niu, Y. Liu, T. Fu, J. Fan, C. Deng, and Z. Huang. “The role of rank in mismatched low-rank symmetric matrix estimation”. In: *2025 IEEE Information Theory Workshop (ITW)*. IEEE. 2025, pp. 1–6.
- [10] J. Husson and B. McKenna. “Large deviations for the largest eigenvalue of generalized sample covariance matrices”. In: *Electronic Journal of Probability* 29 (2024), pp. 1–48.
- [11] A. Guionnet and J. Husson. “Large deviations for the largest eigenvalue of Rademacher matrices”. In: *The Annals of Probability* 48.3 (2020), pp. 1436–1465.
- [12] A. Guionnet and J. Husson. “Asymptotics of k dimensional spherical integrals and applications”. In: *ALEA. Latin American Journal of Probability & Mathematical Statistics* 19.1 (2022), pp. 769–797.
- [13] A. Guionnet and M. Maida. “Large deviations for the largest eigenvalue of the sum of two random matrices”. In: *Electronic Journal of Probability* 25 (2020), Paper no. 14.

- [14] J. Husson. “Large deviations for the largest eigenvalue of matrices with variance profiles”. In: *Electronic Journal of Probability* 27 (2022), pp. 1–44.
- [15] F. Augeri, A. Guionnet, and J. Husson. “Large deviations for the largest eigenvalue of sub-Gaussian matrices”. In: *Communications in Mathematical Physics* 383.2 (2021), pp. 997–1050.
- [16] A. Guionnet and O. Zeitouni. “Large deviations asymptotics for spherical integrals”. In: *Journal of Functional Analysis* 188.2 (2002), pp. 461–515.
- [17] A. Guionnet and O. Zeitouni. “Addendum to: large deviations asymptotics for spherical integrals”. In: *Journal of Functional Analysis* 216.1 (2004), pp. 230–241.
- [18] J. Bun, J.-P. Bouchaud, S. N. Majumdar, and M. Potters. “Instanton approach to large N Harish-Chandra-Itzykson-Zuber integrals”. In: *Physical Review Letters* 113.7 (2014), p. 070201.
- [19] M. Potters and J.-P. Bouchaud. *A first course in random matrix theory: for physicists, engineers and data scientists*. Cambridge University Press, 2020.
- [20] A. Maillard. “Average-case matrix discrepancy: satisfiability bounds”. In: *Random Structures & Algorithms* 67.3 (2025), e70033.
- [21] A. T. James. “Distributions of matrix variates and latent roots derived from normal samples”. In: *The Annals of Mathematical Statistics* 35.2 (1964), pp. 475–501.
- [22] A. Guionnet. “Large deviations and stochastic calculus for large random matrices”. In: *Probability Surveys* 1 (2004), pp. 72–172.
- [23] E. Marinari, G. Parisi, and F. Ritort. “Replica field theory for deterministic models. II. A non-random spin glass with glassy behaviour”. In: *Journal of Physics A: Mathematical and General* 27.23 (1994), pp. 7647–7668.
- [24] G. Parisi and M. Potters. “Mean-field equations for spin models with orthogonal interaction matrices”. In: *Journal of Physics A: Mathematical and General* 28.18 (1995), pp. 5267–5285.
- [25] A. Maillard, L. Foini, A. L. Castellanos, F. Krzakala, M. Mézard, and L. Zdeborová. “High-temperature expansions and message passing algorithms”. In: *Journal of Statistical Mechanics: Theory and Experiment* 2019.11 (2019), p. 113301.
- [26] A. Maillard and D. Kunisky. “Fitting an ellipsoid to random points: predictions using the replica method”. In: *IEEE Transactions on Information Theory* 70.10 (2024), pp. 7273–7296.
- [27] R. Dandekar, P. L. Krapivsky, and K. Mallick. “Current fluctuations in the Dyson gas”. In: *Physical Review E* 110.6 (2024), p. 064153.
- [28] J. Grella, S. N. Majumdar, and G. Schehr. “Non-intersecting Brownian bridges in the flat-to-flat geometry”. In: *Journal of Statistical Physics* 183.3 (2021), p. 49.
- [29] F. Pourkamali and N. Macris. “Bayesian extensive-rank matrix factorization with rotational invariant priors”. In: *Advances in Neural Information Processing Systems* 36 (2023), pp. 24025–24073.
- [30] J. Barbier, F. Camilli, M. Mondelli, and M. Sáenz. “Fundamental limits in structured principal component analysis and how to reach them”. In: *Proceedings of the National Academy of Sciences* 120.30 (2023), e2302028120.
- [31] A. Maillard, F. Krzakala, M. Mézard, and L. Zdeborová. “Perturbative construction of mean-field equations in extensive-rank matrix factorization and denoising”. In: *Journal of Statistical Mechanics: Theory and Experiment* 2022.8 (2022), p. 083301.

- [32] F. Pourkamali and N. Macris. “Rectangular rotational invariant estimator for general additive noise matrices”. In: *2023 IEEE International Symposium on Information Theory (ISIT)*. IEEE. 2023, pp. 2081–2086.
- [33] F. Pourkamali, J. Barbier, and N. Macris. “Matrix inference in growing rank regimes”. In: *IEEE Transactions on Information Theory* 70.11 (2024), pp. 8133–8163.
- [34] F. Pourkamali and N. Macris. “Rectangular rotational invariant estimator for high-rank matrix estimation”. In: *Information and Inference: A Journal of the IMA* 14.3 (2025), iaaf015.
- [35] E. Troiani, V. Erba, F. Krzakala, A. Maillard, and L. Zdeborová. “Optimal denoising of rotationally invariant rectangular matrices”. In: *Mathematical and Scientific Machine Learning*. PMLR. 2022, pp. 97–112.
- [36] J. Barbier, F. Camilli, M.-T. Nguyen, M. Pastore, and R. Skerk. “Statistical physics of deep learning: Optimal learning of a multi-layer perceptron near interpolation”. In: *Physical Review X* 16.3 (2026), p. 031014.
- [37] F. Boncoraglio, E. Troiani, V. Erba, and L. Zdeborová. “Bayes optimal learning of attention-indexed models”. In: *Advances in Neural Information Processing Systems* 38 (2025), pp. 105029–105074.
- [38] V. Erba, E. Troiani, L. Zdeborová, and F. Krzakala. “The nuclear route: Sharp asymptotics of ERM in overparameterized quadratic networks”. In: *Advances in Neural Information Processing Systems* 38 (2025), pp. 88862–88901.
- [39] L. Defilippis, Y. Xu, J. Girardin, E. Troiani, V. Erba, L. Zdeborová, B. Loureiro, and F. Krzakala. “Scaling laws and spectra of shallow neural networks in the feature learning regime”. In: *The Fourteenth International Conference on Learning Representations*. 2026.
- [40] Y. Xu, A. Maillard, L. Zdeborová, and F. Krzakala. “Fundamental limits of matrix sensing: Exact asymptotics, universality, and applications”. In: *The Thirty Eighth Annual Conference on Learning Theory*. PMLR. 2025, pp. 5757–5823.
- [41] V. Erba, E. Troiani, L. Biggio, A. Maillard, and L. Zdeborová. “Bilinear sequence regression: A model for learning from long sequences of high-dimensional tokens”. In: *Physical Review X* 15.2 (2025), p. 021092.
- [42] A. Maillard, E. Troiani, S. Martin, L. Zdeborová, and F. Krzakala. “Bayes-optimal learning of an extensive-width neural network from quadratically many samples”. In: *Advances in Neural Information Processing Systems* 37 (2024), pp. 82085–82132.
- [43] A. Guionnet and M. Maida. “A Fourier view on the R-transform and related asymptotics of spherical integrals”. In: *Journal of Functional Analysis* 222.2 (2005), pp. 435–490.
- [44] B. Collins and P. Śniady. “New scaling of Itzykson–Zuber integrals”. In: *Annales de l’Institut Henri Poincaré (B) Probabilités et Statistiques* 43.2 (2007), pp. 139–146.
- [45] A. Matytsin. “On the large- N limit of the Itzykson–Zuber integral”. In: *Nuclear Physics B* 411.2-3 (1994), pp. 805–820.
- [46] B. Collins, A. Guionnet, and E. Maurel-Segala. “Asymptotics of unitary and orthogonal matrix integrals”. In: *Advances in Mathematics* 222.1 (2009), pp. 172–215.
- [47] A. Guionnet and J. Huang. *Large deviations asymptotics of rectangular spherical integral*. 2021. arXiv: [2106.07146](https://arxiv.org/abs/2106.07146) [math.PR].
- [48] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Second. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2008, pp. x+334.

