On optimal transport maps between $\frac{1}{d}$-concave densities

Guillaume Carlier, Alessio Figalli, and Filippo Santambrogio

Abstract. In this paper we extend the scope of Caffarelli's contraction theorem, which provides a measure of the Lipschitz constant for optimal transport maps between log-concave probability densities in $\mathbb{R}^d$. Our focus is on a broader category of densities, specifically those that are $\frac{1}{d}$-concave and can be represented as V^{-d}, where V is convex. By setting appropriate conditions, we derive linear or sublinear limitations for the optimal transport map. This leads us to a comprehensive Lipschitz estimate that aligns with the principles established in Caffarelli's theorem.

1. Introduction

Given two probability densities f and g on $\mathbb{R}^d$ with finite second moments, the quadratic optimal transport problem between $\mu(\mathrm{d}x) = f(x)\,\mathrm{d}x$ and $\nu(\mathrm{d}y) = g(y)\,\mathrm{d}y$ consists in finding the change of variables T between μ and ν that minimizes the mean squared displacement,

$$\int_{\mathbb{R}^d} |x - T(x)|^2 f(x)\,\mathrm{d}x$$

subject to the constraint $T_{\#}\mu = \nu$, requiring that T transports μ to ν, i.e.,

$$\int_{T^{-1}(A)} f(x)\,\mathrm{d}x = \int_A g(y)\,\mathrm{d}y \quad \text{for every Borel subset } A \text{ of } \mathbb{R}^d.$$

A seminal work of Brenier [2] states that there exists a unique (up to negligible sets for μ) solution T to this quadratic optimal transport problem and that T is characterized by the fact that it has a convex potential i.e., it is of the form $T = \nabla\varphi$ with φ convex. Without assuming that μ and ν have finite second moments, McCann [15] extended Brenier's result (with a different strategy based on cyclical monotonicity arguments rather than on optimal transport). The Brenier–McCann map $T = \nabla\varphi$ with φ convex, seen as the monotone change of variables between the absolutely continuous measures μ and ν on $\mathbb{R}^d$, is by now considered a fundamental object which solves, in particular, in some suitable weak sense the Monge–Ampère equation,

$$\det(D^2\varphi)g(\nabla\varphi) = f. \tag{1.1}$$

The regularity theory for the monotone transport and the corresponding Monge–Ampère equation, initiated by Caffarelli [3], has stimulated an intensive line of research in the last 30 years (see the survey [9] and the textbook [11]). We shall consider the case of densities f and g which are (at least) $C^{0,\alpha}_{\text{loc}}(\mathbb{R}^d)$ with $1/f, 1/g \in L^\infty_{\text{loc}}(\mathbb{R}^d)$; it follows from [8, 12] (see also [1]) that $\varphi \in C^{2,\alpha}_{\text{loc}}(\mathbb{R}^d)$ and solves the Monge–Ampère equation in the classical sense. Regarding global estimates, a groundbreaking result of Caffarelli [4] states that if $f = e^{-V}$ and $g = e^{-W}$ for some $C^{1,1}(\mathbb{R}^d)$ functions V and W with $D^2V \leq \Lambda \text{Id}$ and $D^2W \geq \lambda \text{Id}$ ($\lambda > 0, 0 \leq \Lambda < +\infty$), then T is globally Lipschitz with the explicit bound $\|DT\|_{L^\infty(\mathbb{R}^d)} \leq \sqrt{\Lambda/\lambda}$. To see the importance of Caffarelli's result, it is worth recalling that having a globally Lipschitz transport between f and g enables one to directly transfer various functional inequalities, in particular Poincaré inequalities, from f to g. We refer the interested reader to [5–7, 10, 14] and the references therein for extensions, alternative proofs, and applications.

Since Caffarelli's argument uses in a crucial way the concavity of the logarithm of the determinant, it is really tempting to try to exploit the stronger fact that the determinant to the power $\frac{1}{d}$ is also a concave function on the space of symmetric positive definite matrices, and to see whether this can be useful to go beyond the case of log-concave densities. More precisely, we shall consider densities of the form

$$f = V^{-d}, \quad g = W^{-d}$$

with W typically convex (in which case, we shall say that g is $\frac{1}{d}$-concave) and V semi-concave. Throughout the paper, we shall always assume that V and W are of class $C^{1,1}_{\text{loc}}$, bounded from below away from 0, and of course we impose that V^{-d} and W^{-d} are probability densities

$$\int_{\mathbb{R}^d} \frac{1}{V^d} = \int_{\mathbb{R}^d} \frac{1}{W^d} = 1.$$

Under suitable additional assumptions, we shall indeed obtain a bound on the Lipschitz constant of the monotone transport between such measures, in the spirit of Caffarelli's result. A similar strategy and somehow similar computations were already present in the recent paper [13], which instead focuses on the case of moment measures (i.e., one of the two densities is related to the potential φ itself through a negative power law). However, besides the fact that the moment measure case provides the Monge–Ampère equation with a particular rigid structure, the claimed result of [13] is rather different from ours.

As a first step of independent interest, we establish some linear growth estimates on the monotone transport.

Theorem 1.1. *Assume that, for some $p > 1$, $V^{1/p}$ is Lipschitz and*

$$\liminf_{|y|\to\infty} \frac{W(y)}{|y|^p} > 0, \quad \liminf_{|y|\to\infty} \frac{\nabla W(y) \cdot y}{W(y)} > 1. \tag{1.2}$$

Then the monotone transport T from V^{-d} to W^{-d} satisfies

$$|T(x)| \leq C(1 + |x|), \quad \forall x \in \mathbb{R}^d, \tag{1.3}$$

for some explicit constant $C > 0$.

Using the previous result together with the concavity of $\det^{\frac{1}{d}}$ will enable us to obtain global Lipschitz bounds à la Caffarelli. When V, W, and their derivatives satisfy suitable power-like growth conditions (which imply that the assumptions of Theorem 1.1 are satisfied, as will be explained in Section 3.4), we can bound the Lipschitz constant of the monotone transport. More precisely, let

$$\langle x \rangle := \sqrt{1 + |x|^2}, \quad \forall x \in \mathbb{R}^d.$$

Then we have the following theorem.

Theorem 1.2. *Assume that there exist exponents $p \geq q > 1$ and constants $\lambda > 0$ and $\Lambda \geq 0$ such that*

$$D^2 W \geq \lambda \langle \cdot \rangle^{p-2} \mathrm{Id} \quad and \quad D^2 V \leq \Lambda \langle \cdot \rangle^{q-2} \mathrm{Id}, \quad a.e. \text{ on } \mathbb{R}^d, \tag{1.4}$$

and that

$$\langle \cdot \rangle^{1-p} |\nabla W|, \quad \langle \cdot \rangle^{1-q} |\nabla V|, \quad \frac{\langle \cdot \rangle^q}{V}, \quad and \quad \frac{\langle \cdot \rangle^{q-1}}{1 + |\nabla V|} \quad belong \text{ to } L^\infty(\mathbb{R}^d). \tag{1.5}$$

Then the monotone transport T from V^{-d} to W^{-d} is K-Lipschitz for some explicit constant K.

Note that the previous result covers the case of heavy-tailed densities (which do not necessarily have finite second moments) like

$$f(x) = \frac{a_f}{\langle M_f x \rangle^{\alpha_f d}}, \quad g(y) = \frac{a_g}{\langle M_g y \rangle^{\alpha_g d}}, \quad x \in \mathbb{R}^d, \ y \in \mathbb{R}^d$$

where $\alpha_g \geq \alpha_f > 1$, a_f, a_g are positive normalizing constants, and M_f, M_g nonsingular matrices.

