The Coulomb flow: from optimisation to a PDE

Propagation of a Wasserstein Polyak–Łojasiewicz inequality and its relation to well-posedness

Minimising the Coulomb discrepancy $\mathcal{E}_\nu(\mu) = \frac{1}{2}\|\mu - \nu\|^2_{H^{-1}}$ over probability measures is a convex problem in the linear structure, but its Wasserstein gradient flow is a nonlinear transport equation, and the desired exponential rate does not follow from displacement convexity. Instead, it follows from a Polyak–Łojasiewicz inequality whose constant is the essential infimum of the current density, which may in principle degenerate along the flow. Assuming a sufficiently smooth solution, the argument below proves propagation of this constant and exponential convergence with an explicit rate. The regularity theory itself (Schauder estimates for $\log \mu_0, \log \nu \in C^{0,\alpha}$) is quoted, not proved; see [2].

Part I — Principles recalled

Throughout, M is a smooth closed (compact, boundary-less) Riemannian manifold with volume measure vol, normalised to $\mathrm{vol}(M) < \infty$, and $\mu, \nu \in \mathcal{P}(M)$ are probability measures with densities also written μ, ν. We set $\sigma := \mu - \nu$, so $\int_M \sigma \,\mathrm{d}\,\mathrm{vol} = 0$.

Principle 1 (Coulomb discrepancy as an H^{-1} norm)**.** Let G be the zero-mean Green kernel of $-\Delta$ on M and $u_\sigma := (-\Delta)^{-1}\sigma$, i.e. $-\Delta u_\sigma = \sigma$ with $\int u_\sigma \mathrm{d}\,\mathrm{vol} = 0$. Then

$$2\mathcal{E}_\nu(\mu) := \iint_{M\times M} G(x,y)\,\mathrm{d}\sigma(x)\mathrm{d}\sigma(y) = \langle \sigma, (-\Delta)^{-1}\sigma\rangle = \int_M |\nabla u_\sigma|^2 \,\mathrm{d}\,\mathrm{vol} = \|\sigma\|^2_{H^{-1}}.$$

This is a maximum mean discrepancy with the singular kernel G; the energy space is $H^1(M)/\mathbb{R}$, and $\mathcal{E}_\nu(\mu) = 0$ iff $\mu = \nu$.

Principle 2 (First variation and the flow)**.** $\delta\mathcal{E}_\nu/\delta\mu = u_\sigma$, so the Wasserstein gradient flow $\partial_t\mu_t = \mathrm{div}\big(\mu_t \nabla \frac{\delta\mathcal{E}_\nu}{\delta\mu}\big)$ reads

$$\boxed{\partial_t \mu_t = \mathrm{div}\big(\mu_t \nabla u_t\big), \qquad -\Delta u_t = \mu_t - \nu} \tag{1}$$

— a transport equation whose velocity field $v_t := -\nabla u_t$ is produced from the unknown by an elliptic solve. Total mass is preserved, so $\int \sigma_t \mathrm{d}\,\mathrm{vol} = 0$ for all t.

Principle 3 (Metric slope and energy dissipation)**.** The Wasserstein gradient of $\mathcal{E}_\nu$ at μ is ∇u_σ, measured in $L^2(\mu)$:

$$|\partial\mathcal{E}_\nu|^2(\mu) = \int_M |\nabla u_\sigma|^2 \,\mathrm{d}\mu, \qquad \text{and along (1)} \qquad \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}_\nu(\mu_t) = -|\partial\mathcal{E}_\nu|^2(\mu_t).$$

The energy integrates $|\nabla u|^2$ against vol, whereas the slope integrates the same quantity against μ.