- [49] A. Guionnet. “First order asymptotics of matrix integrals; a rigorous approach towards the understanding of matrix models”. In: *Communications in Mathematical Physics* 244.3 (2004), pp. 527–569.
- [50] S. Belinschi, A. Guionnet, and J. Huang. “Large deviation principles via spherical integrals”. In: *Probability and Mathematical Physics* 3.3 (2022), pp. 543–625.
- [51] J.-B. Zuber. “The large- N limit of matrix integrals over the orthogonal group”. In: *Journal of Physics A: Mathematical and Theoretical* 41.38 (2008), p. 382001.
- [52] B. Collins. “Moments and cumulants of polynomial random variables on unitary groups, the Itzykson-Zuber integral, and free probability”. In: *International Mathematics Research Notices* 2003.17 (2003), pp. 953–982.
- [53] G. W. Anderson, A. Guionnet, and O. Zeitouni. *An introduction to random matrices*. Vol. 118. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2010.
- [54] H. C. Schmidt. “Statistical physics of sparse and dense models in optimization and inference”. PhD thesis. Université Paris Saclay (COMUE), 2018.
- [55] V. A. Marchenko and L. A. Pastur. “Distribution of eigenvalues for some sets of random matrices”. In: *Matematicheskii Sbornik* 72(114) (1967), pp. 507–536.
- [56] J.-D. Benamou and Y. Brenier. “A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem”. In: *Numerische Mathematik* 84.3 (2000), pp. 375–393.
- [57] J.-D. Benamou and G. Carlier. “Augmented Lagrangian methods for transport optimization, mean field games and degenerate elliptic equations”. In: *Journal of Optimization Theory and Applications* 167.1 (2015), pp. 1–26.
- [58] G. Peyré and M. Cuturi. “Computational optimal transport: With applications to data science”. In: *Foundations and Trends® in Machine Learning* 11.5-6 (2019), pp. 355–607.
- [59] W. Gander and G. H. Golub. “Cyclic reduction-history and applications”. In: *Scientific Computing: Proceedings of the Workshop, 10-12 March 1997, Hong Kong*. Springer Science & Business Media. 1998, p. 73.
- [60] L. Vandenberghe and S. Boyd. *Convex optimization*. Vol. 1. Cambridge university press Cambridge, 2004.
- [61] A. Maillard and J.-C. Mourrat. https://github.com/AnMaillard/extensive_rank_HC_IZ_solver_public. 2026.
- [62] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, et al. “PyTorch: An imperative style, high-performance deep learning library”. In: *Advances in Neural Information Processing Systems* 32 (2019), pp. 8024–8035.
- [63] L. F. Richardson. “The approximate arithmetical solution by finite differences of physical problems involving differential equations, with an application to the stresses in a masonry dam”. In: *Philosophical Transactions of the Royal Society of London. Series A, containing papers of a mathematical or physical character* 210 (1911), pp. 307–357.
- [64] D. P. Kingma and J. Ba. “Adam: A method for stochastic optimization”. In: *The Third International Conference on Learning Representations*. 2015.
- [65] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu. “A limited memory algorithm for bound constrained optimization”. In: *SIAM Journal on Scientific Computing* 16.5 (1995), pp. 1190–1208.
- [66] J. Nocedal and S. J. Wright. *Numerical optimization*. Second. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2006.

- [67] C. Villani. *Topics in optimal transportation*. Vol. 58. Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, 2003, pp. xvi+370.
- [68] T. Dominguez and J.-C. Mourrat. *Statistical mechanics of mean-field disordered systems: a Hamilton-Jacobi approach*. Zurich Lectures in Advanced Mathematics. European Mathematical Society, Zürich, 2024.
- [69] N. L. Carothers. *Real analysis*. Cambridge University Press, Cambridge, 2000, pp. xiv+401.
- [70] R. Barrett, M. Berry, T. F. Chan, J. Demmel, J. Donato, J. Dongarra, V. Eijkhout, R. Pozo, C. Romine, and H. Van der Vorst. *Templates for the solution of linear systems: building blocks for iterative methods*. SIAM, 1994.

A Additional properties and technicalities

A.1 Scaling and invariance properties of the HCIZ integral

We make here two simple remarks on properties of the HCIZ integral and the hydrodynamical system of Theorem 1.1.

A.1.1 Scaling parameter $\theta \geq 0$

If we introduce the scaling parameter $\theta \geq 0$, eq. (6) becomes

$$I_{\text{HCIZ}}(\theta, \mu, \nu) = -\frac{3}{4} - \frac{1}{2} \log \theta + \frac{\theta}{2} \left[\int \mu(dx) x^2 + \int \nu(dx) x^2 \right] - \frac{1}{2} [\Sigma(\mu) + \Sigma(\nu)] \quad (87)$$

$$- \frac{1}{2} \inf_{(\rho, v)} \int_0^1 dt \int dy \rho(t, y) \left[v(t, y)^2 + \frac{\pi^2}{3} \rho(t, y)^2 \right],$$

where the conservation of mass equation of eq. (5) is unchanged, and the boundary conditions read:

$$\lim_{t \rightarrow 0} \rho(t, y) dy = \sqrt{\theta} \# \mu \quad \text{and} \quad \lim_{t \rightarrow 1} \rho(t, y) dy = \sqrt{\theta} \# \nu. \quad (88)$$

Notice that we can rescale variables as:

$$\begin{cases} \tilde{\rho}(u, x) := \sqrt{\theta} \rho(\theta u, \sqrt{\theta} x), \\ \tilde{v}(u, x) := \sqrt{\theta} v(\theta u, \sqrt{\theta} x), \end{cases} \quad (89)$$

which yields

$$\int_0^1 dt \int dy \rho(t, y) \left[v(t, y)^2 + \frac{\pi^2}{3} \rho(t, y)^2 \right] = \int_0^{\theta^{-1}} du \int dx \tilde{\rho}(u, x) \left[\tilde{v}(u, x)^2 + \frac{\pi^2}{3} \tilde{\rho}(u, x)^2 \right], \quad (90)$$

The continuity equation (5) is still satisfied by $(\tilde{\rho}, \tilde{v})$, and this rescaling changes the boundary conditions to:

$$\lim_{u \rightarrow 0} \tilde{\rho}(u, x) dx = \mu \quad \text{and} \quad \lim_{u \rightarrow \theta^{-1}} \tilde{\rho}(u, x) dx = \nu. \quad (91)$$

Informally, eq. (90) allows accounting for θ by rescaling either the interpolation time or the space variable.