In Section 2, we prove Theorem 1.1 and explore several of its variants. Note that (possibly nonlinear) growth estimates were obtained by Colombo and Fathi [6], using alternative methods based on concentration inequalities. Section 3 is devoted to establishing limits on the Lipschitz constant of the monotone transport map. This discussion includes a range of relevant assumptions that cover Theorem 1.2 as a special case.

2. Pointwise bounds on T

2.1. Proof of Theorem 1.1

For $R \geq 1$, let us replace W by W_R:

$$W_R(y) = \begin{cases} C_R W(y) & \text{if } |y| \leq R, \\ +\infty & \text{otherwise,} \end{cases}$$

where $C_R := (\int_{B_R} W^{-d})^{\frac{1}{d}}$ is the normalizing constant making W_R^{-d} a probability density, so that $C_R \to 1$ as $R \to \infty$. Then the optimal transport map sending $f = V^{-d}$ to W_R^{-d} is globally bounded. Our aim is to prove a linear estimate on such a map, independent of R.

To simplify the notation, we shall still denote by $T = \nabla \varphi$ the optimal transport map sending $f = V^{-d}$ to W_R^{-d}. Also, we remove the subscript R from W_R, which therefore will be denoted by W.

Normalize φ so that $\min \varphi = 1$. Since $\varphi(x) \to \infty$ as $|x| \to \infty$ and $|\nabla \varphi|$ is bounded by R, the positive function $\frac{|\nabla \varphi|^2}{2\varphi}$ tends to 0 as $|x| \to \infty$, hence achieves its maximum[1] at some $\bar{x}$. Setting

$$u := \frac{|\nabla \varphi|^2}{2}, \quad M = \max \frac{u}{\varphi},$$

then, at a maximum point $\bar{x}$, we have

$$\nabla u = M \nabla \varphi \quad \text{and} \quad D^2 u \le M D^2 \varphi. \tag{2.1}$$

Writing the Monge–Ampère equation (1.1) in log form, namely

$$\log \det D^2 \varphi = d \log W \circ \nabla \varphi - d \log V,$$

we differentiate it in the direction e_k (note that it follows from [8, Theorem 1.1] that $\varphi \in C_{\text{loc}}^{3,\alpha}$). Then, with the convention of summation over repeated indices, we get

$$B^{ij} \varphi_{ijk} = d \frac{W_i}{W} \circ \nabla \varphi \varphi_{ik} - d \frac{V_k}{V}, \quad B := (D^2 \varphi)^{-1}. \tag{2.2}$$

Now we multiply (2.2) by φ_k and sum over $k = 1, \ldots, d$. Using

$$\varphi_{ijk} \varphi_k = u_{ij} - \varphi_{ik} \varphi_{jk}, \quad B^{ij} \varphi_{ik} \varphi_{jk} = \text{tr } D^2 \varphi = \Delta \varphi,$$

and noting that, at $\bar{x}$, thanks to (2.1), we have

$$\varphi_{ik} \varphi_k = M \varphi_i, \quad B^{ij} u_{ij} \le M B^{ij} \varphi_{ij} = M \text{ tr Id} = M d,$$

we deduce the following inequality:

$$M d - \Delta \varphi \ge d \left(M \frac{\nabla W}{W} \circ \nabla \varphi \cdot \nabla \varphi - \frac{\nabla V}{V} \cdot \nabla \varphi \right).$$

We now remark that, by the arithmetic-geometric mean inequality and again using the Monge–Ampère equation (1.1), we have

$$\Delta \varphi \ge d (\det D^2 \varphi)^{\frac{1}{d}} = d \frac{W \circ \nabla \varphi}{V}.$$

---

[1] This is the only reason in the proof to approximate W with W_R, so that the density W_R^{-d} is supported inside B_R.

Thus, we arrive at

$$M - \frac{W \circ \nabla \varphi}{V} \geq M \frac{\nabla W}{W} \circ \nabla \varphi \cdot \nabla \varphi - \frac{\nabla V}{V} \cdot \nabla \varphi,$$

i.e.,

$$M \left[\frac{\nabla W(\nabla \varphi)}{W(\nabla \varphi)} \cdot \nabla \varphi - 1 \right] \leq \frac{\nabla V \cdot \nabla \varphi - W(\nabla \varphi)}{V} \leq \frac{|\nabla V|}{V} |\nabla \varphi|. \tag{2.3}$$

Now our assumption (1.2) implies that there exist positive constants R_0, $\delta_0 > 0$, and C_0 such that whenever $|y| \geq R_0$, one has

$$\frac{\nabla W(y) \cdot y}{W(y)} \geq 1 + \delta_0, \quad W(y) \geq C_0 |y|^p. \tag{2.4}$$

Note that, if $|\nabla \varphi(\bar{x})| \leq R_0$, using $\varphi \geq 1$, we have $M \leq \frac{R_0^2}{2}$.

If, on the contrary, $|\nabla \varphi(\bar{x})| \geq R_0$, the first inequality in (2.4) implies that the left-hand side of (2.3) is positive. In particular, the second term in (2.3) must be positive, namely $W(\nabla \varphi(\bar{x})) < \nabla V(\bar{x}) \cdot \nabla \varphi(\bar{x})$. This, combined with the second inequality in (2.4), yields

$$C_0 |\nabla \varphi(\bar{x})|^p \leq W(\nabla \varphi(\bar{x})) \leq |\nabla V(\bar{x})| |\nabla \varphi(\bar{x})|,$$

hence

$$|\nabla \varphi(\bar{x})| \leq \frac{|\nabla V(\bar{x})|^{\frac{1}{p-1}}}{C_0^{\frac{1}{p-1}}}.$$

Combining all these bounds with (2.3), we finally obtain

$$\delta_0 M \leq \frac{|\nabla V(\bar{x})|^{\frac{p}{p-1}}}{C_0^{\frac{1}{p-1}} V(\bar{x})}.$$

Finally, since $V^{1/p}$ is Lipschitz,

$$\frac{|\nabla V(\bar{x})|^{\frac{p}{p-1}}}{V(\bar{x})} \leq (p \operatorname{Lip} V^{1/p})^{\frac{p}{p-1}}.$$

In conclusion, we have found a universal bound on M:

$$M \leq M_0 := \max \left\{ \frac{R_0^2}{2}, \frac{(p \operatorname{Lip} V^{1/p})^{\frac{p}{p-1}}}{\delta_0 C_0^{\frac{1}{p-1}}} \right\}.$$

Now, using

$$|\nabla \sqrt{\varphi}| = \frac{|\nabla \varphi|}{2\sqrt{\varphi}} = \sqrt{\frac{u}{2\varphi}},$$

we deduce that $\sqrt{\varphi}$ is $\sqrt{\frac{M_0}{2}}$ Lipschitz. Hence, denoting by x^* the point where φ is minimal (i.e., with our convention, $\varphi(x^*) = 1$), for all $x \in \mathbb{R}^d$ we have

$$\sqrt{\varphi(x)} \leq 1 + \sqrt{\frac{M_0}{2}}|x - x^*| \leq 1 + \sqrt{\frac{M_0}{2}}(|x| + |x^*|),$$

therefore, since $\frac{u}{\varphi} \leq M_0$,

$$|\nabla\varphi(x)| \leq \sqrt{2M_0} + M_0|x^*| + M_0|x|. \tag{2.5}$$

The proof will therefore be complete once we find a universal bound on x^*, which can be done as in [4, 7] by a simple mass balance argument that we recall for the sake of completeness.