Principle 4 (Polyak–Łojasiewicz)**.** A functional satisfies a PL (or gradient-domination) inequality with constant $\lambda > 0$ at μ if $\frac{1}{2}|\partial\mathcal{E}_\nu|^2(\mu) \geq \lambda\,\mathcal{E}_\nu(\mu)$. If this holds *along a whole trajectory*, then $\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}_\nu(\mu_t) \leq -2\lambda\mathcal{E}_\nu(\mu_t)$ and $\mathcal{E}_\nu(\mu_t) \leq e^{-2\lambda t}\mathcal{E}_\nu(\mu_0)$. PL requires neither convexity nor uniqueness of minimisers. Here its constant depends on the state.

Principle 5 (Lagrangian form)**.** If v_t is Lipschitz in space, uniformly on $[0,T]$, the flow map $\dot{X}_t(x) = v_t(X_t(x))$, $X_0 = \mathrm{id}$, is well defined, and $\mu_t = (X_t)_\#\mu_0$. Along a characteristic, a density transported by (1) obeys $\frac{\mathrm{d}}{\mathrm{d}t}\big[\mu_t(X_t)\big] = -\mu_t(X_t)\,\mathrm{div}\,v_t(X_t)$.

Part II — Exercise

Step 1 (the energy, the slope, and the dissipation identity). Prove the three identities of Principles 1–3: that $2\mathcal{E}_\nu(\mu) = \int |\nabla u_\sigma|^2 \mathrm{d}\,\mathrm{vol}$, that $\delta\mathcal{E}_\nu/\delta\mu = u_\sigma$, and that $\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}_\nu(\mu_t) = -\int |\nabla u_t|^2 \mathrm{d}\mu_t$ along (1).

Step 2 (a conditional PL inequality). Suppose $\mu \geq m\,\mathrm{vol}$ for some $m > 0$. Show

$$\tfrac{1}{2}|\partial\mathcal{E}_\nu|^2(\mu) \;\geq\; m\,\mathcal{E}_\nu(\mu).$$

Identify where the hypothesis is used and explain why no such bound holds without it. Show also that $m \leq 1/\operatorname{vol}(M)$, so that the rate obtained below can never exceed $2/\operatorname{vol}(M)$.

Step 3 (state the reduction precisely). Step 2 gives a PL constant $m(\mu_t) := \operatorname{ess\,inf}\mu_t$ that depends on time. Write down the exponential decay one would get, and explain why the argument is *not* yet a proof: give the inequality that actually follows from Steps 1–2 without further information, and identify the single missing statement. Formulate it as: *Claim (P).* $\inf_{t\geq 0}\operatorname{ess\,inf}\mu_t > 0$.

Step 4 (the density along a characteristic solves a logistic ODE). Assume a solution smooth enough for Principle 5. Using $\operatorname{div} v_t = -\Delta u_t$ and the elliptic equation in (1), prove the pointwise identity

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}t}\big[\mu_t(X_t)\big] = \mu_t(X_t)\Big(\nu(X_t) - \mu_t(X_t)\Big).} \tag{2}$$

Identify the property of the Coulomb kernel that turns the divergence of the velocity into an *algebraic* expression in the unknown? Would the same computation work for the MMD flow of a smooth positive-definite kernel?

Step 5 (propagation of positivity, and of a ceiling). Let $y(t) := \mu_t(X_t)$ and $a(t) := \nu(X_t)$, so that $\dot{y} = y(a - y)$. Assume

$$0 < m_\nu \leq \nu \leq M_\nu < \infty, \qquad 0 < m_0 \leq \mu_0 \leq M_0 < \infty.$$

(a) Show $y(t) > 0$ for all t for which the solution exists.

(b) By scalar comparison, show $y(t) \geq m := \min\{m_0, m_\nu\}$ and $y(t) \leq M := \max\{M_0, M_\nu\}$ for all such t.

(c) Conclude Claim (P), with the explicit constant m.

