SHARP STABILITY FOR SOBOLEV AND LOG-SOBOLEV INEQUALITIES, WITH OPTIMAL DIMENSIONAL DEPENDENCE

JEAN DOLBEAULT, MARIA J. ESTEBAN, ALESSIO FIGALLI, RUPERT L. FRANK, AND MICHAEL LOSS

Abstract. We prove a quantitative version of the Sobolev inequality with explicit constants. Moreover, the constants have the correct behavior in the limit of large dimensions, which allows us to deduce an optimal quantitative stability estimate for the Gaussian log-Sobolev inequality with an explicit dimension-free constant. Our proofs rely on several ingredients such as, competing symmetries, a flow based on continuous Steiner symmetrization that interpolates continuously between a function and its symmetric decreasing rearrangement and refined estimates on the Sobolev-functional in the neighborhood of the optimal Aubin-Talenti functions.

Contents

(J. Dolbeault, M. J. Esteban) CEREMADE (CNRS UMR No. 7534), PSL University, Université Paris-Dauphine, Place de Lattre de Tassigny, 75775 Paris 16, France.
E-mails: dolbeaul@ceremade.dauphine.fr, esteban@ceremade.dauphine.fr

(A. Figalli) Mathematics Department, ETH Zürich, Ramistrasse 101, 8092 Zürich, Switzerland. *E-mail:* alessio.figalli@math.ethz.ch

(R. L. Frank) Department of Mathematics, LMU Munich, Theresienstr. 39, 80333 München, Germany, and Munich Center for Quantum Science and Technology, Schellingstr. 4, 80799 München, Germany, and Mathematics 253-37, Caltech, Pasadena, CA 91125, USA. *E-mail:* r.frank@lmu.de

(M. Loss) School of Mathematics, Georgia Institute of Technology Atlanta, GA 30332, United States of America. *E-mail:* loss@math.gatech.edu

Date: December 28, 2022.

2020 *Mathematics Subject Classification.* Primary: 49J40; Secondary: 26D10, 35A23.

Key words and phrases. Sobolev inequality, logarithmic Sobolev inequality, stability, rearrangement, Steiner symmetrization.

Acknowledgements. Partial support through US National Science Foundation grants DMS-1954995 (R.L.F.) and DMS-2154340 (M.L.), as well as through the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Germany's Excellence Strategy EXC-2111-390814868 (R.L.F.), the French National Research Agency (ANR) projects EFI (ANR-17-CE40-0030, J.D.) and molQED (ANR-17-CE29-0004, M.J.E.) and the European Research Council under the Grant Agreement No. 721675 (RSPDE) *Regularity and Stability in Partial Differential Equations* (A.F.).

1. Introduction and main results

On $\mathbb{R}^d$ with $d \geq 3$, let us consider the Sobolev inequality

$$\|\nabla f\|^2_{\mathrm{L}^2(\mathbb{R}^d)} \geq S_d \, \|f\|^2_{\mathrm{L}^{2^*}(\mathbb{R}^d)} \quad \forall\, f \in \dot{\mathrm{H}}^1(\mathbb{R}^d)$$

where $2^* = 2\,d/(d-2)$ is the 'Sobolev exponent', $S_d = \frac{1}{4}\,d\,(d-2)\,|\mathbb{S}^d|^{2/d}$ is the sharp Sobolev constant, and $\mathbb{S}^d$ denotes the d-dimensional unit ball. In [12] Brezis and Lieb posed the question whether it is possible to bound the 'Sobolev deficit'

$$\|\nabla f\|^2_{\mathrm{L}^2(\mathbb{R}^d)} - S_d \, \|f\|^2_{\mathrm{L}^{2^*}(\mathbb{R}^d)}$$

on $\dot{\mathrm{H}}^1(\mathbb{R}^d)$ from below in terms of some natural distance from the set of optimizers. The homogeneous Sobolev space $\dot{\mathrm{H}}^1(\mathbb{R}^d)$ is the space of the functions $f \in \mathrm{L}^1_{\mathrm{loc}}(\mathbb{R}^d)$, whose distributional gradient is a square-summable function that vanish at infinity, in the sense that $|\{x \in \mathbb{R}^d : |f(x)| > \epsilon\}| < \infty$ for all $\epsilon > 0$. Here $|A|$ denotes the Lebesgue measure of a measurable set A. Throughout this paper we deal with real-valued functions. With minor additional effort our arguments can be extended to the case of complex-valued functions.

Rodemich [58], Aubin [4] and Talenti [63] (see also [60]) proved that the Sobolev deficit is nonnegative. Moreover, it was shown by Lieb [53], Gidas, Ni and Nirenberg [45] and Caffarelli, Gidas and Spruck [18] that the deficit vanishes if and only if the function f is of the form

$$f(x) = c\left(a + |x-b|^2\right)^{-\frac{d-2}{2}}, \tag{1}$$

where $a \in (0,\infty)$, $b \in \mathbb{R}^d$, and $c \in \mathbb{R}$ are constants. These functions are often called 'Aubin–Talenti functions'. Let $\mathcal{M}$ denote the $(d+2)$-dimensional manifold of functions of the form (1).

The question of Brezis and Lieb was answered by Bianchi and Egnell [6]: for any $d \geq 3$ there is a strictly positive constant c_{BE} such that, for any $f \in \dot{\mathrm{H}}^1(\mathbb{R}^d) \setminus \mathcal{M}$,

$$\mathcal{E}(f) := \frac{\|\nabla f\|^2_{\mathrm{L}^2(\mathbb{R}^d)} - S_d\,\|f\|^2_{\mathrm{L}^{2^*}(\mathbb{R}^d)}}{\inf_{g\in\mathcal{M}} \|\nabla f - \nabla g\|^2_{\mathrm{L}^2(\mathbb{R}^d)}} \geq c_{\mathrm{BE}}\,. \tag{2}$$

We denote by c_{BE} the optimal, that is, largest possible constant in (2).

Lions [55] has shown that if the Sobolev deficit is small for some function f, then f has to be close to the set $\mathcal{M}$ of Sobolev optimizers. The closeness is measured in the strongest possible sense, namely with respect to the norm in $\dot{\mathrm{H}}^1(\mathbb{R}^d)$. The Bianchi–Egnell inequality (2) makes the qualitative result of Lions more quantitative. In particular, it shows that the distance to the manifold vanishes at least like the square root of the Sobolev deficit. Such 'stability' estimates have been established in other contexts as well, e.g., for the isoperimetric inequality or for classical inequalities in real and harmonic analysis. In fact, stability has attracted a lot of attention in recent years and we refer to [44, 26, 39, 27, 23, 30, 21, 24, 38, 25, 61, 42, 43, 40, 10] and the references within for a list of works in this direction. In several of them the strategy of Bianchi and Egnell or its generalizations play an important role.

An interesting point about (2) and other inequalities obtained by this method is that nothing seems to be known about the optimal value of the constant c_{BE} except for the fact that it is strictly positive. The proof in [6] proceeds by a spectral estimate combined with a compactness argument and hence cannot give any information about c_{BE}. Explicit quantitative estimates are known only for a distance to $\mathcal{M}$ measured by a weaker norm than (2), functions of $\dot{\mathrm{H}}^1(\mathbb{R}^d)$ satisfying additional constraints or superquadratic estimates of the distance which degenerate in a neighbourhood of $\mathcal{M}$, while much more is known for subcritical interpolation inequalities than for Sobolev-type inequalities: see [9, 3, 31, 29, 10, 41, 13] for some references.

Stability results for the Sobolev inequality. It is the aim of this article to address the question of proving (2) with an explicit lower bound on $c_{\rm BE}$. Not only will we obtain a computable expression for the constant, we will also determine its asymptotic behavior as $d \to +\infty$. To explain the latter, we recall that

$$c_{\rm BE} \le \frac{4}{d+4} \,. \tag{3}$$

This follows from the proof of [6, Lemma 1]; see also [23, Introduction]. The constant on the right side comes from the spectral gap inequality mentioned before. Thus, $c_{\rm BE}$ decays at least like d^{-1} as $d \to +\infty$. Our main result shows that it does not decay faster than d^{-1}. More precisely, we prove the following theorem.

Theorem 1. *There is a constant $\beta > 0$ with an explicit lower estimate such that for all $d \ge 3$ and all $f \in \mathrm{H}^1(\mathbb{R}^d) \setminus \mathcal{M}$ we have*

$$\mathcal{E}(f) \ge \frac{\beta}{d} \,.$$

We refer to Identity (44) and Remark 32 for an explicit expression of β and some comments on a simpler bound deduced from Proposition 7.

Stability results for the logarithmic Sobolev inequality. Consistently with the optimal d^{-1} behavior of the constant in Theorem 1 we obtain a quantitative version of the stability for the sharp logarithmic Sobolev inequality, which we state next. On $\mathbb{R}^N$, $N \ge 1$, we consider the Gaussian measure

$$d\gamma(x) = e^{-\pi\,|x|^2}\,dx \,.$$

We abbreviate $\mathrm{L}^2(\gamma) = \mathrm{L}^2(\mathbb{R}^N, d\gamma)$ and denote by $\mathrm{H}^1(\gamma)$ the space of all $u \in \mathrm{L}^2(\gamma)$ with distributional gradient in $\mathrm{L}^2(\gamma)$. The logarithmic Sobolev inequality states that for all functions $u \in \mathrm{H}^1(\gamma)$ one has

$$\int_{\mathbb{R}^N} |\nabla u|^2\,d\gamma \ge \pi \int_{\mathbb{R}^N} |u|^2 \ln\left(\frac{|u|^2}{\|u\|^2_{\mathrm{L}^2(\gamma)}}\right) d\gamma \,.$$

The constant π is optimal and equality holds if and only if

$$u(x) = c\,e^{a\cdot x} \tag{4}$$

for some $a \in \mathbb{R}^N$ and $c \in \mathbb{R}$. The corresponding stability result for the logarithmic Sobolev inequality goes as follows.

Theorem 2. *There is an explicit constant $\kappa > 0$ such that for all $N \in \mathbb{N}$ and all $u \in \mathrm{H}^1(\gamma)$,*

$$\int_{\mathbb{R}^N} |\nabla u|^2\,d\gamma - \pi \int_{\mathbb{R}^N} u^2 \ln\left(\frac{|u|^2}{\|u\|^2_{\mathrm{L}^2(\gamma)}}\right) d\gamma \ge \kappa \inf_{a\in\mathbb{R}^N,\,c\in\mathbb{R}} \int_{\mathbb{R}^N} (u - c\,e^{a\cdot x})^2\,d\gamma \,.$$

The *logarithmic Sobolev inequality* on a finite dimensional Euclidean space (with either Gaussian or Lebesgue measures) can be seen as a *large dimensional limit* of the Sobolev inequality, for instance by considering Sobolev's inequality on a sphere of radius $\sqrt{d}$ applied to a function depending only on N real variables as in [5, p. 4818]. The classical versions of the logarithmic Sobolev inequality are usually attributed to Stam [62], Federbush [36], Gross [46], and also Weissler [66] for a scale-invariant form. There is a huge literature on logarithmic

Sobolev inequalities and we refer to [47] for a survey on many early results. Equality cases in the logarithmic Sobolev inequality have been characterized by Carlen in [20, Theorem 5], even with a remainder term, see [20, Theorem 6]. Other remainder terms are given in [8, 35, 32] and, using weaker notions of distances, in [8, 49, 35, 37, 48] while some obstructions to stability results involving strong notions of distance are given in [50, 34]. However, as far as we know, the Bianchi–Egnell strategy has so far not been applied to the logarithmic Sobolev inequality, probably because $u \mapsto |u|^2 \ln |u|^2$ is not twice differentiable at the origin. Here we overcome this issue. In fact, as will be clear from our proof, this non-twice-differentiability issue is closely related to that of obtaining a d^{-1} decay in the setting of Theorem 1.

We have stated the logarithmic Sobolev inequality in its version with respect to the normalized Gaussian measure. It has an equivalent version with respect to the Euclidean measure. We set $u = e^{\pi\,|x|^2/2} v$ and obtain by a simple integration by parts from Theorem 2

$$\int_{\mathbb{R}^N} |\nabla v|^2\,dx - \pi \int_{\mathbb{R}^N} v^2 \ln\left(\frac{v^2}{\|v\|^2_{\mathrm{L}^2(\mathbb{R}^N)}}\right) dx - N\,\pi\,\|v\|^2_{\mathrm{L}^2(\mathbb{R}^N)} \geq \kappa \inf_{b\in\mathbb{R}^N,\,c\in\mathbb{R}} \int_{\mathbb{R}^N} \left|v - c\,e^{-\frac{\pi}{2}|x-b|^2}\right|^2 dx\,.$$

Writing $v(x) = \lambda^{N/2}\,w(\lambda\,x)$ with a parameter $\lambda > 0$, we obtain equivalently

$$\lambda^2 \int_{\mathbb{R}^N} |\nabla w|^2\,dy - \pi \int_{\mathbb{R}^N} w^2 \ln\left(\frac{w^2}{\|w\|^2_{\mathrm{L}^2(\mathbb{R}^N)}}\right) dy - N\,\pi\,(1+\ln\lambda)\|w\|^2_{\mathrm{L}^2(\mathbb{R}^N)}$$
$$\geq \kappa \inf_{b\in\mathbb{R}^N,\,c\in\mathbb{R}} \int_{\mathbb{R}^N} \left|w - c\,e^{-\frac{\pi}{2\lambda^2}\,|y-b|^2}\right|^2 dy\,.$$

We bound the right side from below by extending the infimum over all $\lambda > 0$ and then we optimize the left side with respect to $\lambda > 0$. In this way we obtain the following stability version of the Euclidean logarithmic Sobolev inequality.

Corollary 3. *With $\kappa > 0$ as in Theorem 2 we have for all $N \in \mathbb{N}$ and all $w \in \mathrm{H}^1(\mathbb{R}^N)$,*

$$\|w\|^2_{\mathrm{L}^2(\mathbb{R}^N)} \ln\left(\frac{2}{N\,\pi\,e}\,\frac{\int_{\mathbb{R}^N} |\nabla w|^2\,dx}{\|w\|^2_{\mathrm{L}^2(\mathbb{R}^N)}}\right) - \frac{2}{N} \int_{\mathbb{R}^N} w^2 \ln\left(\frac{w^2}{\|w\|^2_{\mathrm{L}^2(\mathbb{R}^N)}}\right) dx$$
$$\geq \frac{2\,\kappa}{N\,\pi} \inf_{\lambda>0,\,b\in\mathbb{R}^N,\,c\in\mathbb{R}} \int_{\mathbb{R}^N} \left|w - c\,e^{-\frac{\pi}{2\lambda^2}\,|y-b|^2}\right|^2 dy\,.$$

Ideas of the proof of Theorem 1. Let us describe the strategy of the proof of Theorem 1. It consists of three independent parts, corresponding to Sections 3, 4 and 5, respectively. The first and second parts concern nonnegative functions, while in the third part we deduce the inequality for arbitrary functions from that for nonnegative functions. The latter argument uses a certain concavity inherent in the problem. Potentially this argument comes with a loss in the constant, but we show that it does not destroy the d^{-1} behavior that we need to prove Theorem 2.

We now discuss the first and the second parts in more detail. Superficially, the proof is analogous to that by Bianchi and Egnell [6], namely, one splits the problem into two regions, one where f is close to the set of Sobolev optimizers and the other where it is far away. These regions are defined in terms of the quantity $\inf_{g\in\mathcal{M}} \|\nabla f - \nabla g\|^2_{\mathrm{L}^2(\mathbb{R}^d)}/\|\nabla f\|^2_{\mathrm{L}^2(\mathbb{R}^d)}$, specifically by requiring that this quantity is either less or equal than δ, or bigger than δ. Here $\delta > 0$ is a free parameter that will be chosen appropriately at the end. Note that, since $\inf_{g\in\mathcal{M}} \|\nabla f - \nabla g\|^2_{\mathrm{L}^2(\mathbb{R}^d)} \leq \|\nabla f\|^2_{\mathrm{L}^2(\mathbb{R}^d)}$, we may always assume that $\delta \leq 1$ and even $\delta < 1$.

The first part of the proof of Theorem 1 in Section 3 is concerned with a nonnegative function f that is close to the set of optimizers. The basic strategy is to expand the quantity $\|f\|^2_{\mathrm{L}^q(\mathbb{R}^d)}$ with the main term given by the quantity with f replaced by the closest optimizer g. By this choice there will be no linear term in this expansion, and for the quadratic term one uses a spectral gap inequality. A first version of this argument appears in the proof of Proposition 7. Such a naive expansion, however, is not good enough to reproduce the correct d^{-1} behavior of the constant c_{BE}. Instead, a refined argument is needed where we cut the function f/g in its range and treat the different parts by different arguments. The spectral gap inequality is only used for an L^∞-bounded part of the perturbation.

Parenthetically we point out that we actually prove something stronger. Namely, we assume a decomposition $f = g + r$ with $g \in \mathcal{M}$ and a perturbation r satisfying certain orthogonality conditions. These orthogonality conditions for r are guaranteed when g realizes the infimum $\inf_{g' \in \mathcal{M}} \|\nabla f - \nabla g'\|^2_{\mathrm{L}^2(\mathbb{R}^d)}$, but our argument does not make use of this minimality of g. This observation turns out to be convenient when deducing the stability version of the logarithmic Sobolev inequality in Theorem 2.

In the second part of the proof of Theorem 1, described in Section 4, we obtain a lower bound on $\mathcal{E}(f)$ for nonnegative functions f satisfying $\inf_{g\in\mathcal{M}} \|\nabla f - \nabla g\|^2_{\mathrm{L}^2(\mathbb{R}^d)} > \delta\, \|\nabla f\|^2_{\mathrm{L}^2(\mathbb{R}^d)}$. Bianchi and Egnell [6] handle this part by a compactness argument and this is the reason why so far there does not exist a quantitative lower bound on c_{BE}. One can replace this argument by a constructive procedure using an idea taken from a paper by Michael Christ [25], in which he establishes a quantitative error term for the Riesz rearrangement inequality. To implement this idea in our context we construct, using competing symmetries [22] and continuous rearrangement [14], a family of functions $f_\tau, 0 \le \tau < \infty$, such that $f_0 = f$, $\|f_\tau\|_{2^*} = \|f\|_{2^*}$, $\tau \mapsto \|\nabla f_\tau\|_2$ is non-increasing and $\inf_{b\in\mathcal{B}} \|\nabla(f_\tau - b)\|_2^2 \to 0$ as $\tau \to \infty$. Clearly,

$$\mathcal{E}(f) \ge \frac{\|\nabla f\|^2_{\mathrm{L}^2(\mathbb{R}^d)} - S_d\, \|f\|^2_{\mathrm{L}^{2^*}(\mathbb{R}^d)}}{\|\nabla f\|^2_{\mathrm{L}^2(\mathbb{R}^d)}} = 1 - S_d\, \frac{\|f\|^2_{\mathrm{L}^{2^*}(\mathbb{R}^d)}}{\|\nabla f\|^2_{\mathrm{L}^2(\mathbb{R}^d)}} \ge \frac{\|\nabla f_\tau\|^2_{\mathrm{L}^2(\mathbb{R}^d)} - S_d\, \|f_\tau\|^2_{\mathrm{L}^{2^*}(\mathbb{R}^d)}}{\|\nabla f_\tau\|^2_{\mathrm{L}^2(\mathbb{R}^d)}}\,.$$

Starting with $\inf_{g\in\mathcal{M}} \|\nabla f - \nabla g\|^2_{\mathrm{L}^2(\mathbb{R}^d)} > \delta\, \|\nabla f\|^2_{\mathrm{L}^2(\mathbb{R}^d)}$, one would like to run the flow until at a certain point τ_0 one has

$$\inf_{g\in\mathcal{M}} \|\nabla(f_{\tau_0} - g)\|^2_{\mathrm{L}^2(\mathbb{R}^d)} = \delta\, \|\nabla f_{\tau_0}\|^2_{\mathrm{L}^2(\mathbb{R}^d)} \tag{5}$$

and one would conclude that

$$\mathcal{E}(f) \ge \frac{\|\nabla f_{\tau_0}\|^2_{\mathrm{L}^2(\mathbb{R}^d)} - S_d\, \|f_{\tau_0}\|^2_{\mathrm{L}^{2^*}(\mathbb{R}^d)}}{\|\nabla f_{\tau_0}\|^2_{\mathrm{L}^2(\mathbb{R}^d)}} = \delta\, \frac{\|\nabla f_{\tau_0}\|^2_{\mathrm{L}^2(\mathbb{R}^d)} - S_d\, \|f_{\tau_0}\|^2_{\mathrm{L}^{2^*}(\mathbb{R}^d)}}{\inf_{g\in\mathcal{M}} \|\nabla(f_{\tau_0} - g)\|^2_{\mathrm{L}^2(\mathbb{R}^d)}}\,.$$

We could then apply the first part of the proof to the function f_{τ_0} and obtain the desired bound. The details of this argument are more involved than presented here, mostly because the function $\tau \mapsto \|\nabla f_\tau\|_{\mathrm{L}^2(\mathbb{R}^d)}$ need not be continuous, so the existence of a τ_0 as in (5) is not guaranteed.

Continuous rearrangement flows in the setting of Steiner symmetrizations have been used by Pólya–Szegő [56, Note B], Brock [14, 15] and others. In the setting of symmetric decreasing rearrangements of sets it was used by Bucur–Henrot [16] and we will generalize this to functions. Additional results on this flow, which might be useful in other contexts as well, are given in Appendix 7 at the end of the paper.

On the proof of Theorem 2. The proof of Theorem 2 is given in Section 7. The underlying idea is that the logarithmic Sobolev inequality on $\mathbb{R}^N$ can be obtained by taking an appropriate limit in the Sobolev inequalities in dimension d, in the limiting regime as $d \to +\infty$, and that the same property should also be true for the stability inequality, except that for scaling reasons, the $\dot{\mathrm{H}}^1(\mathbb{R}^d)$ distance gives rise only to a stability estimate in $\mathrm{L}^2(\mathbb{R}^N)$ for the logarithmic Sobolev inequality. At a formal level, this suggests that the constant κ in Theorem 2 can be explicitly estimated in terms of the constant β in Theorem 1, but to justify such a limit further estimates on optimal Aubin-Talenti functions are needed. Instead of following this path, we bypass these difficulties by reducing the problem to orthogonality conditions in a neighbourhood of $\mathcal{M}$, see Theorem 33.

Additional observations. Before ending this introduction, let us mention some further progress on the optimal constant c_{BE} that has been made since the first version of this paper appeared on the arXiv. In that first version we had asked whether the upper bound (3) on c_{BE} is strict and whether there is a function f that minimizes $\mathcal{E}(f)$. Both questions have been answered in an original way by T. König. In [51] he shows that the upper bound in (3) is strict and in [52] that the infimum defining c_{BE} is attained. This is reminiscent of the planar isoperimetric inequality, where the constant in the quantitative isoperimetric inequality with Frankel asymmetry is strictly smaller than the constant in the corresponding spectral gap inequality and where one can prove the existence of an optimizing domain; see [7]. For further studies under an additional convexity assumption, see [19, 2, 28].