Given $\alpha \in \mathbb{S}^{d-1}$, define the cone $K_\alpha := \{y \in \mathbb{R}^d : \alpha \cdot y \geq \frac{1}{2}|y|\}$. Since $T(x^*) = 0$, the monotonicity of T implies the inequality $T(x) \cdot (x - x^*) \geq 0$ for every x. In particular, if $T(x) \in K_{x^*/|x^*|}$ we have

$$T(x) \cdot x \geq T(x) \cdot x^* \geq \frac{1}{2}|x^*||T(x)|,$$

which implies $|x| \geq \frac{1}{2}|x^*|$. Hence, we obtain $T^{-1}(K_{x^*/|x^*|} \setminus \{0\}) \subset \mathbb{R}^d \setminus B_{\frac{|x^*|}{2}}$. Thanks to $T_\# f = g$, this yields

$$a(g) := \inf_{\alpha \in \mathbb{S}^{d-1}} \int_{K_\alpha} g \leq \int_{\mathbb{R}^d \setminus B_{\frac{|x^*|}{2}}} f.$$

Since $\int_{\mathbb{R}^d \setminus B_r} f \to 0$ as $r \to +\infty$, we get a bound on $|x^*|$ that only depends on f and g:

$$|x^*| \leq 2 \sup\left\{r > 0 : \int_{\mathbb{R}^d \setminus B_r} f \geq a(g)\right\}.$$

Using this bound[2] in (2.5) completes the proof of Theorem 1.1.

Remark 2.1. Another way to bound M from above is to come back to (2.3) and bound its right-hand side in terms of W^*, the Legendre transform of W:

$$M\left[\frac{\nabla W(\nabla\varphi)}{W(\nabla\varphi)} \cdot \nabla\varphi - 1\right] \leq \frac{\nabla V \cdot \nabla\varphi - W(\nabla\varphi)}{V} \leq \frac{W^*(\nabla V)}{V}.$$

Under our assumption, it is possible to prove the existence of a constant $C > 0$ such that $W^*(\nabla V) \leq CV$. Then, combining this bound with the second condition in (1.2), we reach the same conclusion as in Theorem 1.1.

While in this case both arguments give the same result, the method presented in the proof of Theorem 1.1 seems more flexible than the one discussed here. In particular, we shall use a variation of it in the proofs of Theorems 2.2 and 2.4.

---

[2]Note that, since $1/g \in L^\infty_{\text{loc}}$, one can bound $a(g)$ from below, regardless of the radius $R \geq 1$ we used in the beginning to truncate g.

2.2. Variants

We want to stress the fact that a similar proof to that of Theorem 1.1 can lead to sublinear estimates on the optimal transport map. The idea is to consider a maximum point of $\frac{|\nabla\varphi|^2}{2H(\varphi)}$ for some well-chosen positive, increasing, and smooth function H such that $H(t) \to +\infty$ as $t \to +\infty$. Up to replacing W by W_R as we did above (and normalizing φ again by fixing its minimum to 1), one can assume the maximum is achieved at some point $\bar{x}$ and the goal is to estimate

$$M := \max \frac{|\nabla\varphi|^2}{2H(\varphi)} = \frac{|\nabla\varphi(\bar{x})|^2}{2H(\varphi(\bar{x}))}.$$

At $\bar{x}$, one has

$$D^2\varphi\nabla\varphi = MH'(\varphi)\nabla\varphi, \quad \varphi_{ijk}\varphi_k + \varphi_{ik}\varphi_{jk} \le M[H'(\varphi)\varphi_{ij} + H''(\varphi)\varphi_i\varphi_j].$$

Hence, as in the proof of Theorem 1.1, multiplying (2.2) by φ_k, and summing over k gives (recall the relation $B = D^2\varphi^{-1}$)

$$
\begin{aligned}
d\left(MH'(\varphi)\frac{\nabla W(\nabla\varphi)\cdot\nabla\varphi}{W(\nabla\varphi)} - \frac{\nabla V}{V}\cdot\nabla\varphi\right) &= B^{ij}\varphi_{ijk}\varphi_k \\
&\le B^{ij}(M[H'(\varphi)\varphi_{ij} + H''(\varphi)\varphi_i\varphi_j] - \varphi_{ik}\varphi_{jk}) \\
&= dMH'(\varphi) + MH''(\varphi)B\nabla\varphi\cdot\nabla\varphi - \Delta\varphi \\
&\le dMH'(\varphi) + \frac{H''(\varphi)}{H'(\varphi)}|\nabla\varphi|^2 - d\frac{W(\nabla\varphi)}{V},
\end{aligned}
$$

where, in the last line, we have used the arithmetic-geometric mean inequality and that, at $\bar{x}$, one has $M(D^2\varphi)^{-1}\nabla\varphi = \frac{\nabla\varphi}{H'(\varphi)}$. Since $|\nabla\varphi|^2 = 2MH(\varphi)$ at the point $\bar{x}$, we arrive at the inequality

$$M\left[\frac{\nabla W(\nabla\varphi)\cdot\nabla\varphi}{W(\nabla\varphi)} - 1 - \frac{2H''(\varphi)H(\varphi)}{dH'(\varphi)^2}\right] \le \frac{\nabla V\cdot\nabla\varphi - W(\nabla\varphi)}{H'(\varphi)V}. \tag{2.6}$$

An easy consequence of (2.6) is the following.

Theorem 2.2. *Assume that for some $p, q > 1$ such that*

$$d(q-1)(p-1) > q - p \tag{2.7}$$

one has

$$\liminf_{|y|\to\infty}\frac{W(y)}{|y|^p} > 0, \quad \liminf_{|y|\to\infty}\frac{\nabla W(y)\cdot y}{W(y)} \ge p, \tag{2.8}$$

and, for some $A > 0$,

$$\frac{|\nabla V(x)|}{V(x)} \le \frac{A}{1+|x|}, \quad |\nabla V(x)| \le A(1+|x|)^{q-1}, \quad \forall x \in \mathbb{R}^d. \tag{2.9}$$

Then the monotone transport T from V^{-d} to W^{-d} satisfies

$$|T(x)| \le C(1+|x|)^{\frac{q-1}{p-1}} \quad \forall x \in \mathbb{R}^d,$$

for some explicit constant $C > 0$.

Proof. Let

$$\alpha = \frac{q-1}{p-1}, \quad \theta = \frac{2\alpha}{1+\alpha} = \frac{2q-2}{p+q-2} < 2.$$

Our aim is to bound

$$M := \max \frac{|\nabla\varphi|^2}{2\varphi^\theta}.$$

For this, we choose the function $H(\varphi) = \varphi^\theta$ in (2.6) so that, at the maximum point $\bar{x}$, we get

$$M\left[\theta\left(\frac{\nabla W(\nabla\varphi)}{W(\nabla\varphi)} \cdot \nabla\varphi - 1\right) + \frac{2-2\theta}{d}\right] \leq \varphi^{1-\theta}\frac{\nabla V \cdot \nabla\varphi - W(\nabla\varphi)}{V}. \tag{2.10}$$

As in the proof of Theorem 1.1, it suffices to consider the case when $|\nabla\varphi(\bar{x})|$ is large. In particular, by (2.8) and (2.7), the postfactor after M in the left-hand side of (2.10) is strictly positive. This implies the nonnegativity of the right-hand side, which combined with (2.8) and (2.9) gives

$$|\nabla\varphi(\bar{x})| \lesssim |\nabla V(\bar{x})|^{\frac{1}{p-1}} \lesssim (1+|\bar{x}|)^\alpha, \tag{2.11}$$