A.1.2 Galilean transformation

The equations above are also invariant under Galilean transformations. If (ρ, v) satisfy eqs. (5) and (8), then so do

$$\begin{cases} \tilde{\rho}(t, x) &:= \rho(t, x + \alpha t), \\ \tilde{v}(t, x) &:= v(t, x + \alpha t) - \alpha, \end{cases}$$

for any $\alpha \in \mathbb{R}$. Plugging this relation into eq. (87), we recover that, if $\nu_\alpha(x) := \nu(x - \alpha)$,

$$I_{\text{HCIZ}}(\theta, \mu, \nu_\alpha) = I_{\text{HCIZ}}(\theta, \mu, \nu) + \alpha \theta \mathbb{E}_\mu[X],$$

which is immediate from the finite-dimensional expression of eq. (1).

A.2 Remarks on time discretization

We briefly detail here why the discretization scheme in time that we use in eq. (34) creates a time discretization error that we expect to be of order $\mathcal{O}(\|\Delta t\|_\infty^2) = \mathcal{O}(1/T^2)$, as validated by the numerical experiments in Appendix B.

We consider a single trajectory $x \in \mathcal{C}^2([0, 1], \mathbb{R})$. Let F be twice continuously differentiable on a neighborhood of $x([0, 1])$, then the trapezoidal rule yields

$$\left| \int_0^1 F(x(t)) dt - \sum_{k=0}^T \omega_k F(x_k) \right| \leq C \|\Delta t\|_\infty^2,$$

where the weights ω_k are defined in eq. (33), and $C > 0$ only depends on $\{\|x^{(p)}\|_\infty, \|F^{(p)}\|_\infty\}_{0 \leq p \leq 2}$. The same convergence order holds for the discretized kinetic-energy term in eq. (22), as shown by the following elementary statement.

Lemma A.1. *Let $x \in \mathcal{C}^2([0, 1], \mathbb{R})$ and let $0 = t_0 < t_1 < \dots < t_T = 1$. For $k \in [T]$, set $\Delta t_k := t_k - t_{k-1}$, $x_k := x(t_k)$, and*

$$v_k := \frac{x_k - x_{k-1}}{\Delta t_k}.$$

Then

$$0 \leq \int_0^1 |\partial_t x(t)|^2 dt - \sum_{k=1}^T \Delta t_k v_k^2 \leq \frac{\|\Delta t\|_\infty^2}{\pi^2} \int_0^1 |\partial_t^2 x(t)|^2 dt.$$

In particular, as $\|\Delta t\|_\infty \downarrow 0$, the discrete kinetic energy converges to its continuous counterpart with error $\mathcal{O}(\|\Delta t\|_\infty^2)$.

Proof of Lemma A.1. Denote $I_k := (t_{k-1}, t_k)$. Expanding the square and using that $\int_{I_k} \partial_t x dt = \Delta t_k v_k$ yields

$$\int_{I_k} (\partial_t x(t) - v_k)^2 dt = \int_{I_k} (\partial_t x(t))^2 dt - \Delta t_k v_k^2.$$

And summing over k :

$$\int_0^1 |\partial_t x(t)|^2 dt - \sum_{k=1}^T \Delta t_k v_k^2 = \sum_{k=1}^T \int_{I_k} (\partial_t x(t) - v_k)^2 dt \geq 0. \quad (92)$$

We next apply the one-dimensional Poincaré–Wirtinger inequality on each interval: for any f such that $f, \partial_t f \in L^2(I_k)$, with $\bar{f} := \Delta t_k^{-1} \int_{I_k} f(t) dt$, we have

$$\int_{I_k} (f - \bar{f})^2 dt \leq \frac{\Delta t_k^2}{\pi^2} \int_{I_k} |\partial_t f|^2 dt. \quad (93)$$

Indeed, using the Fourier expansion in $L^2(I_k)$:

$$g(t) := f(t) - \bar{f} = \sqrt{\frac{2}{\Delta t_k}} \sum_{n \geq 1} a_n \cos\left(\frac{n\pi(t - t_{k-1})}{\Delta t_k}\right),$$

Parseval's identity then gives

$$\int_{I_k} g(t)^2 dt = \sum_{n \geq 1} a_n^2, \quad \int_{I_k} |\partial_t g(t)|^2 dt = \sum_{n \geq 1} \left(\frac{n\pi}{\Delta t_k}\right)^2 a_n^2.$$

Since $n^2 \geq 1$, eq. (93) follows. Applying it to eq. (92) then yields immediately the result. $\square$

A.3 Derivation of eq. (39)

We detail here the derivation of the identity

$$\begin{aligned} & \min_{(x_{1,k})_{1 \leq k \leq T-1}} \left\{ \frac{1}{N} \sum_{k=1}^T \frac{1}{\Delta t_k} \sum_{i=1}^N \left[x_{1,k} - x_{1,k-1} + \frac{\Delta t_k}{N+1} \sum_{j=1}^{i-1} d\delta_{j,k} \right]^2 \right\} \\ &= [\bar{x}(0) - \bar{x}(1)]^2 + \frac{1}{(N+1)^2} \sum_{k=1}^T \Delta t_k (d\delta_k^\top) L_{N-1}^{-1} (d\delta_k). \end{aligned} \quad (94)$$

As mentioned in the main text, this is a simple convex quadratic minimization problem. Recall that $x_{i,k} := x_{1,k} + (N+1)^{-1} \sum_{j=1}^{i-1} \delta_{j,k}$. The stationarity condition over $(x_{1,k})$ gives for all $1 \leq k \leq T-1$,

$$\frac{1}{N} \sum_{i=1}^N \left[\frac{x_{i,k} - x_{i,k-1}}{\Delta t_k} - \frac{x_{i,k+1} - x_{i,k}}{\Delta t_{k+1}} \right] = 0.$$

Letting $\bar{x}_k := (1/N) \sum_{i=1}^N x_{i,k}$, this can be rewritten as

$$\frac{\bar{x}_k - \bar{x}_{k-1}}{\Delta t_k} = c,$$

for some constant $c \in \mathbb{R}$ and all $k \in \{1, \dots, T\}$. Thus we have

$$\bar{x}_k = \bar{x}_0 + \sum_{l=1}^k \Delta t_l \left(\frac{\bar{x}_l - \bar{x}_{l-1}}{\Delta t_l} \right) = \bar{x}_0 + ct_k,$$

with $t_k := \sum_{l=1}^k \Delta t_l$; using the boundary conditions, we obtain

$$\bar{x}_k = (1 - t_k)\bar{x}_0 + t_k\bar{x}_T. \quad (95)$$

Notice that

$$\bar{x}_k = x_{1,k} + \frac{1}{N(N+1)} \sum_{i=1}^N \sum_{j=1}^{i-1} \delta_{j,k} = x_{1,k} + \frac{1}{N+1} \sum_{j=1}^{N-1} \left(1 - \frac{j}{N}\right) \delta_{j,k}.$$

Therefore, by eq. (95), the minimizer of eq. (94) satisfies

$$x_{1,k} = (1 - t_k)\bar{x}_0 + t_k\bar{x}_T - \frac{1}{N+1} \sum_{j=1}^{N-1} \left(1 - \frac{j}{N}\right) \delta_{j,k}.$$