In order to make notations lighter, we will write $\|\cdot\|_q = \|\cdot\|_{\mathrm{L}^q(\mathbb{R}^d)}$ whenever the space is $\mathbb{R}^d$ with Lebesgue measure.

2. The Sobolev inequality on the sphere

It is well known that the Sobolev inequality on $\mathbb{R}^d$ has an equivalent formulation on $\mathbb{S}^d$, the unit sphere in $\mathbb{R}^{d+1}$. It will be convenient for us at several steps of our proof to carry out the arguments in the setting of $\mathbb{S}^d$. In this brief preliminary section, let us give some details.

We denote by $\omega = (\omega_1, \omega_2, \ldots, \omega_{d+1})$ the coordinates in $\mathbb{R}^{d+1}$. Then the unit sphere $\mathbb{S}^d \subset \mathbb{R}^{d+1}$ can be parametrized in terms of stereographic coordinates by

$$\omega_j = \frac{2\,x_j}{1+|x|^2}\,, \quad j = 1, \ldots, d\,, \quad \omega_{d+1} = \frac{1-|x|^2}{1+|x|^2}\,.$$

To a function f on $\mathbb{R}^d$ we associate a function F on $\mathbb{S}^d$ via

$$F(\omega) = \left(\frac{1+|x|^2}{2}\right)^{\frac{d-2}{2}} f(x) \quad \forall\, x \in \mathbb{R}^d\,. \tag{6}$$

Then, since $(2/(1+|x|^2))^d$ is the Jacobian of the inverse stereographic projection $x \mapsto \omega$,

$$|\mathbb{S}^d| \int_{\mathbb{S}^d} |F(\omega)|^{2^*}\, d\mu(\omega) = \int_{\mathbb{R}^d} |f(x)|^{2^*}\, dx\,,$$

where μ denotes the uniform probability measure on $\mathbb{S}^d$. Moreover, $F \in \mathrm{H}^1(\mathbb{S}^d)$ if and only if $f \in \dot{\mathrm{H}}^1(\mathbb{R}^d)$, and in this case

$$|\mathbb{S}^d| \int_{\mathbb{S}^d} \left(|\nabla F|^2 + \tfrac14\, d\,(d-2)\,|F|^2\right) d\mu(\omega) = \int_{\mathbb{R}^d} |\nabla f|^2\, dx\,.$$

Therefore the sharp Sobolev inequality on $\mathbb{R}^d$ is equivalent to the following sharp Sobolev inequality on $\mathbb{S}^d$,

$$\int_{\mathbb{S}^d} \left(|\nabla F|^2 + \tfrac14\, d\,(d-2)\,|F|^2\right) d\mu \geq \tfrac14\, d\,(d-2) \left(\int_{\mathbb{S}^d} |F|^{2^*}\, d\mu\right)^{2/2^*} \quad \forall\, F \in \mathrm{H}^1(\mathbb{S}^d, d\mu)\,,$$

with equality exactly for the functions

$$G(\omega) = c\,\left(a + b\cdot\omega\right)^{-\frac{d-2}{2}}\,,$$

and $a > 0$, $b \in \mathbb{R}^d$ and $c \in \mathbb{R}$ are constants. We denote the corresponding set of functions by $\mathscr{M}$. Then the above equivalence shows that

$$\mathcal{E}(f) = \frac{\|\nabla f\|_2^2 - S_d\,\|f\|_{2^*}^2}{\inf_{g\in\mathscr{M}} \|\nabla f - \nabla g\|_2^2} = \frac{\|\nabla F\|^2_{\mathrm{L}^2(\mathbb{S}^d)} + \frac14\, d\,(d-2)\,\|F\|^2_{\mathrm{L}^2(\mathbb{S}^d)} - S_d\,\|F\|^2_{\mathrm{L}^{2^*}(\mathbb{S}^d)}}{\inf_{G\in\mathscr{M}} \left\{\|\nabla F - \nabla G\|^2_{\mathrm{L}^2(\mathbb{S}^d)} + \frac14\, d\,(d-2)\,\|F-G\|^2_{\mathrm{L}^2(\mathbb{S}^d)}\right\}}\,.$$

3. Functions close to the manifold of optimizers

Our goal in this section is to prove a quantitative stability inequality for nonnegative functions close to the manifold of optimizers. It is convenient to prove this result in the equivalent setting of the sphere. We recall that μ denotes the uniform probability measure on $\mathbb{S}^d$. In order to simplify the notation, we write in this section

$$q = 2^* = \frac{2\,d}{d-2}\,, \quad \theta = q - 2 = \frac{4}{d-2}\,, \quad A = \frac14\, d\,(d-2)\,.$$

Theorem 4. *There are explicit constants $\epsilon_0 > 0$ and $\tilde\delta \in (0,1)$ such that for all $d \geq 3$ and for all $-1 \leq r \in \mathrm{H}^1(\mathbb{S}^d)$ satisfying*

$$\left(\int_{\mathbb{S}^d} |r|^q\, d\mu\right)^{2/q} \leq \tilde\delta \tag{7}$$

and

$$\int_{\mathbb{S}^d} r\, d\mu = 0 = \int_{\mathbb{S}^d} \omega_j\, r\, d\mu\,, \quad j = 1, \ldots, d+1\,, \tag{8}$$

one has

$$\int_{\mathbb{S}^d} \left(|\nabla r|^2 + A\,(1+r)^2\right) d\mu - A \left(\int_{\mathbb{S}^d} (1+r)^q\, d\mu\right)^{2/q} \geq \theta\,\epsilon_0 \int_{\mathbb{S}^d} \left(|\nabla r|^2 + A\, r^2\right) d\mu\,.$$

The key feature of this theorem is that the constant $\theta\,\epsilon_0$ behaves like $4\,\epsilon_0\, d^{-1}$ for large d. This d^{-1} behavior leads to a corresponding lower bound on the behavior of c_{BE}, which in view of (3) is optimal.

Remark 5. *In fact, we show that for every $0 < \epsilon_0 < \frac13$ there is a $\tilde\delta > 0$ such that the assertion in the theorem holds for all $d \geq 6$. The same argument also gives that for every $0 < \epsilon_0 < \frac12$ there is a D and a $\tilde\delta > 0$ such that the assertion of the theorem holds for all $d \geq D$. The explicit expression for $\tilde\delta > 0$ can be found in the proofs of Theorem 4, Proposition 21 and in (21).*

The proof of Theorem 4 will take up the rest of this section.

3.1. **The spectral gap inequality.** Of crucial importance in our analysis, just like in that of Bianchi and Egnell [6], is the following spectral bound. It appears, for instance, in Rey's paper [57, Appendix D] slightly before the work of Bianchi and Egnell.

Lemma 6. *Let $d \geq 3$ and assume that $r \in \mathrm{H}^1(\mathbb{S}^d)$ satisfies (8). Then*

$$\int_{\mathbb{S}^d} \left(|\nabla r|^2 - d\, r^2\right) d\mu \geq \frac{4}{d+4} \int_{\mathbb{S}^d} \left(|\nabla r|^2 + A\, r^2\right) d\mu\,.$$

Proof. We recall that the Laplace–Beltrami operator on $\mathbb{S}^d$ is diagonal in the basis of spherical harmonics and that its eigenvalue on spherical harmonics of degree ℓ is $\ell\,(\ell+d-1)$.

Conditions (8) mean that r is orthogonal to spherical harmonics of degrees $\ell \leq 1$. Diagonalizing the Laplace–Beltrami operator, the claimed inequality becomes

$$\ell\,(\ell+d-1) - d \geq \frac{4}{d+4} \left(\ell\,(\ell+d-1) + \tfrac14\, d\,(d-2)\right) \quad \text{for all} \quad \ell \geq 2\,.$$

This is elementary to check. □

3.2. **Warm-up: A bound with suboptimal dimension dependence.** In this subsection we prove a preliminary version of Theorem 4 where the constant $\theta\,\epsilon_0$ on the right side is replaced by some d-dependent constant, which decreases much faster than d^{-1} as d increases.

The motivation for proving this preliminary version is threefold. First, it explains the basic strategy of the proof without the additional difficulty of tracking the dependence on d. The latter will require some rather elaborate additional arguments. Second, this more involved proof works nicely when the exponent $q = 2^*$ is ≤ 3, which means $d \geq 6$. (It is, however, not difficult to adjust it to arbitrary d.) Therefore our chosen proof of Theorem 4 will combine the inequality proved in this subsection for $d = 3$, 4, 5 with the inequality proved in the next subsection for $d \geq 6$. Third, the simpler argument in this subsection gives simpler expressions for the relevant constants, which might be preferable in certain applications where the values of these constants play a role.

Proposition 7. *For all $\tilde\delta > 0$ and for all $-1 \leq r \in \mathrm{H}^1(\mathbb{S}^d)$ satisfying (7) and (8) one has*

$$\int_{\mathbb{S}^d} \left(|\nabla r|^2 + A\,(1+r)^2\right) d\mu - A \left(\int_{\mathbb{S}^d} (1+r)^q\, d\mu\right)^{2/q} \geq \mathsf{m}(\tilde\delta^{1/2}) \int_{\mathbb{S}^d} \left(|\nabla r|^2 + A\, r^2\right) d\mu$$

where $d\mu$ is the uniform probability measure, with

$$\begin{aligned} &\mathsf{m}(\nu) := \tfrac{4}{d+4} - \tfrac{2}{q}\,\nu^{q-2} && \text{if} \quad d \geq 6\,,\\ &\mathsf{m}(\nu) := \tfrac{4}{d+4} - \tfrac13\,(q-1)\,(q-2)\,\nu - \tfrac{2}{q}\,\nu^{q-2} && \text{if} \quad d = 4\,,\,5\,,\\ &\mathsf{m}(\nu) := \tfrac47 - \tfrac{20}{3}\,\nu - 5\,\nu^2 - 2\,\nu^3 - \tfrac13\,\nu^4 && \text{if} \quad d = 3\,. \end{aligned} \tag{9}$$

We note that for any $d \geq 3$ there is a ν_d such that $\mathsf{m}(\nu) > 0$ for $\nu < \nu_d$. Thus, for $\tilde\delta < \nu_d^2$ we obtain a stability inequality.

We begin the proof of Proposition 7 with some elementary inequalities.

Lemma 8. *If $q \geq 2$, then, for all $x \geq 0$,*

$$(1+x)^{\frac2q} \leq 1 + \tfrac2q\, x\,.$$

This is well known and we omit its simple proof.

Lemma 9. *We have the following bounds.*

- *If $2 \le q \le 3$, then, for all $x \ge -1$,*
$$(1+x)^q \le 1 + q\,x + \tfrac{1}{2}\,q\,(q-1)\,x^2 + x_+^q\,.$$
- *If $3 \le q \le 4$, then, for all $x \ge -1$,*
$$(1+x)^q \le 1 + q\,x + \tfrac{1}{2}\,q\,(q-1)\,x^2 + \tfrac{1}{6}\,q\,(q-1)\,(q-2)\,x^3 + |x|^q\,.$$

Similar bounds can also be derived for real $q \in (4,\infty)$. They become increasingly more complicated as q passes an integer. The only bound for $q > 4$ that we shall need corresponds to the critical exponent $q = 6$ when $d = 3$. In that case, we rely on the binomial expansion $(1+x)^6 = 1 + 6\,x + 15\,x^2 + 20\,x^3 + 15\,x^4 + 6\,x^5 + x^6$.

Proof. We begin with the case $2 \le q \le 3$ and set
$$f(x) := (1+x)^q - 1 - q\,x - \tfrac{1}{2}\,q\,(q-1)\,x^2 - x_+^q\,.$$
For any $x \ge -1$, we compute
$$f'(x) = q\left((1+x)^{q-1} - 1 - (q-1)\,x - x_+^{q-1}\right),$$
$$f''(x) = q\,(q-1)\left((1+x)^{q-2} - 1 - x_+^{q-2}\right).$$
For $-1 \le x \le 0$ we clearly have $(1+x)^{q-2} - 1 - x_+^{q-2} = (1-|x|)^{q-2} - 1 \le 0$. For $x \ge 0$ we have, by a well-known elementary inequality, $(1+x)^{q-2} - 1 - x_+^{q-2} = (1+x)^{q-2} - 1 - x^{q-2} \le 0$. To summarize, f is concave on $[-1,\infty)$. We conclude that, for all $x \ge -1$,
$$f(x) \le f(0) - f'(0)\,x\,.$$
Since $f(0) = f'(0) = 0$, this is the claimed inequality.

We now turn to the case $3 \le q \le 4$ and set this time
$$f(x) := (1+x)^q - 1 - q\,x - \tfrac{1}{2}\,q\,(q-1)\,x^2 - \tfrac{1}{6}\,q\,(q-1)\,(q-2)\,x^3 - |x|^q\,.$$
Again, we compute
$$f'(x) = q\left((1+x)^{q-1} - 1 - (q-1)\,x - \tfrac{1}{2}\,(q-1)\,(q-2)\,x^2 - |x|^{q-2}\,x\right),$$
$$f''(x) = q\,(q-1)\Big((1+x)^{q-2} - 1 - (q-2)\,x - |x|^{q-2}\Big)\,.$$
Since again $f(0) = f'(0) = 0$, the claimed inequality will follow if we can show concavity of f on $[-1,\infty)$, that is, $g \le 0$ on $[-1,\infty)$ where
$$g(x) := (1+x)^{q-2} - 1 - (q-2)\,x - |x|^{q-2}\,.$$
We compute
$$g'(x) = (q-2)\left((1+x)^{q-3} - 1 - |x|^{q-4}\,x\right),$$
$$g''(x) = (q-2)\,(q-3)\left((1+x)^{q-4} - |x|^{q-4}\right).$$
We discuss g separately on $[-1,0]$ and on $(0,\infty)$.

- We begin with the second case. For $x > 0$ we have, by the same elementary inequality as before, $(1+x)^{q-3} - 1 - x^{q-3} < 0$. Thus, $g' < 0$ on $(0,\infty)$. Since $g(0) = 0$, we deduce $g < 0$ on $(0,\infty)$.
- Now let us consider the interval $[-1,0]$. We see that $g'' > 0$ on $(-1,-1/2)$ and $g'' < 0$ on $(-1/2,0)$. Therefore g' is increasing on $(-1,-1/2)$ and decreasing on $(-1/2,0)$. Since $g'(-1) = g'(0) = 0$, we conclude that $g' > 0$ on $(-1,0)$ and therefore g is increasing on $(-1,0)$. Since $g(0) = 0$ we conclude that $g < 0$ on $[-1,0)$, as claimed.

This completes the proof of the lemma. □

From these lemmas we easily obtain the following inequalities.

Proposition 10. *Let $(X, d\mu)$ be a measure space and u, $r \in \mathrm{L}^q(X, d\mu)$ for some $q \geq 2$ with $u \geq 0$ and $u + r \geq 0$. Assume also that $\int_X u^{q-1}\, r\, d\mu = 0$.*

- *If $2 \leq q \leq 3$, then*
$$\|u+r\|_q^2 \leq \|u\|_q^2 + \|u\|_q^{2-q}\left((q-1)\int_X u^{q-2}\, r^2\, d\mu + \frac{2}{q}\int_X r_+^q\, d\mu\right).$$
- *If $3 \leq q \leq 4$, then*
$$\|u+r\|_q^2 \leq \|u\|_q^2 + \|u\|_q^{2-q}\left((q-1)\int_X u^{q-2}\, r^2\, d\mu + \tfrac{1}{3}\,(q-1)\,(q-2)\int_X u^{q-3}\, r^3\, d\mu + \tfrac{2}{q}\int_X |r|^q\, d\mu\right).$$
- *If $q = 6$, then*
$$\|u+r\|_q^2 \leq \|u\|_q^2 + \|u\|_q^{2-q}\left(5\int_X u^{q-2}\, r^2\, d\mu + \tfrac{20}{3}\int_X u^{q-3}\, r^3\, d\mu + 5\int_X u^{q-4}\, r^4\, d\mu + 2\int_X u^{q-5}\, r^5\, d\mu + \tfrac{1}{3}\int_X r^6\, d\mu\right).$$

Proof of Proposition 10. For $2 \leq q \leq 3$ we have, by Lemma 9, almost everywhere on X,
$$(u+r)^q \leq u^q + q\, u^{q-1}\, r + \tfrac{1}{2}\, q\,(q-1)\, u^{q-2}\, r^2 + r_+^q\,.$$
Integrating this and using the assumed orthogonality condition, we obtain
$$\int_X (u+r)^q\, d\mu \leq \int_X u^q\, d\mu + \tfrac{1}{2}\, q\,(q-1)\int_X u^{q-2}\, r^2\, d\mu + \int_X r_+^q\, d\mu\,.$$
Applying Lemma 8, we obtain
$$\left(\int_X (u+r)^q\, d\mu\right)^{\frac{2}{q}} \leq \left(\int_X u^q\, d\mu\right)^{\frac{2}{q}} + \left(\int_X u^q\, d\mu\right)^{\frac{2-q}{q}}\left((q-1)\int_X u^{q-2}\, r^2\, d\mu + \tfrac{2}{q}\int_X r_+^q\, d\mu\right).$$
This is the claimed inequality for $2 \leq q \leq 3$. The proof for $3 < q \leq 4$ is similar and the inequality for $q = 6$ follows from expanding the polynomial. □

Proof of Proposition 7. Let r be as in Theorem 4. Because of the mean-zero condition we can apply Proposition 10 with $u = 1$ on $X = \mathbb{S}^d$ and $d\mu$ the uniform probability measure. We simplify the resulting term using Hölder and Sobolev, which imply for $2 < t \leq q$,
$$\int_{\mathbb{S}^d} |r|^t\, d\mu \leq \left(\int_{\mathbb{S}^d} |r|^q\, d\mu\right)^{t/q} \leq \tilde{\delta}^{(t-2)/2}\, A^{-1}\int_{\mathbb{S}^d}\left(|\nabla r|^2 + A\, r^2\right) d\mu\,.$$
In this way, we obtain
$$\left(\int_{\mathbb{S}^d} (1+r)^q\, d\mu\right)^{2/q} \leq 1 + (q-1)\int_{\mathbb{S}^d} r^2\, d\mu + \mathsf{n}(\tilde{\delta}^{1/2})\, A^{-1}\int_{\mathbb{S}^d}\left(|\nabla r|^2 + A\, r^2\right) d\mu\,,$$

where

$$\begin{aligned}
&\mathsf{n}(\nu) := \tfrac{2}{q}\,\nu^{q-2} && \text{if} \quad d \ge 6\,,\\
&\mathsf{n}(\nu) := \tfrac{1}{3}\,(q-1)\,(q-2)\,\nu + \tfrac{2}{q}\,\nu^{q-2} && \text{if} \quad d = 4\,,\,5\,,\\
&\mathsf{n}(\nu) := \tfrac{20}{3}\,\nu + 5\,\nu^2 + 2\,\nu^3 + \tfrac{1}{3}\,\nu^4 && \text{if} \quad d = 3\,.
\end{aligned}$$

Using $A\,(q-2) = d$, we deduce that

$$\int_{\mathbb{S}^d} \left(|\nabla r|^2 + A\,(1+r)^2\right) d\mu - A \left(\int_{\mathbb{S}^d} (1+r)^q\,d\mu\right)^{2/q} \\ \ge \int_{\mathbb{S}^d} \left(|\nabla r|^2 - d r^2\right) d\mu - \mathsf{n}(\tilde\delta^{1/2}) \int_{\mathbb{S}^d} \left(|\nabla r|^2 + A\,r^2\right) d\mu\,.$$

Using the spectral gap inequality in Lemma 6 and noting that $\mathsf{m}(\nu) = \frac{4}{d+4} - \mathsf{n}(\nu)$, we obtain the claimed inequality. □

Remark 11. *The estimates of Proposition 7 are good enough for proving Theorem 4 for d finite, but fail for proving that the stability constant is of the order of $\theta\,\epsilon_0$ in the large d limit, for some positive ϵ_0 independent of d and $\theta = q - 2 = 4/(d-2)$. Indeed, if we write that $\mathsf{m}(\nu) \ge \theta\,\epsilon_0$, we obtain*

$$\nu^{q-2} \le \frac{q}{2} \left(\frac{4}{d+4} - (q-2)\,\epsilon_0\right) \le \frac{q}{2}\,\frac{4}{d+4} = \frac{4\,d}{(d-2)\,(d+4)} \le \frac{4}{d-2}\,,$$

which means $\nu \le \left(\frac{d-2}{4}\right)^{-\frac{d-2}{4}} < \sqrt{\tilde\delta}$ for d large enough, for any given $\tilde\delta > 0$. Theorem 4 cannot be deduced from Proposition 7 as $d \to +\infty$ and this is why we need better estimates.

3.3. **Cutting r into pieces.** We turn now to the proof of Theorem 4 with the optimal dependence of the constant on the dimension. Thus, until the end of Section 3 we will assume that r satisfies the assumptions of Theorem 4. The following proposition gives an upper bound on

$$(1+r)^q - 1 - q\,r$$

for real numbers r in terms of three numbers

$$r_1 := \min\{r, \gamma\}\,, \quad r_2 := \min\{(r-\gamma)_+, M-\gamma\} \quad \text{and} \quad r_3 := (r-M)_+ \tag{10}$$

where γ and M are parameters such that $0 < \gamma < M$. We will later apply this when r is a function. Our goal is to obtain a bound in terms of

$$\theta := q-2 \quad \text{where} \quad q = 2^* = \frac{2\,d}{d-2}\,. \tag{11}$$

We have in mind to let $d \to +\infty$ so that $\theta \to 0_+$.

Proposition 12. *Given $M \in (0,+\infty)$ and $\overline{M} \in [\sqrt{e}, +\infty)$, there are two positive constants C_M and $C_{M,\overline{M}}$ depending respectively only on M and $\{M, \overline{M}\}$ such that, for any $\gamma \in (0, M]$, $q \in [2,3]$ and $r \in [-1,\infty)$, we have*

$$(1+r)^q - 1 - q\,r \le \tfrac{1}{2}\,q\,(q-1)\,(r_1+r_2)^2 + 2\,(r_1+r_2)\,r_3 + \left(1 + C_M\,\theta\,\overline{M}^{-1}\ln \overline{M}\right) r_3^q \\ + \left(\tfrac{3}{2}\,\gamma\,\theta\,r_1^2 + C_{M,\overline{M}}\,\theta\,r_2^2\right) \mathbb{1}_{\{r \le M\}} + C_{M,\overline{M}}\,\theta\,M^2\,\mathbb{1}_{\{r > M\}} \tag{12}$$

with r_1, r_2, r_3 and θ given by (10) and (11).

For the proof of Proposition 12, we need two elementary lemmas.