(here, the symbol $\lesssim$ is just hiding a positive multiplicative constant). Then, using (2.10), (2.9), and the definition of M, we have

$$M \lesssim \frac{1}{1+|\bar{x}|}\frac{|\nabla\varphi(\bar{x})|}{\varphi^{\frac{\theta}{2}}(\bar{x})}\varphi^{1-\frac{\theta}{2}}(\bar{x}) \lesssim \frac{M^{\frac{1}{2}}}{1+|\bar{x}|}\frac{|\nabla\varphi(\bar{x})|^{\frac{2-\theta}{\theta}}}{M^{\frac{1}{\theta}-\frac{1}{2}}}$$

$$\lesssim M^{1-\frac{1}{\theta}}\frac{(1+|\bar{x}|)^{\frac{\alpha(2-\theta)}{\theta}}}{1+|\bar{x}|} \lesssim M^{1-\frac{1}{\theta}},$$

where, in the second line, we used (2.11) and $\frac{2-\theta}{\theta} = \frac{1}{\alpha}$. This gives a universal bound on M, and therefore on the Lipschitz constant of $\varphi^{1-\frac{\theta}{2}}$. So, arguing as at the end of the proof of Theorem 1.1 (using a bound on the norm of the minimum point of φ), we obtain

$$|T| = |\nabla\varphi| \lesssim \varphi^{\frac{\theta}{2}} \lesssim (1+|\cdot|)^{\frac{\theta}{2-\theta}} = (1+|\cdot|)^\alpha,$$

as desired. ∎

Remark 2.3. It is unclear to us whether the condition (2.7) (that is automatically satisfied when $p \geq q$ or when d is large) is really necessary to obtain a superlinear bound on $|T|$ when $p < q$.

Using the same method, a similar sublinear estimate can be obtained when the target is Gaussian.

Theorem 2.4. *Assume that g is a Gaussian and that there exists $A > 0$ such that*

$$\frac{|\nabla V(x)|}{V(x)} \leq \frac{A}{1+|x|} \quad \forall x \in \mathbb{R}^d. \tag{2.12}$$

Then the monotone transport T from V^{-d} to g satisfies

$$|T(x)| \le C \sqrt{1 + \log(1 + |x|)} \quad \forall x \in \mathbb{R}^d,$$

for some explicit constant $C > 0$.

Proof. Performing a translation and a suitable change of coordinates if necessary, we may assume that W is of the form $W(y) = c_d e^{\frac{|y|^2}{2d}}$ (so that $g = W^{-d}$ is a standard Gaussian). This time, we take $H(\varphi) := \log(e + \varphi)$ and consider

$$M := \max \frac{|\nabla\varphi|^2}{2\log(e + \varphi)}.$$

Note that the function H is concave, so that (2.6) implies in this case

$$M\left[\frac{\nabla W(\nabla\varphi) \cdot \nabla\varphi}{W(\nabla\varphi)} - 1\right] \le \frac{\nabla V \cdot \nabla\varphi - W(\nabla\varphi)}{H'(\varphi)V}. \tag{2.13}$$

At a maximum point $\bar{x}$ (again such a point exists provided we replace W by W_R as in the proof of Theorem 1.1), (2.13) becomes

$$M\left[\frac{2M}{d}\log(e + \varphi) - 1\right] = M\left[\frac{1}{d}|\nabla\varphi|^2 - 1\right] \le (e + \varphi)\frac{\nabla V \cdot \nabla\varphi - W(\nabla\varphi)}{V}.$$

Again, we look for an explicit bound on M. If $M \le d$ there is nothing to prove. If $M > d$, the left-hand side is larger than $M\log(e + \varphi(\bar{x}))$ (remember $\varphi \ge 1 > 0$) and therefore, at the point $\bar{x}$, we have

$$M \le \frac{(e + \varphi)}{\log(e + \varphi)}\frac{\nabla V \cdot \nabla\varphi - W(\nabla\varphi)}{V}. \tag{2.14}$$

As in the proof of Theorem 1.1, the positivity of the right-hand side implies

$$c_d e^{\frac{|\nabla\varphi(\bar{x})|^2}{2d}} \le |\nabla V(x)|\,|\nabla\varphi(\bar{x})|.$$

Thus, assuming without loss of generality that $|\nabla\varphi(\bar{x})|$ is large (because if not, we automatically get a bound on M), the left-hand side is larger than $|\nabla\varphi(\bar{x})|e^{\frac{|\nabla\varphi(\bar{x})|^2}{4d}}$ and we get

$$|\nabla\varphi(\bar{x})| \le 2d\sqrt{\log(|\nabla V(\bar{x})|)} \lesssim \sqrt{1 + \log(1 + |\bar{x}|)}$$

where, in the second inequality, we used the fact that (2.12) implies a power growth condition on ∇V.

Denoting by x^* the minimum point of φ and arguing as in the last step of the proof of Theorem 1.1 in order to bound $|x^*|$, we get

$$\varphi(\bar{x}) \le 1 + |\nabla\varphi(\bar{x})|(|\bar{x}| + |x^*|) \le C(1 + |\bar{x}|)\sqrt{1 + \log(1 + |\bar{x}|)}. \tag{2.15}$$

Also, by (2.14), the definition of M, and (2.12), we have

$$M \leq \frac{A(e + \varphi(\bar{x}))}{\log(e + \varphi(\bar{x}))(1 + |\bar{x}|)} |\nabla\varphi(\bar{x})| = \frac{A\sqrt{2M}}{(1 + |\bar{x}|)} \frac{(e + \varphi(\bar{x}))}{\sqrt{\log(e + \varphi(\bar{x}))}}.$$

Hence, combining this bound with (2.15) and the fact that $t \in [1, +\infty) \mapsto \frac{e+t}{\sqrt{\log(e+t)}}$ is increasing, we get

$$\sqrt{M} \leq \frac{A\sqrt{2}}{(1 + |\bar{x}|)} \frac{(e + \varphi(\bar{x}))}{\sqrt{\log(e + \varphi(\bar{x}))}}$$

$$\lesssim \frac{\sqrt{1 + \log(1 + |\bar{x}|)}}{\sqrt{1 + \log((1 + |\bar{x}|)\sqrt{1 + \log(1 + |\bar{x}|)})}} \lesssim 1,$$

which gives the desired bound on M.

Now defining

$$\Phi(t) := \int_1^t \frac{1}{\sqrt{\log(e + s)}} \, ds, \quad t \geq 1,$$

we see that the function $\Phi \circ \varphi$ is $\sqrt{2M}$ Lipschitz. Hence

$$\Phi(\varphi(x)) \leq \sqrt{2M}(|x| + |x^*|) \leq C(1 + |x|), \quad \forall x \in \mathbb{R}^d,$$

which implies

$$|\nabla\varphi(x)| \leq \sqrt{2M} \sqrt{\log(e + \Phi^{-1}(C(1 + |x|)))}. \tag{2.16}$$

Noticing that $\Phi^{-1}(0) = 1$ and $(\Phi^{-1})' = \sqrt{\log(e + \Phi^{-1})} \leq \sqrt{\Phi^{-1}}$, one easily gets the subquadratic bound $\Phi^{-1}(t) \leq (1 + \frac{1}{2}t)^2$, which together with (2.16) gives the desired estimate on $|T|$. ∎