Step 6 (close the loop). Combine Steps 2 and 5 to obtain

$$\mathcal{E}_\nu(\mu_t) \leq e^{-2mt}\,\mathcal{E}_\nu(\mu_0), \qquad \text{equivalently} \qquad \|\mu_t - \nu\|_{H^{-1}} \leq e^{-mt}\|\mu_0 - \nu\|_{H^{-1}}.$$

Step 7 (what has been assumed, and why the reduction is not circular). List precisely the properties of the solution used in Steps 4–6. Then explain the logical structure: Steps 4–5 produce *a priori* bounds $m \leq \mu_t \leq M$ that are uniform in t; explain why such bounds are exactly what a continuation argument needs, so that *local* existence in a smooth class plus the above yields *global* existence, rather than the two hypotheses being circular. State the well-posedness result of [2] that is being quoted, without proving it.

Step 8 (why a closed manifold). Show that on $\mathbb{R}^d$ no probability density satisfies $\mu \geq m > 0$ everywhere, so that the PL constant of Step 2 is necessarily 0. Name a second, independent obstruction on $\mathbb{R}^d$. Finally, comment on the fact that $\mathcal{E}_\nu$ is convex along linear interpolations but that this exercise never used convexity.

Part III — Solutions

Solution. **Step 1.** With $-\Delta u_\sigma = \sigma$ and M closed, integrating by parts,

$$\langle \sigma, (-\Delta)^{-1}\sigma\rangle = \int_M u_\sigma\,(-\Delta u_\sigma)\,\mathrm{d}\,\mathrm{vol} = \int_M |\nabla u_\sigma|^2\,\mathrm{d}\,\mathrm{vol}\,.$$

For the first variation, $\mathcal{E}_\nu(\mu) = \frac{1}{2}\langle \mu-\nu, (-\Delta)^{-1}(\mu-\nu)\rangle$ is a quadratic form with the self-adjoint operator $(-\Delta)^{-1}$, so $\frac{\mathrm{d}}{\mathrm{d}\epsilon}\mathcal{E}_\nu(\mu+\epsilon\chi)|_{\epsilon=0} = \langle \chi, (-\Delta)^{-1}\sigma\rangle = \int \chi\, u_\sigma$, i.e. $\delta\mathcal{E}_\nu/\delta\mu = u_\sigma$ (defined up to the additive constant fixed by the zero-mean normalisation, which is irrelevant since $\int \chi = 0$). Finally, along (1),

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}_\nu(\mu_t) = \int_M u_t\,\partial_t\mu_t\,\mathrm{d}\,\mathrm{vol} = \int_M u_t\,\operatorname{div}(\mu_t\nabla u_t)\,\mathrm{d}\,\mathrm{vol} = -\int_M |\nabla u_t|^2\,\mathrm{d}\mu_t = -|\partial\mathcal{E}_\nu|^2(\mu_t).$$

Solution. **Step 2.** Since $|\nabla u_\sigma|^2 \ge 0$ and $\mu \ge m$ vol as measures,

$$\tfrac{1}{2}|\partial\mathcal{E}_\nu|^2(\mu) = \tfrac{1}{2}\int_M |\nabla u_\sigma|^2\,\mathrm{d}\mu \;\ge\; \tfrac{m}{2}\int_M |\nabla u_\sigma|^2\,\mathrm{d}\,\mathrm{vol} \;=\; m\,\mathcal{E}_\nu(\mu).$$

The hypothesis converts the reference measure of the slope (μ) into that of the energy (vol). Without a lower bound the inequality fails: if μ has a vacuum region V where ∇u_σ is large, the slope does not see it while the energy does. Concretely, concentrating μ away from the support of ∇u_σ makes $|\partial\mathcal{E}_\nu|^2(\mu)/\mathcal{E}_\nu(\mu)$ arbitrarily small.

For the last point, $1 = \int_M \mu\,\mathrm{d}\,\mathrm{vol} \ge m\,\mathrm{vol}(M)$, so $m \le 1/\mathrm{vol}(M)$: a uniform density has $m = 1/\mathrm{vol}(M)$ and no measure does better. The largest rate obtainable by this argument is therefore $2/\mathrm{vol}(M)$.