Lemma 13. *If $2\le q\le 3$, then for all $r\in[-1,\infty)$,*

$$(1+r)^q\le 1+q\,r+\tfrac12\,q\,(q-1)\,r^2+(q-2)\,r_+^3\,.$$

Proof. The inequality for $-1\le r\le 0$ follows from Lemma 9. Let now $r\ge 0$. Then

$$(1+r)^q-1-q\,r-\tfrac12\,q\,(q-1)\,r^2=q\,(q-1)\,(q-2)\int_0^r\int_0^s\int_0^t(1+u)^{q-3}\,du\,dt\,ds\,.$$

Since $q\le 3$ we have $(1+u)^{q-3}\le 1$ and therefore

$$q\,(q-1)\,(q-2)\int_0^r\int_0^s\int_0^t(1+u)^{q-3}\,du\,dt\,ds\le q\,(q-1)\,(q-2)\int_0^r\int_0^s\int_0^t du\,dt\,ds$$
$$=\tfrac q3\,\tfrac{q-1}2\,(q-2)\,r^3\le(q-2)\,r^3\,,$$

as claimed. □

Lemma 14. *For all $q\ge 2$ and all $v\ge\overline M\ge\sqrt e$ we have*

$$q\,v^{q-1}-2\,v\le\frac{1+2\ln\overline M}{\overline M}\,(q-2)\,v^q\quad\textit{and}\quad\tfrac12\,q\,(q-1)\,v^{q-2}-1\le\frac{\frac{1+q}2+\ln\overline M}{\overline M^2}\,(q-2)\,v^q\,.$$

Proof. Let

$$v_*^{(1)}:=\big(2\,\tfrac{q-1}q\big)^{\frac1{q-2}}\quad\text{and}\quad v_*^{(2)}:=\big(\tfrac1{q-1}\big)^{\frac1{q-2}}\,.$$

Then an elementary computation shows that $v\mapsto q\,v^{-1}-2\,v^{1-q}$ is increasing on $\big(0,v_*^{(1)}\big]$ and decreasing on $\big[v_*^{(1)},\infty\big)$. Similarly $v\mapsto\frac12\,q\,(q-1)\,v^{-2}-v^{-q}$ is increasing on $\big(0,v_*^{(2)}\big]$ and decreasing on $\big[v_*^{(2)},\infty\big)$. Thus,

$$q\,v^{q-1}-2\,v\le\Big(q\,\overline M^{-1}-2\,\overline M^{1-q}\Big)\,v^q\quad\text{for all}\quad v\ge\overline M\ge v_*^{(1)}$$

and

$$\tfrac12\,q\,(q-1)\,v^{q-2}-1\le\Big(\tfrac12\,q\,(q-1)\,\overline M^{-2}-\overline M^{-q}\Big)_+\,v^q\quad\text{for all}\quad v\ge\overline M\ge v_*^{(2)}\,.$$

One has $v_*^{(1)}\ge 1\ge v_*^{(2)}$ and, using $\ln t\le t-1$ for all $t>0$, we find

$$\ln v_*^{(1)}\le\tfrac1q\le\tfrac12\,,\quad\text{that is,}\quad v_*^{(1)}\le\sqrt e\,.$$

Thus, the above inequality hold, in particular, for $v\ge\overline M\ge\sqrt e$.

Moreover, using $1-t^{-1}\le\ln t$ for $t>1$ we can bound

$$q\,\overline M^{-1}-2\,\overline M^{1-q}=(q-2)\,\overline M^{-1}+2\,\Big(\overline M^{-1}-\overline M^{1-q}\Big)\le(q-2)\,\overline M^{-1}\,\big(1+2\,\ln\overline M\big)$$

and

$$\tfrac12\,q\,(q-1)\,\overline M^{-2}-\overline M^{-q}=\big(\tfrac12\,q\,(q-1)-1\big)\,\overline M^{-2}+\Big(\overline M^{-2}-\overline M^{-q}\Big)\le(q-2)\,\overline M^{-2}\,\big(\tfrac{1+q}2+\ln\overline M\big)\,.$$

This proves the assertion. □

Proof of Proposition 12. We now turn to the proof of (12). Assume first that $r \leq M$. We apply Lemma 13 and obtain

$$(1+r)^q - 1 - q\,r \leq \tfrac{1}{2}\,q\,(q-1)\,(r_1+r_2)^2 + \theta\,(r_1+r_2)_+^3\,.$$

If $r \leq \gamma$, then $r_2 = 0$ and (12) follows from $(r_1)_+^3 \leq \gamma\,r_1^2 \leq \frac{3}{2}\,\gamma\,r_1^2$. If $\gamma < r \leq M$, we have, since $r_1 = \gamma$ and $3\,r_1\,r_2 \leq \frac{1}{2}\,r_1^2 + \frac{9}{2}\,r_2^2$, we have

$$(r_1+r_2)_+^3 = \gamma\,r_1^2 + 3\,\gamma\,r_1\,r_2 + 3\,\gamma\,r_2^2 + r_2^3 \leq \tfrac{3}{2}\,\gamma\,r_1^2 + \left(\tfrac{15}{2}\,\gamma + M\right) r_2^2\,.$$

Since $\gamma \leq M$ this proves (12) with $C_{M,\overline{M}} \geq \frac{17}{2}\,M$.

From here on, let us consider the case $r > M$. Using $r = M + r_3$ we can write

$$(1+r)^q - 1 - q\,r = (1+r)^q - (1+r)^2 + (1+M)^2 - 1 - q\,M - (q-2)\,r_3 + r_3^2 + 2\,M\,r_3\,.$$

We use

$$(1+M)^2 - 1 - q\,M - \tfrac{1}{2}\,q\,(q-1)\,M^2 = -\tfrac{1}{2}\,(q-2)\,M\,\big(2 + (q+1)\,M\big) \leq 0$$

as well as $-\,(q-2)\,r_3 \leq 0$, to get

$$(1+r)^q - 1 - q\,r \leq \tfrac{1}{2}\,q\,(q-1)\,M^2 + 2\,M\,r_3 + r_3^2 + (1+r)^q - (1+r)^2\,. \tag{13}$$

Note that the terms $2\,M\,r_3 = 2\,(r_1+r_2)\,r_3$ and $\frac{1}{2}\,q\,(q-1)\,M^2 = \frac{1}{2}\,q\,(q-1)\,(r_1+r_2)^2$ are already of the form required in (12). In the following we bound the remaining terms $r_3^2 + (1+r)^q - (1+r)^2$. We do this separately in the cases $M < r \leq M + \overline{M}$ and $r > M + \overline{M}$, where $\overline{M} \geq 0$ is an additional parameter.

If $M < r \leq M + \overline{M}$, we have

$$(1+r)^q - (1+r)^2 \leq C_{M,\overline{M}}^{(1)}\,\theta \quad \text{and} \quad r_3^2 - r_3^q \leq C_{\overline{M}}^{(1)}\,\theta\,.$$

Inserting this into (13), we have for $M < r \leq M + \overline{M}$

$$(1+r)^q - 1 - q\,r \leq 2\,M\,r_3 + r_3^q + \left(\tfrac{1}{2}\,q\,(q-1) + C_{M,\overline{M}}\,\theta\right) M^2\,,$$

provided

$$C_{M,\overline{M}} \geq M^{-2}\left(C_{M,\overline{M}}^{(1)} + C_{\overline{M}}^{(1)}\right)\,.$$

This is a bound of the form (12), since $r_1 + r_2 = M$ for $r > M$.

Next, we consider the case $r > M + \overline{M}$, that is $r_3 = r - M > \overline{M}$. By Lemma 13 we have

$$\begin{aligned}
(1+r)^q = (1+M+r_3)^q &= r_3^q\left(1 + \tfrac{1+M}{r_3}\right)^q \\
&\leq r_3^q + q\,r_3^{q-1}\,(1+M) + \tfrac{1}{2}\,q\,(q-1)\,r_3^{q-2}\,(1+M)^2 + \theta\,r_3^{q-3}\,(1+M)^3 \\
&\leq r_3^q + q\,r_3^{q-1}\,(1+M) + \tfrac{1}{2}\,q\,(q-1)\,r_3^{q-2}\,(1+M)^2 + \theta\,\overline{M}^{q-3}\,(1+M)^3 \\
&= r_3^q + q\,r_3^{q-1}\,(1+M) + \tfrac{1}{2}\,q\,(q-1)\,r_3^{q-2}\,(1+M)^2 + C_{M,\overline{M}}^{(2)}\,\theta\,.
\end{aligned}$$

In the last inequality, we used $q \leq 3$ and $r_3 > \overline{M}$. This, together with

$$(1+r)^2 = (1+M+r_3)^2 = r_3^2 + 2\,r_3\,(1+M) + (1+M)^2\,,$$

gives

$$\begin{aligned}\tfrac12\,q\,(q-1)\,M^2+2\,M\,r_3+r_3^2+(1+r)^q-(1+r)^2 \\ \le 2\,M\,r_3+r_3^q+\left(q\,r_3^{q-1}-2\,r_3\right)(1+M) \\ +\left(\tfrac12\,q\,(q-1)\,r_3^{q-2}-1\right)(1+M)^2+C^{(2)}_{M,\overline{M}}\,\theta+\tfrac12\,q\,(q-1)\,M^2\,.\end{aligned}$$

We now assume that $\overline{M}\ge\sqrt{e}$. Then, by Lemma 14,

$$q\,r_3^{q-1}-2\,r_3\le\frac{1+2\,\ln\overline{M}}{\overline{M}}\,\theta\,r_3^q\quad\text{and}\quad\tfrac12\,q\,(q-1)\,r_3^{q-2}-1\le\frac{2+\ln\overline{M}}{\overline{M}^2}\,\theta\,r_3^q\,.$$

Thus,

$$\begin{aligned}\tfrac12\,q\,(q-1)\,M^2+2\,M\,r_3+r_3^2+(1+r)^q-(1+r)^2 \\ \le 2\,M\,r_3+\left(1+\frac{C_M\,\ln\overline{M}}{\overline{M}}\,\theta\right)r_3^q+C^{(2)}_{M,\overline{M}}\,\theta+\tfrac12\,q\,(q-1)\,M^2\end{aligned}$$

where C_M is a constant satisfying

$$\frac{1+2\,\ln\overline{M}}{\overline{M}}\,(1+M)+\frac{2+\ln\overline{M}}{\overline{M}^2}\,(1+M)^2\le\frac{C_M\,\ln\overline{M}}{\overline{M}}\quad\text{for all}\quad\overline{M}\ge\sqrt{e}\,.$$

Combining this with (13) we obtain a bound of the form (12), provided the constant $C_{M,\overline{M}}$ there satisfies

$$C_{M,\overline{M}}\ge M^{-2}\,C^{(2)}_{M,\overline{M}}\,.$$

This concludes the proof with $C_{M,\overline{M}}=M^{-2}\,\max\left\{C^{(1)}_{M,\overline{M}}+C^{(1)}_{\overline{M}},\,C^{(2)}_{M,\overline{M}}\right\}$. □

Corollary 15. *Given $\epsilon>0$, $M>0$, and $\gamma\in(0,M/2)$, there is a constant $C_{\gamma,\epsilon,M}>0$ with the following property: if $2\le q\le3$, $r\in[-1,\infty)$, then*

$$\begin{aligned}(1+r)^q-1-q\,r\le\left(\tfrac12\,q\,(q-1)+2\,\gamma\,\theta\right)r_1^2+\left(\tfrac12\,q\,(q-1)+C_{\gamma,\epsilon,M}\,\theta\right)r_2^2 \\ +2\,r_1\,r_2+2\,(r_1+r_2)\,r_3+(1+\epsilon\,\theta)\,r_3^q\end{aligned}\tag{14}$$

with r_1, r_2, r_3 and θ given by (10) and (11).

Proof. Since

$$q\,(q-1)\,r_1\,r_2=2\,r_1\,r_2+(3+\theta)\,\theta\,r_1\,r_2\le2\,r_1\,r_2+4\,\theta\,r_1\,r_2\le2\,r_1\,r_2+\tfrac{\gamma}{2}\,\theta\,r_1^2+\tfrac{8}{\gamma}\,\theta\,r_2^2$$

and

$$C_{M,\overline{M}}\,M^2\,\mathbb{1}_{\{r>M\}}\le4\,C_{M,\overline{M}}\,(M-\gamma)^2\,\mathbb{1}_{\{r>M\}}\le4\,C_{M,\overline{M}}\,r_2^2\,,$$

we deduce from (12) that

$$\begin{aligned}(1+r)^q-1-q\,r\le\left(\tfrac12\,q\,(q-1)+2\,\gamma\,\theta\right)r_1^2+\left(\tfrac12\,q\,(q-1)+\tfrac{8}{\gamma}\,\theta+5\,C_{M,\overline{M}}\,\theta\right)r_2^2 \\ +2\,r_1\,r_2+2\,(r_1+r_2)\,r_3+\left(1+C_M\,\theta\,\overline{M}^{-1}\ln\overline{M}\right)r_3^q\,.\end{aligned}$$

Given any $M\ge2\,\gamma$, we choose $\overline{M}$ such that $\overline{M}\ge\sqrt{e}$ and $C_M\,\overline{M}^{-1}\ln\overline{M}\le\epsilon$. Then (14) follows with $C_{\gamma,\epsilon,M}=\frac{8}{\gamma}+5\,C_{M,\overline{M}}$. □

We will apply Corollary 15 for q close to 2 and the main point is how the constants depend on q. Apart from the 'natural' terms $\frac{1}{2}\,q\,(q-1)\,r_1^2$, $\frac{1}{2}\,q\,(q-1)\,r_2^2$, $2\,r_1\,r_2$ and $2\,(r_1+r_2)\,r_3$, all other terms are multiplied by θ, which is small in our application. Moreover, we have the freedom to choose γ and ϵ as small as we please (independent of q) and so the prefactors of the terms r_1^2 and r_3^q are almost the natural ones. The price to be paid is a rather large constant in front of the error term involving r_2^2. In order to have better estimates as $d\to+\infty$, more work is needed.

3.4. **A detailed estimate of the deficit.** We assume that $-1\le r\in\mathrm{H}^1(\mathbb{S}^d)$ satisfies the orthogonality conditions (8) as well as the smallness condition (7) with some $\tilde\delta$, and we show that, if this $\tilde\delta$ is small enough, given $\epsilon_0\in(0,\frac{1}{3})$, we obtain the claimed inequality.

Given two parameters ϵ_1, $\epsilon_2>0$ we apply Corollary 15 with

$$\gamma=\epsilon_1/2\,,\quad \epsilon=\epsilon_2\quad\text{and}\quad C_{\gamma,\epsilon,M}=C_{\epsilon_1,\epsilon_2}\,. \tag{15}$$

In terms of these parameters, we decompose $r=r_1+r_2+r_3$. We obtain

$$\int_{\mathbb{S}^d}|\nabla r|^2\,d\mu=\int_{\mathbb{S}^d}|\nabla r_1|^2\,d\mu+\int_{\mathbb{S}^d}|\nabla r_2|^2\,d\mu+\int_{\mathbb{S}^d}|\nabla r_3|^2\,d\mu$$

and, since r has mean zero,

$$\int_{\mathbb{S}^d}(1+r)^2\,d\mu=1+\int_{\mathbb{S}^d}r^2\,d\mu\,.$$

Moreover,

$$\int_{\mathbb{S}^d}r^2\,d\mu=\int_{\mathbb{S}^d}r_1^2\,d\mu+\int_{\mathbb{S}^d}r_2^2\,d\mu+\int_{\mathbb{S}^d}r_3^2\,d\mu+2\int_{\mathbb{S}^d}r_1\,r_2\,d\mu+2\int_{\mathbb{S}^d}(r_1+r_2)\,r_3\,d\mu\,.$$

According to Corollary 15 and using again the fact that r has mean zero, we have

$$\begin{aligned}\int_{\mathbb{S}^d}(1+r)^q\,d\mu\le 1+\big(\tfrac{1}{2}\,q\,(q-1)+\epsilon_1\,\theta\big)\int_{\mathbb{S}^d}r_1^2\,d\mu+\big(\tfrac{1}{2}\,q\,(q-1)+C_{\epsilon_1,\epsilon_2}\,\theta\big)\int_{\mathbb{S}^d}r_2^2\,d\mu\\ +2\int_{\mathbb{S}^d}r_1\,r_2\,d\mu+2\int_{\mathbb{S}^d}(r_1+r_2)\,r_3\,d\mu+(1+\epsilon_2\,\theta)\int_{\mathbb{S}^d}r_3^q\,d\mu\,.\end{aligned}$$

Using $(1+x)^{2/q}\le 1+\frac{2}{q}\,x$, we obtain

$$\begin{aligned}\left(\int_{\mathbb{S}^d}(1+r)^q\,d\mu\right)^{2/q}&\le 1+\big(q-1+\tfrac{2}{q}\,\epsilon_1\,\theta\big)\int_{\mathbb{S}^d}r_1^2\,d\mu+\big(q-1+\tfrac{2}{q}\,C_{\epsilon_1,\epsilon_2}\,\theta\big)\int_{\mathbb{S}^d}r_2^2\,d\mu\\ &\quad+\tfrac{4}{q}\int_{\mathbb{S}^d}r_1\,r_2\,d\mu+\tfrac{4}{q}\int_{\mathbb{S}^d}(r_1+r_2)\,r_3\,d\mu+\tfrac{2}{q}\,(1+\epsilon_2\,\theta)\int_{\mathbb{S}^d}r_3^q\,d\mu\\ &\le 1+(q-1+\epsilon_1\,\theta)\int_{\mathbb{S}^d}r_1^2\,d\mu+(q-1+C_{\epsilon_1,\epsilon_2}\,\theta)\int_{\mathbb{S}^d}r_2^2\,d\mu\\ &\quad+2\int_{\mathbb{S}^d}r_1\,r_2\,d\mu+2\int_{\mathbb{S}^d}(r_1+r_2)\,r_3\,d\mu+\tfrac{2}{q}\,(1+\epsilon_2\,\theta)\int_{\mathbb{S}^d}r_3^q\,d\mu\,.\end{aligned}$$

In the last inequality we used $\frac{2}{q} \le 1$. For the final term, however, it is vital that we keep $\frac{2}{q}$. We thus have, for any $0 < \epsilon_0 \le \theta^{-1}$,

$$\begin{aligned}\int_{\mathbb{S}^d} \left(|\nabla r|^2 + A\,(1+r)^2\right) d\mu - A\left(\int_{\mathbb{S}^d} (1+r)^q\, d\mu\right)^{2/q}\\ \ge \theta\,\epsilon_0 \int_{\mathbb{S}^d} \left(|\nabla r|^2 + A\, r^2\right) d\mu\\ + (1-\theta\,\epsilon_0)\int_{\mathbb{S}^d} \left(|\nabla r_1|^2 + A\, r_1^2\right) d\mu - A\,(q-1+\epsilon_1\,\theta)\int_{\mathbb{S}^d} r_1^2\, d\mu\\ + (1-\theta\,\epsilon_0)\int_{\mathbb{S}^d} \left(|\nabla r_2|^2 + A\, r_2^2\right) d\mu - A\,(q-1+C_{\epsilon_1,\epsilon_2}\,\theta)\int_{\mathbb{S}^d} r_2^2\, d\mu\\ + (1-\theta\,\epsilon_0)\int_{\mathbb{S}^d} \left(|\nabla r_3|^2 + A\, r_3^2\right) d\mu - \tfrac{2}{q}\,A\,(1+\epsilon_2\,\theta)\int_{\mathbb{S}^d} r_3^q\, d\mu\,.\end{aligned}$$

With another parameter $\sigma_0 > 0$ we define

$$\begin{aligned}I_1 &:= (1-\theta\,\epsilon_0)\int_{\mathbb{S}^d} \left(|\nabla r_1|^2 + A\, r_1^2\right) d\mu - A\,(q-1+\epsilon_1\,\theta)\int_{\mathbb{S}^d} r_1^2\, d\mu + A\,\sigma_0\,\theta\int_{\mathbb{S}^d} (r_2^2 + r_3^2)\, d\mu\,,\\ I_2 &:= (1-\theta\,\epsilon_0)\int_{\mathbb{S}^d} \left(|\nabla r_2|^2 + A\, r_2^2\right) d\mu - A\left(q-1+(\sigma_0 + C_{\epsilon_1,\epsilon_2})\,\theta\right)\int_{\mathbb{S}^d} r_2^2\, d\mu\,,\\ I_3 &:= (1-\theta\,\epsilon_0)\int_{\mathbb{S}^d} \left(|\nabla r_3|^2 + A\, r_3^2\right) d\mu - \tfrac{2}{q}\,A\,(1+\epsilon_2\,\theta)\int_{\mathbb{S}^d} r_3^q\, d\mu - A\,\sigma_0\,\theta\int_{\mathbb{S}^d} r_3^2\, d\mu\,.\end{aligned}$$

We recall that $A = \frac{1}{4}\,d\,(d-2)$. For later purposes, we note that $A\,\theta = A\,(q-2) = d$ and

$$\begin{aligned}I_1 &= (1-\theta\,\epsilon_0)\int_{\mathbb{S}^d} |\nabla r_1|^2\, d\mu - d\,(1+\epsilon_0+\epsilon_1)\int_{\mathbb{S}^d} r_1^2\, d\mu + d\,\sigma_0\int_{\mathbb{S}^d} (r_2^2 + r_3^2)\, d\mu\,,\\ I_2 &= (1-\theta\,\epsilon_0)\int_{\mathbb{S}^d} |\nabla r_2|^2\, d\mu - d\,(1+\epsilon_0+\sigma_0+C_{\epsilon_1,\epsilon_2})\int_{\mathbb{S}^d} r_2^2\, d\mu\,.\end{aligned}$$

To summarize, we have

$$\int_{\mathbb{S}^d} \left(|\nabla r|^2 + A\,(1+r)^2\right) d\mu - A\left(\int_{\mathbb{S}^d} (1+r)^q\, d\mu\right)^{2/q} \ge \theta\,\epsilon_0\int_{\mathbb{S}^d} \left(|\nabla r|^2 + A\, r^2\right) d\mu + \sum_{k=1}^{3} I_k\,.$$

In the following we will show that I_1, I_3 and I_2 are nonnegative, in this order.

3.5. **Bound on I_1.** The intuition here is the same as in the proof of the spectral gap inequality in Lemma 6. Namely, the lowest L^2-eigenvalue of $\int_{\mathbb{S}^d} |\nabla u|^2\, d\mu$ on functions orthogonal to spherical harmonics of degree less or equal than 1 is $2\,(d+1)$, while the term that we are subtracting corresponds to a component that is multiplied by a number only slightly larger than d. Therefore, there is space to accomodate the errors coming from ϵ_0 and ϵ_1. Another source of an error comes from the fact that, while r is orthogonal to spherical harmonics of degree less or equal than 1, r_1 need not be. However, as we will see, it nearly is. To control the corresponding error from orthogonality we need the positive terms involving σ_0.

Proposition 16. *For any $0 < \epsilon_0 < \frac{1}{3}$, there is a constant $\overline{\sigma}_0(\gamma, \epsilon_0, \tilde{\delta}) > 0$ depending explicitly on γ, ϵ_0 and $\tilde{\delta}$ such that for all $d \ge 6$ and all $r \in \mathrm{H}^1(\mathbb{S}^d)$ such that $r \ge -1$ and satisfying (7)* *and* (8) *as in Theorem 4, with* θ *given by* (11),

$$\epsilon_1=\tfrac12\,(1-3\,\epsilon_0) \tag{16}$$

and $\sigma_0\ge\overline{\sigma}_0(\gamma,\epsilon_0,\tilde\delta)$, *one has*

$$I_1\ge0\,.$$

Notice that $\theta=q-2\le1$ with $q=2\,d/(d-2)$ means $d\ge6$. An expression of $\overline{\sigma}_0$ is given below in (20).