3. Bounding the Lipschitz constant of T

3.1. A formal argument

Assume that the optimal transport $T = \nabla\varphi$ from V^{-d} to W^{-d} has at most linear growth, i.e., it satisfies (1.3) for some constant C (as is the case for instance under the assumptions of Theorem 1.1). Also assume that there are constants $A, B, \lambda > 0$, and $\Lambda \geq 0$ such that for every x and y in $\mathbb{R}^d$,

$$\frac{1 + |\nabla V(x)|^2}{V(x)^2} \leq \frac{A^2}{\langle x \rangle^2}, \quad \frac{W(y)^2}{1 + |\nabla W(y)|^2} \leq B^2 \langle y \rangle^2, \tag{3.1}$$

and, for a.e. $x \in \mathbb{R}^d$, $y \in \mathbb{R}^d$,

$$D^2 V(x) \leq \frac{\Lambda(1 + |\nabla V(x)|^2)\mathrm{Id}}{V(x)}, \quad D^2 W(y) \geq \frac{\lambda(1 + |\nabla W(y)|^2)\mathrm{Id}}{W(y)}. \tag{3.2}$$

Remark 3.1. Since we have

$$D^2\left(-\frac{1}{V}\right) = \frac{D^2 V}{V^2} - 2\frac{\nabla V \otimes \nabla V}{V^3},$$

we obtain

$$\frac{D^2 V}{V^2} \geq D^2\left(-\frac{1}{V}\right) \geq \frac{D^2 V}{V^2} - 2\frac{|\nabla V|^2}{V^3}\mathrm{Id}.$$

Thus, the first condition in (3.2) amounts to a semiconcavity condition on $-\frac{1}{V}$ bounding from above its Hessian by a multiple of $\frac{1+|\nabla V|^2}{V^3}\mathrm{Id}$. This explains the first condition that will appear in the sequel in (3.15), which may look strange at first glance but is simply an incremental ratio version of this semiconcavity assumption.

We now show how, under assumptions (3.1)–(3.2), one can (at least formally) derive a bound on the Lipschitz constant of T, i.e., the largest eigenvalue of $D^2\varphi$ by an argument à la Caffarelli. Then, in the next section, we shall show how to make the argument rigorous.

Let $(\bar{x}, \bar{e})$ maximize the function $\mathbb{R}^d \times \mathbb{S}^{d-1} \ni (x, e) \mapsto \langle D^2\varphi(x)e, e\rangle$, and assume, without loss of generality, $\bar{e} = e_1$. Then

$$\nabla\varphi_{11} = 0, \quad D^2\varphi_{11} \leq 0, \quad \varphi_{i1} = 0 \text{ for } i \neq 1, \tag{3.3}$$

where the last condition follows from the fact that e_1 has to be an eigenvector of $D^2\varphi(\bar{x})$. Writing the Monge–Ampère equation in the form

$$F(D^2\varphi) = \frac{W \circ \nabla\varphi}{V}, \quad \text{where } F = \det^{\frac{1}{d}}, \tag{3.4}$$

and differentiating it first once and then twice with respect to the x_1 variable, we get

$$F'(D^2\varphi)D^2\varphi_1 = \frac{W_i\varphi_{i1}}{V} - \frac{V_1 W}{V^2},$$

and

$$\langle F''(D^2\varphi)D^2\varphi_1, D^2\varphi_1\rangle + F'(D^2\varphi)D^2\varphi_{11}$$
$$= \frac{W_{ik}\varphi_{i1}\varphi_{k1}}{V} + \frac{W_i\varphi_{i11}}{V} - 2\frac{W_i\varphi_{i1}V_1}{V^2} - \frac{V_{11}W}{V^2} + 2\frac{V_1^2 W}{V^3}, \tag{3.5}$$

where W and its derivatives are evaluated at $\nabla\varphi$. Since F is concave on the space of symmetric definite positive matrices, the first term in the left-hand side is nonpositive. Since, at the point $\bar{x}$, $D^2\varphi_{11} \leq 0$ and F is nondecreasing (in the sense of matrices) on the space of symmetric semidefinite positive matrices, the second term on the left-hand side of (3.5) is nonpositive at the point $\bar{x}$. Thanks to (3.3), at the point $\bar{x}$ the right-hand side of (3.5) simplifies to

$$\frac{W_{11}\varphi_{11}^2}{V} - 2\frac{W_1 V_1\varphi_{11}}{V^2} - \frac{V_{11}W}{V^2} + 2\frac{V_1^2 W}{V^3}.$$

Thus, setting $\bar{y} := \nabla\varphi(\bar{x})$, we deduce

$$
\begin{aligned}
W_{11}(\bar{y})\varphi_{11}(\bar{x})^2 &\leq \frac{V_{11}(\bar{x})W(\bar{y})}{V(\bar{x})} + \frac{2V_1(\bar{x})W_1(\bar{y})\varphi_{11}(\bar{x})}{V(\bar{x})} - \frac{2V_1(\bar{x})^2 W(\bar{y})}{V(\bar{x})^2} \\
&\leq \frac{V_{11}(\bar{x})W(\bar{y})}{V(\bar{x})} + \frac{W_{11}(\bar{y})\varphi_{11}(\bar{x})^2}{2} + \frac{2V_1(\bar{x})^2 W_1(\bar{y})^2}{V(\bar{x})^2 W_{11}(\bar{y})},
\end{aligned}
\tag{3.6}
$$

where we have used Young's inequality in the last line and discarded the last (negative) term of the previous one. Therefore, we arrive at

$$
\frac{W_{11}(\bar{y})\varphi_{11}(\bar{x})^2}{2} \leq \frac{V_{11}(\bar{x})W(\bar{y})}{V(\bar{x})} + \frac{2V_1(\bar{x})^2 W_1(\bar{y})^2}{V(\bar{x})^2 W_{11}(\bar{y})}
$$

which, thanks to (3.2), yields

$$
\varphi_{11}(\bar{x})^2 \leq \left(\frac{2\Lambda}{\lambda} + \frac{4}{\lambda^2}\right) \frac{1 + |\nabla V(\bar{x})|^2}{V(\bar{x})^2} \frac{W(\bar{y})^2}{1 + |\nabla W(\bar{y})|^2}.
$$

Thus, (3.1) and (1.3) imply that, denoting by $M = M(A, B, \lambda, \Lambda)$ the explicit constant

$$
M = \sqrt{\frac{2\Lambda}{\lambda} + \frac{4}{\lambda^2}} AB,
$$

one has

$$
\max_{x,e}\langle D^2\varphi(x)e, e\rangle = \varphi_{11}(\bar{x}) \leq M\frac{1 + |\bar{y}|}{1 + |\bar{x}|} = M\frac{1 + |\nabla\varphi(\bar{x})|}{1 + |\bar{x}|} \leq M(1 + C),
$$

where C is the constant in (1.3).

This shows that $T = \nabla\varphi$ is $M(1 + C)$ Lipschitz, as desired. Unfortunately, this argument is only formal not only because it assumes φ to be four times differentiable (which anyhow could be assumed by approximation) but, more importantly, because maximizing sequences for the second derivatives of φ need not be bounded. Hence, to give rigorous statements and proofs, we shall use incremental ratios (similarly to [7]) as explained in the next section.