Solution. **Step 3.** What Steps 1–2 give is the differential inequality

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}_\nu(\mu_t) \;\le\; -2\,m(\mu_t)\,\mathcal{E}_\nu(\mu_t), \qquad m(\mu_t) = \operatorname{ess\,inf}\mu_t,$$

hence by Grönwall $\mathcal{E}_\nu(\mu_t) \le \exp\bigl(-2\int_0^t m(\mu_s)\mathrm{d}s\bigr)\mathcal{E}_\nu(\mu_0)$. This does not imply exponential decay if $m(\mu_s)$ tends rapidly to zero. The missing statement is

$$\textit{Claim (P).}\ \inf_{t\ge 0}\operatorname{ess\,inf}\mu_t =: m > 0.$$

The optimisation problem has thereby been reduced to a purely PDE question: does the flow (1) preserve a uniform lower bound on the density?

Solution. **Step 4.** By Principle 5, along a characteristic $\frac{\mathrm{d}}{\mathrm{d}t}\mu_t(X_t) = -\mu_t(X_t)\operatorname{div} v_t(X_t)$ with $v_t = -\nabla u_t$. Now

$$\operatorname{div} v_t = -\Delta u_t = \mu_t - \nu,$$

by the elliptic equation in (1). Substituting gives (2).

The Coulomb kernel is the Green function of $-\Delta$, so applying the divergence recovers the source: $\operatorname{div}\nabla = \Delta$ and $-\Delta u_t = \mu_t - \nu$. The velocity divergence is therefore a pointwise algebraic function of the unknown. For a smooth positive-definite kernel k, the velocity is $-\nabla(k * \sigma)$ and $\operatorname{div} v = -(\Delta k)*\sigma$, a *nonlocal* expression that cannot be compared pointwise with μ; the comparison argument below therefore does not apply.

Solution. **Step 5.** Write $\dot{y} = y(a-y) =: f(t,y)$ with $a(t) = \nu(X_t) \in [m_\nu, M_\nu]$.

(a) $y \equiv 0$ solves the ODE, so by uniqueness (the right-hand side is locally Lipschitz in y) a solution with $y(0) = \mu_0(x) > 0$ cannot reach 0 in finite time. Equivalently $\frac{\mathrm{d}}{\mathrm{d}t} \log y = a - y$ stays finite as long as y does.

(b) *Lower bound.* Let $m = \min\{m_0, m_\nu\}$ and let z solve $\dot{z} = z(m_\nu - z)$, $z(0) = m$. Since $m \leq m_\nu$, z is nondecreasing, so $z(t) \geq m$ for all t. Moreover for $y \geq 0$

$$f(t,y) = y(a(t) - y) \;\geq\; y(m_\nu - y),$$

and $y(0) \geq m_0 \geq m = z(0)$; the scalar comparison theorem gives $y(t) \geq z(t) \geq m$.

Upper bound. Symmetrically, let $M = \max\{M_0, M_\nu\}$ and Z solve $\dot{Z} = Z(M_\nu - Z)$, $Z(0) = M$. Since $M \geq M_\nu$, Z is nonincreasing, so $Z \leq M$; and $f(t,y) \leq y(M_\nu - y)$ with $y(0) \leq M_0 \leq M$, whence $y(t) \leq Z(t) \leq M$.

(c) The bounds hold at every $x \in M$, uniformly, since m and M do not depend on the characteristic. As $\mu_t = (X_t)_{\#}\mu_0$ and X_t is a diffeomorphism, this is exactly $m \leq \mu_t \leq M$, i.e. Claim (P) with the constant $m = \min\{m_0, m_\nu\}$.

Remark. Equation (2) is a logistic equation with carrying capacity $\nu(X_t)$. Consequently, the interval

$$[\min\{m_0, m_\nu\}, \max\{M_0, M_\nu\}]$$

is invariant along characteristics.