Notice that this expression remains valid for the fixed boundary variables, i.e. for $k \in \{0, T\}$. In particular, we have for any $1 \leq k \leq T$

$$x_{1,k} - x_{1,k-1} = \Delta t_k (\bar{x}_T - \bar{x}_0) - \frac{\Delta t_k}{N+1} \sum_{j=1}^{N-1} \left(1 - \frac{j}{N}\right) d\delta_{j,k}. \quad (96)$$

We now plug back eq. (96) into eq. (94):

$$\begin{aligned} & \min_{(x_{1,k})_{1 \leq k \leq T-1}} \left\{ \frac{1}{N} \sum_{k=1}^T \frac{1}{\Delta t_k} \sum_{i=1}^N \left[x_{1,k} - x_{1,k-1} + \frac{\Delta t_k}{N+1} \sum_{j=1}^{i-1} d\delta_{j,k} \right]^2 \right\} \\ &= \frac{1}{N} \sum_{k=1}^T \Delta t_k \sum_{i=1}^N \left[\bar{x}_T - \bar{x}_0 + \frac{1}{N+1} \sum_{j=1}^{N-1} \left(\mathbb{1}\{j \leq i-1\} - 1 + \frac{j}{N} \right) d\delta_{j,k} \right]^2, \end{aligned}$$

$$\begin{aligned}
&\stackrel{(a)}{=} [\bar{x}_T - \bar{x}_0]^2 + \frac{1}{N} \sum_{k=1}^T \Delta t_k \sum_{i=1}^N \left[\frac{1}{N+1} \sum_{j=1}^{N-1} \left(\mathbb{1}\{j \leq i-1\} - 1 + \frac{j}{N} \right) d\delta_{j,k} \right]^2, \\
&= [\bar{x}_T - \bar{x}_0]^2 + \frac{1}{(N+1)^2} \sum_{k=1}^T \Delta t_k \sum_{j,l=1}^{N-1} (d\delta_{j,k})(d\delta_{l,k}) M_{jl},
\end{aligned}$$

using in (a) that $\sum_{i=1}^N (\mathbb{1}\{j \leq i-1\} - 1 + j/N) = 0$ for all $1 \leq j \leq N-1$, and with

$$\begin{aligned}
M_{jl} &:= \left[\frac{1}{N} \sum_{i=1}^N \left(\mathbb{1}\{j \leq i-1\} - 1 + \frac{j}{N} \right) \left(\mathbb{1}\{l \leq i-1\} - 1 + \frac{l}{N} \right) \right], \\
&= \frac{1}{N} \left[\min(j, l) - \frac{jl}{N} \right].
\end{aligned}$$

This completes the derivation of eq. (94), since one can easily check that $M = L_{N-1}^{-1}$, as defined in eq. (40).

A.4 Minimization algorithm with a rescaling parameter θ

Recall the discussion in Section 2.4.

Objective function – We let $t := \theta u \in (0, 1)$. We still consider time steps $\Delta t = (\Delta t_k)_{k=1}^T$, with $\sum_{k=1}^T \Delta t_k = 1$. It is a straightforward computation to check that eq. (42) becomes, under rescaling:

$$\begin{aligned}
&G_{N, \Delta t, \varepsilon}^{(\theta)}(\delta) \\
&= \frac{\theta}{2} [\bar{x}(0) - \bar{x}(1)]^2 + \frac{\pi^2}{6\theta N} \sum_{k=1}^T \sum_{i=1}^{N-1} \omega_k \sigma_\varepsilon(\delta_{i,k}) + \frac{\theta}{2(N+1)^2} \sum_{k=1}^T \Delta t_k (d\delta_k^\top) L_{N-1}^{-1} (d\delta_k),
\end{aligned} \tag{97}$$

The objective is still to minimize this functional over all $(\delta_{i,k})$ for $1 \leq i \leq N-1$ and $1 \leq k \leq T-1$, under the constraint $\delta_{i,k} \geq 0$ and the same boundary conditions as in eq. (38).

Time schedule – We use an argument similar to the one sketched in Section 2.2. At time $u \in (0, \theta^{-1})$, one checks easily from eq. (43) and Appendix A.1 that $\rho(u, x) dx$ is the law of a free Brownian bridge

$$X_u := (1-t)X_0 + tX_1 + \sqrt{\frac{t(1-t)}{\theta}} S, \tag{98}$$

where $t := \theta u \in (0, 1)$, and recall that X_0 has distribution μ , X_1 has distribution ν , and S is a semicircular variable, free with (X_0, X_1) . We fix an even integer $T \geq 2$ denoting the number of discretization points, and we focus on the regime $t \in [0, 1/2]$. We let $\tilde{T} := T/2$.

Eq. (98) shows that the evolution of X_u happens on two timescales: a timescale $t \sim \theta$, and a second timescale $t = \Theta(1)$. This is particularly important for $\theta \ll 1$, as we might miss relevant dynamics if we use a naive discretization with too few points close to $t = 0$. Moreover, for $t \downarrow 0$, we have $X_u \simeq X_0 + \sqrt{t/\theta} S$. This motivates a discretization of $t \in [0, 1/2]$ as $t_k = g_\theta(k/\tilde{T})$ for $k \in \{0, \dots, \tilde{T}\}$, with g_θ chosen such that:

- (i) $g_\theta(0) = 0$, $g_\theta(1) = 1/2$, and g_θ is strictly increasing.
- (ii) $g_{\theta=1}(v) = v^2/2$, to be consistent with the case $\theta = 1$ of Section 2.2.
- (iii) $g_\theta(v) \sim \theta v^2/2$ for $v \downarrow 0$.
- (iv) As $\theta \downarrow 0$, $g_\theta(v) = \Theta(\theta)$ for each fixed $v < 1/2$ and $g_\theta(v) = \Theta(1)$ for each fixed $v > 1/2$. This ensures that, in the limit $\theta \downarrow 0$, half of the discretized points cover the timescale $t \sim \theta$.

An example of a simple function satisfying all the properties above is:

$$g_\theta(v) := \frac{-1 + 4v^2 + \sqrt{(1 - 4v^2)^2 + 4\theta v^2(3 + \theta)}}{4(3 + \theta)}, \quad (99)$$

and this is the choice we make in our numerical simulations. We show the behavior of this function in Fig. 6. All in all, the time steps we use for a generic $\theta \geq 0$ are:

$$t_k := \begin{cases} g_\theta\left(\frac{k}{T/2}\right) & , \quad (k \in \{0, \dots, T/2\}), \\ g_\theta\left(\frac{T-k}{T/2}\right) & , \quad (k \in \{T/2 + 1, \dots, T\}). \end{cases} \quad (100)$$

and we let $\Delta t_k := t_k - t_{k-1}$.

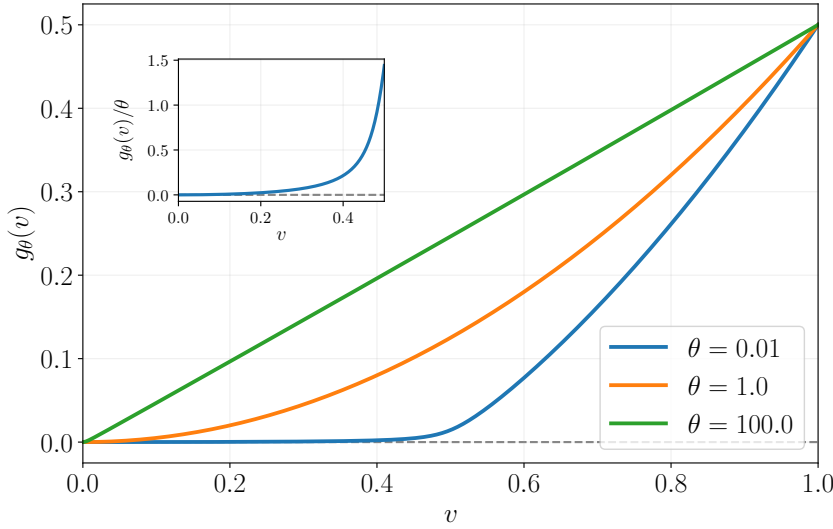


Figure 6: The function $g_\theta(v)$ defined in eq. (99), for different values of θ .