Proof. We split the proof in three simple steps.

Step 1. Let $\tilde r_1$ be the orthogonal projection of r_1 onto the space of spherical harmonics of degree ≥2, that is,

$$\tilde r_1=r_1-\int_{\mathbb S^d}r_1\,d\mu-(d+1)\,\omega\cdot\int_{\mathbb S^d}\omega'\,r_1(\omega')\,d\mu(\omega')$$

as $\sqrt{d+1}\,\omega_j$ is $\mathrm L^2$-normalized with respect to the uniform probability measure on the sphere for any $j=1,\,2,\,\ldots,N+1$. Then

$$\begin{aligned}I_1&=(1-\theta\,\epsilon_0)\int_{\mathbb S^d}|\nabla\tilde r_1|^2\,d\mu-d\,(1+\epsilon_0+\epsilon_1)\int_{\mathbb S^d}\tilde r_1^2\,d\mu+d\,\sigma_0\int_{\mathbb S^d}(r_2^2+r_3^2)\,d\mu\\&\qquad-d\,(1+\epsilon_0+\epsilon_1)\left(\int_{\mathbb S^d}r_1\,d\mu\right)^2-d\,(d+1)\,\big((1+\theta)\,\epsilon_0+\epsilon_1\big)\left|\int_{\mathbb S^d}\omega\,r_1\,d\mu\right|^2\\&\ge\big(2\,(d+1)\,(1-\theta\,\epsilon_0)-d\,(1+\epsilon_0+\epsilon_1)\big)\int_{\mathbb S^d}\tilde r_1^2\,d\mu+d\,\sigma_0\int_{\mathbb S^d}(r_2^2+r_3^2)\,d\mu\\&\qquad-d\,(1+\epsilon_0+\epsilon_1)\left(\int_{\mathbb S^d}r_1\,d\mu\right)^2-d\,(d+1)\,\big((1+\theta)\,\epsilon_0+\epsilon_1\big)\left|\int_{\mathbb S^d}\omega\,r_1\,d\mu\right|^2.\end{aligned}$$

In the equality, we used the fact that the ω_j's are eigenfunctions of the Laplace–Beltrami operator with eigenvalue d. In the inequality, we used the fact that the operator is bounded from below by $2\,(d+1)$ on the orthogonal complement of spherical harmonics of degree less or equal than 1.

Step 2. With ϵ_1 given by (16), it is easy to see that for any $\epsilon_0<\frac13$, using $\theta\le1$, we have

$$2\,(d+1)\,(1-\theta\,\epsilon_0)-d\,(1+\epsilon_0+\epsilon_1)\ge\tfrac d2\,(1-3\,\epsilon_0)+2\,(1-\epsilon_0)>d\,\epsilon_1>0\,. \tag{17}$$

Using

$$\int_{\mathbb S^d}\tilde r_1^2\,d\mu=\int_{\mathbb S^d}r_1^2\,d\mu-\left(\int_{\mathbb S^d}r_1\,d\mu\right)^2-(d+1)\left|\int_{\mathbb S^d}\omega\,r_1\,d\mu\right|^2$$

and $\theta \le 1$, we obtain

$$\begin{aligned}\frac{1}{d}\,I_1 &\ge \epsilon_1\int_{\mathbb{S}^d}\tilde r_1^2\,d\mu+\sigma_0\int_{\mathbb{S}^d}(r_2^2+r_3^2)\,d\mu\\&\qquad-(1+\epsilon_0+\epsilon_1)\left(\int_{\mathbb{S}^d}r_1\,d\mu\right)^2-(d+1)\,\big((1+\theta)\,\epsilon_0+\epsilon_1\big)\left|\int_{\mathbb{S}^d}\omega\,r_1\,d\mu\right|^2\\&\ge \epsilon_1\int_{\mathbb{S}^d}r_1^2\,d\mu+\sigma_0\int_{\mathbb{S}^d}(r_2^2+r_3^2)\,d\mu\\&\qquad-(1+\epsilon_0)\left(\int_{\mathbb{S}^d}r_1\,d\mu\right)^2-2\,(d+1)\,\epsilon_0\left|\int_{\mathbb{S}^d}\omega\,r_1\,d\mu\right|^2\,.\end{aligned}$$

Step 3. Let us take care of the rank one terms coming from the orthogonality conditions. We will show that $I_1 \ge 0$ for an appropriately chosen σ_0 as a consequence of

$$(1+\epsilon_0)\left(\int_{\mathbb{S}^d}r_1\,d\mu\right)^2+2\,(d+1)\,\epsilon_0\left|\int_{\mathbb{S}^d}\omega\,r_1\,d\mu\right|^2\le\epsilon_1\int_{\mathbb{S}^d}r_1^2\,d\mu+\sigma_0\int_{\mathbb{S}^d}(r_2^2+r_3^2)\,d\mu\,. \tag{18}$$

Let Y be one of the functions 1 and $a\cdot\omega$, $a\in\mathbb{R}^{d+1}$. Then, since $\int_{\mathbb{S}^d}Y\,r\,d\mu=0$ by (8),

$$\left(\int_{\mathbb{S}^d}Y\,r_1\,d\mu\right)^2=\left(\int_{\mathbb{S}^d}Y(r_2+r_3)\,d\mu\right)^2\le\|Y\|_{\mathrm{L}^4(\mathbb{S}^d)}^2\;\mu\big(\{r_2+r_3>0\}\big)^{1/2}\;\|r_2+r_3\|_{\mathrm{L}^2(\mathbb{S}^d)}^2\;.$$

Since $\{r_2+r_3>0\}\subset\{r_1\ge\gamma\}$, we have

$$\mu(\{r_2+r_3>0\})\le\mu(\{r_1\ge\gamma\})\le\frac{1}{\gamma^2}\int_{\mathbb{S}^d}r_1^2\,d\mu=\frac{1}{\gamma^2}\;\|r_1\|_{\mathrm{L}^2(\mathbb{S}^d)}^2\;.$$

Thus we have

$$\left(\int_{\mathbb{S}^d}Y\,r_1\,d\mu\right)^2\le\|Y\|_{\mathrm{L}^4(\mathbb{S}^d)}^2\;\frac{\sqrt{2\,\tilde\delta}}{\gamma}\;\|r_1\|_{\mathrm{L}^2(\mathbb{S}^d)}\left(\int_{\mathbb{S}^d}\left(r_2^2+r_3^2\right)d\mu\right)^{1/2} \tag{19}$$

using $\|r_2+r_3\|_{\mathrm{L}^2(\mathbb{S}^d)}^2\le\sqrt{2\,\tilde\delta}\left(\int_{\mathbb{S}^d}(r_2^2+r_3^2)\,d\mu\right)^{1/2}$ because $\|r_2+r_3\|_{\mathrm{L}^2(\mathbb{S}^d)}^2\le 2\int_{\mathbb{S}^d}(r_2^2+r_3^2)\,d\mu$ and

$$\|r_2+r_3\|_{\mathrm{L}^2(\mathbb{S}^d)}\le\|r\|_{\mathrm{L}^2(\mathbb{S}^d)}\le\|r\|_{\mathrm{L}^q(\mathbb{S}^d)}\le\sqrt{\tilde\delta}\,.$$

If $Y=1$, then clearly $\|Y\|_{\mathrm{L}^4(\mathbb{S}^d)}=1$ and (19) gives

$$\left(\int_{\mathbb{S}^d}r_1\,d\mu\right)^2\le\frac{\sqrt{2\,\tilde\delta}}{\gamma}\;\|r_1\|_{\mathrm{L}^2(\mathbb{S}^d)}\left(\int_{\mathbb{S}^d}\left(r_2^2+r_3^2\right)d\mu\right)^{1/2}\;.$$

If $Y=a\cdot\omega$, then a quick computation gives

$$\|Y\|_{\mathrm{L}^4(\mathbb{S}^d)}^4=\frac{\int_0^\pi\cos^4\theta\,\sin^{d-1}\theta\,d\theta}{\int_0^\pi\sin^{d-1}\theta\,d\theta}\,|a|^4=\frac{3\,|a|^4}{(d+3)\,(d+1)}\le\frac{3\,|a|^4}{(d+1)^2}\,.$$

From (19) applied with $a=\int_{\mathbb{S}^d}\omega\,r_1\,d\mu$, we obtain

$$(d+1)\left|\int_{\mathbb{S}^d}\omega\,r_1\,d\mu\right|^2=\frac{d+1}{|a|^2}\left(\int_{\mathbb{S}^d}Y\,r_1\,d\mu\right)^2\le\sqrt3\,\frac{\sqrt{2\,\tilde\delta}}{\gamma}\;\|r_1\|_{\mathrm{L}^2(\mathbb{S}^d)}\left(\int_{\mathbb{S}^d}\left(r_2^2+r_3^2\right)d\mu\right)^{1/2}\,.$$

Summing up, we have

$$\epsilon_1\,\|r_1\|^2_{\mathrm{L}^2(\mathbb{S}^d)}+\sigma_0\int_{\mathbb{S}^d}\big(r_2^2+r_3^2\big)\,d\mu-(1+\epsilon_0)\left(\int_{\mathbb{S}^d}r_1\,d\mu\right)^2-2\,(d+1)\,\epsilon_0\left|\int_{\mathbb{S}^d}\omega\,r_1\,d\mu\right|^2$$
$$\ge\epsilon_1\,\|r_1\|^2_{\mathrm{L}^2(\mathbb{S}^d)}+\sigma_0\int_{\mathbb{S}^d}\big(r_2^2+r_3^2\big)\,d\mu-\big(1+(2\sqrt3+1)\,\epsilon_0\big)\,\frac{\sqrt{2\,\tilde\delta}}{\gamma}\,\|r_1\|_{\mathrm{L}^2(\mathbb{S}^d)}\left(\int_{\mathbb{S}^d}\big(r_2^2+r_3^2\big)\,d\mu\right)^{1/2}$$

and the right-hand side is nonnegative under a nonpositive discriminant condition which is satisfied by $\sigma_0\ge\overline\sigma_0(\gamma,\epsilon_0,\tilde\delta)$ with

$$\overline\sigma_0(\gamma,\epsilon_0,\delta):=\frac{1}{2\,\epsilon_1}\,\big(1+(2\sqrt3+1)\,\epsilon_0\big)^2\,\frac{\delta}{\gamma^2}\,. \tag{20}$$

This choice establishes (18) and allows us to conclude that $I_1\ge0$. □

Let us define

$$\delta_1:=\frac{4\,\epsilon_1\,\epsilon_2\,\gamma^2}{q\,\big(1+(2\sqrt3+1)\,\epsilon_0\big)^2}\,. \tag{21}$$

The condition $\sigma_0\ge\overline\sigma_0(\gamma,\epsilon_0,\tilde\delta)$ of Proposition 16 can be inverted as follows.

Corollary 17. *For any $0<\epsilon_0<\frac13$ and $\sigma_0>0$, for all $d\ge6$ and all $r\in\mathrm{H}^1(\mathbb{S}^d)$ such that $r\ge-1$ and satisfying (7) and (8) as in Theorem 4, with θ, ϵ_1, ϵ_2 and δ_1 respectively given by (11), (16), (15) and (21), if*

$$0<\tilde\delta\le\delta_1\,\frac{q\,\sigma_0}{2\,\epsilon_2}\,,$$

then one has $I_1\ge0$.

Remark 18. *The assumption $\epsilon_0<\frac13$ is used in (16) to guarantee that ϵ_1 takes positive values. A less restrictive condition can be obtained by requesting that the left-hand side in (17) is actually 0. We see that if $\epsilon_0<1$, then a similar bound as in (17), namely with $\frac12\,(1-\epsilon_0)$ on the right side, holds for all sufficiently large d, depending on ϵ_0.*

3.6. **Bound on I_3.** The idea for bounding this term is to use the Sobolev inequality. The extra coefficient $\frac2q<1$ gives us enough room to accomodate all error terms.

Proposition 19. *Assume that $\tilde\delta\in(0,1)$ and $0<\epsilon_0<\frac13$. With*

$$\epsilon_2:=\frac14\,(1-3\,\epsilon_0) \tag{22}$$

and $\sigma_0=\frac2q\,\epsilon_2$, for all $d\ge6$, all $\tilde\delta\le1$ and all r as in Theorem 4, one has

$$I_3\ge0\,.$$

Proof. Taking into account the choice for σ_0, we have

$$I_3=(1-\theta\,\epsilon_0)\int_{\mathbb{S}^d}\big(|\nabla r_3|^2+A\,r_3^2\big)\,d\mu-\tfrac2q\,A\left((1+\epsilon_2\,\theta)\int_{\mathbb{S}^d}r_3^q\,d\mu+\epsilon_2\,\theta\int_{\mathbb{S}^d}r_3^2\,d\mu\right)$$

We have $\|r_3\|_{\mathrm{L}^q(\mathbb{S}^d)}^q \le \|r_3\|_{\mathrm{L}^q(\mathbb{S}^d)}^2$ because $\|r_3\|_{\mathrm{L}^q(\mathbb{S}^d)} \le \|r\|_{\mathrm{L}^q(\mathbb{S}^d)} \le 1$ and $\|r_3\|_{\mathrm{L}^2(\mathbb{S}^d)} \le \|r_3\|_{\mathrm{L}^q(\mathbb{S}^d)}$ by Hölder's inequality. Thus, we obtain

$$\begin{aligned} I_3 &\ge (1-\theta\,\epsilon_0)\int_{\mathbb{S}^d}\left(|\nabla r_3|^2 + A\,r_3^2\right)d\mu - A\,\tfrac{2}{q}(1+2\,\epsilon_2\,\theta)\left(\int_{\mathbb{S}^d} r_3^q\,d\mu\right)^{2/q} \\ &\ge \frac{\theta}{q}\,(1-q\,\epsilon_0-4\,\epsilon_2)\int_{\mathbb{S}^d}\left(|\nabla r_3|^2 + A\,r_3^2\right)d\mu \ge 0\,, \end{aligned}$$

using $\theta = q-2 \le 1$ and Sobolev's inequality: $\|\nabla r_3\|_{\mathrm{L}^2(\mathbb{S}^d)}^2 + A\,\|r_3\|_{\mathrm{L}^2(\mathbb{S}^d)}^2 \ge A\,\|r_3\|_{\mathrm{L}^q(\mathbb{S}^d)}^2$. □

Remark 20. *The restriction $\epsilon_0 < \frac{1}{3}$ can be relaxed to $\epsilon_0 < \frac{1}{2}$ at the expense of having the inequality valid only in sufficiently high dimensions d, depending on ϵ_0. Indeed, ignoring the influence of ϵ_2 and σ_0 for the moment, the inequality at the end of the previous proof requires $1-\frac{q}{2}\,\epsilon_0 > 0$ and this is possible in all sufficiently high dimensions if and only if $\epsilon_0 < \frac{1}{2}$. Since this inequality is strict, the errors from ϵ_2 and σ_0 can then be accomodated as well.*

3.7. **Bound on I_2.** At this point in the proof, for given $0 < \epsilon_0 < \frac{1}{3}$, we have fixed the parameters ϵ_1 and ϵ_2 and we have found a δ_3 such that I_1, $I_3 \ge 0$ under the assumption $\tilde\delta \le \delta_3$. Here we show that, by further decreasing $\tilde\delta$, if necessary, we can ensure that $I_3 \ge 0$. The idea to achieve this is to use that r_2 satisfies an improved spectral gap inequality.

Proposition 21. *For any $0 < \epsilon_0 < \frac{1}{3}$, let $\sigma_0 = \frac{2}{q}\,\epsilon_2$. Then there is a $\delta_2 \in (0,1)$ such that, for all $d \ge 6$, all $\tilde\delta \le \delta_2$ and all r as in Theorem 4, one has*

$$I_2 \ge 0\,.$$

Proof. We first claim that for any L^2-normalized spherical harmonic Y of degree $k \in \mathbb{N}$, we have

$$\left|\int_{\mathbb{S}^d} Y\,r_2\,d\mu\right| \le 3^{\frac{k}{2}}\,\gamma^{-\frac{q}{4}}\,\tilde\delta^{\frac{q}{8}}\,\|r_2\|_{\mathrm{L}^2(\mathbb{S}^d)}\,. \tag{23}$$

Indeed, according to [33, Theorem 1], for any such spherical harmonic and any $p \in [2,\infty)$ we have

$$\|Y\|_{\mathrm{L}^p(\mathbb{S}^d)} \le (p-1)^{\frac{k}{2}}\,.$$

Thus, we can bound

$$\left|\int_{\mathbb{S}^d} Y\,r_2\,d\mu\right| \le \|Y\|_{\mathrm{L}^4(\mathbb{S}^d)}\,\mu\big(\{r_2>0\}\big)^{\frac14}\,\|r_2\|_{\mathrm{L}^2(\mathbb{S}^d)} \le 3^{\frac{k}{2}}\,\mu\big(\{r_2>0\}\big)^{\frac14}\,\|r_2\|_{\mathrm{L}^2(\mathbb{S}^d)}\,.$$

Meanwhile,

$$\mu\big(\{r_2>0\}\big) = \mu\big(\{r>\gamma\}\big) \le \frac{1}{\gamma^q}\,\|r\|_{\mathrm{L}^q(\mathbb{S}^d)}^q \le \frac{\tilde\delta^{q/2}}{\gamma^q}\,.$$

This leads to the claimed bound (23).

If $\pi_k\,r_2$ denotes the projection of r_2 onto spherical harmonics of degree k, from (23) to $Y = \pi_k\,r_2/\,\|\pi_k\,r_2\|_{\mathrm{L}^2(\mathbb{S}^d)}$, it follows that

$$\|\Pi_k\,r_2\|_{\mathrm{L}^2(\mathbb{S}^d)} \le 3^{\frac{k}{2}}\,\gamma^{-\frac{q}{4}}\,\tilde\delta^{\frac{q}{8}}\,\|r_2\|_{\mathrm{L}^2(\mathbb{S}^d)}\,.$$

Next, for any $K \in \mathbb{N}$, if $\Pi_K\,r_2 := \sum_{k<K}\pi_k\,r_2$ denotes the projection of r_2 onto spherical harmonics of degree less than K, then

$$\|\Pi_K\,r_2\|_{\mathrm{L}^2(\mathbb{S}^d)} = \left(\textstyle\sum_{k<K}\|\pi_k\,r_2\|_{\mathrm{L}^2(\mathbb{S}^d)}^2\right)^{1/2} \le \gamma^{-\frac{q}{4}}\,\tilde\delta^{\frac{q}{8}}\,\|r_2\|_{\mathrm{L}^2(\mathbb{S}^d)}\,\sqrt{\textstyle\sum_{k<K}3^k} \le 3^{\frac{K}{2}}\,\gamma^{-\frac{q}{4}}\,\tilde\delta^{\frac{q}{8}}\,\|r_2\|_{\mathrm{L}^2(\mathbb{S}^d)}\,.$$

From this we conclude that

$$\begin{aligned}\int_{\mathbb{S}^d} |\nabla r_2|^2 \, d\mu \geq \int_{\mathbb{S}^d} |\nabla(1-\Pi_K)\, r_2|^2 \, d\mu \\ \geq K\,(K+d-1) \int_{\mathbb{S}^d} |(1-\Pi_K)\, r_2|^2 \, d\mu \\ = K\,(K+d-1) \left(\|r_2\|^2_{\mathrm{L}^2(\mathbb{S}^d)} - \|\Pi_K\, r_2\|^2_{\mathrm{L}^2(\mathbb{S}^d)} \right) \\ \geq K\,(K+d-1) \left(1 - 3^K \, \gamma^{-\frac{q}{2}} \, \tilde{\delta}^{\frac{q}{4}} \right) \|r_2\|^2_{\mathrm{L}^2(\mathbb{S}^d)} \,.\end{aligned}$$

Consequently,

$$I_2 \geq \left((1-\theta\,\epsilon_0)\, K\,(K+d-1) \left(1 - 3^K \, \gamma^{-\frac{q}{2}} \, \tilde{\delta}^{\frac{q}{4}} \right) - d\,(1+\epsilon_0+\sigma_0+C_{\epsilon_1,\epsilon_2}) \right) \|r_2\|^2_{\mathrm{L}^2(\mathbb{S}^d)} \,.$$

We choose $K \in \mathbb{N}$ and $\delta_2 > 0$ such that

$$K := 1 + \left[2\, \frac{1+\epsilon_0+\sigma_0+C_{\epsilon_1,\epsilon_2}}{1-\epsilon_0} \right] \quad \text{and} \quad \delta_2 := \frac{1}{4}\, \frac{\gamma^2}{3^{2K}} \tag{24}$$

where $[x]$ denotes the integer part of $x \in \mathbb{R}$ and δ_3 is given by (22). From the definition of δ_2, if $\tilde{\delta} \leq \delta_2$, we have $1 - 3^K \, \gamma^{-\frac{q}{2}} \, \tilde{\delta}^{\frac{q}{4}} \geq \frac{1}{2}$ and conclude that $I_2 \geq 0$ because $K+d-1 \geq d$. □

3.8. **Proof of Theorem 4.** We assume that $d \geq 6$ and fix some $\epsilon_0 \in (0, 1/3)$. With the choice

$$\gamma = \epsilon_2 = 2\,\epsilon_1 = \tfrac{1}{4}\,(1-3\,\epsilon_0) \quad \text{and} \quad \sigma_0 = \frac{2}{q}\,\epsilon_2$$

according to (15), (16), and (22) on the one hand so that the assumptions of Corollary 17, Proposition 19 and Proposition 21 are fulfilled, and an arbitrary choice of

$$M \geq 2\,\gamma\,, \quad \overline{M} \geq \sqrt{e} \quad \text{and} \quad \epsilon = \gamma$$

which determines $C_{\epsilon_1,\epsilon_2} = C_{\gamma,\epsilon,M}$ according to (15) on the other hand, the condition

$$\tilde{\delta} = \min\left\{\delta_1, \delta_2\right\}$$

with δ_1 and δ_2 given by (21) and (24), we claim that I_1, I_2 and I_3 are nonnegative, which completes the proof of Theorem 4 for $q \leq 3$, that is $d \geq 6$. The assertion for $d = 3$, 4, 5 follows from the result proved in Subsection 3.2. □

4. Functions away from the manifold of optimizers

Our goal in this section is to prove a stability inequality for nonnegative functions that are, in a certain sense, 'far' away from the manifold of optimizers. Let us introduce

$$\mathscr{I}(\delta) := \inf \left\{ \mathcal{E}(f) \,:\, 0 \leq f \in \dot{\mathrm{H}}^1(\mathbb{R}^d) \setminus \mathcal{M}\,,\, \inf_{g \in \mathcal{M}} \|\nabla f - \nabla g\|_2^2 \leq \delta\, \|\nabla f\|_2^2 \right\}. \tag{25}$$

Theorem 22. *Let $\delta \in (0,1)$ and assume that $0 \leq f \in \dot{\mathrm{H}}^1(\mathbb{R}^d) \setminus \mathcal{M}$ satisfies*

$$\inf_{g \in \mathcal{M}} \|\nabla f - \nabla g\|_2^2 \geq \delta\, \|\nabla f\|_2^2 \,.$$

Then, with $\mathscr{I}(\delta)$ defined by (25), we have

$$\mathcal{E}(f) \geq \delta\, \mathscr{I}(\delta)\,.$$

We will prove this theorem by symmetrization. First, we will use a discrete symmetrization procedure to get somewhat close to the manifold, then we will use a further continuous symmetrization procedure to fine tune the distance to the manifold.