3.2. Using incremental ratios

Given a function $h: \mathbb{R}^d \to \mathbb{R}$, a small constant $\varepsilon > 0$, and $(x, e) \in \mathbb{R}^d \times \mathbb{S}^{d-1}$, set

$$
h^\varepsilon(x, e) := h(x + \varepsilon e) + h(x - \varepsilon e) - 2h(x).
\tag{3.7}
$$

For $R > 0$ (chosen large in a way that will be specified later), we replace W by W_R as we did in the beginning of the proof of Theorem 1.1. Then, by an abuse of notation, we denote the monotone transport map from V to W_R by $T = \nabla\varphi$. To obtain a Lipschitz bound on T, our goal is to find an upper bound that does not depend on R on the quantity

$$
\limsup_{\varepsilon \to 0^+} \varepsilon^{-2} \sup_{\mathbb{R}^d \times \mathbb{S}^{d-1}} \varphi^\varepsilon.
$$

Indeed, this will enable us to conclude thanks to the stability of the monotone transport map as $R \to \infty$. Thanks to [7, Lemma 3.1], denoting $\hat{x} := \frac{x}{|x|}$ for $x \in \mathbb{R}^d \setminus \{0\}$, one has

$$T(x) - R\hat{x} \to 0 \quad \text{as } |x| \to +\infty. \tag{3.8}$$

This implies in particular that, for any fixed $\varepsilon > 0$, the positive function $\varphi^\varepsilon(x, e) = \int_0^\varepsilon (T(x + te) - T(x - te)) \cdot e \, dt$ tends to 0 as $|x| \to \infty$, uniformly in $e \in \mathbb{S}^{d-1}$. In particular, it achieves its maximum at some point $(\bar{x}, \bar{e})$ (for the moment, we do not explicitly write the dependence of $(\bar{x}, \bar{e})$ in terms of ε). Setting $\bar{y} := \nabla\varphi(\bar{x})$, the fact that $(\bar{x}, \bar{e})$ maximizes φ^ε over $\mathbb{R}^d \times \mathbb{S}^{d-1}$ yields

$$\nabla\varphi(\bar{x} + \varepsilon\bar{e}) + \nabla\varphi(\bar{x} - \varepsilon\bar{e}) = 2\bar{y}, \quad D^2\varphi(\bar{x} + \varepsilon\bar{e}) + D^2\varphi(\bar{x} - \varepsilon\bar{e}) \leq 2D^2\varphi(\bar{x}) \tag{3.9}$$

and that, for some $\delta \in \mathbb{R}$, one has

$$\nabla\varphi(\bar{x} + \varepsilon\bar{e}) - \nabla\varphi(\bar{x} - \varepsilon\bar{e}) = 2\delta\bar{e}. \tag{3.10}$$

Thus, using the notation $\partial_{\bar{e}}\varphi = \nabla\varphi \cdot \bar{e}$, we have

$$\nabla\varphi(\bar{x} \pm \varepsilon\bar{e}) = \bar{y} \pm \delta\bar{e}, \quad \delta = \frac{1}{2}(\partial_{\bar{e}}\varphi(\bar{x} + \varepsilon\bar{e}) - \partial_{\bar{e}}\varphi(\bar{x} - \varepsilon\bar{e})). \tag{3.11}$$

In particular, it follows from (3.11) that we have

$$\sup_{\mathbb{R}^d \times \mathbb{S}^{d-1}} \varphi^\varepsilon = \varphi^\varepsilon(\bar{x}, \bar{e}) = \int_0^\varepsilon (\partial_{\bar{e}}\varphi(\bar{x} + t\bar{e}) - \partial_{\bar{e}}\varphi(\bar{x} - t\bar{e})) \, dt$$

$$\leq \int_0^\varepsilon (\partial_{\bar{e}}\varphi(\bar{x} + \varepsilon\bar{e}) - \partial_{\bar{e}}\varphi(\bar{x} - \varepsilon\bar{e})) \, dt = 2\varepsilon\delta, \tag{3.12}$$

where we have used the convexity of φ to obtain

$$\partial_{\bar{e}}\varphi(\bar{x} + t\bar{e}) \leq \partial_{\bar{e}}\varphi(\bar{x} + \varepsilon\bar{e}), \quad \partial_{\bar{e}}\varphi(\bar{x} - t\bar{e}) \geq \partial_{\bar{e}}\varphi(\bar{x} - \varepsilon\bar{e}), \quad \forall t \in (0, \varepsilon).$$

Hence, to prove $\varphi^\varepsilon(\bar{x}, \bar{e}) \leq K\varepsilon^2$ for some universal constant K, it will be sufficient to show $\delta \leq K\varepsilon$ for some universal constant K.

Denoting as before $F = \det^{\frac{1}{d}}$, the second condition in (3.9) yields

$$F(D^2\varphi(\bar{x})) \geq F\left(\frac{1}{2}D^2\varphi(\bar{x} + \varepsilon\bar{e}) + \frac{1}{2}D^2\varphi(\bar{x} - \varepsilon\bar{e})\right)$$

$$\geq \frac{1}{2}F\left(D^2\varphi(\bar{x} + \varepsilon\bar{e})\right) + \frac{1}{2}F\left(D^2\varphi(\bar{x} - \varepsilon\bar{e})\right),$$

where we used again the concavity of F. Using the Monge–Ampère equation (3.4) at the points $\bar{x}$, $\bar{x} \pm \varepsilon\bar{e}$, and recalling (3.11), we obtain

$$0 \geq \frac{W(\bar{y} + \delta\bar{e})}{V(\bar{x} + \varepsilon\bar{e})} + \frac{W(\bar{y} - \delta\bar{e})}{V(\bar{x} - \varepsilon\bar{e})} - 2\frac{W(\bar{y})}{V(\bar{x})}$$

$$= \frac{W(\bar{y} + \delta\bar{e}) + W(\bar{y} - \delta\bar{e}) - 2W(\bar{y})}{V(\bar{x})}$$

$$+ \left(W(\bar{y} + \delta\bar{e}) - W(\bar{y})\right)\left(\frac{1}{V(\bar{x} + \varepsilon\bar{e})} - \frac{1}{V(\bar{x})}\right)$$

$$+ \Big(W(\bar{y} - \delta\bar{e}) - W(\bar{y}) \Big) \Big(\frac{1}{V(\bar{x} - \varepsilon\bar{e})} - \frac{1}{V(\bar{x})} \Big)$$
$$+ W(\bar{y}) \Big(\frac{1}{V(\bar{x} + \varepsilon\bar{e})} + \frac{1}{V(\bar{x} - \varepsilon\bar{e})} - \frac{2}{V(\bar{x})} \Big).$$

Thus, multiplying by $V(\bar{x})$ and using the notation in (3.7), we obtain a discrete analog of (3.6):

$$W^\delta(\bar{y}, \bar{e}) \leq \Gamma^+ + \Gamma^- + V(\bar{x}) W(\bar{y}) \Big(-\frac{1}{V} \Big)^\varepsilon (\bar{x}, \bar{e}), \tag{3.13}$$

where

$$\Gamma^\pm := (W(\bar{y} \pm \delta\bar{e}) - W(\bar{y})) \frac{V(\bar{x} \pm \varepsilon\bar{e}) - V(\bar{x})}{V(\bar{x} \pm \varepsilon\bar{e})}.$$

The goal now is to find suitable assumptions ensuring that (3.13) implies $\delta \leq K\varepsilon$. This is the purpose of the next section.

3.3. Assumptions ensuring that T is globally Lipschitz

We aim to find assumptions guaranteeing that (3.13) implies a universal bound $\delta \leq K\varepsilon$. Indeed, thanks to (3.12), this will imply a uniform bound on the eigenvalues of $D^2\varphi$.

We shall assume that V and W satisfy assumptions (3.1) and (3.2) that naturally appeared in the formal discussion in Section 3.1. However, to address the case where the maximizer $\bar{x} = \bar{x}(\varepsilon)$ escapes to ∞ as $\varepsilon \to 0$, we shall need extra hypotheses that are, to some extent, asymptotic incremental ratio versions of (3.1) and (3.2) (recall Remark 3.1). Still, as we shall see, the constants from these extra assumptions do not appear in the final Lipschitz bound.