Preconditioning – The preconditioning procedure of Section 2.3, i.e. the approximation to the Hessian of $G_{N,\Delta t,\varepsilon}^{(\theta)}$, generalizes straightforwardly. For completeness, we state that in this case the Hessian reads:

$$\text{Hess } G_{N,\Delta t,\varepsilon}^{(\theta)} = \frac{\pi^2}{6\theta N} \text{Diag} [w_k \sigma_\varepsilon''(\delta_{i,k})] + \frac{\theta}{(N+1)^2} L_{N-1}^{-1} \otimes L_{T-1}(\Delta t), \quad (101)$$

while the approximation used in the preconditioner is:

$$H^{(\theta)}(\delta) := \mathbf{I}_{N-1} \otimes \text{Diag} \left[\left(\frac{\pi^2 w_k}{6\theta N(N-1)} \sum_{i=1}^{N-1} \sigma_\varepsilon''(\delta_{i,k}) \right)_{k=1}^{T-1} \right] \mathbf{I}_{T-1} + \frac{\theta}{(N+1)^2} L_{N-1}^{-1} \otimes L_{T-1}(\Delta t). \quad (102)$$

A.5 Preconditioned conjugate gradient

We recall here the preconditioned conjugate gradient procedure (see e.g. [70]) in Algorithm 2. The classical conjugate gradient method corresponds to $H = \mathbf{I}$. However, this method can suffer from slow convergence if the condition number $\kappa(A) := \lambda_{\max}(A)/\lambda_{\min}(A)$ is too large. To remedy this issue, if $H = EE^T$ is the Cholesky decomposition, Algorithm 2 effectively applies the classical conjugate gradient method to the linear system $E^{-1}A(E^{-1})^T y = E^{-1}b$, where $y = E^T x$. One thus aims to choose a preconditioner H such that both the following hold.

Algorithm 2: Preconditioned conjugate gradient (PCG).

Result: The unique solution $x = \text{PCG}(H; A, b)$ to the linear system $Ax = b$

Input: A positive-definite matrix A , a vector b , a positive-definite matrix H (the *preconditioner*), an initial guess x_0 , and a precision target $\varepsilon > 0$;

Initialize $r_0 = b - Ax_0$, $z_0 = H^{-1}r_0$, $p_0 = z_0$;

while $\|r_k\|_2 > \varepsilon\|b\|_2$ **do**

$$\left\{ \begin{array}{lcl} \alpha_k & = & \frac{r_k^\top z_k}{p_k^\top A p_k}, \\ x_{k+1} & = & x_k + \alpha_k p_k, \\ r_{k+1} & = & b - Ax_{k+1} = r_k - \alpha_k A p_k, \\ z_{k+1} & = & H^{-1} r_{k+1}, \\ p_{k+1} & = & z_{k+1} + \frac{r_{k+1}^\top z_{k+1}}{r_k^\top z_k} p_k, \end{array} \right.$$

$k = k + 1$;

end

(i) $\kappa(E^{-1}A(E^{-1})^\top)$ is much smaller than $\kappa(A)$, i.e. H is a “good approximation” to A .

(ii) Solving linear systems of the type $Hx = z$ can be done at a low computational cost.

B Runtime and convergence analysis experiments

We report in this section results of experiments on the total runtime of our algorithm (Appendix B.1) and its convergence properties (Appendix B.2).

B.1 Runtime analysis

We analyze here the total runtime of Algorithm 1. We take $\mu = \nu = \sigma_{\text{s.c.}}$ as boundary densities, and vary the parameter $\theta \in \{0.01, 1.0\}$. Runtime includes optimization time but does not include the time required to set up the problem or the time required to certify convergence of the last step.

Scaling with the time discretization parameter T – In Fig. 7 we show how the total runtime scales for large T , at fixed number of particles N . Notice that the number of operations per iteration is $\mathcal{O}(T)$ (for a fixed value of N and increasing T). The observed scaling is even slower, exhibiting a very favorable runtime scaling as T increases.

Scaling with the number of particles N – In Fig. 8 we show how the total runtime scales for large N , at fixed time grid size T . We find two clear behaviors: for $\theta = 1$ the total runtime scales as $\mathcal{O}(N)$, while for $\theta = 0.01$ it appears to scale rather as $\mathcal{O}(N^{3/2})$. Notice that the number of operations per PCG iteration is $\mathcal{O}(N \log N)$ (for a fixed value of T). The observed scalings therefore suggest that the total number of PCG iterations should increase with N for smaller θ , signaling a more ill-conditioned problem. This is compatible with what we observe on the total number of PCG iterations below (see Fig. 10).

Joint scaling in N, T – In Fig. 9 we show the runtime scaling as a function of a single parameter, either $N^{3/2}T$ (for $\theta = 0.01$) or NT (for $\theta = 1.0$). In both cases we find that these single parameters represent relatively well the scaling of the total runtime, compatible with the fixed- N and fixed- T experiments shown in Figs 7 and 8.

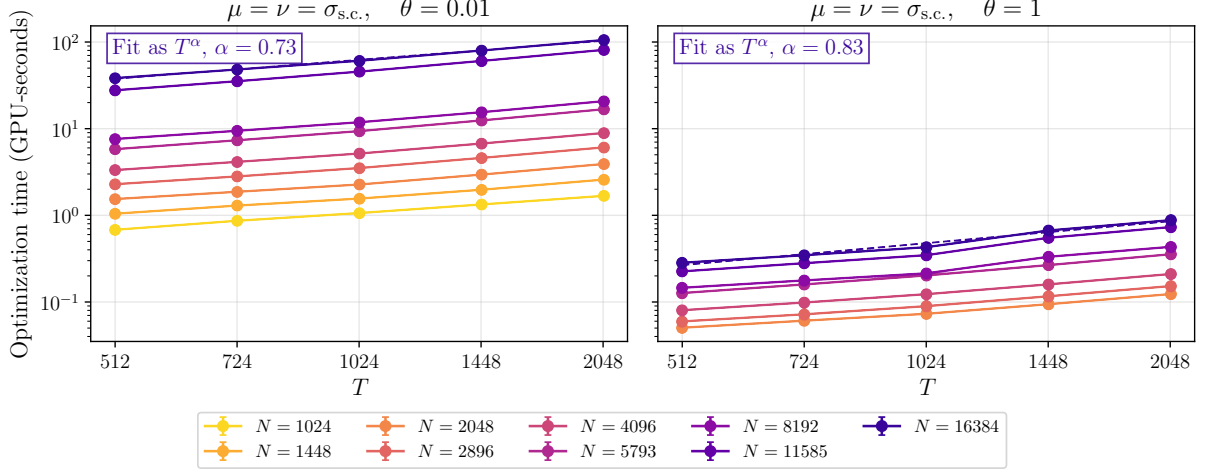


Figure 7: Runtime scaling analysis for increasing T , on a single RTX8000 GPU. Each point is averaged over 5 runs, error bars are very small (almost invisible).