4.1. **Competing symmetries.** The functional $\mathcal{E}(f)$ is conformally invariant in the sense that if $C : \mathbb{R}^d \cup \{\infty\} \to \mathbb{R}^d \cup \{\infty\}$ is a conformal map, the function

$$f_C(x) = |\det DC(x)|^{1/2^*} f\big(C(x)\big)$$

satisfies

$$\mathcal{E}(f_C) = \mathcal{E}(f)\,.$$

In order to verify this, we recall that any conformal map is a composition of scalings, translations, rotations and inversions. For scalings, translations and rotations in $\mathbb{R}^d$ the claimed invariance is easy to see. The additional map to consider is the inversion $I(x) = \frac{x}{|x|^2}$ and a straightforward change of variables shows that

$$\|\nabla f_I\|_2^2 = \|\nabla f\|_2^2\,, \quad \|f_I\|_{2^*}^2 = \|f\|_{2^*}^2\,.$$

The equality

$$\inf_{g\in\mathcal{M}} \|\nabla(f_I - g)\|_2^2 = \inf_{g\in\mathcal{M}} \|\nabla f - \nabla g\|_2^2$$

follows from

$$\inf_{g\in\mathcal{M}} \|\nabla(f_I - g)\|_2^2 = \inf_{g\in\mathcal{M}} \|\nabla(f - g_I)\|_2^2 = \inf_{g\in\mathcal{M}} \|\nabla f - \nabla g\|_2^2$$

since $I^2 = I$ and $g \to g_I$ maps the set $\mathcal{M}$ to itself in a one-to-one and onto fashion.

Another and perhaps easier way to see the conformal invariance is to pull the problem up to the sphere via the stereographic projection, as discussed in Section 2. On the sphere the inversion I takes the form of the reflection $(s_1, \ldots, s_d, s_{d+1}) \to (s_1, \ldots, s_d, -s_{d+1})$, which clearly leaves the functional on the sphere unchanged.

A second ingredient for the construction of the discrete symmetrization flow is the technique of 'competing symmetries', invented in [22]. Consider any nonnegative function $f \in \dot{\mathrm{H}}^1(\mathbb{R}^d)$ and its counterpart $F \in \mathrm{H}^1(\mathbb{S}^d)$ given by (6). Set

$$(UF)(\omega) = F(\omega_1, \omega_2, \ldots, \omega_{d+1}, -\omega_d)\,,$$

which corresponds to a rotation by $\pi/2$ that maps the 'north pole' axis $(0, 0, \ldots, 1)$ to $(0, \ldots, 1, 0)$. Reversing (6) the function on $\mathbb{R}^d$ that corresponds to UF is given by

$$(Uf)(x) = \left(\frac{2}{|x - e_d|^2}\right)^{\frac{d-2}{2}} f\left(\frac{x_1}{|x - e_d|^2}, \ldots, \frac{x_{d-1}}{|x - e_d|^2}, \frac{|x|^2 - 1}{|x - e_d|^2}\right), \tag{26}$$

where $e_d = (0, \ldots, 0, 1) \in \mathbb{R}^d$. It follows that

$$\mathcal{E}(Uf) = \mathcal{E}(f)\,.$$

The operation U is obviously linear, invertible and an isometry on $\mathrm{L}^{2^*}(\mathbb{R}^d)$.

We also consider the symmetric decreasing rearrangement

$$\mathcal{R}f(x) = f^*(x)\,.$$

The most important properties are that f and f^* are equimeasurable and that $\|\nabla f^*\|_2 \le \|\nabla f\|_2$. For elementary properties of rearrangements the reader may consult [54]. Being equimeasurable, this map is also an isometry on $\mathrm{L}^{2^*}(\mathbb{R}^d)$. It is when using the decreasing rearrangement that we use the fact that f is a nonnegative function. For functions that change sign one conventionally defines their rearrangement as the rearrangement of their absolute value. Passing from a function to its absolute value does not alter the numerator of $\mathcal{E}(f)$ but may decrease the denominator so that other arguments are needed.

On $\mathbb{R}^d$, let

$$g_*(x) := |\mathbb{S}^d|^{-\frac{d-2}{2\,d}} \left(\frac{2}{1+|x|^2}\right)^{\frac{d-2}{2}} . \tag{27}$$

Note that $\|g_*\|_{2^*} = 1$ because it is obtained as the stereographic projection of the constant function on $\mathbb{S}^d$ with 2^*-norm equal to 1. The following theorem was proved in [22].

Theorem 23. *Let $f \in \mathrm{L}^{2^*}(\mathbb{R}^d)$ be a nonnegative function. Consider the sequence $(f_n)_{n\in\mathbb{N}}$ of functions*

$$f_n = (\mathcal{R}U)^n f \quad \forall\, n \in \mathbb{N}\,. \tag{28}$$

Then

$$\lim_{n\to\infty} \|f_n - h_f\|_{2^*} = 0$$

where $h_f = \|f\|_{2^*}\, g_* \in \mathcal{M}$. Moreover, if $f \in \dot{\mathrm{H}}^1(\mathbb{R}^d)$, then $(\|\nabla f_n\|_2^2)_{n\in\mathbb{N}}$ is a nonincreasing sequence.

It does not seem clear whether the functional $\mathcal{E}(f)$ decreases or increases under rearrangement. The next lemma helps to explain this point. Define $\mathcal{M}_1$ to be the set of the elements in $\mathcal{M}$ with 2^*-norm equal to 1.

Lemma 24. *For any $f \in \dot{\mathrm{H}}^1(\mathbb{R}^d)$, we have*

$$\inf_{g\in\mathcal{M}} \|\nabla f - \nabla g\|_2^2 = \|\nabla f\|_2^2 - S_d \sup_{g\in\mathcal{M}_1} \left(f, g^{2^*-1}\right)^2 .$$

Here $(\cdot,\cdot)$ is the $\mathrm{L}^2(\mathbb{R}^d)$ inner product or, more precisely, the duality pairing between $\mathrm{L}^{2^*}(\mathbb{R}^d)$ and $\mathrm{L}^{(2^*)'}(\mathbb{R}^d)$.

Proof. Let g be any Aubin–Talenti function with $\|g\|_{2^*} = 1$. The function g is an optimizer of the Sobolev inequality, i.e., $\|\nabla g\|_2^2 = S_d\,\|g\|_{2^*}^2 = S_d$ and is a solution of the Sobolev equation

$$-\Delta g = S_d\,\frac{g^{2^*-1}}{\|g\|_{2^*}^{2^*-2}} = S_d\, g^{2^*-1}\,.$$

Hence for any nonnegative constant c we find

$$\|\nabla(f - c\,g)\|_2^2 = \|\nabla f\|_2^2 - 2\,c\,(\nabla f, \nabla g) + c^2\,\|\nabla g\|_2 = \|\nabla f\|_2^2 - 2\,c\,S_d\,\left(f, g^{2^*-1}\right) + S_d\, c^2$$

and minimizing with respect to c we find the lower bound $\|\nabla f\|_2^2 - S_d\,\left(f, g^{2^*-1}\right)^2$, which proves the lemma. □

Under the decreasing rearrangement, the term $\|\nabla f\|_2^2$ does not increase whereas the term $\sup_{g\in\mathcal{M}_1}\left(f, g^{2^*-1}\right)^2$ increases. To see this, note that the supremum is attained at some Aubin–Talenti function of the form (1), which is a strictly symmetric decreasing function about some point $b \in \mathbb{R}^d$. Replacing f by its symmetric decreasing rearrangement about that point increases $\left(f, g^{2^*-1}\right)^2$, in fact strictly unless f is already symmetric decreasing about the point b. Thus, while the numerator in $\mathcal{E}(f)$ decreases under rearrangements so does the denominator and there are no direct conclusions to be drawn from this. The next lemma summarizes what we have shown.

Lemma 25. *For the sequence $(f_n)_{n\in\mathbb{N}}$ in Theorem 23 we have that $n\mapsto \sup_{g\in\mathcal{M}_1}\left(f_n, g^{2^*-1}\right)^2$ is strictly increasing, $n\mapsto \inf_{g\in\mathcal{M}}\|\nabla f_n-\nabla g\|_{2^*}^2$ is strictly decreasing and*

$$\lim_{n\to\infty}\inf_{g\in\mathcal{M}}\|\nabla f_n-\nabla g\|_2^2=\lim_{n\to\infty}\|\nabla f_n\|_2^2-S_d\,\|h_f\|_{2^*}^2=\lim_{n\to\infty}\|\nabla f_n\|_2^2-S_d\,\|f\|_{2^*}^2\,.$$

Proof. From

$$\inf_{g\in\mathcal{M}}\|\nabla f_n-\nabla g\|_2^2=\|\nabla f_n\|_2^2-S_d\sup_{g\in\mathcal{M}_1}\left(f_n, g^{2^*-1}\right)^2$$

we see that the first term converges since $(\|\nabla f_n\|_2^2)_{n\in\mathbb{N}}$ is a nonincreasing sequence. For the second term, which is strictly increasing, we have by Hölder's inequality

$$\sup_{g\in\mathcal{M}_1}\left(f_n, g^{2^*-1}\right)^2\le\|f_n\|_{2^*}^2=\|f\|_{2^*}^2$$

and since g_* as defined in (27) is in $\mathcal{M}_1$ we have

$$\liminf_{n\to\infty}\sup_{g\in\mathcal{M}_1}\left(f_n, g^{2^*-1}\right)^2\ge\liminf_{n\to\infty}\left(f_n, g_*\right)^2=\|f\|_{2^*}^2$$

by Theorem 23. □

Lemma 26. *Assume that $0\le f\in\dot{\mathrm{H}}^1(\mathbb{R}^d)\setminus\mathcal{M}$ satisfies*

$$\inf_{g\in\mathcal{M}}\|\nabla f-\nabla g\|_2^2\ge\delta\,\|\nabla f\|_2^2$$

and let $(f_n)_{n\in\mathbb{N}}$ be the sequence defined by (28). Then one of the following alternatives holds:

(a) *for all $n=0,1,2\ldots$ we have*

$$\inf_{g\in\mathcal{M}}\|\nabla f_n-\nabla g\|_2^2\ge\delta\,\|\nabla f_n\|_2^2$$

(b) *there is a natural number n_0 such that*

$$\inf_{g\in\mathcal{M}}\|\nabla f_{n_0}-\nabla g\|_2^2\ge\delta\,\|\nabla f_{n_0}\|_2^2$$

and

$$\inf_{g\in\mathcal{M}}\|\nabla f_{n_0+1}-\nabla g\|_2^2<\delta\,\|\nabla f_{n_0+1}\|_2^2\,.$$

Proof. Assume that alternative (a) does not hold. Then there is a largest value $n_0\ge0$ such that $\inf_{g\in\mathcal{M}}\|\nabla f_{n_0}-\nabla g\|_2^2\ge\delta\,\|\nabla f_{n_0}\|_2^2$. □

Lemma 27. *Assume that $0\le f\in\dot{\mathrm{H}}^1(\mathbb{R}^d)\setminus\mathcal{M}$ satisfies*

$$\inf_{g\in\mathcal{M}}\|\nabla f-\nabla g\|_2^2\ge\delta\,\|\nabla f\|_2^2$$

and suppose that in Lemma 26 alternative (a) *holds for the sequence $(f_n)_{n\in\mathbb{N}}$ defined by* (28)*. Then*

$$\mathcal{E}(f)\ge\delta\,.$$

Proof. We have

$$\mathcal{E}(f)=\frac{\|\nabla f\|_2^2-S_d\,\|f\|_{2^*}^2}{\inf_{g\in\mathcal{M}}\|\nabla f-\nabla g\|_2^2}\ge\frac{\|\nabla f\|_2^2-S_d\,\|f\|_{2^*}^2}{\|\nabla f\|_2^2}\ge\frac{\|\nabla f_n\|_2^2-S_d\,\|f\|_{2^*}^2}{\|\nabla f_n\|_2^2}\,,\tag{29}$$

where the second inequality is a consequence of $\|\nabla f_n\|_2^2 \leq \|\nabla f\|_2^2$ for all $n = 0, 1, 2, \ldots$ proved in Theorem 23. By the assumption that alternative (a) holds and by Lemma 25, we learn that

$$\lim_{n\to\infty} \|\nabla f_n\|_2^2 \leq \frac{1}{\delta} \lim_{n\to\infty} \inf_{g\in\mathcal{M}} \|\nabla f_n - \nabla g\|_2^2 = \frac{1}{\delta}\left(\lim_{n\to\infty} \|\nabla f_n\|_2^2 - S_d \|f\|_{2^*}^2\right).$$

Since

$$\lim_{n\to\infty} \|\nabla f_n\|_2^2 - S_d \|f\|_{2^*}^2 \geq \delta \lim_{n\to\infty} \|\nabla f_n\|_2^2 \geq \delta\, S_d \lim_{n\to\infty} \|f_n\|_{2^*}^2 = \delta\, S_d \|f\|_{2^*}^2 > 0\,,$$

we can take the limit as $n \to \infty$ on the right side of (29) and compute the limit of the quotient as the quotient of the limits. This proves the lemma. □

4.2. **Continuous rearrangement.** Next, we analyze the case where the alternative (b) in Lemma 26 holds. We recall that $\mathscr{I}(\delta)$ was defined in (25).

Lemma 28. *For any $\delta \in (0, 1]$, we have $\mathscr{I}(\delta) \leq 1$.*

Proof. By Lemma 24, we have

$$\inf_{g\in\mathcal{M}} \|\nabla f - \nabla g\|_2^2 = \|\nabla f\|_2^2 - S_d \sup_{g\in\mathcal{M}_1} \left(f, g^{2^*-1}\right)^2$$

and it follows from Hölder's inequality that

$$\sup_{g\in\mathcal{M}_1} \left(f, g^{2^*-1}\right)^2 \leq \|f\|_{2^*}^2\,.$$

Thus, the denominator in $\mathcal{E}(f)$ that enters the definition of $\mathscr{I}(\delta)$ is at least as large as the numerator, so the quotient is at most 1. □

Our goal in this subsection is to prove the following lower bound on $\mathcal{E}(f)$.

Lemma 29. *Assume that $0 \leq f \in \dot{\mathrm{H}}^1(\mathbb{R}^d) \setminus \mathcal{M}$ satisfies*

$$\inf_{g\in\mathcal{M}} \|\nabla f - \nabla g\|_2^2 \geq \delta \|\nabla f\|_2^2$$

for some $\delta \in (0, 1)$ and suppose that in Lemma 26 alternative (b) *holds for the sequence $(f_n)_{n\in\mathbb{N}}$ of Theorem 23 defined by* (28)*. Then, with $\mathscr{I}(\delta)$ defined by* (25)*, we have*

$$\mathcal{E}(f) \geq \delta\, \mathscr{I}(\delta)\,.$$

For the proof of this lemma we introduce a continuous rearrangement flow that interpolates between a function and its symmetric decreasing rearrangement. The basic ingredient for this flow is similar to a flow that Brock introduced [14, 15] and that interpolates between a function and its Steiner symmetrization with respect to a given hyperplane. Brock's construction, in turn, is based on ideas of Rogers [59] and Brascamp–Lieb–Luttinger [11]. Our flow is obtained by glueing together infinitely many copies of Brock's flows with respect to a sequence of judiciously chosen hyperplanes. A similar construction was performed by Bucur and Henrot [16]; see also [25].

More specifically, for a given hyperplane H, Brock's flow interpolates between a given function f and f^{*H}, the Steiner symmetrized function with respect to H. The family that interpolates between f and f^{*H} is denoted by $f_\tau^H, \tau \in [0, \infty]$, and we have

$$f_0 = f\,, \quad f_\infty^H = f^{*H}\,.$$

Further, for any τ, f_τ^H and f are equimeasurable, i.e.,

$$\left|\left\{x \in \mathbb{R}^d : f_\tau^H(x) > t\right\}\right| = \left|\left\{x \in \mathbb{R}^d : f(x) > t\right\}\right| \quad \forall\, t > 0\,.$$

Moreover, if $f \in \mathrm{L}^p(\mathbb{R}^d)$ for some $1 \le p < \infty$, then $\tau \mapsto f_\tau^H$ is continuous in $\mathrm{L}^p(\mathbb{R}^d)$.

By choosing a sequence of hyperplanes we construct another flow $\tau \mapsto f_\tau$ that has the same properties but interpolates between f and f^*, the symmetric decreasing rearrangement. In Appendix 7 we explain this in more detail and prove the following properties that are important for our proof, assuming $f \in \dot{\mathrm{H}}^1(\mathbb{R}^d)$. From the $\mathrm{L}^{2^*}(\mathbb{R}^d)$ continuity of the flow we will deduce that

$$\lim_{\tau \to \tau_0} \sup_{g \in \mathcal{M}_1} (f_\tau, g)^2 = \sup_{g \in \mathcal{M}_1} (f_{\tau_0}, g)^2 \,. \tag{30}$$

Concerning the gradient we prove the monotonicity

$$\|\nabla f_{\tau_2}\|_2 \le \|\nabla f_{\tau_1}\|_2 \,, \quad 0 \le \tau_1 \le \tau_2 \le \infty \,,$$

and the right continuity

$$\lim_{\tau_2 \to \tau_1^+} \|\nabla f_{\tau_2}\|_2 = \|\nabla f_{\tau_1}\|_2 \,, \quad 0 \le \tau_1 < \infty \,. \tag{31}$$

Proof of Lemma 29. We begin by motivating and explaining the strategy of the proof. As before, we bound

$$\mathcal{E}(f) = \frac{\|\nabla f\|_2^2 - S_d \, \|f\|_{2^*}^2}{\inf_{g \in \mathcal{M}} \|\nabla f - \nabla g\|_2^2} \ge \frac{\|\nabla f\|_2^2 - S_d \, \|f\|_{2^*}^2}{\|\nabla f\|_2^2} \ge \frac{\|\nabla f_{n_0}\|_2^2 - S_d \, \|f_{n_0}\|_{2^*}^2}{\|\nabla f_{n_0}\|_2^2} \,. \tag{32}$$

We could bound the right side further from below by replacing f_{n_0} by f_{n_0+1}. This bound, however, might be too crude for our purposes and we proceed differently. The move from f_{n_0} to f_{n_0+1} consists of two steps, namely first applying a conformal rotation and second applying symmetric decreasing rearrangement. The first step leaves all terms on the right side invariant and we do carry out this step. The second step leaves the 2^*-norm invariant, while the gradient term does not go up. In fact, the gradient term might go down too far. Therefore, we replace the application of the rearrangement by a continuous rearrangement flow. In order to make the notation less cumbersome we shall denote $U f_{n_0}$ by f_0 where U denotes the conformal rotation (26). We denote by f_τ, $0 \le \tau \le \infty$, the continuous rearrangement starting at f_0 and let

$$\mathsf{f}_\infty = f_{n_0+1} \,. \tag{33}$$

Ideally, we would like to find $\tau_0 \in [0, \infty)$ such that

$$\inf_{g \in \mathcal{M}} \|\nabla \mathsf{f}_{\tau_0} - \nabla g\|_2^2 = \delta \, \|\nabla \mathsf{f}_{\tau_0}\|_2^2 \,.$$

Then the right side of (32) is equal to

$$1 - S_d \, \frac{\|\mathsf{f}_0\|_{2^*}^2}{\|\nabla \mathsf{f}_0\|_2^2} \ge 1 - S_d \, \frac{\|\mathsf{f}_{\tau_0}\|_{2^*}^2}{\|\nabla \mathsf{f}_{\tau_0}\|_2^2} = \delta \, \frac{\|\nabla \mathsf{f}_{\tau_0}\|_2^2 - S_d \, \|\mathsf{f}_{\tau_0}\|_{2^*}^2}{\inf_{g \in \mathcal{M}} \|\nabla \mathsf{f}_{\tau_0} - \nabla g\|_2^2} \,,$$

which can be bounded from below by $\delta \, \mathscr{I}(\delta)$, since f_{τ_0} is admissible in the infimum (25). This would prove the desired bound.

The problem with this argument is that the existence of such a $\tau_0 \in [0, \infty)$ is in general not clear, since neither of the terms $\inf_{g \in \mathcal{M}} \|\nabla \mathsf{f}_\tau - \nabla g\|_2^2$ and $\|\nabla \mathsf{f}_\tau\|_2^2$ needs to be continuous in τ. Nevertheless, we will be able to adapt the above argument to yield the same conclusion.

We now turn to the details of the argument. Recalling that

$$\inf_{g \in \mathcal{M}} \|\nabla \mathsf{f}_0 - \nabla g\|_2^2 \ge \delta \, \|\nabla \mathsf{f}_0\|_2^2 \,,$$

we define

$$\tau_0 := \inf\left\{\tau \geq 0 : \inf_{g\in\mathcal{M}} \|\nabla \mathsf{f}_\tau - \nabla g\|_2^2 < \delta \,\|\nabla \mathsf{f}_\tau\|_2^2\right\}$$

with the convention that $\inf\emptyset = \infty$. If $\tau < \tau_0 \in (0,\infty]$, similarly as before, the right side of (32) is equal to

$$\frac{\|\nabla \mathsf{f}_0\|_2^2 - S_d\,\|\mathsf{f}_0\|_{2^*}^2}{\|\nabla \mathsf{f}_0\|_2^2} = 1 - S_d\,\frac{\|\mathsf{f}_0\|_{2^*}^2}{\|\nabla \mathsf{f}_0\|_2^2} \geq \frac{\|\nabla \mathsf{f}_\tau\|_2^2 - S_d\,\|\mathsf{f}_{\tau_0}\|_{2^*}^2}{\|\nabla \mathsf{f}_\tau\|_2^2} \geq \delta\,\frac{\|\nabla \mathsf{f}_\tau\|_2^2 - S_d\,\|\mathsf{f}_{\tau_0}\|_{2^*}^2}{\inf_{g\in\mathcal{M}}\|\nabla \mathsf{f}_\tau - \nabla g\|_2^2}\,,$$

where the last inequality arises from $\inf_{g\in\mathcal{M}}\|\nabla \mathsf{f}_\tau - \nabla g\|_2^2 \geq \delta\,\|\nabla \mathsf{f}_\tau\|_2^2$ for any $\tau\in[0,\tau_0)$. Taking the limit inferior as $\tau\to\tau_0^-$, we obtain

$$\frac{\|\nabla \mathsf{f}_0\|_2^2 - S_d\,\|\mathsf{f}_0\|_{2^*}^2}{\|\nabla \mathsf{f}_0\|_2^2} \geq \delta\,\frac{\lim_{\tau\to\tau_0^-}\|\nabla \mathsf{f}_\tau\|_2^2 - S_d\,\|\mathsf{f}_{\tau_0}\|_{2^*}^2}{\liminf_{\tau\to\tau_0^-}\inf_{g\in\mathcal{M}}\|\nabla \mathsf{f}_\tau - \nabla g\|_2^2}\,. \tag{34}$$

Note that the denominator appearing here does not vanish. Indeed, we have

$$\inf_{g\in\mathcal{M}}\|\nabla \mathsf{f}_\tau - \nabla g\|_2^2 \geq \delta\,\|\nabla \mathsf{f}_\tau\|_2^2 \geq \delta\,S_d\,\|\mathsf{f}_\tau\|_{2^*}^2 = \delta\,S_d\,\|f\|_{2^*}^2 > 0 \quad \forall\,\tau\in[0,\tau_0)$$

and, as a consequence,

$$\liminf_{\tau\to\tau_0^-}\inf_{g\in\mathcal{M}}\|\nabla \mathsf{f}_\tau - \nabla g\|_2^2 \geq \delta\,S_d\,\|f\|_{2^*}^2 > 0\,.$$

The same inequality (34) remains valid if $\tau_0 = 0$ and if we interpret $\lim_{\tau\to\tau_0^-}$ and $\liminf_{\tau\to\tau_0^-}$ as evaluating at $\tau_0 = 0$.