These extra assumptions are as follows: There exist $A_0, B_0, \lambda_0, \Lambda_0, R_0, \alpha_0 > 0$ such that, for every $z \in \mathbb{R}^d \setminus B_{R_0}$, $e \in \mathbb{S}^{d-1}$, and $\alpha \in (0, \alpha_0)$, the following holds:

$$\frac{|V(z + \alpha e) - V(z)|}{\alpha V(z + \alpha e)} \leq \frac{A_0}{1 + |z|}, \quad \frac{|W(z + \alpha e) - W(z)|W(z)}{\alpha(1 + |\nabla W(z)|^2)} \leq B_0(1 + |z|), \tag{3.14}$$

$$\alpha^{-2} \Big(-\frac{1}{V} \Big)^\alpha (z, e) \leq \frac{\Lambda_0(1 + |\nabla V(z)|^2)}{V(z)^3}, \quad \frac{W^\alpha(z, e)}{\alpha^2} \geq \frac{\lambda_0(1 + |\nabla W(z)|^2)}{W(z)}. \tag{3.15}$$

We are now in position to prove the following theorem.

Theorem 3.2. *In addition to the assumptions of Theorem 1.1, let V and W satisfy (3.1)–(3.2)–(3.15)–(3.14) for some constants $A, B, \lambda, \Lambda > 0$, and $A_0, B_0, \lambda_0, \Lambda_0, R_0, \alpha_0$. Then the monotone transport map $T = \nabla\varphi$ from V^{-d} to W^{-d} satisfies*

$$DT(x) = D^2\varphi(x) \leq K\mathrm{Id}, \quad \forall x \in \mathbb{R}^d, \tag{3.16}$$

with $K := \sqrt{\frac{2\Lambda}{\lambda} + \frac{4}{\lambda^2}} AB(1 + C)$, where C is the constant in (1.3) obtained in the proof of Theorem 1.1.

Proof. Recall that we have truncated the target measure g by replacing W with W_R. This slightly affects the value of constants λ and B in the assumptions concerning W, but these values will converge to those for W when in the end one lets $R \to \infty$. Note also that, since W^{-d} is supported inside B_R, the arguments of W and its derivatives will also belong to B_R. In addition, as in the proof of Theorem 1.1, we have a linear bound on $T = T_R$, with a constant which converges to the value C in (1.3) as $R \to \infty$. Hence, to prove (3.16), it is enough to show the same inequality for T_R with these modified constants depending on R, and then to let $R \to +\infty$. For this reason and to simplify the notation, in what follows we omit the dependence on R. On the contrary, we shall from now on explicitly write the dependence of $\bar{x}$, $\bar{e}$, $\bar{y} = \nabla\varphi(\bar{x})$, and δ with respect to ε, denoting them, respectively, $\bar{x}(\varepsilon)$, $\bar{e}(\varepsilon)$, $\bar{y}(\varepsilon)$, and $\delta(\varepsilon)$. We shall need to consider two cases.

• *First case: $\bar{x}(\varepsilon)$ has a limit point as $\varepsilon \to 0^+$.* In this case, up to regularizing V and W (which only slightly changes that value of the constants in the assumptions, and that anyhow will converge to the original values as the regularizing parameter goes to zero), we can assume that V and W are both of class $C^{2,\alpha}_{\text{loc}}$ for some $\alpha > 0$. This guarantees, in particular, that the function φ is at least of class C^4.

Now, up to a suitable extraction, we may assume that $\bar{x}(\varepsilon)$ converges to some $\bar{x}$ and that $\bar{e}(\varepsilon)$ converges to a unit vector, say e_1. Then it is easy to check that

$$\varphi_{11}(\bar{x}) = \max_{(x,e)\in\mathbb{R}^d\times\mathbb{S}^{d-1}} \langle D^2\varphi(x)e, e\rangle.$$

Thus, in this case, the formal argument described in Section 3.1 can be applied verbatim to deduce that

$$\varphi_{11}(\bar{x}) \leq \sqrt{\frac{2\Lambda}{\lambda} + \frac{4}{\lambda^2}} AB(1 + C),$$

proving the desired estimate.

• *Second case: $|\bar{x}(\varepsilon)| \to \infty$ as $\varepsilon \to 0^+$.* Note first that, thanks to (3.8) and (3.10),

$$\delta(\varepsilon) = \frac{1}{2}|T(\bar{x}(\varepsilon) + \varepsilon\bar{e}(\varepsilon)) - T(\bar{x}(\varepsilon) - \varepsilon\bar{e}(\varepsilon))| \to 0 \quad \text{as } \varepsilon \to 0^+.$$

Also, again by (3.8),

$$|\bar{y}(\varepsilon)| = |T(\bar{x}(\varepsilon))| \to R \quad \text{as } \varepsilon \to 0^+.$$

Hence, up to choosing R large enough, we may assume that for ε small enough one can use the asymptotic assumptions (3.14)–(3.15) concerning W (resp. V) at $(z, \alpha, e) = (\bar{y}(\varepsilon), \delta(\varepsilon), \pm\bar{e}(\varepsilon))$ (resp. $(\bar{x}(\varepsilon), \varepsilon, \pm\bar{e}(\varepsilon))$). Thus, for $\varepsilon > 0$ small, it follows from (3.13), (3.14), and (3.15), that

$$\lambda_0\delta(\varepsilon)^2 \leq 2A_0B_0 \frac{1 + |\bar{y}(\varepsilon)|}{1 + |\bar{x}(\varepsilon)|}\varepsilon\delta(\varepsilon) + \Lambda_0 \frac{1 + |\nabla V(\bar{x}(\varepsilon))|^2}{V(\bar{x}(\varepsilon))^2} \frac{W(\bar{y}(\varepsilon))^2}{1 + |\nabla W(\bar{y}(\varepsilon))|^2}\varepsilon^2.$$

Combining this bound with (3.1), we obtain

$$\lambda_0 \delta(\varepsilon)^2 \leq 2 A_0 B_0 \frac{1 + |\bar{y}(\varepsilon)|}{1 + |\bar{x}(\varepsilon)|} \varepsilon \delta(\varepsilon) + \Lambda_0 A^2 B^2 \frac{1 + |\bar{y}(\varepsilon)|^2}{1 + |\bar{x}(\varepsilon)|^2} \varepsilon^2.$$

Since $|\bar{x}(\varepsilon)| \to \infty$ and $|\bar{y}(\varepsilon)| \to R$ as $\varepsilon \to 0^+$, this implies

$$\delta(\varepsilon) = o(\varepsilon).$$

Recalling (3.12), this implies that $D^2 \varphi \equiv 0$, which would prove that the map T is constant, and is a contradiction (since g is not a Dirac mass). Hence, this second case cannot happen, which concludes the proof. ∎

3.4. Proof of Theorem 1.2.

It suffices to check that the assumptions of Theorem 1.2 imply those of Theorem 3.2 (and in particular those of Theorem 1.1).

We start from some preliminary considerations. First, note that, by (1.4), W is convex. Thus, combining (1.4) and the assumption $|\nabla W| \lesssim \langle \cdot \rangle^{p-1}$, it follows that $W \langle \cdot \rangle^{-p}$ is bounded both from above and away from 0 (which we shall simply denote as $W \sim \langle \cdot \rangle^p$).

Using Taylor's formula with an integral remainder, (1.4) yields

$$\langle y \rangle |\nabla W(y)| \geq \nabla W(y) \cdot y = W(y) - W(0) + \int_0^1 s \langle D^2 W(sy) y, y \rangle \, ds$$

$$\geq W(y) - W(0) + \frac{\lambda |y|^p}{p} \qquad (3.17)$$

from which we deduce $(1 + |\nabla W|) \sim \langle \cdot \rangle^{p-1}$.