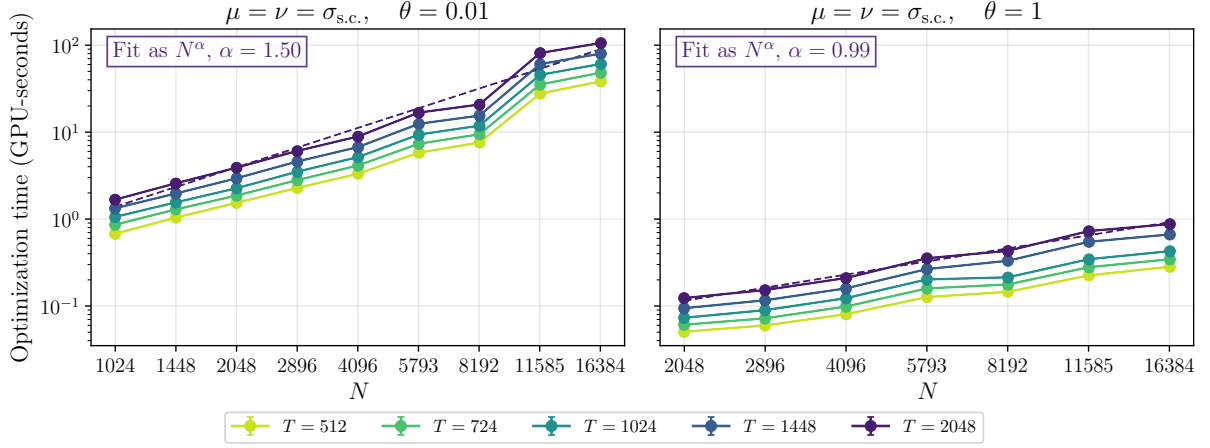


Figure 8: Runtime scaling analysis for increasing N , on a single RTX8000 GPU. Average over 5 runs each time, error bars are very small (almost invisible).

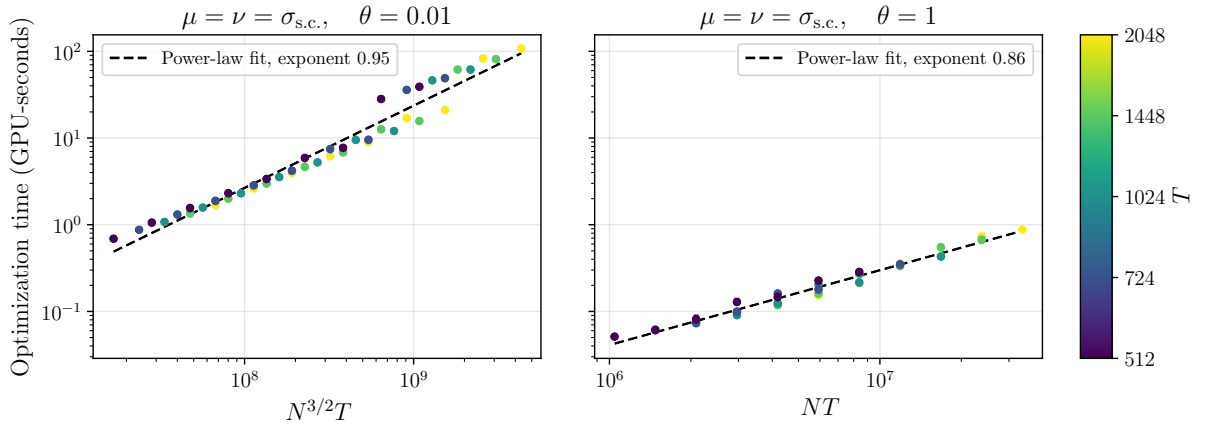


Figure 9: Runtime scaling analysis for increasing both N and T , on a single RTX8000 GPU. Average over 5 runs each time, error bars are very small (almost invisible).

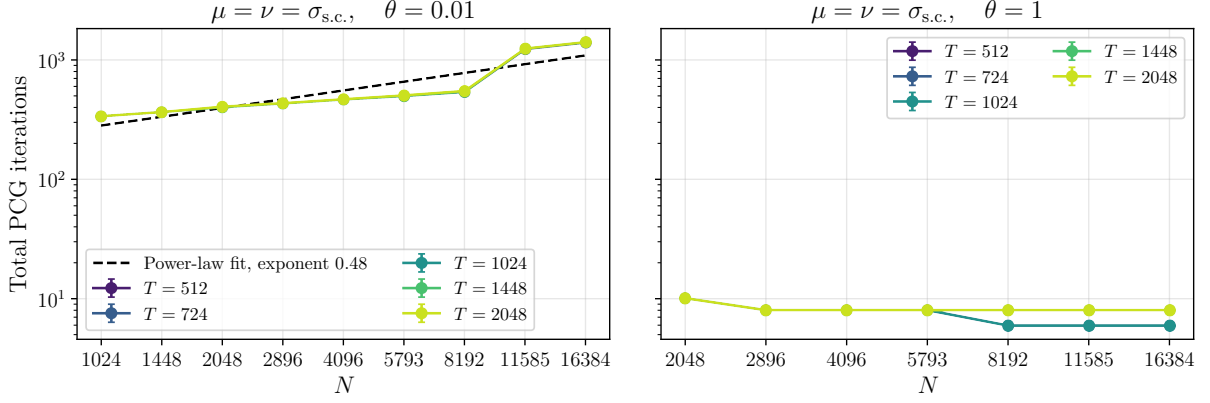


Figure 10: Total number of PCG iterations performed (across all Newton updates), as a function of N , for a given value of T .

Total number of PCG iterations – Finally, in Fig. 10 we show the total number of PCG iterations across all Newton updates, as a function of N and for a range of values of T . One can make two remarks: first, the number appears independent of T in both cases studied, suggesting that refining the time discretization does not worsen the conditioning of the problem. Secondly, for $\theta = 1$ the problem is (after pre-conditioning) relatively well-conditioned even as we increase N , since the total number of PCG iterations is roughly constant. However, for $\theta = 0.01$, the number of iterations appears to scale approximately as $N^{1/2}$, which is likely due to a Hessian near the global minimum which is more ill-conditioned despite the preconditioning. Recall that since each PCG update requires $\mathcal{O}(NT \log N)$ iterations, multiplying the number of PCG iterations by $NT \log N$ gives a rough estimate of the computation time, compatible with the figures shown above.

B.2 Convergence analysis

In this section, we analyze the convergence of the output of Algorithm 1 $J_{N,\Delta t,\varepsilon}$ (see eq. (35)) to its asymptotic limit as $N, T \rightarrow \infty$, i.e. $J(\theta, \mu, \nu)$ of eq. (7). Throughout this section, we have $\varepsilon = 10^{-7}$ fixed, we denote $J_{N,\Delta t,\varepsilon}(\theta, \mu, \nu)$ as $J_{N,T}$ to highlight its dependency on N, T , and denote simply $J(\theta, \mu, \nu)$ as J . For concreteness, we perform the analysis in two cases for the boundary measures:

- (C1) We set $\mu = \nu = \sigma_{\text{s.c.}}$, with the scaling $\theta = 0.01$. In this case the value of J is known analytically, so that both the successive differences and the signed total error $J_{N,T} - J$ are accessible.
- (C2) We use the same setting as Fig. 4: $\mu = \mu_{\text{sMP},\kappa}$ the symmetrized Marchenko-Pastur law (see eq. (26)) with $\kappa = 2$, $\nu = \sigma_{\text{s.c.},0.5}$, and $\theta = 1$. Here J is not known analytically, and we rely on successive differences only to estimate the convergence rates.