At this point we find it convenient to apply Lemma 24 and use the representation

$$\inf_{g\in\mathcal{M}}\|\nabla \mathsf{f}_\tau - \nabla g\|_2^2 = \|\nabla \mathsf{f}_\tau\|_2^2 - S_d\,\sup_{g\in\mathcal{M}_1}\left(\mathsf{f}_\tau, g^{2^*-1}\right)^2\,.$$

Using (30), that is, the continuity of $\tau\mapsto\sup_{g\in\mathcal{M}_1}\left(\mathsf{f}_\tau, g^{2^*-1}\right)^2$, we see that

$$\liminf_{\tau\to\tau_0^-}\inf_{g\in\mathcal{M}}\|\nabla \mathsf{f}_\tau - \nabla g\|_2^2 = \lim_{\tau\to\tau_0^-}\|\nabla \mathsf{f}_\tau\|_2^2 - S_d\,\sup_{g\in\mathcal{M}_1}\left(\mathsf{f}_{\tau_0}, g^{2^*-1}\right)^2\,.$$

Thus, the relevant quotient is equal to

$$\frac{\lim_{\tau\to\tau_0^-}\|\nabla \mathsf{f}_\tau\|_2^2 - S_d\,\|\mathsf{f}_{\tau_0}\|_{2^*}^2}{\lim_{\tau\to\tau_0^-}\|\nabla \mathsf{f}_\tau\|_2^2 - S_d\,\sup_{g\in\mathcal{M}_1}\left(\mathsf{f}_{\tau_0}, g^{2^*-1}\right)^2}\,. \tag{35}$$

Our goal in the remainder of this proof is to show that this quotient is larger or equal than $\mathscr{I}(\delta)$. We will use the fact that

$$\sup_{g\in\mathcal{M}_1}\left(\mathsf{f}_{\tau_0}, g^{2^*-1}\right)^2 \leq \|f_{\tau_0}\|_{2^*}^2\,, \tag{36}$$

which follows from Hölder's inequality. We also note that equality holds here if and only if $\mathsf{f}_{\tau_0}\in\mathcal{M}$.

Let us first handle the case where $\mathsf{f}_{\tau_0}\in\mathcal{M}$. Then by (4.2) and because of equality in (36), the quotient (35) is equal to 1, which by Lemma 28 can be further bounded from below by $\mathscr{I}(\delta)$, leading to the claimed bound. This completes the proof in the case $\mathsf{f}_{\tau_0}\in\mathcal{M}$ and in what follows we assume

$$\mathsf{f}_{\tau_0}\notin\mathcal{M}\,.$$

As a consequence of this assumption and (36), we have

$$\|\nabla \mathsf{f}_{\tau_0}\|_2^2 > S_d \, \|\mathsf{f}_{\tau_0}\|_{2^*}^2 \geq S_d \sup_{g\in\mathcal{M}_1} \left(\mathsf{f}_{\tau_0}, g^{2^*-1}\right)^2 . \tag{37}$$

Next, we observe that for $\alpha > \beta$ the function $x \mapsto (x-\alpha)/(x-\beta)$ is monotone increasing on the interval (β,∞). This, together with the strict inequality in (37), implies that the quotient (35) can be bounded from below by

$$\frac{\lim_{\tau\to\tau_0^-} \|\nabla \mathsf{f}_\tau\|_2^2 - S_d \, \|\mathsf{f}_{\tau_0}\|_{2^*}^2}{\lim_{\tau\to\tau_0^-} \|\nabla \mathsf{f}_\tau\|_2^2 - S_d \sup_{g\in\mathcal{M}_1} \left(\mathsf{f}_{\tau_0}, g^{2^*-1}\right)^2} \geq \frac{\|\nabla \mathsf{f}_{\tau_0}\|_2^2 - S_d \, \|\mathsf{f}_{\tau_0}\|_{2^*}^2}{\|\nabla \mathsf{f}_{\tau_0}\|_2^2 - S_d \sup_{g\in\mathcal{M}_1} \left(\mathsf{f}_{\tau_0}, g^{2^*-1}\right)^2} . \tag{38}$$

We now claim that

$$\inf_{g\in\mathcal{M}} \|\nabla \mathsf{f}_{\tau_0} - \nabla g\|_2^2 \leq \delta \, \|\nabla \mathsf{f}_{\tau_0}\|_2^2 . \tag{39}$$

Once this is proved, we can bound the right side of (38) from below by $\mathscr{I}(\delta)$. This inequality is the claimed inequality after taking into account (34).

To prove (39), we first note that it is verified if $\tau_0 = \infty$. Indeed, $\mathsf{f}_\infty = f_{n_0+1}$ by (33) and therefore, by assumption of alternative (b), $\inf_{g\in\mathcal{M}} \|\nabla \mathsf{f}_\infty - \nabla g\|_2^2 < \delta \, \|\nabla \mathsf{f}_\infty\|_2^2$.

Now let $\tau_0 < \infty$. We argue by contradiction and assume that

$$\inf_{g\in\mathcal{M}} \|\nabla \mathsf{f}_{\tau_0} - \nabla g\|_2^2 > \delta \, \|\nabla \mathsf{f}_{\tau_0}\|_2^2 . \tag{40}$$

Because of this strict inequality and the definition of τ_0 there are $\sigma_k \in (\tau_0,\infty)$ for any $k\in\mathbb{N}$ with $\lim_{k\to\infty} \sigma_k = \tau_0$ such that $\inf_{g\in\mathcal{M}} \|\nabla \mathsf{f}_{\sigma_k} - \nabla g\|_2^2 < \delta \, \|\nabla \mathsf{f}_{\sigma_k}\|_2^2$, that is,

$$\|\nabla \mathsf{f}_{\sigma_k}\|_2^2 - S_d \sup_{g\in\mathcal{M}_1} \left(\mathsf{f}_{\sigma_k}, g^{2^*-1}\right)^2 < \delta \, \|\nabla \mathsf{f}_{\sigma_k}\|_2^2 \quad \forall\, k\in\mathbb{N} .$$

Letting $k\to\infty$ and using (30) as well as the right continuity of $\|\nabla \mathsf{f}_\tau\|_2^2$, see (31), we deduce that

$$\|\nabla \mathsf{f}_{\tau_0}\|_2^2 - S_d \sup_{g\in\mathcal{M}_1} \left(\mathsf{f}_{\tau_0}, g^{2^*-1}\right)^2 \leq \delta \, \|\nabla \mathsf{f}_{\tau_0}\|_2^2 .$$

This is the same as $\inf_{g\in\mathcal{M}} \|\nabla \mathsf{f}_{\tau_0} - \nabla g\|_2^2 \leq \delta \, \|\nabla \mathsf{f}_{\tau_0}\|_2^2$ and contradicts (40). This proves (39) and completes the proof of the lemma. □

Remark 30. *The above argument would be simpler if $\tau \mapsto \|\nabla \mathsf{f}_\tau\|_2^2$ were continuous for an appropriate choice of hyperplanes (see Appendix 7) in the definition of the flow. Since the flow is weakly continuous in $\dot{\mathrm{H}}^1(\mathbb{R}^d)$, continuity of the norm is equivalent to (strong) continuity of the flow in $\dot{\mathrm{H}}^1(\mathbb{R}^d)$. Thus, for continuity of the norm for an appropriate choice of hyperplanes, it is necessary that there is such a choice for which the Steiner symmetrizations approximate f^* in $\dot{\mathrm{H}}^1(\mathbb{R}^d)$. According to a theorem of Burchard [17] this holds if and only if f is co-area regular, i.e, if and only if the distribution function*

$$h \mapsto |\{x\in\mathbb{R}^d : f(x) > h,\ \nabla f(x) = 0\}|$$

has no absolutely continuous component. As shown by Almgren and Lieb [1], both co-area regular and co-area irregular functions are dense for $d \geq 2$.

4.3. **Proof of Theorem 22.** It is now easy to prove the main result of this section, Theorem 22. Let $\delta \in (0,1)$ and assume that $0 \le f \in \dot{\mathrm{H}}^1(\mathbb{R}^d) \setminus \mathcal{M}$ satisfies

$$\inf_{g \in \mathcal{M}} \|\nabla f - \nabla g\|_2^2 \ge \delta \, \|\nabla f\|_2^2 \,.$$

By Lemma 26 either alternative (a) or (b) holds. In the first case, we apply Lemmas 27 and 28, and in the second case, we apply Lemma 29. This completes the proof. □

5. From nonnegative functions to arbitrary functions

We recall that c_{BE} denotes the optimal constant in (2). Similarly, we denote by $c_{\mathrm{BE}}^{\mathrm{pos}}$ the optimal constant in (2) when restricted to nonnegative functions f. Thus, $c_{\mathrm{BE}}^{\mathrm{pos}} \ge c_{\mathrm{BE}}$. We do not know whether these two constants coincide or not. The main result in this section will be to prove the following lower bound on c_{BE} in terms of $c_{\mathrm{BE}}^{\mathrm{pos}}$.

Proposition 31. *For any $d \ge 3$,*

$$c_{\mathrm{BE}} \ge \min\left\{ \tfrac{1}{2}\, c_{\mathrm{BE}}^{\mathrm{pos}}, 1 - 2^{-\frac{2}{d}} \right\}.$$

Proof. To simplify the notation, given a function $v \in \dot{\mathrm{H}}^1(\mathbb{R}^d)$, we define the deficit

$$\mathrm{D}(v) := \|\nabla v\|_{\mathrm{L}^2(\mathbb{R}^d)}^2 - S_d \, \|v\|_{\mathrm{L}^{2^*}(\mathbb{R}^d)}^2 \,.$$

Also, we set $\alpha_d := \frac{2}{2^*} = 1 - \frac{2}{d} < 1$,

$$h(p) := p^{\alpha_d} + (1-p)^{\alpha_d} - 1 \,, \quad \text{and} \quad h_d := h(\tfrac{1}{2}) = 2^{1-\alpha_d} - 1 = 2^{\frac{2}{d}} - 1 \,.$$

Let us consider a function $u \in \dot{\mathrm{H}}^1(\mathbb{R}^d)$. By homogeneity we can assume that $\|u\|_{\mathrm{L}^{2^*}(\mathbb{R}^d)} = 1$. Let $u_\pm$ denote the positive and negative parts of u, set

$$m := \|u_-\|_{\mathrm{L}^{2^*}(\mathbb{R}^d)}^{2^*} \,,$$

and assume (without loss of generality) that

$$m \in [0, 1/2] \,. \tag{41}$$

Note that $\|u_+\|_{\mathrm{L}^{2^*}(\mathbb{R}^d)}^{2^*} = 1 - m$ and $\|\nabla u\|_{\mathrm{L}^2(\mathbb{R}^d)}^2 = \|\nabla u_-\|_{\mathrm{L}^2(\mathbb{R}^d)}^2 + \|\nabla u_+\|_{\mathrm{L}^2(\mathbb{R}^d)}^2$. Hence, we have

$$\mathrm{D}(u) = \|\nabla u\|_{\mathrm{L}^2(\mathbb{R}^d)}^2 - S_d = \mathrm{D}(u_+) + \mathrm{D}(u_-) + S_d \, h(m). \tag{42}$$

Since the function $p \mapsto h(p)$ is monotone increasing and concave on $[0, 1/2]$, we have

$$2\, h_d \, p \le h(p) \,. \tag{43}$$

Also, if we set $\xi_d := 2\,(1 - 2^{-\alpha_d})$, the function $f(p) := (1-p)^{\alpha_d} - 1 + \xi_d \, p$ satisfies $f(0) = f(1/2) = 0$ and $f''(p) \le 0$, so that $f(p) \ge 0$ for all $m \in [0, 1/2]$. Hence, by (41), we have

$$(1-p)^{\alpha_d} \ge 1 - \xi_d \, p \,,$$

which, by the definition of $h(p)$, yields

$$h(p) \ge p^{\alpha_d} - \xi_d \, p \,.$$

Combining this bound with (43), this gives

$$\left(1 + \frac{\xi_d}{2\, h_d}\right) h(p) \ge p^{\alpha_d} \,.$$

Therefore, recalling (42) and noticing that $\mathrm{D}(u_-) + S_d\, m^{\alpha_d} = \|\nabla u_-\|^2_{\mathrm{L}^2(\mathbb{R}^d)}$, we get

$$\mathrm{D}(u) \geq \mathrm{D}(u_+) + \mathrm{D}(u_-) + S_d\, \frac{2\,h_d}{2\,h_d + \xi_d}\, m^{\alpha_d} \geq \mathrm{D}(u_+) + \frac{2\,h_d}{2\,h_d + \xi_d}\, \|\nabla u_-\|^2_{\mathrm{L}^2(\mathbb{R}^d)} \,.$$

By definition, we have

$$\mathrm{D}(u_+) \geq c_{\mathrm{BE}}^{\mathrm{pos}} \inf_{g \in \mathcal{M}} \|\nabla u_+ - \nabla g\|^2_{\mathrm{L}^2(\mathbb{R}^d)} \,.$$

As a consequence, if $g_+ \in \mathcal{M}$ is optimal for u_+, we obtain

$$\begin{aligned}
\mathrm{D}(u) &\geq c_{\mathrm{BE}}^{\mathrm{pos}}\, \|\nabla u_+ - \nabla g_+\|^2_{\mathrm{L}^2(\mathbb{R}^d)} + \frac{2\,h_d}{2\,h_d + \xi_d}\, \|\nabla u_-\|^2_{\mathrm{L}^2(\mathbb{R}^d)} \\
&\geq \min\left\{ c_{\mathrm{BE}}^{\mathrm{pos}}, \frac{2\,h_d}{2\,h_d + \xi_d} \right\} \Big(\|\nabla u_+ - \nabla g_+\|^2_{\mathrm{L}^2(\mathbb{R}^d)} + \|\nabla u_-\|^2_{\mathrm{L}^2(\mathbb{R}^d)} \Big) \\
&\geq \frac{1}{2} \min\left\{ c_{\mathrm{BE}}^{\mathrm{pos}}, \frac{2\,h_d}{2\,h_d + \xi_d} \right\} \|\nabla u - \nabla g_+\|^2_{\mathrm{L}^2(\mathbb{R}^d)} \,.
\end{aligned}$$

Since $2\,h_d + \xi_d = 2 \cdot 2^{\frac{2}{d}} - 2 + 2 - 2^{1-\alpha_d} = 2^{\frac{2}{d}}$ we get

$$\frac{h_d}{2\,h_d + \xi_d} = 2^{-\frac{2}{d}} \left(2^{\frac{2}{d}} - 1 \right) = 1 - 2^{-\frac{2}{d}} \,,$$

which concludes the proof. □

6. Stability of the Sobolev inequality: Proof of Theorem 1

We now combine the results from the previous three sections and deduce in this way the main result of this paper.

Proof. We recall that the constant $c_{\mathrm{BE}}^{\mathrm{pos}}$ was defined in the previous subsection and that $\mathscr{I}(\delta)$ was defined in (25). Then, as a consequence of Theorem 22, we have

$$c_{\mathrm{BE}}^{\mathrm{pos}} \geq \sup_{0<\delta\leq 1} \delta\, \mathscr{I}(\delta) \,.$$

(Indeed, for any $\delta \in (0,1)$, if f satisfies $\|\nabla f - \nabla g\|_2^2 \geq \delta\, \|\nabla f\|^2$, then $\mathcal{E}(f) \geq \delta\, \mathscr{I}(\delta)$, while if $\|\nabla f - \nabla g\|_2^2 \leq \delta\, \|\nabla f\|^2$, then $\mathcal{E}(f) \geq \mathscr{I}(\delta) \geq \delta\, \mathscr{I}(\delta)$.) Thus, it remains to bound $\mathscr{I}(\delta)$ for a suitable $\delta \in (0,1)$.

We let $\epsilon_0, \tilde\delta > 0$ be as in Theorem 4. We will bound $\mathscr{I}(\delta)$ with $\delta = \frac{\tilde\delta}{1+\tilde\delta}$. Thus, let $0 \leq f \in \dot{\mathrm{H}}^1(\mathbb{R}^d)$ with

$$\inf_{g \in \mathcal{M}} \|\nabla g - \nabla f\|_2^2 \leq \tfrac{\tilde\delta}{1+\tilde\delta}\, \|\nabla f\|_2^2 \,.$$

It is easy to see that the infimum on the left side is attained. After a translation, a dilation and multiplication by a constant, we may assume that it is attained at $g = (2/(1+|x|^2))^{(d-2)/2}$. We now pass to the sphere using the stereographic projection as in Section 2. Let $0 \leq u \in \mathrm{H}^1(\mathbb{S}^d)$ be the function associated to f. The function 1 is associated to g and we set $r := u - 1$. The fact that the distance is attained at 1 implies that r satisfies the orthogonality conditions (8). Moreover, we have

$$\int_{\mathbb{S}^d} \left(|\nabla r|^2 + A\, r^2 \right) d\mu \leq \tfrac{\tilde\delta}{1+\tilde\delta} \int_{\mathbb{S}^d} \left(|\nabla u|^2 + A\, u^2 \right) d\mu = \tfrac{\tilde\delta}{1+\tilde\delta} \left(A + \int_{\mathbb{S}^d} \left(|\nabla r|^2 + A\, r^2 \right) d\mu \right),$$

so

$$\int_{\mathbb{S}^d} \left(|\nabla r|^2 + A\, r^2 \right) d\mu \leq \tilde\delta\, A \,.$$

By the Sobolev inequality, this implies

$$\left(\int_{\mathbb{S}^d} r^q \, d\mu\right)^{2/q} \leq \tilde{\delta} \,,$$

and therefore we are in the situation of Theorem 4. We deduce that

$$\int_{\mathbb{S}^d} \left(|\nabla u|^2 + A\, u^2\right) d\mu - A \left(\int_{\mathbb{S}^d} u^q \, d\mu\right)^{2/q} \geq \theta\, \epsilon_0 \int_{\mathbb{S}^d} \left(|\nabla r|^2 + A\, r^2\right) d\mu \,.$$

Translating this result back to $\mathbb{R}^d$, we have shown that

$$\mathscr{I}\left(\tfrac{\tilde{\delta}}{1+\tilde{\delta}}\right) \geq \theta\, \epsilon_0 = \tfrac{4\,\epsilon_0}{d-2} \,,$$

and therefore

$$c_{\mathrm{BE}}^{\mathrm{pos}} \geq \tfrac{\tilde{\delta}}{1+\tilde{\delta}}\, \tfrac{4\,\epsilon_0}{d-2} \,,$$

where we recall that $0 < \epsilon_0 < \frac{1}{3}$ is fixed and $\tilde{\delta}$ depends on ϵ_0, but not on d. This constant has the claimed d^{-1} behavior.

We turn now to the case of general, not necessarily nonnegative functions. By Proposition 31

$$c_{\mathrm{BE}} \geq \min\left\{\tfrac{1}{2}\, c_{\mathrm{BE}}^{\mathrm{pos}}, 1 - 2^{-\frac{2}{d}}\right\} \,.$$

Using $1 - 2^{-\frac{2}{d}} \geq (2\,\ln 2)/d$ together with the result for $c_{\mathrm{BE}}^{\mathrm{pos}}$ we obtain also in the general case the claimed d^{-1} behavior. As constant in Theorem 4 we get

$$\beta = \min\left\{\tfrac{2\,\epsilon_0\,\tilde{\delta}}{1+\tilde{\delta}}, 2\,\ln 2\right\} \,, \tag{44}$$

which is computable, since $\tilde{\delta}$ depends in a complicated, yet explicit way on ϵ_0. ☐

Remark 32. *The constant given by (44) is a lower estimate of $d\, c_{\mathrm{BE}}$, which is of the same order as the strict upper estimate obtained from (3). If we apply Proposition 7 instead of Theorem 4 in the above argument, we obtain*

$$c_{\mathrm{BE}}^{\mathrm{pos}} \geq \sup_{0<\delta\leq 1} \delta \mathscr{I}(\delta) \geq \sup_{\tilde{\delta}>0} \tfrac{\tilde{\delta}}{1+\tilde{\delta}}\, \mathsf{m}(\tilde{\delta}^{1/2}) = \sup_{0<\delta<1} \delta\, \mathsf{m}\left(\left(\tfrac{\delta}{1-\delta}\right)^{1/2}\right)$$

with m given by (9). As explained in Remark 11, this lower bound is not very good for larges dimensions. In the above expression, it corresponds to a right-hand side of the order of $2^{-d}\, d^{-(d+2)/2}$ as $d \to +\infty$, but for $d = 3, 4, 5, 6$ it gives decent numerical lower bounds on $c_{\mathrm{BE}}^{\mathrm{pos}}$.

7. Stability of the logarithmic Sobolev inequality: Proof of Theorem 2

Just like the quantitative version of the sharp Sobolev inequality, we prove the quantitative version of the sharp logarithmic Sobolev inequality in two steps, one close and one far from the set of optimizers. Let us start with the result that replaces Theorem 4.

Theorem 33. *There are explicit constants $\eta > 0$ and $\tilde{\delta} > 0$ such that for all $N \in \mathbb{N}$ and for all $-1 \leq r \in \mathrm{H}^1(\gamma)$ satisfying*

$$\int_{\mathbb{R}^N} r^2 \, d\gamma \leq \tilde{\delta} \tag{45}$$

and

$$\int_{\mathbb{R}^N} r \, d\gamma = 0 = \int_{\mathbb{R}^N} x_j\, r \, d\gamma \,, \quad j = 1 \ldots, N \,, \tag{46}$$

one has

$$\int_{\mathbb{R}^N} |\nabla r|^2\,d\gamma-\pi\int_{\mathbb{R}^N}(1+r)^2\ln\left(\frac{(1+r)^2}{\|1+r\|^2_{\mathrm{L}^2(\gamma)}}\right)\,d\gamma\ge\eta\int_{\mathbb{R}^N} r^2\,d\gamma\,.$$

Remark 34. *The constant $\tilde\delta$ coincides with the corresponding constant in Theorem 4 and $\eta=2\,\pi\,\epsilon_0$. Indeed, Remark 5 together with the proof below implies that one can fix $\eta\in(0,\pi)$ and then the bound holds with some $\tilde\delta$ depending on η.*

Proof. Notice that x is in $\mathrm{L}^2(\gamma)$ so that orthogonality constraints raise no integration issues.