We list here below the different assumptions that we need to prove, and explain how to obtain them:

- Lipschitz condition on $V^{1/p}$ (an assumption in Theorem 1.1): The fact that $V^{\frac{1}{q}}$ (hence also $V^{\frac{1}{p}}$, since $p \geq q$ and V is bounded away from 0) is Lipschitz directly follows from the second and third conditions in (1.5).

- First condition in (1.2) (an assumption of Theorem 1.1): We have already pointed out that we have $W \sim \langle \cdot \rangle^p$; in particular, W satisfies the first condition in (1.2).

- Second condition in (1.2) (an assumption in Theorem 1.1): This is a consequence of (3.17), dividing it by $W(y)$.

- Upper bound on $\frac{W^2}{1+|\nabla W|^2}$ (from (3.1)): We saw that from (3.17) we obtain $(1 + |\nabla W|) \sim \langle \cdot \rangle^{p-1}$; this, together with $W \sim \langle \cdot \rangle^p$, provides the desired bound.

- Upper bound on $\frac{1+|\nabla V|^2}{V^2}$ (from (3.1)): This is a consequence of the second and third conditions in (1.5).

- Upper bound on $D^2 V$ (first condition in (3.2)): One deduces from (1.5) that $V \sim \langle \cdot \rangle^q$; therefore, thanks to the last condition in (1.5), we have $\langle \cdot \rangle^{q-2} \lesssim \frac{1+|\nabla V|^2}{V}$. The result then follows by assumption (1.4).

- Lower bound on D^2W (second condition in (3.2)): This is a consequence of the assumption (1.4), together with $(1 + |\nabla W|) \sim \langle \cdot \rangle^{p-1}$ and $W \sim \langle \cdot \rangle^p$.

- Upper bound on the second-order incremental ratio of $-1/V$ (first condition in (3.15)): Let $e \in \mathbb{S}^{d-1}$, $\alpha \in [-1, 1]$, and $z \in \mathbb{R}^d$ with $|z|$ large. Since $-D^2(\frac{1}{V}) \leq \frac{D^2V}{V^2}$, using (1.4) and $V \sim \langle \cdot \rangle^q$, we get

$$\left(-\frac{1}{V}\right)^\alpha (z, e) \leq \Lambda \int_0^\alpha \int_{-s}^s \frac{\langle z + \tau e \rangle^{q-2}}{V^2(z + \tau e)} \, d\tau \, ds \lesssim \langle z \rangle^{-q-2}\alpha^2,$$

  which implies the first condition in (3.15) since $\langle \cdot \rangle^{-q-2} \lesssim \frac{1+|\nabla V|^2}{V^3}$.

- Lower bound on the second-order incremental ratio of W (second condition in (3.15)): This is essentially the same argument as the previous one (note, however, that the constants Λ and λ in (3.15) may differ from those in (1.4)).

- Upper bounds on the incremental ratios of V and W (the two conditions in (3.14)): For large $|z|$, these two conditions can be deduced from upper bounds on $|\nabla V|/V$ and $|\nabla W|W/(1 + |\nabla W|^2)$, respectively, and these two quantities are bounded thanks to the assumptions (1.5) (in what concerns V) and to the fact that (3.17) provides $(1 + |\nabla W|) \sim \langle \cdot \rangle^{p-1}$ and $W \sim \langle \cdot \rangle^p$.

References

[1] S. Alesker, S. Dar, and V. Milman, A remarkable measure preserving diffeomorphism between two convex bodies in $\mathbf{R}^n$. *Geom. Dedicata* **74** (1999), no. 2, 201–212 Zbl 0927.52007 MR 1674116

[2] Y. Brenier, Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.* **44** (1991), no. 4, 375–417 Zbl 0738.46011 MR 1100809

[3] L. A. Caffarelli, The regularity of mappings with a convex potential. *J. Amer. Math. Soc.* **5** (1992), no. 1, 99–104 Zbl 0753.35031 MR 1124980

[4] L. A. Caffarelli, Monotonicity properties of optimal transportation and the FKG and related inequalities. *Comm. Math. Phys.* **214** (2000), no. 3, 547–563 Zbl 0978.60107 MR 1800860

[5] S. Chewi and A.-A. Pooladian, An entropic generalization of Caffarelli's contraction theorem via covariance inequalities. *C. R. Math. Acad. Sci. Paris* **361** (2023), 1471–1482 Zbl 1534.49041 MR 4683324

[6] M. Colombo and M. Fathi, Bounds on optimal transport maps onto log-concave measures. *J. Differential Equations* **271** (2021), 1007–1022 Zbl 1456.49036 MR 4154935

[7] M. Colombo, A. Figalli, and Y. Jhaveri, Lipschitz changes of variables between perturbations of log-concave measures. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)* **17** (2017), no. 4, 1491–1519 Zbl 1383.60006 MR 3752535

[8] D. Cordero-Erausquin and A. Figalli, Regularity of monotone transport maps between unbounded domains. *Discrete Contin. Dyn. Syst.* **39** (2019), no. 12, 7101–7112 Zbl 1448.35188 MR 4026183

[9]   G. De Philippis and A. Figalli, The Monge–Ampère equation and its link to optimal trans-
      portation. *Bull. Amer. Math. Soc. (N.S.)* **51** (2014), no. 4, 527–580 Zbl 1515.35005
      MR 3237759

[10] M. Fathi, N. Gozlan, and M. Prod'homme, A proof of the Caffarelli contraction theorem via
      entropic regularization. *Calc. Var. Partial Differential Equations* **59** (2020), no. 3, article no. 96
      Zbl 1458.49038 MR 4098037

[11] A. Figalli, *The Monge–Ampère equation and its applications*. Zur. Lect. Adv. Math., European
      Mathematical Society (EMS), Zürich, 2017 Zbl 1435.35003 MR 3617963

[12] A. Figalli and Y. Jhaveri, Regularity properties of monotone measure-preserving maps. *Adv.
      Nonlinear Stud.* **23** (2023), no. 1, article no. 20220057 Zbl 1518.35424 MR 4568475

[13] H. Khanh, A variant of Caffarelli's contraction theorem for probability distributions of negative
      powers. *Nonlinear Anal.* **239** (2024), article no. 113417 Zbl 1530.49043 MR 4658531

[14] A. V. Kolesnikov, Global Hölder estimates for optimal transportation (in Russian). *Mat.
      Zametki* **88** (2010), no. 5, 708–728 MR 2868394. English translation: *Math. Notes* **88** (2010),
      no. 5-6, 678–695 Zbl 1236.49087

[15] R. J. McCann, Existence and uniqueness of monotone measure-preserving maps. *Duke
      Math. J.* **80** (1995), no. 2, 309–323 Zbl 0873.28009 MR 1369395

Guillaume Carlier
CEREMADE, UMR CNRS 7534, Université Paris Dauphine, PSL, Pl. de Lattre de Tassigny,
75775 Paris; Mokaplan, INRIA, 75013 Paris, France; carlier@ceremade.dauphine.fr

Alessio Figalli
Department of Mathematics, ETH Zürich, Rämistrasse 101, 8092 Zürich, Switzerland;
alessio.figalli@math.ethz.ch

Filippo Santambrogio
Universite Claude Bernard Lyon 1, ICJ UMR5208, CNRS, Ecole Centrale de Lyon, INSA Lyon,
Université Jean Monnet, 69622 Villeurbanne, France; santambrogio@math.univ-lyon1.fr