Anticipating on the results of this section, the experiments below are all consistent with the two leading discretization errors being of different orders *and* essentially decoupled, i.e. with an error model

$$J_{N,T} = J + \frac{a}{T^2} - \frac{b}{N} + o(T^{-2}) + o(N^{-1}), \quad (103)$$

where $a, b > 0$ depend on (θ, μ, ν) but not on (N, T) . Note in particular the opposite signs: the time discretization overestimates J , while the particle discretization underestimates it.

B.2.1 Convergence as $T \rightarrow \infty$ and $\|\Delta t\|_\infty \rightarrow 0$

We first study the error introduced by the time discretization, for fixed values of the number of particles N . We analyze (C1) in Fig. 11, and (C2) in Fig. 12. In both cases the

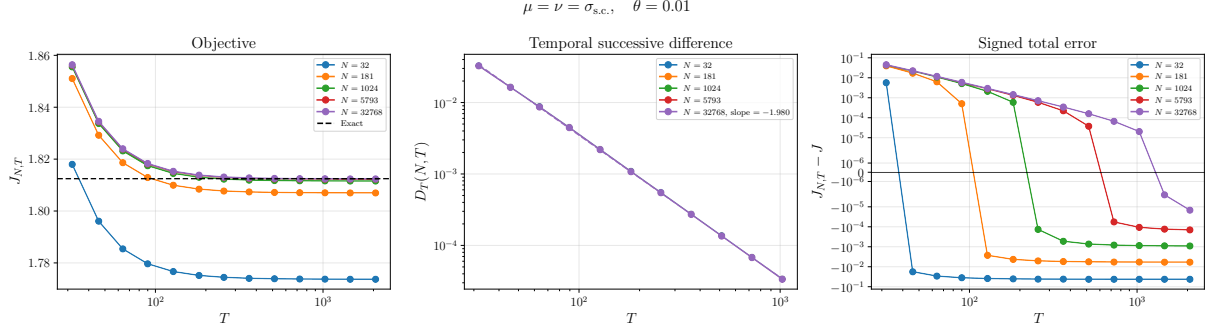


Figure 11: Convergence analysis as the number of time discretization steps T grows, for (C1).

For different values of N we show the value of the objective alongside its exact asymptotic value (left), the successive difference $D_T(N, T) := |J_{N,2T} - J_{N,T}|$ (middle), and the signed total error $J_{N,T} - J$ (right). The curves in the middle plot are superimposed: the time discretization error is essentially N -independent, and a least-squares fit is consistent with a $1/T^2$ decay. In the right plot, the error changes sign at an N -dependent value of T : it is positive as long as the time discretization dominates, and saturates at the (negative) finite- N bias, in agreement with eq. (103).

$$\mu = \text{sMP}(\kappa = 2), \quad \nu = \sigma_{s.c.}(0.5), \quad \theta = 1$$

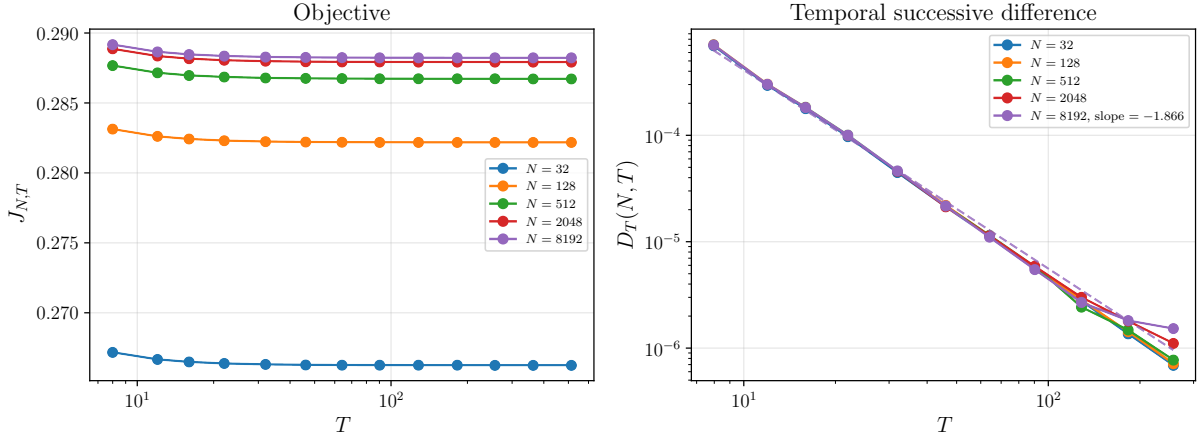


Figure 12: Convergence analysis as the number of time discretization steps T grows, for (C2). For different values of N we show the value of the objective (left) and the successive difference $D_T(N, T) := |J_{N,2T} - J_{N,T}|$ (right). The objective is already within $\sim 10^{-4}$ of its large- T limit for $T \simeq 30$, which is why we restrict to $T \leq 512$ here. The fitted slope is again compatible with a $1/T^2$ decay.

time-discretization error is compatible with a $1/T^2$ decay, in accordance with the order of our discretization scheme, and is only weakly dependent on N . We also note that the objective is already very close to its large- T limit for moderate values of $T \sim 10^2$, so that in practice the accuracy of Algorithm 1 is often limited by N rather than by T .

B.2.2 Convergence as $N \rightarrow \infty$

We then study the convergence rate of the objective as $N \rightarrow \infty$, for fixed values of the time discretization grid size T . We analyze (C1) in Fig. 13, and (C2) in Fig. 14. The finite- N error

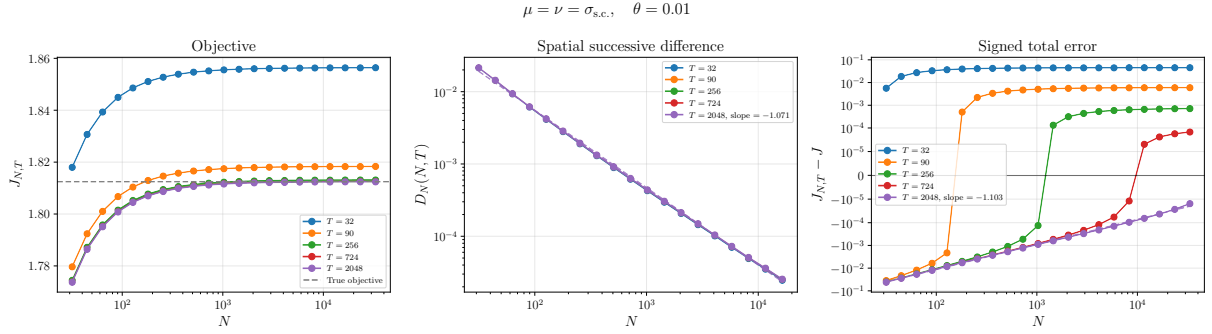


Figure 13: Convergence analysis as the number of particles N grows, for (C1). For different values of T we show the value of the objective alongside its exact asymptotic value (left), the successive difference $D_N(N, T) := |J_{2N,T} - J_{N,T}|$ (middle), and the signed total error $J_{N,T} - J$ (right). The curves in the middle plot are superimposed, so that the finite- N error is essentially T -independent, and is compatible with eq. (103). The fitted slope is compatible with a $1/N$ leading-order decay.