We denote $\Sigma_d:=\{x\in\mathbb{R}^{d+1}\,:\,|x|=\rho_d\}$ with $\rho_d:=\sqrt{d/(2\,\pi)}$. (The factor of $1/(2\,\pi)$ in the definition of ρ_d is necessary to get the π in the exponent of the Gaussian density.) We integrate on Σ_d with respect to the uniform probability measure, which we denote by $d\mu_d$. By rescaling our result in Theorem 4 we find that

$$\int_{\Sigma_d}|\nabla R|^2\,d\mu_d-\pi\,\tfrac{d-2}{2}\left(\left(\int_{\Sigma_d}(1+R)^{\frac{2d}{d-2}}\,d\mu_d\right)^{\frac{d-2}{d}}-\int_{\Sigma_d}(1+R)^2\,d\mu_d\right)$$
$$\ge 2\,\pi\,\epsilon_0\int_{\Sigma_d}\left(\tfrac{1}{\pi}\,\tfrac{2}{d-2}\,|\nabla R|^2+R^2\right)\,d\mu_d\,. \tag{47}$$

(Note that rescaling results in a factor ρ_d^{-2} in front of all terms except the term $|\nabla R|^2$.) This inequality is valid for all $R\in\mathrm{H}^1(\Sigma_d)$ such that

$$\left(\int_{\Sigma_d}R^{\frac{2d}{d-2}}\,d\mu_d\right)^{\frac{d-2}{d}}\le\tilde\delta \tag{48}$$

and

$$\int_{\Sigma_d}R\,d\mu_d=0=\int_{\Sigma_d}x_j\,R\,d\mu_d\,,\quad j=1,\ldots,d+1\,. \tag{49}$$

Given a function $r\in\mathrm{H}^1(\gamma)$ and a $d>N$, we apply this inequality to the function

$$R_d(x):=r(x_1,\ldots,x_N)-\int_{\Sigma_d}r\,d\mu_d-2\,\pi\,\tfrac{d+1}{d}\sum_{n=1}^N x_n\int_{\Sigma_d}y_n\,r(y_1,\ldots,y_N)\,d\mu_d(y)$$

for $x\in\Sigma_d$. This function satisfies the orthogonality conditions (49). Note here that the functions $\sqrt{2\,\pi}\,\sqrt{(d+1)/d}\,x_j$ are L^2-normalized on Σ_d.

We now use the well-known fact that, as $d\to+\infty$, the marginal of $d\mu_d$ corresponding to the first N coordinates converges to $d\gamma$. Thus,

$$\lim_{d\to+\infty}\int_{\Sigma_d}|\nabla r|^2\,d\mu_d=\int_{\mathbb{R}^N}|\nabla r|^2\,d\gamma\,,\quad \lim_{d\to+\infty}\int_{\Sigma_d}r^2\,d\mu_d=\int_{\mathbb{R}^N}r^2\,d\gamma\,,$$
$$\lim_{d\to+\infty}\int_{\Sigma_d}r\,d\mu_d=\int_{\mathbb{R}^N}r\,d\gamma=0\,,\qquad \lim_{d\to+\infty}\int_{\Sigma_d}y_n\,r(y_1,\ldots,y_N)\,d\mu_d(y)=\int_{\mathbb{R}^N}y_n\,r\,d\gamma=0\,.$$

From this we conclude easily that

$$\lim_{d\to+\infty}\int_{\Sigma_d}|\nabla R_d|^2\,d\mu_d=\int_{\mathbb{R}^N}|\nabla r|^2\,d\gamma\,,\quad \lim_{d\to+\infty}\int_{\Sigma_d}R_d^2\,d\mu_d=\int_{\mathbb{R}^N}r^2\,d\gamma\,.$$

With some modest amount of effort one also finds that

$$\lim_{d\to+\infty}\int_{\Sigma_d}R_d^{\frac{2d}{d-2}}\,d\mu_d=\int_{\mathbb{R}^N}r^2\,d\gamma\,.$$

In particular, since we assume that the right side is less than $\tilde\delta$, the same is true for the left side when d is sufficiently large, and consequently the smallness condition (48) holds when d is sufficiently large. Thus, inequality (47) is valid for all sufficiently large d.

We drop the gradient term on the right side. Passing to the limit $d\to+\infty$, we infer that

$$\int_{\mathbb R^N}|\nabla r|^2\,d\gamma-\pi\,\limsup_{d\to+\infty}\tfrac{d-2}{2}\left(\left(\int_{\Sigma_d}(1+R_d)^{\frac{2d}{d-2}}\,d\mu_d\right)^{\frac{d-2}{d}}-\int_{\Sigma_d}(1+R_d)^2\,d\mu_d\right)\ge 2\,\pi\,\epsilon_0\int_{\mathbb R^N}r^2\,d\gamma\,.$$

Finally, we verify that

$$\limsup_{d\to+\infty}\frac{d-2}{2}\left(\left(\int_{\Sigma_d}(1+R_d)^{\frac{2d}{d-2}}\,d\mu_d\right)^{\frac{d-2}{d}}-\int_{\Sigma_d}(1+R_d)^2\,d\mu_d\right)$$
$$=\int_{\mathbb R^N}(1+r)^2\,\ln\left(\frac{(1+r)^2}{\|1+r\|^2_{\mathrm L^2(\gamma)}}\right)\,d\gamma\,.$$

In fact, if the orthogonality conditions were not present and the marginals would already be equal to their limit, this would follow from the fact that

$$\lim_{p\to1^+}\frac{1}{p-1}\left(\left(\int_{\mathbb R^N}h^p\,d\gamma\right)^{1/p}-\int_{\mathbb R^N}h\,d\gamma\right)=\int_{\mathbb R^N}h\ln\left(\frac{h}{\int_{\mathbb R^N}h\,d\gamma}\right)d\gamma\,,$$

valid on any measure space for any nonnegative function h that satisfies $h\in\mathrm L^1\cap\mathrm L^{p_0}(\gamma)$ for some $p_0>1$. Proving the latter fact is simple, as well as including the effect of the orthogonality conditions and the convergence of the marginals and we shall omit it. These remarks complete the proof Theorem 33. □

We emphasize that in the previous proof we did not use Theorem 1, but rather Theorem 4. In this way we avoid having to control the distance to the set of optimizers in the high-dimensional limit, which seems harder than verifying the orthogonality conditions.

Proof of Theorem 2. As in the proof of Theorem 1, we first prove the result for nonnegative functions and then extend it to sign changing solutions. Let us denote by $\kappa^{\rm pos}$ the stability constant in the stability inequality restricted to nonnegative functions.

Step 1. Let η and $\tilde\delta$ be as in Theorem 33. For $0\le u\in\mathrm H^1(\gamma)$ we distinguish two cases.
• The first case is where

$$\inf_{a\in\mathbb R^N,\,c\in\mathbb R}\int_{\mathbb R^N}(u-c\,e^{a\cdot x})^2\,d\gamma\le\frac{\tilde\delta}{1+\tilde\delta}\int_{\mathbb R^N}u^2\,d\gamma\,.$$

The infimum on the left-hand side is attained at some $a\in\mathbb R^N$ and $c\in\mathbb R$ as can be checked by optimizing $\int_{\mathbb R^N}\left|v-c\,e^{|a|^2/(2\,\pi)-\pi\,|x-a/\pi|^2/2}\right|^2dx$ where $v(x):=u(x)\,e^{-\,\pi\,|x|^2/2}$. Let

$$\tilde u(y):=e^{-\,y\cdot a-\frac{|a|^2}{2\,\pi}}\,u\big(y+\tfrac{a}{\pi}\big)\,.$$

Then, by a simple computation involving an integration by parts and a change of variables,

$$\int_{\mathbb R^N}|\nabla\tilde u|^2\,d\gamma-\pi\int_{\mathbb R^N}\tilde u^2\ln\left(\frac{\tilde u^2}{\|\tilde u\|^2_{\mathrm L^2(\gamma)}}\right)\,d\gamma=\int_{\mathbb R^N}|\nabla u|^2\,d\gamma-\pi\int_{\mathbb R^N}u^2\ln\left(\frac{u^2}{\|u\|^2_{\mathrm L^2(\gamma)}}\right)\,d\gamma\,.$$

Therefore, the deficit of $\tilde u$ coincides with that of u, while the infimum for $\tilde u$ among all functions of the form (4) is attained at the constant $c\,e^{|a|^2/(2\,\pi)}$. Finally, by multiplying $\tilde u$ with a constant, we may assume that this constant is equal to one. To summarize, we may assume without loss of generality that the infimum in the theorem is attained at $a = 0$ and $c = 1$.

Let us set $r := u - 1$. Then the minimality implies that r satisfies the orthogonality conditions (46). Moreover, we have

$$\int_{\mathbb{R}^N} r^2\, d\gamma \le \frac{\tilde\delta}{1+\tilde\delta} \int_{\mathbb{R}^N} u^2\, d\gamma = \frac{\tilde\delta}{1+\tilde\delta} \left(1 + \int_{\mathbb{R}^N} r^2\, d\gamma\right),$$

so

$$\int_{\mathbb{R}^N} r^2\, d\gamma \le \tilde\delta\,.$$

Thus, the smallness condition (45) is satisfied and we can apply Theorem 33. This yields the inequality in the theorem with a stability constant η.

• Next, we consider the case where

$$\inf_{a\in\mathbb{R}^N,\, c\in\mathbb{R}} \int_{\mathbb{R}^N} (u - c\, e^{x\cdot a})^2\, d\gamma > \frac{\tilde\delta}{1+\tilde\delta} \int_{\mathbb{R}^N} u^2\, d\gamma\,.$$

We argue similarly as we did in Section 4 concerning the Sobolev inequality, but there are some simplifications in this case.

For $f \in \mathrm{L}^2(\gamma)$ we denote by Uf its Gaussian rearrangement, that is, the function on $\mathbb{R}^N$ whose superlevel sets have the form $\{x \in \mathbb{R}^N : x_1 < \mu\}$ for some $\mu \in \mathbb{R}$ and have the same γ-measure as the corresponding superlevel sets of f. Moreover, we denote

$$Vf := e^{\frac{\pi}{2}\,|x|^2}\, \mathcal{R}\left(e^{-\frac{\pi}{2}\,|x|^2}\, f\right),$$

where $\mathcal{R}$ is, as before, the Euclidean rearrangement. Then, as shown in [22, Theorem 4.1], for any $0 \le f \in \mathrm{L}^2(\gamma)$ one has

$$f_n := (VU)^n f \to \|f\|_{\mathrm{L}^2(\gamma)} \quad \text{in} \quad \mathrm{L}^2(\gamma)\,.$$

Moreover, $\|f_n\|_{\mathrm{L}^2(\gamma)} = \|f\|_{\mathrm{L}^2(\gamma)}$ and

$$n \mapsto \int_{\mathbb{R}^N} |\nabla f_n|^2\, d\gamma - \pi \int_{\mathbb{R}^N} f_n^2 \ln\left(\frac{f_n^2}{\|f_n\|^2_{\mathrm{L}^2(\gamma)}}\right) d\gamma$$

is nonincreasing.

We apply this procedure to our function u and obtain a sequence of functions u_n with constant $\mathrm{L}^2(\gamma)$-norm. Moreover, since

$$\inf_{c,a} \|u_n - c\, e^{a\cdot x}\|_{\mathrm{L}^2(\gamma)} \le \big\|u_n - \|u\|_{\mathrm{L}^2(\gamma)}\big\|_{\mathrm{L}^2(\gamma)} \to 0\,,$$

there is an $n_0 \in \mathbb{N}$ such that

$$\inf_{c,a} \|u_{n_0} - c\, e^{a\cdot x}\|^2_{\mathrm{L}^2(\gamma)} \ge \frac{\tilde\delta}{1+\tilde\delta}\, \|u\|^2_{\mathrm{L}^2(\gamma)} > \inf_{c,a} \|u_{n_0+1} - c\, e^{a\cdot x}\|^2_{\mathrm{L}^2(\gamma)}\,.$$

We have

$$\frac{\int_{\mathbb{R}^N} |\nabla u|^2 \, d\gamma - \pi \int_{\mathbb{R}^N} u^2 \ln\left(\frac{u^2}{\|u\|^2_{\mathrm{L}^2(\gamma)}}\right) d\gamma}{\inf_{c,a} \|u - c\, e^{a\cdot x}\|^2_{\mathrm{L}^2(\gamma)}} \geq \frac{\int_{\mathbb{R}^N} |\nabla u|^2 \, d\gamma - \pi \int_{\mathbb{R}^N} u^2 \ln\left(\frac{u^2}{\|u\|^2_{\mathrm{L}^2(\gamma)}}\right) d\gamma}{\|u\|^2_{\mathrm{L}^2(\gamma)}}$$

$$\geq \frac{\int_{\mathbb{R}^N} |\nabla u_{n_0}|^2 \, d\gamma - \pi \int_{\mathbb{R}^N} u_{n_0}^2 \ln\left(\frac{u_{n_0}^2}{\|u_{n_0}\|^2_{\mathrm{L}^2(\gamma)}}\right) d\gamma}{\|u\|^2_{\mathrm{L}^2(\gamma)}} .$$

We now use a continuous rearrangement flow to connect u_{n_0} to u_{n_0+1}. More precisely, we a consider a family of functions u_τ, $\tau \in [0,\infty]$, where $\mathsf{u}_0 := U u_{n_0}$ and $\mathsf{u}_\infty := u_{n_0+1}$. We define u_τ as $e^{\pi\,|x|^2/2}$ times the continuous (Euclidean) rearrangement of $e^{-\,\pi\,|x|^2/2}\, U u_{n_0}$. In the same way as in Lemma 36 one sees that

$$\tau \mapsto \inf_{c,a} \|\mathsf{u}_\tau - c\, e^{a\cdot x}\|^2_{\mathrm{L}^2(\gamma)}$$

is continuous, and therefore there is a $\tau_0 \in [0,\infty)$ such that

$$\inf_{c,a} \|\mathsf{u}_{\tau_0} - c\, e^{a\cdot x}\|^2_{\mathrm{L}^2(\gamma)} = \frac{\tilde\delta}{1+\tilde\delta}\, \|u\|^2_{\mathrm{L}^2(\gamma)} .$$

It follows that

$$\frac{\int_{\mathbb{R}^N} |\nabla u_{n_0}|^2 \, d\gamma - \pi \int_{\mathbb{R}^N} u_{n_0}^2 \ln\left(\frac{u_{n_0}^2}{\|u_{n_0}\|^2_{\mathrm{L}^2(\gamma)}}\right) d\gamma}{\|u\|^2_{\mathrm{L}^2(\gamma)}}$$

$$\geq \frac{\int_{\mathbb{R}^N} |\nabla \mathsf{u}_{\tau_0}|^2 \, d\gamma - \pi \int_{\mathbb{R}^N} \mathsf{u}_{\tau_0}^2 \ln\left(\frac{\mathsf{u}_{\tau_0}^2}{\|\mathsf{u}_{\tau_0}\|^2_{\mathrm{L}^2(\gamma)}}\right) d\gamma}{\|u\|^2_{\mathrm{L}^2(\gamma)}}$$

$$= \frac{\tilde\delta}{1+\tilde\delta}\, \frac{\int_{\mathbb{R}^N} |\nabla \mathsf{u}_{\tau_0}|^2 \, d\gamma - \pi \int_{\mathbb{R}^N} \mathsf{u}_{n_0}^2 \ln\left(\frac{\mathsf{u}_{\tau_0}^2}{\|\mathsf{u}_{\tau_0}\|^2_{\mathrm{L}^2(\gamma)}}\right) d\gamma}{\inf_{c,a} \|\mathsf{u}_{\tau_0} - c\, e^{a\cdot x}\|^2_{\mathrm{L}^2(\gamma)}} .$$

According to the first case, the right side is larger or equal than $\kappa^{\mathrm{pos}} := \frac{\tilde\delta}{1+\tilde\delta}\,\eta$. This concludes the proof in the case of nonnegative functions.

Step 2. Finally, we prove the theorem in the general case. This is a variation of the argument in Proposition 31.

We shall use the notation

$$D(v) := \int_{\mathbb{R}^N} |\nabla v|^2 \, d\gamma - \pi \int_{\mathbb{R}^N} v^2 \ln\left(\frac{v^2}{\|v\|^2_{\mathrm{L}^2(\gamma)}}\right) d\gamma \quad \text{for } v \in \mathrm{H}^1(\gamma) .$$

Let $u \in \mathrm{H}^1(\gamma)$. By homogeneity we can assume $\|u\|_{\mathrm{L}^2(\gamma)} = 1$. Replacing u by $-\,u$ if necessary, we can also assume that

$$m := \|u_-\|^2_{\mathrm{L}^2(\gamma)} \in [0, \tfrac12] .$$

Then

$$D(u) = D(u_+) + D(u_-) + \pi\, h(m)$$

with

$$h(p) := -\big(p\ln p + (1-p)\ln(1-p)\big)\,.$$

Since the function $p \mapsto h(p)$ is monotone increasing and concave on $[0,\frac12]$, it holds that

$$h(p) \geq (2\ln 2)\,p \quad \text{for all} \quad p\in[0,\tfrac12]\,.$$

Thus, with $\kappa^{\rm pos}$ denoting the constant from Step 1,

$$\begin{aligned} D(u) &\geq D(u_+) + (2\,\pi\,\ln 2)\,m \geq \kappa^{\rm pos}\inf_{a,c}\|u_+ - c\,e^{a\cdot x}\|^2_{\mathrm{L}^2(\gamma)} + (2\,\pi\,\ln 2)\,\|u_-\|^2_{\mathrm{L}^2(\gamma)} \\ &\geq \tfrac12\,\min\big\{\kappa^{\rm pos},\,2\,\pi\,\ln 2\big\}\,\inf_{a,c}\|u - c\,e^{a\cdot x}\|^2_{\mathrm{L}^2(\gamma)}\,. \end{aligned}$$

This proves the inequality in the general case, with $\kappa = \frac12\,\min\big\{\kappa^{\rm pos},\,2\,\pi\,\ln 2\big\}$. □

Appendix. Some remarks about continuous rearrangement

In this appendix we discuss some aspects of the continuous rearrangement and prove some of its properties.

Brock's continuous Steiner rearrangement is based on the following operation for functions of one real variable that are finite union of disjoint characteristic functions $\sum_{k=1}^N \chi_{(-a_k,a_k)}(x-b_k)$. Replace this function by $\sum_{k=1}^N \chi_{(-a_k,a_k)}\big(x - e^{-t}\,b_k\big)$ where t varies from 0 to ∞. As t increases, the intervals start moving closer and as soon as any two intervals touch one stops the process and redefines the set of intervals by joining the two that touched. Then one restarts the process and keeps repeating it until all of them are joined into one. The movement stops once this interval is centered at the origin. By the outer regularity of Lebesgue measure the level sets of a measurable function can be approximated by open sets and, since in one dimension this is a countable union of open intervals, one can further approximate the level set by a finite number of open disjoint intervals for which one uses the sliding argument explained above.

As mentioned before, this procedure can be generalized to higher dimensions by considering Steiner symmetrization with respect to a hyperplane. One considers any hyperplane H through the origin and then rearranges the function symmetrically about the hyperplane along each line perpendicular to H, resulting in a function denoted by f^{*H}. For more information see [54]. In this fashion one obtains a continuous rearrangement $f \to \mathsf{f}^H_\tau, \tau\in[0,\infty]$, which was studied in detail by Brock [14, 15]. We shall refer to the statements in those papers.

To pass from Steiner symmetrization to the symmetric decreasing rearrangement we consider a sequence of continuous Steiner symmetrizations and chain them with a new continous parameter à la Bucur–Henrot. Inspired by [16, 25], we proceed as follows. Given a function $f\in \mathrm{L}^p(\mathbb{R}^d)$ for some $1\leq p<\infty$ there is a sequence $(H_n)_{n\in\mathbb{N}}$ of hyperplanes such that, defining recursively with $f_0 = f$,

$$f_n := f_{n-1}^{*H_n}\,,\quad n = 1,2,\ldots\,,$$

we have

$$f_n \to f^* \quad \text{in}\,\mathrm{L}^p(\mathbb{R}^d) \quad \text{as} \quad n\to\infty\,.$$

In fact, it is shown in [65, Theorem 4.3] that this holds for 'almost every' (in an appropriate sense) choice of hyperplanes. It is also of interest that this sequence can actually be chosen in a universal fashion (that is, independent of f and p); see [64, Theorem 5.2].

Given f and the sequence $(f_n)_{n\in\mathbb{N}}$ as above, we set for any $n = 0,1,2,\ldots$

$$\phi_n(\tau) := e^{\frac{\tau-n}{n+1-\tau}} - 1\,,\quad \tau\in[n,n+1]\,,$$

and define

$$\mathsf{f}_\tau := \mathsf{f}_{n,\phi_n(\tau)}\,, \tag{50}$$

where the right side denotes Brock's continuous Steiner symmetrization with respect to the hyperplane H_n with parameter $\phi_n(\tau)$ applied to f_n. As τ runs from n to $n+1$, $\phi_n(\tau)$ runs from 0 to ∞, so f_τ is well defined even for $\tau \in \mathbb{N}_0$.

From the properties of Brock's flow, see, in particular, [15, Lemma 4.1], we obtain the following properties for our flow.

Proposition 35. *Let $d \geq 1$, $1 \leq p < \infty$ and let $0 \leq f \in \mathrm{L}^p(\mathbb{R}^d)$. Then, for any $\tau \in [0,\infty]$, the function f_τ defined by (50) is in $\mathrm{L}^p(\mathbb{R}^d)$ and $\|\mathsf{f}_\tau\|_p = \|\mathsf{f}\|_p$. Moreover, for any $\tau \in [0,\infty]$ and any sequence $(\tau_n)_{n\in\mathbb{N}}$ with $\lim_{n\to\infty} \tau_n = \tau$,*

$$\lim_{n\to\infty} \|\mathsf{f}_{\tau_n} - \mathsf{f}_\tau\|_p = 0\,.$$

The following fact is important for us.

Lemma 36. *Let $d \geq 3$ and $0 \leq f \in \mathrm{L}^{2^*}(\mathbb{R}^d)$. The function*

$$\tau \mapsto \sup_{u\in\mathcal{M}_1} \left(\mathsf{f}_\tau, u^{2^*-1}\right)^2$$

with f_τ defined by (50) is continuous.