Solution. **Step 6.** By Step 5, $\operatorname{ess\,inf} \mu_t \geq m$ for every t, so Step 2 applies at every time with the *same* constant. With Step 1,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}_\nu(\mu_t) = -|\partial\mathcal{E}_\nu|^2(\mu_t) \leq -2m\,\mathcal{E}_\nu(\mu_t) \quad\Longrightarrow\quad \mathcal{E}_\nu(\mu_t) \leq e^{-2mt}\mathcal{E}_\nu(\mu_0).$$

Since $2\mathcal{E}_\nu = \|\mu - \nu\|^2_{H^{-1}}$ (Principle 1), taking square roots gives the stated decay of $\|\mu_t - \nu\|_{H^{-1}}$. Because $\|\cdot\|_{H^{-1}}$ metrises weak convergence against H^1 test functions, this is genuine convergence $\mu_t \to \nu$, at an explicit exponential rate.

Solution. **Step 7.** The properties used are: (i) a solution μ_t of (1) on $[0,T]$ with μ_t continuous and positive; (ii) sufficient spatial regularity, for example $u_t \in C^{2,\alpha}$, so that $v_t = -\nabla u_t$ is Lipschitz and the flow map X_t exists, is unique, and is a diffeomorphism, with $\mu_t = (X_t)_{\#}\mu_0$; and (iii) sufficient temporal regularity to differentiate $\mathcal{E}_\nu(\mu_t)$ and integrate by parts. No regularity result is proved here.

The structure is a bootstrap, not a circle:

$$\begin{array}{rcl} \text{local existence in a smooth class} & \longrightarrow & \text{a priori bounds } m \leq \mu_t \leq M \\ & \longrightarrow & \text{global existence} \\ & \longrightarrow & \text{exponential decay.} \end{array}$$

Steps 4–5 use only a solution on $[0,T]$ and produce bounds independent of T. These bounds prevent density degeneration at a finite maximal time. In the quoted local theory, they combine with the Hölder and elliptic estimates for $-\Delta u_t = \mu_t - \nu$ to control the regularity required by the continuation criterion. The result of [2] is: on a closed manifold, if $\log \mu_0$ and $\log \nu$ are $C^{0,\alpha}$, the flow (1) has a global solution in that class, remains uniformly positive, and satisfies $\mathcal{E}_\nu(\mu_t) \leq e^{-2mt}\mathcal{E}_\nu(\mu_0)$. The present exercise is the second and fourth arrows; the first and third are the content of that paper.

Solution. **Step 8.** If $\mu \geq m > 0$ on $\mathbb{R}^d$ then $\int_{\mathbb{R}^d} \mu = +\infty$ unless $m = 0$: no probability density is uniformly positive on an unbounded domain, so $\operatorname{ess\,inf} \mu = 0$ and the PL constant of Step 2 vanishes identically. A second obstruction is the loss of compactness: mass may escape to infinity, and the compact continuation estimates used above are unavailable. One may instead seek local estimates on bounded regions and finite time intervals.

Finally: $\mathcal{E}_\nu$ is a positive quadratic form of $\mu - \nu$, hence convex along *linear* interpolations $t \mapsto (1-t)\mu_0 + t\mu_1$. That convexity is with respect to the wrong geometry — the flow (1) is a gradient flow for W_2, and displacement convexity of $\mathcal{E}_\nu$ is not available. The exponential rate instead follows from a Polyak–Łojasiewicz inequality propagated by the pointwise logistic structure (2); the optimisation question is thereby reduced to a comparison principle.

References

[1] L. Ambrosio, N. Gigli and G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*, 2nd ed., Birkhäuser, 2008.

[2] S. Boufadène and F.-X. Vialard, *On the global convergence of Wasserstein gradient flow of the Coulomb discrepancy*, arXiv:2312.00800; SIAM Journal on Mathematical Analysis (2025).

[3] H. Karimi, J. Nutini and M. Schmidt, *Linear convergence of gradient and proximal-gradient methods under the Polyak–Łojasiewicz condition*, ECML PKDD, 2016.

[4] R. J. McCann, *A convexity principle for interacting gases*, Advances in Mathematics **128** (1997), 153–179.