$$\mu = \text{sMP}(\kappa = 2), \quad \nu = \sigma_{s.c.}(0.5), \quad \theta = 1$$

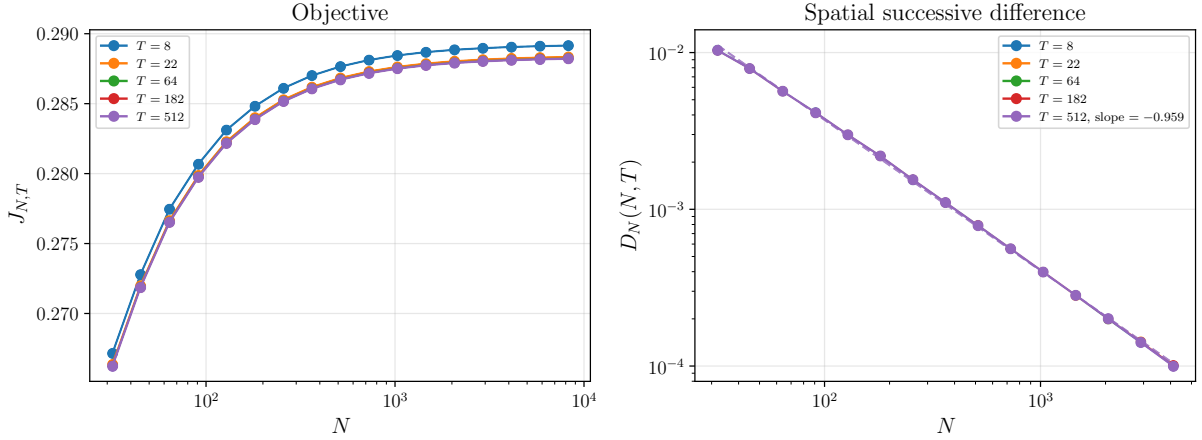


Figure 14: Convergence analysis as the number of particles N grows, for (C2). For different values of T we show the value of the objective (left) and the successive difference $D_N(N, T) := |J_{2N,T} - J_{N,T}|$ (right). As in Fig. 13 the successive differences for different T are superimposed, and the fitted slope is compatible with a $1/N$ decay.

is thus compatible with a $1/N$ decay in both cases, and nearly independent of T , consistently with the prediction of eq. (103).

B.2.3 Richardson extrapolation in N

The $1/N$ decay observed above motivates the use of Richardson extrapolation [63]. The idea can be summarized as follows. If, at fixed T :

$$J_{N,T} = J_{N=\infty,T} + \frac{a_1}{N} + \frac{a_2}{N^\alpha} + o(N^{-\alpha}),$$

for some $\alpha > 1$ and some coefficients $a_1, a_2 \in \mathbb{R}$, then letting

$$J_{N,T}^{(R)} := 2J_{N,T} - J_{N/2,T},$$

one has

$$J_{N,T}^{(R)} = J_{N=\infty,T} - \frac{2(2^{\alpha-1} - 1)a_2}{N^\alpha} + o(N^{-\alpha}), \quad (104)$$

i.e. the leading $1/N$ term is eliminated. We apply this procedure for (C1) in Fig. 15, and to (C2)

$$\mu = \nu = \sigma_{s.c.}, \quad \theta = 0.01$$

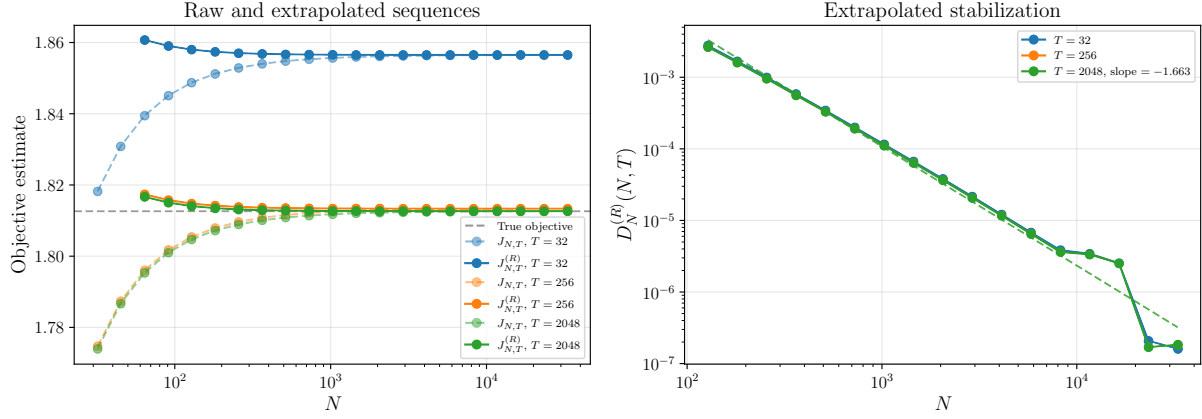


Figure 15: Richardson extrapolation (see eq. (104)) for (C1). We plot the raw and extrapolated objective estimates (left), as well as the successive difference

$$D_N^{(R)}(N, T) := |J_{2N,T}^{(R)} - J_{N,T}^{(R)}| \text{ (right).}$$

in Fig. 16. We observe a clear acceleration of the convergence, compatible with eq. (104) with

$$\mu = \text{sMP}(\kappa = 2), \quad \nu = \sigma_{s.c.}(0.5), \quad \theta = 1$$

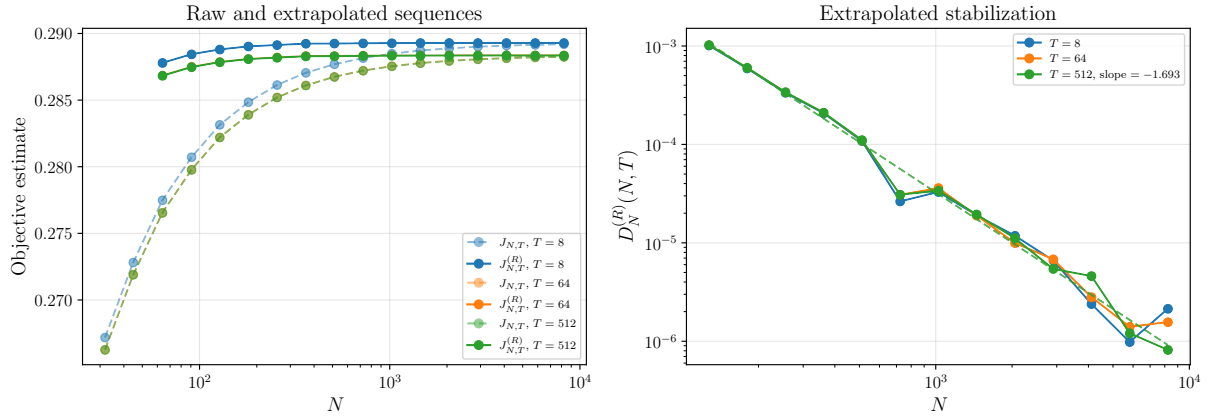


Figure 16: Richardson extrapolation (see eq. (104)) for (C2). We plot the raw and extrapolated objective estimates (left), as well as the successive difference

$$D_N^{(R)}(N, T) := |J_{2N,T}^{(R)} - J_{N,T}^{(R)}| \text{ (right).}$$

an exponent $\alpha \in [1.6, 1.7]$. In practice the gain is substantial: reaching an error of 10^{-4} in (C1) requires $N \simeq 10^4$ for the raw sequence against $N \simeq 10^3$ for the extrapolated one.