Proof. We use the fact, shown in Proposition 35, that

$$\lim_{\tau_1\to\tau_2} \|\mathsf{f}_{\tau_1} - \mathsf{f}_{\tau_2}\|_{2^*} = 0\,.$$

Fix $\varepsilon > 0$. There exists $u_1 \in \mathcal{M}_1$ such that $\sup_{u\in\mathcal{M}_1} \left|\left(\mathsf{f}_{\tau_1}, u^{2^*-1}\right)\right| \leq \left|\left(\mathsf{f}_{\tau_1}, u_1^{2^*-1}\right)\right| + \varepsilon$ and hence

$$\begin{aligned}\sup_{u\in\mathcal{M}_1} \left|\left(\mathsf{f}_{\tau_1}, u^{2^*-1}\right)\right| - \sup_{u\in\mathcal{M}_1} \left|\left(\mathsf{f}_{\tau_2}, u^{2^*-1}\right)\right| &\leq \left|\left(\mathsf{f}_{\tau_1}, u_1^{2^*-1}\right)\right| + \varepsilon - \left|\left(\mathsf{f}_{\tau_2}, u_1^{2^*-1}\right)\right| \\ &\leq \left|\left(\mathsf{f}_{\tau_1}, u_1^{2^*-1}\right) - \left(\mathsf{f}_{\tau_2}, u_1^{2^*-1}\right)\right| + \varepsilon\,,\end{aligned}$$

which by Hölder's inequality is bounded above by

$$\|\mathsf{f}_{\tau_1} - \mathsf{f}_{\tau_2}\|_{2^*}\, \|u_1^{2^*-1}\|_q + \varepsilon = \|\mathsf{f}_{\tau_1} - \mathsf{f}_{\tau_2}\|_{2^*} + \varepsilon$$

with $q = \frac{2^*}{2^*-1}$. Hence

$$\limsup_{\tau_2\to\tau_1} \left(\sup_{u\in\mathcal{M}_1} \left|\left(\mathsf{f}_{\tau_1}, u^{2^*-1}\right)\right| - \sup_{u\in\mathcal{M}_1} \left|\left(\mathsf{f}_{\tau_2}, u^{2^*-1}\right)\right|\right) \leq \varepsilon\,.$$

There exists $u_2 \in \mathcal{M}_1$ such that $\sup_{u\in\mathcal{M}_1} \left|\left(\mathsf{f}_{\tau_2}, u^{2^*-1}\right)\right| \leq \left|\left(\mathsf{f}_{\tau_2}, u_2^{2^*-1}\right)\right| + \varepsilon$ and hence

$$\sup_{u\in\mathcal{M}_1} \left|\left(\mathsf{f}_{\tau_1}, u^{2^*-1}\right)\right| - \sup_{u\in\mathcal{M}_1} \left|\left(\mathsf{f}_{\tau_2}, u^{2^*-1}\right)\right| \geq \left|\left(\mathsf{f}_{\tau_1}, u_2^{2^*-1}\right)\right| - \left|\left(\mathsf{f}_{\tau_2}, u_2^{2^*-1}\right)\right| - \varepsilon\,,$$

which is greater or equal to

$$-\left|\left(\mathsf{f}_{\tau_1}, u_2^{2^*-1}\right) - \left(\mathsf{f}_{\tau_2}, u_2^{2^*-1}\right)\right| - \varepsilon \geq -\|\mathsf{f}_{\tau_1} - \mathsf{f}_{\tau_2}\|_{2^*} - \varepsilon\,.$$

Hence

$$\liminf_{\tau_2\to\tau_1} \left(\sup_{u\in\mathcal{M}_1} \left|\left(\mathsf{f}_{\tau_1}, u^{2^*-1}\right)\right| - \sup_{u\in\mathcal{M}_1} \left|\left(\mathsf{f}_{\tau_2}, u^{2^*-1}\right)\right|\right) \geq -\,\varepsilon\,.$$

This proves the claimed continuity. ◻

We now consider the behavior of the gradient under the rearrangement flow. The following proposition is closely related to [15, Theorems 3.2 and 4.1], but there inhomogeneous Sobolev spaces are considered, which leads to some minor changes. For the sake of simplicity we provide the details.

Proposition 37. *Let* $0 \leq f \in \dot{\mathrm{H}}^1(\mathbb{R}^d)$*. Then* f_τ *defined by* (50) *is in* $\dot{\mathrm{H}}^1(\mathbb{R}^d)$ *and* $\tau \mapsto \|\nabla \mathsf{f}_\tau\|_2$ *is a nonincreasing, right-continuous function.*

Proof. By construction, it suffices to prove these properties for Brock's flow. Since the latter has the semigroup property $(\mathsf{f}_\sigma)_\tau = \mathsf{f}_{\sigma+\tau}$ for all σ, $\tau \geq 0$, it suffices to prove monotonicity and right-continuity at $\tau = 0$.

We begin with the proof of monotonicity, which we first prove under the additional assumption that $f \in \mathrm{L}^2(\mathbb{R}^d)$. This is shown in [15, Theorem 3.2], but we give an alternative proof. We proceed as in the proof of [54, Lemma 1.17]. Extending [14, Corollary 2] to the sequence of Steiner symmetrizations we find for three nonnegative functions f, g, h that

$$\iint_{\mathbb{R}^d\times\mathbb{R}^d} \mathsf{f}_\tau(x)\, g_\tau(x-y)\, h_\tau(y)\, dx\, dy \geq \iint_{\mathbb{R}^d\times\mathbb{R}^d} f(x)\, g(x-y)\, h(y)\, dx\, dy\,.$$

If we choose $g(x-y)$ to be the standard heat kernel, i.e., $g(x-y) = e^{\Delta t}(x-y)$, then $g_\tau(x-y) = g(x-y)$ and hence

$$\iint_{\mathbb{R}^d\times\mathbb{R}^d} \mathsf{f}_\tau(x)\, e^{\Delta t}(x-y)\, \mathsf{f}_\tau(y)\, dx\, dy \geq \iint_{\mathbb{R}^d\times\mathbb{R}^d} f(x)\, e^{\Delta t}(x-y)\, f(y)\, dx\, dy\,.$$

Since $\|\mathsf{f}_\tau\|_2 = \|f\|_2$ by the equimeasurability of rearrangement,

$$\frac{1}{t}\left(\|\mathsf{f}_\tau\|_2^2 - \left(\mathsf{f}_\tau, e^{\Delta t}\,\mathsf{f}_\tau\right)\right) \leq \frac{1}{t}\left(\|f\|_2^2 - \left(f, e^{\Delta t} f\right)\right)$$

and letting $t \to 0$ yields the first claim under the additional assumption $f \in \mathrm{L}^2(\mathbb{R}^d)$.

For general $0 \leq f \in \dot{\mathrm{H}}^1(\mathbb{R}^d)$ we apply the above argument to the functions $(f-\epsilon)_+$, $\epsilon > 0$. They belong to $\mathrm{L}^2(\mathbb{R}^d)$ since f vanishes at infinity and belongs to $\mathrm{L}^{2^*}(\mathbb{R}^d)$. We obtain

$$\left\|\nabla\big((f-\epsilon)_+\big)_\tau\right\|_2 \leq \|\nabla(f-\epsilon)_+\|_2 \leq \|\nabla f\|_2\,. \tag{51}$$

We claim that $\mathsf{f}_\tau \in \dot{\mathrm{H}}^1(\mathbb{R}^d)$ and $\nabla\big((f-\epsilon)_+\big)_\tau \rightharpoonup \nabla \mathsf{f}_\tau$ in $\mathrm{L}^2(\mathbb{R}^d)$ as $\epsilon \to 0^+$. Once this is shown, the claimed inequality follows from (51) by the weak lower semicontinuity of the L^2 norm.

To prove the claimed weak convergence, note that by (51), $\nabla\big((f-\epsilon)_+\big)_\tau$ is bounded in $\mathrm{L}^2(\mathbb{R}^d)$ as $\epsilon \to 0^+$ and therefore has a weak limit point. Let $F \in \mathrm{L}^2(\mathbb{R}^d)$ be any such limit point. Since $(f-\epsilon)_+ \to f$ in $\mathrm{L}^{2^*}(\mathbb{R}^d)$, the nonexpansivity of the rearrangement [14, Lemma 3] implies that $\big((f-\epsilon)_+\big)_\tau \to \mathsf{f}_\tau$ in $\mathrm{L}^{2^*}(\mathbb{R}^d)$. Thus, for any $\Phi \in C^1_c(\mathbb{R}^d)$,

$$\int_{\mathbb{R}^d} (\nabla\cdot\Phi)\, \mathsf{f}_\tau\, dx \leftarrow \int_{\mathbb{R}^d} (\nabla\cdot\Phi)\, \big((f-\epsilon)_+\big)_\tau\, dx = -\int_{\mathbb{R}^d} \Phi\cdot\nabla\big((f-\epsilon)_+\big)_\tau\, dx \to -\int_{\mathbb{R}^d} \Phi\cdot F\, dx$$

as $\epsilon \to 0^+$. This proves that f_τ is weakly differentiable with $\nabla \mathsf{f}_\tau = F$. In particular, $\mathsf{f}_\tau \in \dot{\mathrm{H}}^1(\mathbb{R}^d)$ (note that f_τ vanishes at infinity since f does and since these functions are equimeasurable) and the limit point F is unique. This concludes the proof of the first part of the proposition.

Let us now show the right-continuity at $\tau = 0$. It follows from Proposition 35 that $\mathsf{f}_\tau \to f$ in $\mathrm{L}^{2^*}(\mathbb{R}^d)$ as $\tau \to 0^+$. This implies that $\nabla \mathsf{f}_\tau \rightharpoonup \nabla f$ in $\mathrm{L}^2(\mathbb{R}^d)$ as $\tau \to 0^+$. (Indeed, the argument is similar to the one used in the first part of the proof. The family $\nabla \mathsf{f}_\tau$ is bounded in $\mathrm{L}^2(\mathbb{R}^d)$

as $\tau \to 0^+$ and, if F denotes any weak limit point in $\mathrm{L}^2(\mathbb{R}^d)$, then the convergence in $\mathrm{L}^{2^*}(\mathbb{R}^d)$ and the definition of weak derivatives implies that $F = \nabla f$.) By weak lower semicontinuity, we deduce that

$$\|\nabla f\|_2 \le \liminf_{\tau \to 0^+} \|\nabla \mathsf{f}_\tau\|_2 \,.$$

This, together with the reverse inequality, which was established in the first part of the proof, proves the claimed right continuity. □

We note that the proposition remains valid for $0 \le f \in \dot{\mathrm{W}}^{1,p}(\mathbb{R}^d)$ with $1 \le p < d$. If $p \neq 2$, the monotonicity for the gradient for $f \in \mathrm{W}^{1,p}(\mathbb{R}^d)$ is proved in [15, Theorem 3.2]. The remaining arguments above carry over to $p \neq 2$.

References

[1] Frederick J. Almgren, Jr. and Elliott H. Lieb. Symmetric decreasing rearrangement is sometimes continuous. *J. Amer. Math. Soc.*, 2(4):683–773, 1989.

[2] Angelo Alvino, Vincenzo Ferone, and Carlo Nitsch. A sharp isoperimetric inequality in the plane. *J. Eur. Math. Soc. (JEMS)*, 13(1):185–206, 2011.

[3] Anton Arnold, Jean-Philippe Bartier, and Jean Dolbeault. Interpolation between logarithmic Sobolev and Poincaré inequalities. *Commun. Math. Sci.*, 5(4):971–979, 2007.

[4] Thierry Aubin. Équations différentielles non linéaires et problème de Yamabe concernant la courbure scalaire. *J. Math. Pures Appl. (9)*, 55(3):269–296, 1976.

[5] William Beckner. Sobolev inequalities, the Poisson semigroup, and analysis on the sphere $\mathbb{S}^n$. *Proc. Nat. Acad. Sci. U.S.A.*, 89(11):4816–4819, 1992.

[6] Gabriele Bianchi and Henrik Egnell. A note on the Sobolev inequality. *J. Funct. Anal.*, 100(1):18–24, 1991.

[7] Chiara Bianchini, Gisella Croce, and Antoine Henrot. On the quantitative isoperimetric inequality in the plane. *ESAIM Control Optim. Calc. Var.*, 23(2):517–549, 2017.

[8] Sergey G. Bobkov, Nathael Gozlan, Cyril Roberto, and Paul-Marie Samson. Bounds on the deficit in the logarithmic Sobolev inequality. *J. Funct. Anal.*, 267(11):4110–4138, dec 2014.

[9] Matteo Bonforte, Jean Dolbeault, Gabriele Grillo, and Juan Luis Vázquez. Sharp rates of decay of solutions to the nonlinear fast diffusion equation via functional inequalities. *Proc. Natl. Acad. Sci. USA*, 107(38):16459–16464, 2010.

[10] Matteo Bonforte, Jean Dolbeault, Bruno Nazaret, and Nikita Simonov. Stability in Gagliardo-Nirenberg-Sobolev inequalities: flows, regularity and the entropy method. arXiv: 2007.03674, to appear in *Memoirs of the AMS*.

[11] Herm Jan Brascamp, Elliott H. Lieb, and Joaquin M. Luttinger. A general rearrangement inequality for multiple integrals. *J. Funct. Anal.*, 17:227–237, 1974.

[12] Haïm Brezis and Elliott H. Lieb. Sobolev inequalities with remainder terms. *J. Funct. Anal.*, 62(1):73–86, 1985.

[13] Giovanni Brigati, Jean Dolbeault, and Nikita Simonov. Logarithmic Sobolev and interpolation inequalities on the sphere: constructive stability results. arXiv: 2211.13180, 2022.

[14] Friedemann Brock. Continuous Steiner-symmetrization. *Math. Nachr.*, 172:25–48, 1995.

[15] Friedemann Brock. Continuous rearrangement and symmetry of solutions of elliptic problems. *Proc. Indian Acad. Sci. Math. Sci.*, 110(2):157–204, 2000.

[16] Dorin Bucur and Antoine Henrot. Stability for the Dirichlet problem under continuous Steiner symmetrization. *Potential Anal.*, 13(2):127–145, 2000.

[17] Almut Burchard. Steiner symmetrization is continuous in $W^{1,p}$. *Geom. Funct. Anal.*, 7(5):823–860, 1997.

[18] Luis A. Caffarelli, Basilis Gidas, and Joel Spruck. Asymptotic symmetry and local behavior of semilinear elliptic equations with critical Sobolev growth. *Comm. Pure Appl. Math.*, 42(3):271–297, 1989.

[19] Stefano Campi. Isoperimetric deficit and convex plane sets of maximum translative discrepancy. *Geom. Dedicata*, 43(1):71–81, 1992.

[20] Eric A. Carlen. Superadditivity of Fisher's information and logarithmic Sobolev inequalities. *J. Funct. Anal.*, 101(1):194–211, 1991.

[21] Eric A. Carlen, Rupert L. Frank, and Elliott H. Lieb. Stability estimates for the lowest eigenvalue of a Schrödinger operator. *Geom. Funct. Anal.*, 24(1):63–84, 2014.

[22] Eric A. Carlen and Michael Loss. Extremals of functionals with competing symmetries. *J. Funct. Anal.*, 88(2):437–456, 1990.

[23] Shibing Chen, Rupert L. Frank, and Tobias Weth. Remainder terms in the fractional Sobolev inequality. *Indiana Univ. Math. J.*, 62(4):1381–1397, 2013.

[24] Michael Christ. A sharpened Hausdorff-Young inequality. arXiv:1406.1210, 2014.

[25] Michael Christ. A sharpened Riesz-Sobolev inequality. arXiv: 1706.02007, 2017.

[26] Andrea Cianchi, Nicola Fusco, Francesco Maggi, and Aldo Pratelli. The sharp Sobolev inequality in quantitative form. *J. Eur. Math. Soc. (JEMS)*, 11(5):1105–1139, 2009.

[27] Marco Cicalese and Gian Paolo Leonardi. A selection principle for the sharp quantitative isoperimetric inequality. *Arch. Ration. Mech. Anal.*, 206(2):617–643, 2012.

[28] Marco Cicalese and Gian Paolo Leonardi. Best constants for the isoperimetric inequality in quantitative form. *J. Eur. Math. Soc. (JEMS)*, 15(3):1101–1129, 2013.

[29] Jean Dolbeault, Maria J. Esteban, and Michael Loss. Interpolation inequalities on the sphere: linear vs. nonlinear flows. *Ann. Fac. Sci. Toulouse Math. (6)*, 26(2):351–379, 2017.

[30] Jean Dolbeault and Gaspard Jankowiak. Sobolev and Hardy–Littlewood–Sobolev inequalities. *J. Differential Equations*, 257(6):1689–1720, 2014.

[31] Jean Dolbeault and Giuseppe Toscani. Improved interpolation inequalities, relative entropy and fast diffusion equations. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 30(5):917–934, 2013.

[32] Jean Dolbeault and Giuseppe Toscani. Stability results for logarithmic Sobolev and Gagliardo-Nirenberg inequalities. *Int. Math. Res. Not. IMRN*, 2016(2):473–498, 2016.

[33] Javier Duoandikoetxea. Reverse Hölder inequalities for spherical harmonics. *Proc. Amer. Math. Soc.*, 101(3):487–491, 1987.

[34] Ronen Eldan, Joseph Lehec, and Yair Shenfeld. Stability of the logarithmic Sobolev inequality via the Föllmer process. *Ann. Inst. Henri Poincaré Probab. Stat.*, 56(3):2253–2269, 2020.

[35] Max Fathi, Emanuel Indrei, and Michel Ledoux. Quantitative logarithmic Sobolev inequalities and stability estimates. *Discrete Contin. Dyn. Syst.*, 36(12):6835–6853, 2016.

[36] Paul Federbush. Partially alternate derivation of a result of Nelson. *J. Math. Phys.*, 10:50–52, 1969.

[37] F. Feo, E. Indrei, M. R. Posteraro, and C. Roberto. Some remarks on the stability of the log-Sobolev inequality for the Gaussian measure. *Potential Anal.*, 47(1):37–52, 2017.

[38] Alessio Figalli, Nicola Fusco, Francesco Maggi, Vincent Millot, and Massimiliano Morini. Isoperimetry and stability properties of balls with respect to nonlocal energies. *Comm. Math. Phys.*, 336(1):441–507, 2015.

[39] Alessio Figalli, Francesco Maggi, and Aldo Pratelli. A mass transportation approach to quantitative isoperimetric inequalities. *Invent. Math.*, 182(1):167–211, 2010.

[40] Alessio Figalli and Yi Ru-Ya Zhang. Sharp gradient stability for the Sobolev inequality. *Duke Math. J.*, 171(12), Sep 2022.

[41] Rupert L. Frank. Degenerate stability of some Sobolev inequalities. *Ann. Inst. H. Poincaré Anal. Non Linéaire (Online first). Doi: 10.4171/AIHPC/35*, Jun 2022.

[42] Rupert L. Frank and Elliott H. Lieb. A note on a theorem of M. Christ. arXiv:1909.04598, 2019.

[43] Rupert L. Frank and Elliott H. Lieb. Proof of spherical flocking based on quantitative rearrangement inequalities. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)*, 22(3):1241–1263, 2021.

[44] Nicola Fusco, Francesco Maggi, and Aldo Pratelli. The sharp quantitative Sobolev inequality for functions of bounded variation. *J. Funct. Anal.*, 244(1):315–341, 2007.

[45] Basilis Gidas, Wei Ming Ni, and Louis Nirenberg. Symmetry of positive solutions of nonlinear elliptic equations in $\mathbb{R}^n$. In *Mathematical analysis and applications, Part A*, volume 7 of *Adv. in Math. Suppl. Stud.*, pages 369–402. Academic Press, New York-London, 1981.

[46] Leonard Gross. Logarithmic Sobolev inequalities. *Amer. J. Math.*, 97(4):1061–1083, 1975.

[47] Leonard Gross. Hypercontractivity, logarithmic Sobolev inequalities, and applications: a survey of surveys. In *Diffusion, quantum theory, and radically elementary mathematics*, volume 47 of *Math. Notes*, pages 45–73. Princeton Univ. Press, Princeton, NJ, 2006.

[48] Emanuel Indrei and Daesung Kim. Deficit estimates for the logarithmic Sobolev inequality. *Differential Integral Equations*, 34(7-8):437–466, 2021.

[49] Emanuel Indrei and Diego Marcon. A quantitative log-Sobolev inequality for a two parameter family of functions. *Int. Math. Res. Not. IMRN*, 2014(20):5563–5580, 2014.

[50] Daesung Kim. Instability results for the logarithmic Sobolev inequality and its application to related inequalities. *Discrete Contin. Dyn. Syst.*, 42(9):4297–4320, 2022.

[51] Tobias König. On the sharp constant in the Bianchi-Egnell stability inequality. arXiv:2210.08482, 2022.

[52] Tobias König. Stability for the Sobolev inequality: existence of a minimizer. arXiv:2211.14185, 2022.

[53] Elliott H. Lieb. Sharp constants in the Hardy-Littlewood-Sobolev and related inequalities. *Ann. of Math. (2)*, 118(2):349–374, 1983.

[54] Elliott H. Lieb and Michael Loss. *Analysis*, volume 14 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, second edition, 2001.

[55] P.-L. Lions. The concentration-compactness principle in the calculus of variations. The limit case. I. *Rev. Mat. Iberoamericana*, 1(1):145–201, 1985.

[56] G. Pólya and G. Szegö. *Isoperimetric Inequalities in Mathematical Physics*. Annals of Mathematics Studies, No. 27. Princeton University Press, Princeton, N. J., 1951.

[57] Olivier Rey. The role of the Green's function in a nonlinear elliptic equation involving the critical Sobolev exponent. *J. Funct. Anal.*, 89(1):1–52, 1990.

[58] Eugene R. Rodemich. The Sobolev inequalities with best possible constants. *Analysis Seminar Caltech*, 1966.

[59] C. Ambrose Rogers. A single integral inequality. *J. London Math. Soc.*, 32:102–108, 1957.

[60] Gerald Rosen. Minimum value for c in the Sobolev inequality $\|\phi^3\| \leq c\,\|\nabla\phi\|^3$. *SIAM J. Appl. Math.*, 21:30–32, 1971.

[61] Francis Seuffert. An extension of the Bianchi-Egnell stability estimate to Bakry, Gentil, and Ledoux's generalization of the Sobolev inequality to continuous dimensions. *J. Funct. Anal.*, 273(10):3094–3149, 2017.

[62] A. J. Stam. Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Information and Control*, 2:101–112, 1959.

[63] Giorgio Talenti. Best constant in Sobolev inequality. *Ann. Mat. Pura Appl. (4)*, 110:353–372, 1976.

[64] Jean van Schaftingen. Universal approximation of symmetrizations by polarizations. *Proc. Amer. Math. Soc.*, 134(1):177–186, 2006.

[65] Aljoša Volčič. Random Steiner symmetrizations of sets and functions. *Calc. Var. Partial Differential Equations*, 46(3-4):555–569, 2013.

[66] Fred B. Weissler. Logarithmic Sobolev inequalities for the heat-diffusion semigroup. *Trans. Amer. Math. Soc.*, 237:255–269, 1978.